Pre-Analysis Plan for Phase II of 'The Effects of Media Messages on Social Attitudes in Uganda'

Susanne Baltes, Jasper Cooper, Donald Green, Anna Wilke

October 16, 2016

Contents

1 Pre-Analysis Plan

This document summarizes key features of Phase II of the Media and Social Norms in Uganda Randomized Trial and describes the authors' pre-analysis plan (PAP). The design of Phase I of this project has been registered under the following ID: 20160119AA. This document predates endline data collection, and our plan is therefore blind to outcomes. Any contingency not accounted for in this PAP will be dealt with according to the Standard Operating Procedures for Don Green's lab at Columbia as of June 7, 2016.

2 Motivation

The objective of this study is to estimate the impact of short video vignettes on the attitudes of viewers and those in their local community. Specifically, the project evaluates the impact of dramatized messages on views on three issues: stigma against women who have had abortions, domestic violence against women, and teacher absenteeism. Three short videos for each issue were produced and screened in rural Uganda, where these three issues are prominent social concerns. The videos themselves, which were produced by Peripheral Vision International in the local Lugandan language using Ugandan actors, may be found in the digital materials that supplement this PAP.

3 Intervention Description

Screenwriters for each video were instructed to emphasize prescriptive social norms: authoritative statements from leaders, professionals, or sympathetic characters about proper behavior. In the case of abortion, the prescriptive norm in question is that we are obliged to help those who suffer from complications resulting from an abortion regardless of our personal views about abortion. For domestic violence, the norm is that those who become aware of domestic violence should intervene to prevent violence from spiraling out of control. For

absenteeism, the norm is that parents and guardians should take an active role in monitoring their children's education and take collective action in order to address problems if they arise. The overarching hypothesis is that viewers will become sensitized to these norms and diffuse them throughout the community. The videos can be viewed at this address: http://tiny.cc/uganda_media.

The video vignettes were inserted in commercial breaks during six films that were presented during a film festival in 112 Ugandan villages which took place between July 30 and September 4, 2016. The film festival consisted of the screening of a different movie including the inserted video vignettes every week on either Saturday or Sunday for six consecutive weeks. Screenings were free of charge and the films were feature-length movies unrelated to the three topics of interest in this study. The attached map shows the location of the villages in this study. Villages were blocked by geographic location and randomly assigned to one of seven treatment arms: placebo (no vignettes), vignettes on abortion only, vignettes on domestic violence only, vignettes on teacher absenteeism only, vignettes on abortion and domestic violence, vignettes on abortion and absenteeism, and vignettes on domestic violence and absenteeism. The film festival and randomly assigned video messages were administered nearly flawlessly in 112 villages. In two villages, one scheduled film did not air. In one case, this was due to the video hall owner suspecting the film of spreading black magic; in another case, a local leader sought to prevent the screening apparently in an effort to extract a gratuity. In neither case, was this due to the treatment vignette featured in the film.

4 Design and Randomization

4.1 Site selection and blocking

As the intervention requires the presence of a video hall in all study sites, we conducted an initial census of all Ugandan trading centers (TC) with video halls in the districts of Mubende, Mityana, Masaka and Lwengo, which led to the identification of 342 video halls in approximately 300 candidate trading centers. During this census, potentially problematic sites were identified according to a number of pragmatic criteria and non-randomly excluded, narrowing down the candidate trading centers to 247.[1]

We sought to create a sample of 112 sites (trading center clusters) to include in the study. With seven treatment conditions, this allows us to organize the sample into 16 blocks of 7 clusters each.

Having non-randomly excluded some clusters from our census, we sought to choose the largest set possible while imposing the constraint of all clusters being at least 4.5 km apart. Using a random walk algorithm in `C++`, we were able to choose a set consisting of 125 clusters meeting this spatial criterion, imposed to address concerns about spillovers. However, we were unable to use all of the clusters in this set: some trading centers were either too big (making it hard to find respondents likely to have seen the film) or too small (prohibiting the requisite number of respondents being interviewed); some did not have satellite imagery available (inhibiting our sampling strategy); at others the owner of the video hall was impossible to contact; and still yet others did not appear to be located in normal townships, but instead were located at mines or military barracks. Therefore, we ended up replacing 19 clusters in the initially selected set by hand-selecting other clusters sufficiently distant from the remaining set. The respondents in the clusters selected in this manner comprise our sample.

4.2 Random assignment code

Before assigning units to treatment we organized the sample into 16 blocks of 7 clusters each using the `blockTools` package and the following code:

[1] We excluded video halls that operated seasonally, could not be rented on weekends, for which the GPS coordinates were missing, that only screened football matches instead of movies, and where enumerators indicated it may be difficult to successfully screen movies.

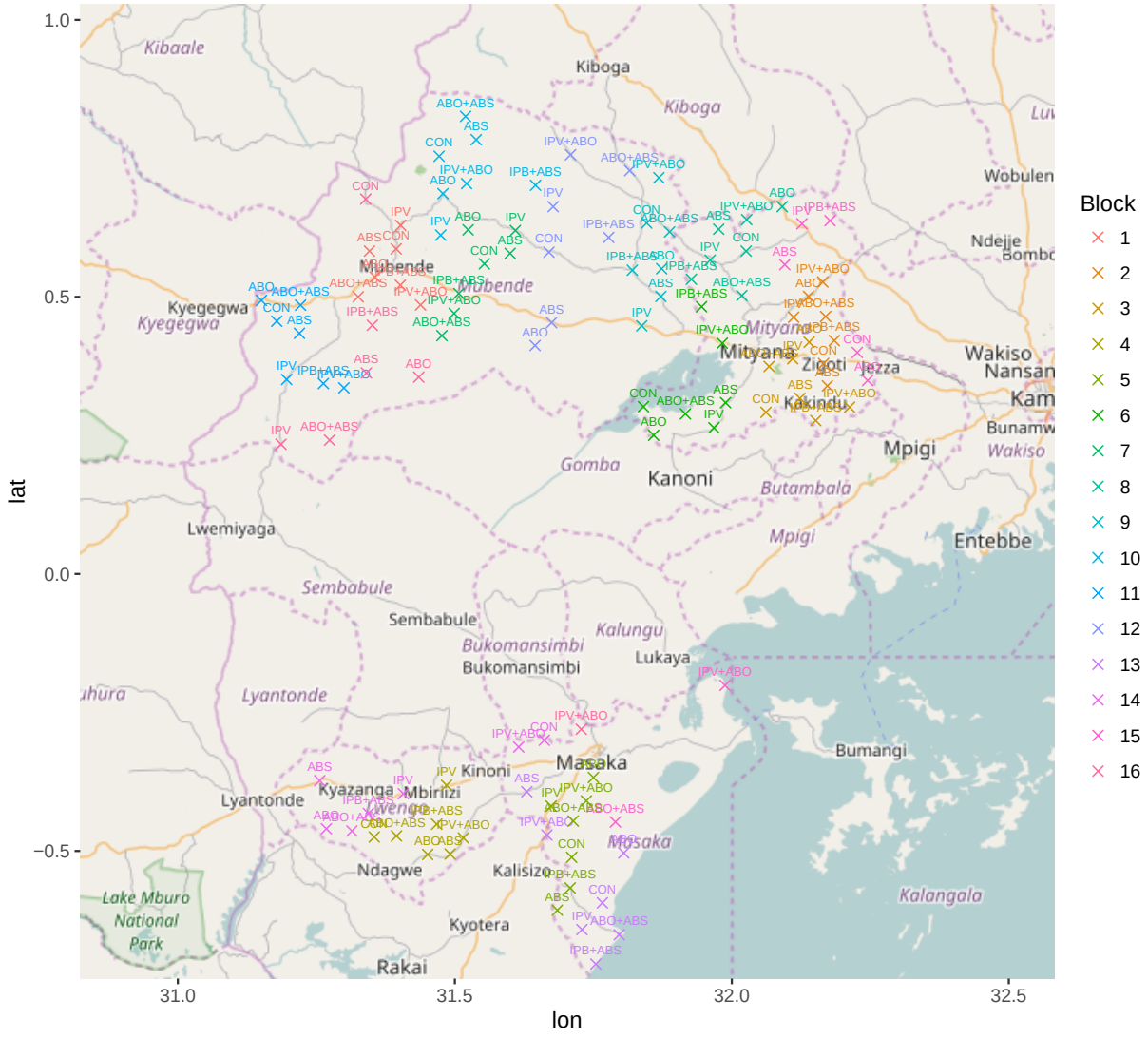

Figure 1: Randomization Map

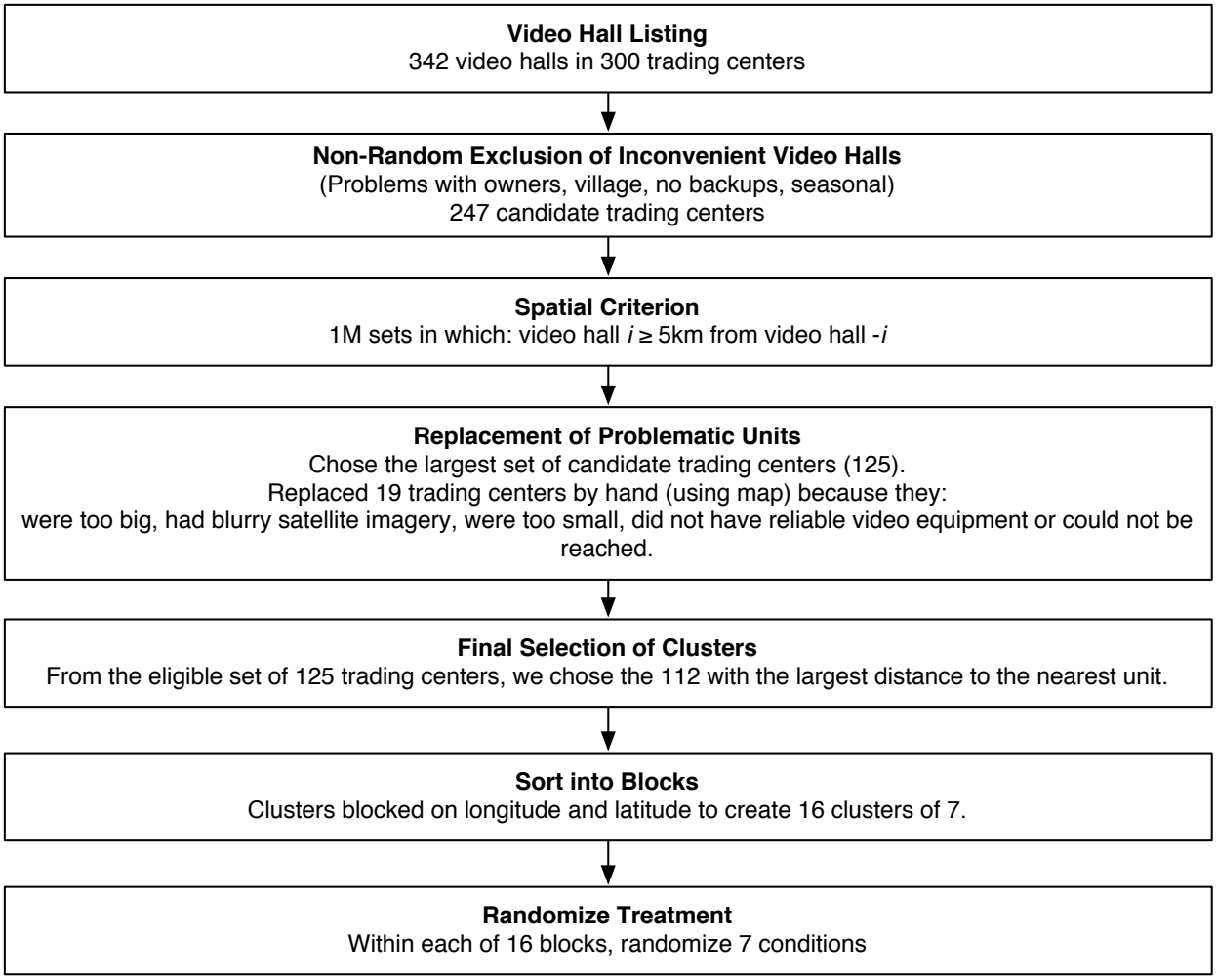

Figure 2: Population Selection

```
library(blockTools)
data <- read.csv(file = "final_set.csv")
block_vars <- c("longitude","latitude")
block_object <- block(data = data,
                      n.tr = 7,
                      id.vars = "tc",
                      block.vars = block_vars)
data$block_id <- createBlockIDs(obj = block_object, data = data, id.var = "tc")
```

We then randomly assigned the seven treatment conditions within each block using the following function.

```
assign_treatment <- function(
  block_vector, condition_names, unit_ID
){
  # Find unique blocks
  unique_blocks <- unique(block_vector)
  # Check condition_names length == number within blocks
  block_lengths <- table(block_vector)
  if(any(block_lengths != length(condition_names)))
    stop("There must be as many condition names as units within each block.")
  # Create assignment vector
  assignment_vector <- rep(NA, length(block_vector))
  # Permute treatment within blocks
  for(block_j in unique_blocks){
    assignment_vector[block_vector == block_j] <- sample(condition_names)
  }
  # Create data.frame without backup units
  unit_data <- data.frame(id = unit_ID,
                          block_vector = block_vector,
                          treatment = assignment_vector)
  return(unit_data)
}
```

We set a random number seed, randomly assigned once, and merged this into the data.

```
set.seed(123456789)
condition_names <- c(
  "control",
  "IPV",
  "abortion",
  "absenteeism",
  "IPV_abortion",
  "IPV_absenteeism",
  "abortion_absenteeism"
)
assignment_matrix <- with(
  data,
  assign_treatment(block_vector = block_id,
                   condition_names = condition_names,
                   unit_ID = tc)
)
assignment_data <- merge(data,
                         assignment_matrix,
```

```
                        by.x = "tc",
                        by.y = "id",
                        all.x = TRUE)
```

4.3 Sampling and subjects

For the endline survey, we plan to conduct 4480 interviews with respondents between 18 and 65 years of age (40 respondents in each of 112 trading centers), plus a follow-up sample of respondents between 18-35 (described below). Additionally, we plan to interview 2240 teenagers between 15 and 17 years of age (20 teenagers in each of 112 trading centers). Apart from the household surveys, we also conduct interviews with all members of the Village Health Team (VHTs) in each of the 112 trading centers (approximately 560 respondents, 5 VHTS in each of 112 trading centers).

Within each trading center, respondents for the household surveys are sampled from a circular area around the video hall that was used to screen the treatment messages. Enumerators receive a map for each trading center that shows a circle around this video hall with a radius of between 200 and 800 meters.[2] The radius was chosen based on the population density of the given trading center as judged from satellite images. Enumerators work with village leaders (LC1 chairpersons) or other knowledgeable members of the community to produce a list of all households that reside within the circle indicated on the map. The households on this list will be shuffled so that they are in random order. The first 40 households on this randomly ordered list of households will be selected. 20 of these are randomly chosen as households within which a female respondent will be interviewed. A male respondent will be interviewed in each of the other 20 households. Upon entering a household, enumerators list all adult household members (aged 18-65) of the relevant gender and randomly select one of them as the respondent. If no respondent of the relevant gender resides in the selected household, another household is randomly chosen from the list of households within the circle around the video hall. Enumerators and adult respondents will be matched by gender. In each of the 40 households, we will also interview a teenager if a teenager in the relevant age range and of the relevant gender resides in this household: Female teenagers will be interviewed by female enumerators (i.e. in households with female adult respondents), while male teenagers will be interviewed by both male and female enumerators (i.e. in households with either female or male adult respondents). We anticipate that around half of the 40 households will have a teenage member of the relevant gender and within the relevant age range so that we end up with approximately 20 teenager respondents per trading center.

We will do a second round of sampling to increase the number of direct Compliers in our sample. The sampling for the second round will proceed in the following way: We will identify the two clusters in each of the 7 treatment conditions with the fewest Direct Compliers. If more than two clusters have the same lowest number of Compliers, we randomly select two clusters from the set of clusters with the lowest number of Compliers. If there is a unique cluster with the lowest number of Compliers but multiple clusters with the same second lowest number of Compliers, the cluster with the lowest number of Compliers will be selected. Additionally, one cluster will be randomly selected from the set of clusters with the same second lowest number of Compliers. In this second round of sampling for Compliers, households are selected from the same randomly ordered list, continuing the sequence followed in the endline, so that the sampled units are the same units that would have been sampled had we continued endline data collection. In order to over-sample Compliers, the sampling strategy within households will be altered to target respondents between 18 - 35, aiming for a target of 2/3 men.

[2] There is one trading center for which no satellite images were available to create a map. In this case, an enumerator was sent to the village ahead of time to identify landmarks that could be used to approximate a radius.

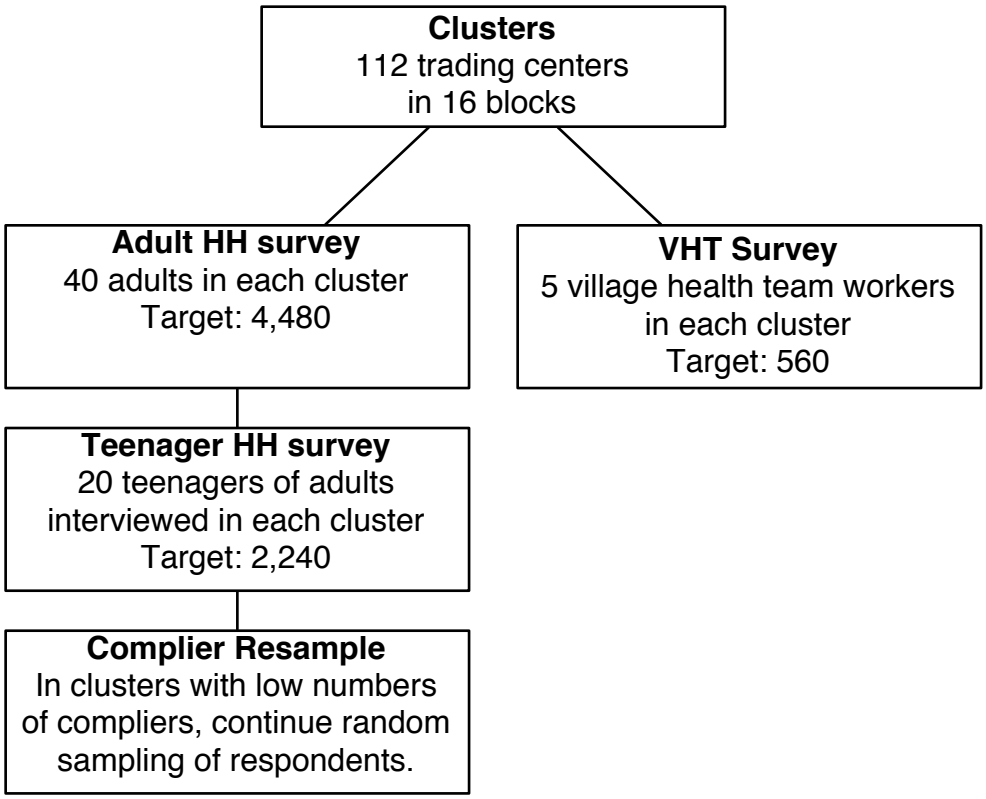

Figure 3: Sampling

5 Covariates

5.1 Variable selection

We will use lasso regression (described in greater detail below) to select the minimal number of covariates that best predict each outcome, and include only these in our estimation. The pool of covariates that we will choose from varies slightly for different models:

1) Individual-level analysis with adult and teenage respondents respectively

- Variables taken from our household surveys with adult and teenage respondents that we consider as unlikely to have been influenced by our treatments. These variables have been flagged as covariates in the attached adult and teenager questionnaires.
- The distance of each respondent's household from the video hall in which our treatment messages were screened.
- The length of the radius of the circle around the video hall from which we are sampling our respondents.
- Information about characteristics of the trading center taken from the survey with VHTs. The corresponding questions have been flagged as cluster level covariates in the attached VHT questionnaire. They will be averaged over all VHTs from the same trading center.
- For models that have the number of incidents in which a women was beaten as an outcome (see item `woman_beaten` in adult questionnaire), we will also include the number of beatings of non-women as a potential covariate in the lasso regression (`beatings` - `women_beaten`, see adult questionnaire).

2) VHT-level or trading center level analysis of outcomes taken from VHT survey

- The length of the radius of the circle around the video hall from which we are sampling our respondents.
- Information about characteristics of the trading center taken from the survey with VHTs. The corresponding questions have been flagged as cluster level covariates in the attached VHT questionnaire. For the trading-center analysis, these variables will be averaged over all VHTs from the same trading center.
- Characteristics of individual VHTs taken from the VHT survey which are plausibly unaffected by the treatment. The corresponding questions have been flagged as individual level covariates in the attached VHT questionnaire. For the trading-center analysis, these variables will be averaged over all VHTs from the same trading center.
- Variables taken from our household surveys with adult respondents that we consider as unlikely to have been influenced by our treatments. These will be averaged across all respondents from the same trading center.

We will perform lasso on covariate data that has item-level missingness removed through multiple chained equations, as described below.

The lasso procedure that we plan to use features a generalized linear model with lasso penalization, and is implemented in the `glmnet` package for `R`. The loss function requires selecting a regularization parameter, λ, that determines the severity of the penalty for including extra covariates. Since this regularization parameter cannot be optimally chosen in advance, we will select it using 10-fold cross-validation. Specifically, for each outcome, we will choose the λ that minimizes the 10-fold cross-validation error averaged over 10 runs (since the folds are chosen at random). Only the covariates retained by the lasso will be included in the specification.

We will perform this lasso variable selection method using the entire list of covariates, including block and resample indicators, among the entire sample. This implies that for each outcome, a different number of covariates may be included on the right-hand side.

In addition to the list of covariates selected by the lasso method, we will also include indicators for block membership, for membership in the Complier resampling round, as well as a variable that stores the average (across the six video screenings) of the number of people that attended a screening in the respondent's trading center. (In trading centers in which less than six screenings took place, the average will be calculated across the five screenings that did take place). The right-hand side will thus always include lasso-selected covariates, block and resample indicators, average audience size and treatment indicators. We condition on average audience size because the sizes of clusters may be heterogeneous, due to the uneven numbers of different compliance subgroups in each cluster. If potential outcomes are correlated with cluster size, this can bias estimates of the average treatment effect in the population from which the respondents were sampled. In the pilot study we found evidence to suggest that the degree of cluster-level heterogeneity in the outcomes we study is small, and so we do not see this issue as a big concern. Nonetheless, we control for the average audience size in order to mitigate issues arising from heterogeneity in cluster size.

6 Outcomes

6.1 Outcome measures

Our outcomes are drawn from endline surveys with adult and teenage villagers as well as from surveys with VHTs. The three questionnaires (adults, teenagers, VHTs) are presented in a supplementary appendix. A large number of individual survey questions concerning abortion, domestic violence, and teacher absenteeism comprise our outcome measures. All questions that will be used as outcome measures have been flagged as such in the attached questionnaires. The attached questionnaires also contain information on what type of orientation (attitudes, norms etc.) each question is capturing. Below, we discuss the types of orientations that we are measuring and how outcomes will be coded.

6.1.1 General Coding Rules

- Where not otherwise indicated, outcome measures will be constructed from single items.
- Items with two answer categories will be coded as indicator variables which take the value 1 if the respondent chooses the answer category that is more in line with the norm portrayed in our videos. In the attached questionnaires, answer options that we consider to be in line with the norm portrayed in the videos are shown in green.
- For items with more answer categories, we will create numerical variables that take higher values the more the respondent's answer is in line with the norm portrayed in the video.
- Indices will be constructed in the following manner: All variables that form part of the index will be standardized to range from 0 to 1 by dividing by their maximum possible value. Subsequently, they will be summed and divided by the number of variables that form part of the index.

6.1.2 Conative Attitudes

Our measures of conative attitudes capture the respondent's proclivity to take actions in line with the norms portrayed in our treatment videos, i.e. to intervene against domestic violence, to help women who suffer from post-abortion complications, and to engage in collective action to solve the problem of teacher absenteeism. Items that measure conative attitudes present respondents with a hypothetical scenario and ask the respondent to choose between two possible courses of action, one of which we interpret as in line with the norm (marked as green in the questionnaire). For each scenario, respondents are asked to make between one to four choices between two hypothetical courses of action.

Explicitly Modeled Versus Unmodeled Actions

We distinguish between items where the action that is in line with the norm is explicitly modeled in our videos and items where it is not. Items that we interpret to contain an explicitly modeled action that is in line with the norm have been flagged as such in the attached questionnaire.

Coding Details

Where respondents are asked to evaluate more than one pair of action for the same scenario, actions have been randomly paired (see attached questionnaire and the attached document randomization outcome questions for details). In these cases, we will create indicator variables that take the value 1 for every respondent that chooses a given option that is in line with the norm regardless of the counter-normative option that this action has been paired with for this particular respondent. For instance, the item `report_police` as it is shown in the questionnaire asks the respondent to choose between the following two actions in response to a request for help from a victim of domestic violence: 'a) I would accompany her to the police to report the incident.' and 'b) I would calm her down and tell her that the situation is bound to get better.' We interpret option a) as the one that is line with the norm that one should take action against domestic violence. Due to the randomized pairing of outcome options, only a randomly chosen subset of respondents will be asked to choose between actions a) and b), the rest will face a choice between a) and one of three other actions that are not in line with the norm of intervening. We will create a indicator variable that will take the value 1 for any respondent who chooses action a) regardless of the counter-normative action that a) was paired with for that respondent. We will build two additive indices for each issue area – one that combines all items with actions that are explicitly modeled in our videos and one that combines all items with actions that are not explicitly modeled in our videos.

6.1.3 Conative Prescriptive Norms

Our measures of conative prescriptive norms capture respondents' views on what other people *should* do in a given scenario. Again, respondents are asked to choose between two possible actions that they might want others to take one of which is in line with the norms portrayed in the videos.

6.1.4 Conative Descriptive Norms

Our measures of conative descriptive norms capture respondents' views on what other people *would* do in a given scenario.

6.1.5 Efficacy

Our efficacy items pertain to whether respondents perceive taking the actions that are in line with the norm portrayed in the videos as being effective.

6.1.6 Discussion

We are interested in whether our treatments affect the extent to which villagers discuss our three topics of interest – also in order to gauge the plausibility of spillover effects from those who saw the videos to others who were not directly exposed to them.

Coding Details

The item `discussion` (see attached questionnaires) asks respondents to indicate whether they discussed each of a number of topics – among them our three topics of interest. We will use this item to create three indicator variables, one for each of our topics. These dummies will take the value 1 if a respondent indicated the he or she has discussed the respective topic in the past week.

6.1.7 Cognitive Attitudes

We also ask respondents whether they approve of domestic violence and abortions under varying scenarios. We do not measure cognitive attitudes with regard to teacher absenteeism, since it is generally perceived as a social bad and as such is not a normatively contested issue in Uganda.

Coding Details

Our outcome measures for cognitive attitudes consist of two indices capturing views on IPV and abortion respectively. For IPV, each respondent will answer the item `disobey` and one randomly chosen item out of the following three: `chatting`, `financial pressure` and `housework`. For each respondent, the index will be an additive combination of the answers to the two items that the respondent answered. For abortion, the index will combine the items `moth_health` and `no_children`.

6.1.8 Cognitive Norms

Our measures of cognitive norms pertain to respondents' views of whether others in their community approve of domestic violence. We do not include measures of cognitive norms with regard to abortion and teacher absenteeism.

Coding Details

Our outcome measures for cognitive norms consist of an index capturing perceptions of others' views on IPV. Each respondent will answer the item `comm_disobeys` and another norm item from the following set: `comm_chatting`, `comm_fmancial_pressure` and `comm_housework` (the one that corresponds to the randomly chosen attitude item). For each respondent, the index will be an additive combination of the answers to the two norm items that the respondent answered.

6.1.9 Priorities

We also measure the extent to which respondents perceive problems related to our treatment messages as priorities.

Coding Details

The item `goals` will be recoded into three dummies. The first indicator will take the value 1 if the respondent chooses 'Reducing violence against girls and women' as a priority. The second indicator will take the value 1 if the respondent chooses 'Reducing illiteracy' as a priority. The third indicator will take the value 1 if the respondent chooses 'Reducing the number of women dying after giving birth' as a priority.

The item `spending` will be recoded into two dummies. The first indicator will will take the value 1 if the respondent chooses 'More health centers and doctors' as a priority. The second indicator will take the value 1 if the respondent chooses 'More schools and teachers' as a priority.

6.1.10 Incidents and Reporting

We obtain measures of incidents and reporting from our household survey with adult respondents and from our VHT survey. From both surveys, we will construct three sets of outcomes;

1) **Count variables** that measure the number of incidents of domestic violence or abortion that the respondent is aware of.

2) **Categorical variables** capturing whether there were no incidents, unreported incidents or reported incidents (and in one case who reported the incidents).

3) **Other variables** such as perceptions of frequency with which incidents happen and whether incidents ever go unreported.

Coding Details

Variables from adult household survey

- The event count variable `women_beating` that contains the number of incidences in which a woman was beaten that a respondent recalls.
- A categorical variable that takes the value 0 if `women_beating = 0`, the value 1 if `women_beating > 0` and `reporting = 'None'` and the value 2 if `women_beating > 0` and `reporting != 'None'`.
- The variable `beating_frequency` will be recoded into a categorical variable that takes the value 0 if `beating_frequency = Less than once a month`, 1 if `beating_frequency = About once a month`, 2 if `beating_frequency = Just about once a week` and 3 if `beating_frequency = Almost every day`.

Variables from VHT survey

- Two count variables that contain the number of times somebody approached the VHT because of violence in the family and complications following an abortion respectively (see the item `reporting_vht`). Both count variables will be aggregated to the trading center level. There are two approaches to this aggregation. The first is to consider each VHT's incidents report as a noisy measure of the true number of incidents in the trading center. This interpretation suggests taking the average over the reports of all VHTs within the same trading center. The second asks what the most extreme scenario is that has been reported for a given trading center which suggests selecting the maximum number of incidents that have been reported in a given trading center. We will reported results using both approaches.

- The items `unreported_abortion`, `unreported_ipv` and `complain_absenteeism` will be recoded into dummies that take the value 1 if the VHT answered `no`. They will be aggregated to the trading center level by taking the average.

We will also create the following two categorical variables which will not be aggregated to the trading center level because we model them as choices using an ordered probit estimator:

- A categorical variable that takes the value 0 if the VHT reported `complications following an abortion = 0` in response to item `reporting_vht` and `incidents_abortion_unreported == 0`, the value 1 if the VHT reported `complications following an abortion = 0` in response to item `reporting_vht` and `incidents_abortion_unreported > 0`, the value 2 if the VHT reported `complications following an abortion > 0` in response to item `reporting_vht` and `person_asking_abortion = 'the woman who had the abortion'` for all reported cases, the value 3 if the VHT reported `complications following an abortion > 0` in response to item `reporting_vht` and `person_asking_abortion != 'the woman who had the abortion'` for at least one reported abortion case.
- A categorical variable that takes the value 0 if the VHT reported `violence in the family = 0` in response to item `reporting_vht` and `incidents_ipv_unreported == 0`, the value 1 if the VHT reported `violence in the family = 0` in response to item `reporting_vht` and `incidents_ipv_unreported > 0`, the value 2 if the VHT reported `violence in the family > 0` in response to item `reporting_vht` and `person_asking_ipv = 'victim'` for all reported cases, the value 3 if the VHT reported `violence in the family > 0` in response to item `reporting_vht` and `person_asking_ipv != 'victim'` for at least one reported violence case.

6.1.11 Qualitative Outcomes

In addition to the outcome measures which are based on closed ended questions, we will also use outcome measures that are created based on the transcriptions of audio recordings of parts of our endline interviews with adults and teenagers. For details of these outcome measures see the coding guide in the supplementary material.

6.2 Multiple Comparisons

We recognize that this assortment of outcome measures creates a multiple comparisons problem, which we address by the following procedure. Estimated t-ratios for the analyses will be sorted in descending order (for all one-tailed hypotheses). A null distribution of t-ratios will be simulated by repeating the random assignment 10,000 times and for each simulation generating a vector of estimated t-ratios. To obtain the p-value for the largest estimated t-ratio, we will ask, what proportion of the simulations generated a largest p-value at least as large as what we obtained empirically? The same procedure will be applied to the second largest p-value, and so forth (Anderson 2008).

6.3 Item-level missingness

Non-response to outcome questions will be dealt with through imputation methods. Specifically, we will use multivariate imputation via chained equations (MICE) as implemented in the `mice` package for `R`. Imputations will be performed within issue areas: missing responses to IPV outcomes will be imputed using observed responses to all other IPV outcomes, and so forth.

7 Definition of Latent Types

7.1 Individual Compliance

Our analysis will zero in on four important subgroups which will be defined based on the following two questions in our adult and teenager questionnaires:

19.1) Recently, a series of six free films (Pirates of the Caribbean, Creed, Fast and Furious, Spy, Slumdog Millionaire, Oz The Great And Powerful) were screened in the kibanda in your trading center. Have you heard about the screenings and if so, how many screenings did you attend?

- 6
- 5
- 4
- 3
- 2
- 1
- 0 but knew about screenings
- 0 did not know about the screenings
- Don't know
- Refuse to answer

19.2) Did your friends or family attend any of the screenings?

- Yes, friends and family
- Yes, friends
- Yes, family
- No
- Don't know
- Refuse to answer

The first group may be called Compliers – those who would attend a film festival regardless of the video messages embedded in its commercial breaks. Under the exclusion restriction, which stipulates that nothing about treatment assignment influences potential outcomes apart from exposure to the messages, we can obtain unbiased estimates of the CACE by restricting the analysis to those subjects who report having attended at least one of the films in response to question 19.1). A second subgroup comprises Indirect Compliers: those who did not attend the film(s) themselves but report that family or friends attended. The third group consists of Apprised Never-Takers who report that they knew about the film(s) but did not attend any screenings and also do not report that friends and family attended. Finally, the never-takers are those who, according to their endline survey reports, neither attended or knew about the screenings nor have direct family or friendship ties to those who attended.

7.2 Intra-Household Combinations of Compliance

Our main analysis draws on the latent types described in the previous section. In a second analysis, we will concentrate on households from which we interviewed both an adult and a teenager in order to identify intra-household spillovers. For this analysis, we will collapse Indirect Compliers, Apprised Never-Takers and Never-Takers into one category of Never-Takers and focus on the following four subgroups:

1) households in which the parent and the teenager are Compliers
2) households in which the parent is a Complier and the teenager is a Never-Taker
3) households in which the teenager is a Complier and the parent is a Never-Taker
4) households in which the parent and teenager are Never-Takers.

Subgroup	Answer to 19.1)	Answer to 19.2)
Compliers	1,2,3,4,5, or 6	anything
Indirect Compliers	0 but knew about screenings 0 did not know about the screenings Don't know Refuse to answer	Yes, friends and family Yes, friends Yes, family
Apprised Never-Takers	0 but knew about screenings	No Don't know Refuse to answer
Never-Takers	0 did not know about the screenings Don't know Refuse to answer	No Don't know Refuse to answer

8 Crossover effects

When designing the study, we expected each of the three messages to affect only views and perceived norms in that issue domain. In other words, we expect messages about domestic violence to affect views and perceived norms regarding domestic violence but not regarding abortion or absenteeism. Thus, our core model is one in which views or norms regarding a given topic are represented as a function of village-level exposure to messages on that topic and that topic only. Correspondingly, the hypotheses that we describe in the following section refer to the effect of village-level exposure to messages on one topic on outcome measures pertaining to this topic. This core model is nested within a more general model that allows for crossover effects from messaging in one domain to views or norms in a different domain as well as for interaction effects between different models. As stated above, we do not expect to see crossover and interaction effects and we did not see evidence for such effects in the first phase of our study. That said, the existence of crossover and interaction effects is not entirely implausible, because all our treatments convey the message that one should take action of some sort. The videos on domestic violence have been changed to put a stronger emphasis on action orientations since phase 1. As described in more detail in the analysis section, we will thus test for the existence of crossover and interaction effects. If crossover and interaction effects exist, we expect them to be positive. We also expect them to be strongest for the abortion and IPV treatments given that both are related to views on gender roles.

9 Hypotheses

Our main analysis will focus on compliers, the subgroup for which we expect effects to be largest. Below, we state our hypotheses with regard to this group. Subsequently, we explain our expectations with regard to other subgroups.

9.1 Main Analysis

The following hypotheses will be tested separately for adults and teenagers. Hypotheses which involve outcomes that have also been included in the VHT survey will also separately be tested among VHTs.

9.1.1 Primary Hypotheses

We expect to see the strongest treatment effects for the following classes of outcomes:

Discussion

Hypothesis: A higher proportion of treated Compliers indicates having discussed the topic of the video messages that they were exposed to over the past week. (One-tailed).

For this hypothesis, we will test the joined significance of the estimates of the three treatment effects (the effect of the domestic violence treatment on discussions of domestic violence, the effect of the abortion treatment on discussions of abortion and the effect of the absenteeism treatment on discussions of teacher absenteeism).

Conative Attitudes (modeled)

Since all of our videos focus on norms related to the need to take action to help a victim or solve a problem, we strongly expect to see effects on conative attitudes. Based on results from phase 1, we especially expect to see our treatments increase the respondents' proclivity to choose actions that were explicitly modeled in our videos. We are less certain that our treatments will increase the respondent's proclivity to choose actions that are in line with the norms promoted in our videos but that were not explicitly modeled in our videos.

Hypothesis: A higher proportion of treated Compliers selects the action that conforms to the norm modeled in the videos. (One-tailed).

Conative Prescriptive Norms

Hypothesis: A higher proportion of treated Compliers states that others believe that one should take the action that conforms to the norm modeled in the videos. (One-tailed).

9.1.2 Secondary Hypotheses

We might see treatment effects for the following classes of outcomes:

Conative Attitudes (not modeled)

Hypothesis: A higher proportion of treated Compliers selects the action that conforms to the norm modeled in the videos. (One-tailed).

Conative Descriptive Norms

Hypothesis: A higher proportion of treated Compliers indicates that others will take the action that conforms to the norm modeled in the videos. (One-tailed).

Efficacy

Hypothesis: A higher proportion of treated Compliers indicates that taking the action that conforms to the norm modeled in the videos will be effective. (One-tailed).

9.1.3 Tertiary Hypotheses

We expect weak treatment effects if any for the following classes of outcomes:

Cognitive Attitudes

Hypothesis: A lower proportion of treated Compliers indicates that they believe domestic violence (abortion) is justified across different scenarios (One-tailed).

Cognitive Norms

Hypothesis: A lower proportion of treated Compliers indicates that other believe that domestic violence is justified across different scenarios (One-tailed).

Priorities

Hypothesis: A higher proportion of treated Compliers chooses problems related to our messages as priorities (One-tailed).

9.2 Analysis of Never-Takers

We do not expect to see any effects for Never-Takers. As a robustness check, we will test all of the above hypotheses among Never-Takers, separately for adults and teenagers. The results will be reported in an appendix.

9.3 Intra-Cluster Spillover Analysis

This analysis will again be conducted separately for adults and teenagers. We will test all of the above hypotheses among the other compliance groups and hypothesize that treatment effects become steadily weaker as we move from Compliers to Indirect Compliers to Apprised Never-Takers. We do not expect to see treatment effects on Indirect Compliers and Apprised Non-Compliers for outcomes where there was no treatment effect on Direct Compliers. We will therefore only test those hypotheses among Indirect Compliers and Apprised Non-Compliers for which we can reject the null hypothesis among Direct Compliers on at least the 10% significance level. As an exception to this rule, we will also test those hypotheses among Indirect Compliers and Apprised Non-Compliers for which we could reject the null hypothesis among Direct Compliers in phase 1 of the study on at least the 10% significance level.

9.4 Intra-Household Spillover Analysis

We expect strong effects for respondents from households where both the parent and the teenager are Compliers. We interpret any effects on teenage respondents in households with an adult Complier and a teenage Never-Takers as intra-household spillovers. Likewise, we interpret any effects on adult respondents in households with an adult Never-Taker and a teenage Complier as intra-household spillovers. Again, we will only look for such intra-household spillovers with regard to outcomes for which we find an effect on Direct Compliers that is significant on at least the 10% level and for those outcomes for which we found such an effect in phase 1 of our study. We do not expect any effects on respondents households where both the parent and the teenager are Never-Takers.

9.5 Incidents and Reporting

With regard to the number of domestic violence and abortion incidents reported by adult villagers and VHTs, the effects could go in two directions: On the one hand, we expect that our treatments will increase the reporting of such incidents and the attentiveness of villagers to such issues. This would imply an increase in the number of reported incidents in response to village-level exposure to our videos. On the other hand, our treatments might decrease the occurrence of such incidents. Even though we do not consider it very likely that our treatments are strong enough to produce the second effect, we will conduct a two-tailed test of the following hypothesis:

Hypothesis: Respondents (VHTs) in treated villages recall more (or less) incidents of domestic violence (abortion) (Two-tailed).

For both the household survey and the VHT outcomes, we will first test the above hypothesis. If we reject the null hypothesis of no treatment effect in favor of a negative effect on the incidents rate, we will refrain from analyzing the categorical variables. If we cannot reject the null hypothesis of no negative effect on the incidents rate (significant positive or insignificant effect), we will test the following hypothesis involving the categorical outcomes:

Hypothesis: Respondents (VHTs) in treated villages are more likely to recall more incidents of domestic violence (abortion), those incidents are more likely to have been reported (and in case of the VHT outcomes, reporting is more likely to have been done by somebody other than the victim) (One-tailed).

We also test the following hypotheses:

Hypothesis: Respondents in treated villages perceive domestic violence to happen more (or less) frequently (Two-tailed).

Hypothesis: A lower proportion of VHTs in treated villages reports that cases of domestic violence (abortion) ever go unreported (One-tailed).

Hypothesis: A higher (or lower) proportion of VHTs in treated villages reports that parents complain to them about the problem of teacher absenteeism (Two-tailed).

All the hypotheses that pertain to incidents and reporting outcomes from our household survey will be tested among all respondents (ITT) as well as among direct compliers (CACE). The reasoning is that our treatments might either affect the actual number of incidents occurring or being reported in a given trading center. Any respondent regardless of his or her compliance type could pick up on such a change. In this case, we are interested in the ITT. Instead, our treatment might affect how much attention respondents pay to incidents of domestic violence. If changes in the incidents and reporting outcomes are driven by this mechanism, then we would expect effects to be strongest among Direct Compliers, i.e., we are interested in the CACE.

9.6 Dosage

We expect treatment effects to be larger the more screenings a Direct Compliers has attended. Please see the analysis section for how we will test for dosage effects.

9.7 Qualitative Outcomes

Our hypotheses for the analysis of our qualitative outcome measures parallel our other hypotheses in the sense that we strongly expect treatment effects on categories related to action orientation, i.e. we expect treated Compliers to become more likely to say that they will take action, show empathy and suggest alternative courses of action. We have fewer expectations for treatment effects on statements relating to cognitive attitudes, but still hypothesize that treated Compliers will be less likely to unconditionally support domestic violence or oppose abortion, more likely to unconditionally oppose domestic violence and support abortion and more likely to qualify any statements of support for domestic violence and opposition to abortion.

9.8 Heterogeneous Treatment Effects

We will look for heterogeneous treatment effects by gender and by whether the respondent's first language is Luganda. In terms of the heterogeneous effects by gender, we consider the direction of the interaction effect as theoretically ambiguous and will hence conduct two-tailed tests. In terms of the heterogeneous effect by language, we expect videos to be more influential among those who can easily understand the dialogue (one-tailed).

We also plan to analyse the transmission of effects from Compliers to others as a function of distance to the bibanda. The details of this analysis will be determined based on data from the first phase of the study.

9.9 Magnitude of Effects in Phase I and Phase II

We will also test how treatment effects of the domestic violence treatment differ between phase I and phase II of the study (two-tailed), because our messaging on domestic violence was changed for phase II. Specifically, the messaging on domestic violence in phase II puts a stronger emphasis on empathy with the victims of domestic violence and on the need for action to stop domestic violence before it spirals out of control. We do not expect treatment effects for the abortion and teacher absenteeism treatments to differ between phase I and phase II, because the messaging on these topics has not been changed.

10 Estimation

10.1 Main Specifications

For most analyses with binary and for all analyses with continuous dependent variables, we estimate treatment effects using least squares regression with clustered standard errors and fixed effects for block and resample membership as well as controlling for average audience size. For some analyses with ordinal dependent variables, we will use an ordered probit model to estimate effects.

As described in the section on variable selection above, we will use lasso regression to select the covariates most predictive of the outcome in each specification. The variables retained by this procedure will be added to the right hand side of the regression equation alongside the block and resample indicators and a measure of average audience size.

10.2 Compliance Subgroup Analysis

Almost all analyses will condition on membership to different latent types in the sample, defined in the manner described above. In the analysis, this will be achieved by subsetting the data used for analysis to the relevant subtype. Thus, when looking at effects among Never-Takers, for example, the analysis will be run only among those defined as Never-Takers.

10.3 Characterization of Uncertainty

Unless otherwise indicated, we will run randomization inference using the random assignment function described above to calculate p-values on quantities of interest, and will consider these the final arbiters on the inference drawn. Standard errors will serve primarily as a heuristic and will not form the basis for inference. Cluster-robust standard errors will be calculated using the `sandwich` package for `R` for all least squares specifications on the individual level, and bootstrapped standard errors will be calculated for ordinal probit models. Given the large number of clusters in this study, these methods for estimating standard errors are expected to generate reliable results.

10.4 Crossover Effects Diagnosis

As described above, we use the term ''crossover effects" to refer to the effects that videos on one substantive topic may have on outcomes associated with another substantive topic. For example, if the IPV videos were to affect attitudes about abortion, that would constitute a crossover effect. For each outcome we analyze, we will run a two-step diagnosis for crossover effects.

1. The first step is only run if the outcome is abortion- or IPV-related; as specified above, these are two outcomes where we do see a positive interaction as plausible. We take the observed difference in F-statistics between the main model that only has the relevant treatment indicator and covariates on the right-hand side, and an augmented model that contains the other treatment indicator (IPV if the outcome is abortion-related, and vice versa) and its interaction with the main treatment indicator. We then permute the treatment assignment 1000 times, and obtain a p-value for the difference in F-statistics.
2. In the second step, which we carry out for all models, we will run the 'saturated crossover' model, which contains all treatment indicators and their pairwise interactions. We follow a similar F-test procedure using randomization inference as described in step 1, and obtain a p-value for the observed difference in F-statistics.

If the p-value obtained in step 2 is less than .05, we report the fully saturated model from step 2. If the p-value obtained from step 2 is greater than .05 but that from step 1 is less than .05, we report the IPV-Abortion interaction model from step 1. If neither p-value is less than .05, we report the main specification with only the relevant treatment indicator and covariates on the right-hand side.

10.5 Analysis of Village Health Team Surveys

The village health team surveys contain outcomes related to the reporting of incidents related to IPV, abortion and teacher absenteeism. As described in the outcome section above, some of these are coded as categorical outcome variables and will be analyzed at the individual level using ordered probit analysis. The others will be analyzed using least squares at the aggregate level using the average or maximum to summarize the cluster-specific outcome.

10.6 Estimation Strategy for Dosage Effects

In order to test for dosage effects among Direct Compliers, we will fit the main models as for the main analysis replacing the respective treatment indicator with a variable that takes the value 0 if the Complier resides in a trading center that did not receive the respective treatment message. If the Complier resides in a trading center that was assigned to the respective treatment message, the variable takes the values 1 to 6 depending on how many screenings the Complier reports to have attended. Additionally, the models will include six indicator variables that take the value 1 if a Complier indicated having attended 1, 2, 3, 4, 5 or 6 screenings respectively.

An additional complication for the dosage analysis arises, because, as mentioned above, only five instead of six screenings occurred in two of our trading centers. Since the failure to screen was limited to one of the six screenings per trading center and seemed to be due to reasons unrelated to the messages, our main analysis will ignore this issue and proceed as if all screenings had been implemented in all trading centers. That said we will run two robustness check: First, we will run the analysis only on the villages in which all screenings were aired excluding the two in which this was not the case. Second, we will run the analysis excluding all blocks that contain a trading center in which less than six screenings happened.

10.7 Estimation Strategy for Heterogeneous Effects

As mentioned above, we plan to analyze heterogeneity in effects by gender and whether the respondent's first language is Luganda. In the estimation, we will do so by including an interaction with the treatment variable on the right hand side of the regression.

10.8 Transparent reporting of Covariate Adjustment

We will report all results using the "bare-bones" version of our model, with only treatment, resample and block indicators, and no covariates.

10.9 Joint hypothesis tests for discussion effects

As stated above, we believe that discussion of the content in the vignettes may constitute a vector for the transmission of the treatment from Direct Compliers to other sub-types in the sample. As a feasibility check for this mechanism, we plan to analyze whether the treatment caused respondents to report discussing any of the topic areas to which they were assigned.

To do this, we generate three indicator variables that indicate 0 if the respondent did not report discussing a given topic and 1 if they did. Among Direct Compliers, we then regress each of these three dummies on the

corresponding treatment condition. This gives three estimates of the effect on discussion, of which we take the average. We then permute a randomization distribution of the three estimates under the sharp null of no effect, computing their average each time. By comparing the randomization distribution of the averages to the observed average, we obtain the probability that one would see an average effect as large as we observe if the true data-generating process featured an effect of 0 for all of the topic areas. If the p-value is less than .1, we conclude that the treatment caused Direct Compliers to discuss the topics depicted in the vignettes for at least one of the topic areas.

10.10 Effects across Phase I and Phase II

We will use a simple t-test to test for differences in treatment effects of the domestic violence treatment across phase I and phase II. For the abortion and absenteeism treatment, we plan to integrate results from phase I and phase II using a Bayesian framework.

References

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 484 (484): 1481–95.