Pre-Analysis plan for Follow-up Survey to 'Encouraging Cooperation with the State – A Field Experiment on Household Connections to the Police'

Anna Wilke

6/13/2019

Contents

1 Overview

This document describes the data collection and analysis strategy for a follow-up survey to a field experiment that randomly assigned 100 out of 250 study households located in South Africa's Northwest Province to receive a closer connection to the police in the form of the MeMeZa alarm system. The follow-up survey starts in mid-June 2019. Key features of the experiment (baseline data collection, intervention, random assignment, endline data collection etc.) have been described in the first pre-analysis plan (PAP) and will not be repeated in this document. Section 2 of this document describes the sample that will be interviewed during the follow-up survey.

Section 3 outlines an updated theoretical framework that differs from the one outlined in the previous pre-analysis plan. This theoretical framework has been developed after seeing results from the endline survey that was conducted in November and December 2018 but prior to the realization of any outcomes from the upcoming follow-up survey. Like the previous framework, the theory focuses on how individuals choose to rely on either state or mob justice to deal with crime. The theory makes two core predictions about the effects of a closer connection to the police. First, the framework implies that a closer connection to the police may reduce the willingness to rely on or participate in mob justice by improving perceptions of police service delivery. Second, the framework implies that a closer connection to the police may reduce the willingness to participate in mob justice by increasing perceptions of police oversight. The framework also has additional implications for the effect of perceptions of police service delivery and police oversight on the demand for a closer connection to the police.

The three subsequent parts of this pre-analysis plan outline three sets of outcome-specific hypotheses pertaining to these theoretical expectations and describe ways in which these will be tested. Section 4 pertains to the direct effects of the alarm treatment among the main sample of households. The focus is on the effect of the alarms on final outcomes (willingness to rely on the police and participate or support mob justice, respectively) as well as intermediate outcomes that reflect the two above-mentioned mechanisms (perceptions of police service delivery and police oversight). Section 5 pertains to indirect effects of the alarm treatment on a sample of neighboring households. Measurements from neighbors serve two purposes. First, they provide a second source of information on the willingness of main households to participate in mob justice and rely on the police that is, possibly, less susceptible to social desirability bias. Second, they allow for an analysis of how effects of an improved connection to the police spill over to the surrounding community. Section 6 pertains to hypotheses that involve two information treatments that will be administered during the follow-up survey. The information treatments serve two purposes. First, they have been designed to provide information on the extent to which the alarm treatment affects the willingness to participate in mob justice through the two above-mentioned mechanisms (improvements of perceptions of police service quality and perceptions of increased police oversight). Second, allow for tests of some of the additional implications regarding the demand for a connection to the police.

Taken together, these sections contain a large number of hypotheses. To deal with the resulting problem of multiple comparisons, section 7 summarizes the three main theoretical claims that this study seeks to test as well as the hypotheses that pertain to each theoretical claim. This section also describes how I will assess the validity of each of these three theoretical claims by testing the global null hypothesis that all the constituent null hypotheses pertaining to the respective claim are true.

Sections 8 and 9 deal with item-level missingness and attrition. The final section describes updates regarding

the behavioral t-shirt measure that was introduced in a previous addendum to the first PAP.

Any details that are not described in this pre-analysis plan can be found in the original PAP. Any contingency not accounted for in the previous PAP, the previous addendum or this addendum will be dealt with according to the Standard Operating Procedures for Don Green's lab at Columbia as of June 7, 2016. This study has received approval from the Columbia University Institutional Review Board (IRB), protocol AAAR6346.

2 Sample

The sample interviewed during the previous endline survey consists of two respondents in each of 250 study households. I aim to re-interview the same respondents in the follow-up survey. In addition to re-interviews with the respondents surveyed in the endline, the follow-up survey will comprise interviews with neighbors. Specifically, one neighboring household will be sampled for each main household in the sample. Enumerators will be instructed to always survey the household to the right of the main household. If there is no neighbor to the right (e.g. because the house is at a corner), enumerators are allowed to replace the house to the right with the house to the left. Within each neighboring household, I will interview one randomly selected adult woman and one randomly selected adult man. In all-male and all-female households, I will interview two men and two women, respectively. In single member households, I will only interview one respondent.

3 Theoretical Framework

The main interest is to understand whether and how a closer connection to the police, here in the form of the MeMeZa alarm system, affects the willingness to draw on the police and the state's justice institutions rather than on mob justice to sanction criminals.

Suppose that individuals care about criminals being sanctioned. We can think about this desire as being driven by various concerns such as a taste for vengeance or a desire to deter future crime. Further suppose that individuals can deal with crimes in one of two ways. First, an individual may report the crime to the police. Denote the utility that the individual expects to derive from the outcome of the case if she reports the case to the authorities by $x_S(c)$, where c is the individual's "connection" to the police. $x'_S(c) > 0$, i.e. a closer connection to the police, say in the form of the alarm system, improves the outcome that can be obtained by reporting the case to the authorities. A helpful way to parametarize $x_S(c)$ is to assume $x_S(c) = p(c)\omega_S$, where ω_S is the individual's expected utility from the sanction placed on a criminal by the criminal justice system once the criminal has been apprehended and $p(c)$ is the probability that a criminal is apprehended by the police in the first place. As described in the previous pre-analysis plan, I expect a closer connection to the police in the form of the alarm system to decrease response times and hence to increase the probability that a criminal is apprehended by the police, i.e. $p'(c) > 0$, which implies $x'_S(c) > 0$.

Second, the individual may get together a group of community members and punish the criminal through mob justice. Let's denote her expected utility from doing so by $x_C(c)$, where $x'_C(c) < 0$. We can parametarize this expected utility as $x_C(c) = \omega_C - q(c)\gamma$, where ω_C is the expected utility that the individual derives from the sanction that the community places on the criminal, $q(c)$ is the probability that someone who participates in mob justice is arrested for doing so and γ is the expected dis-utility of the sanction that the justice system

puts on mob justice perpetrators. The idea is that $q'(c) > 0$ which implies $x'_C(c) < 0$. In other words, having a closer connection to the police increases the likelihood of being arrested for participation in mob justice. One reason could be that, if called, the police is more likely to show up quickly in a street where an alarm is present. A second reason may be that alarm owners feel like they have lost their "anonymity"from the perspective of the justice system. After all, alarm owners' names and addresses are on file at the police station. As a consequence, individuals who live in a household with an alarm may feel like they can be more easily identified, found and arrested if they engage in any illegal activity, including mob justice.

Finally, it will be convenient to also consider how the nature of the crime under consideration may affect the relative attractiveness of state and community justice. Let's assume that there are different kinds of crime and denote the kind of crime by e, where $e \sim U[0,1]$. I assume that the utility of engaging in mob justice for a crime of type e is given by $x_C(c) - e$. The idea is that low e crimes lend themselves well to being dealt with by the community, while high e crimes are very costly for the community to deal with. One way to think about e is as the severity of the crime. Communities may be good at dealing with petty crimes. Organized crime, on the other hand, is costly to deal with by communities, since perpetrators tend to be highly armed and capable of retaliation.

For a given crime of type e, an individual will rely on community justice if and only if

$$x_C(c) - x_S(c) = \omega_C - q(c)\gamma - p(c)\omega_S > e \tag{1}$$

4 Direct Effects of Alarm Treatment

4.1 Theoretical Predictions

This formalization suggests that, by increasing c, the alarm treatment may increase the willingness to rely on the police and reduce the willingness to participate in or rely on mob justice in two ways:

- *Improved Service Delivery.* It may increase the utility $x_S(c)$ of justice provided by the state (since $p'(c) > 0$),
- *Increased Police Oversight.* It may decrease the utility $x_C(c)$ of justice provided by the community (since $q'(c) > 0$).

4.2 Outcome-specific Hypotheses

In the following, I will describe the empirical implications of this theory that I will test using the follow-up data. All main hypotheses focus on intent-to-treat effects (ITT) among respondents from main households in the sample. Throughout, I use A and a to refer to indicators for assignment to the alarm treatment among main households. I provide details on the empirical specification for each hypothesis. More information on notation is provided in the subsequent section on estimation.

4.2.1 Main Outcomes

4.2.1.1 Increased Reliance on Police

4.2.1.1.1 Willingness to reach out to police in emergency

Estimand: $E[Y(A=1) - Y(A=0)]$

Hypothesis: The treatment increases the willingness to reach out to the police in an emergency.

Direction: One-tailed (upper)

Outcome: `alert_police`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.1.1.2 Willingness to cooperate with police

Estimand: $E[Y(A=1) - Y(A=0)]$

Hypothesis: The treatment increases the proclivity to report crimes to the police.

Direction: One-tailed (upper)

Outcome: Index of `report_police` and `report_gbv`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.1.2 Decreased Willingness to Support, Participate in and Rely on Mob justice

4.2.1.2.1 Willingness to participate in mob justice

Estimand: $E[Y(A=1) - Y(A=0)]$

Hypothesis: The treatment decreases the willingness to participate in mob justice.

Direction: One-tailed (lower)

Outcome: Additive index constructed of `join_beating` and `join_beating_2`.

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.1.2.2 Willingness to reach out to community in emergency

Estimand: $E[Y(A=1) - Y(A=0)]$

Hypothesis: The treatment decreases the willingness to reach out to the community in an emergency.

Direction: One-tailed (lower)

Outcome: Additive index constructed of `alert_community` and `alert_neighbors`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

Additional remarks: Reaching out to the community is not illegal. This outcome should therefore only be affected through the improved service delivery mechanism but not the increased oversight mechanism.

4.2.1.2.3 Support for mob justice

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment decreases support for mob justice.

Direction: One-tailed (lower)

Outcome: Additive index constructed of `beat_known_thief`, `arrest_mob`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

Additional remarks: Supporting mob justice is not illegal. This outcome should therefore only be affected through the improved service delivery mechanism but not the increased oversight mechanism.

4.2.2 Intermediate Outcomes

4.2.2.1 Improved Perception of Service Delivery by Police

4.2.2.1.1 Contact with police

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment increases the share of subjects who have recently spoken to someone from the police.

Direction: One-tailed (upper)

Outcome: `speak_to_police`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.2.1.2 Perceived response time

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment decreases the perceived response time of the police to an emergency call.

Direction: One-tailed (lower)

Outcome: `response_time`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.2.1.3 Perceived service quality

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment increases the perceived quality of the police service.

Direction: One-tailed (upper)

Outcome: Additive index constructed of `take_problem_seriously`, `lack_of_effort` and `people_go_free_police`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.2.2 Increased Perception of Police Oversight

4.2.2.2.1 Perceived police response to mob justice

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment increases the perceived likelihood that the police would arrive during an ongoing mob justice incident.

Direction: One-tailed (upper)

Outcome: `police_reaction_mob_justice`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.2.2.2 Perceived police interest in sanctioning mob justice perpetrators

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment increases the perceived likelihood that the police will make an effort to sanction mob justice perpetrators.

Direction: One-tailed (upper)

Outcome: `police_intention_mob_justice`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.2.2.3 Perceived likelihood that police discovers illegal behavior

Estimand: $E[Y(A=1) - Y(A=0)]$

Hypothesis: The treatment increases the perceived likelihood that illegal behavior by the respondent would be discovered by the police.

Direction: One-tailed (upper)

Outcome: Additive index constructed of `find_out_stolen_car` and `find_out_illegal_immigrant`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.2.2.4 Perceived anonymity in eyes of police

Estimand: $E[Y(A=1) - Y(A=0)]$

Hypothesis: The treatment increases the share of respondents who feel that they are known to the police.

Direction: One-tailed (upper)

Outcome: Additive index constructed of `police_know_name` and `police_know_house`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.3 Sub-Group Analyses and Heterogeneous Effects

I expect effects to be heterogeneous along two dimensions that correspond to the two main mechanisms described above.

4.2.3.1 Prior Perceptions of Service Delivery

First, similar to the analysis pre-registered in the last pre-analysis plan, I expect the improvement of perceptions of service delivery to be larger among individuals who had more negative views about the police at baseline. Similar to what was pre-registered in the previous PAP, I will therefore create an index `police_evaluation_follow_up` and will separately estimate treatment effects on this index among subjects from households whose baseline scores on `police_evaluation_bl` (see previous PAP) are lower or equal to the median of this index and subjects from households whose baseline scores fall above the median. The index `police_evaluation_follow_up` will contain the following items:

- `response_time`
- `take_problem_seriously`
- `lack_of_effort`
- `people_go_free_police`

Specifically, I will test the following hypotheses:

4.2.3.1.1 Conditional effect among those with low priors of police service quality

Estimand: $E[Y(A=1) - Y(A=0)|M=0]$

Hypothesis: The alarm treatment improves perceptions of the quality of the police service among respondents from households that, at baseline, fell weakly below the median in terms of their perceptions of police quality.

Direction: One-tailed (upper)

Outcome: `police_evaluation_follow_up`

Moderator: `police_evaluation_bl`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{a} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{a} * \boldsymbol{m} + \delta \boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.3.1.2 Conditional effect among those with high priors of police service quality

Estimand: $E[Y(A=1) - Y(A=0)|M=1]$

Hypothesis: The alarm treatment improves or worsens perceptions of the quality of the police service among respondents from households that, at baseline, fell above the median in terms of their perceptions of police quality.

Direction: Two-tailed

Outcome: `police_evaluation_follow_up`

Moderator: `police_evaluation_bl`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{a} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{a} * \boldsymbol{m} + \delta \boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.3.1.3 Difference in conditional effects by priors of police service quality

Estimand: $E[Y(A=1) - Y(A=0)|M=1] - E[Y(A=1) - Y(A=0)|M=0]$

Hypothesis: The effect of the alarm treatment on perceptions of the quality of the police service is smaller among respondents from households that, at baseline, fell above the median in terms of their perceptions of police quality.

Direction: one-tailed (lower)

Outcome: `police_evaluation_follow_up`

Moderator: `police_evaluation_bl`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{a} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{a} * \boldsymbol{m} + \delta \boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

If I am able to reject this null hypothesis of no difference between conditional treatment effects on the 10% significance level, I will conduct the same sub-group analyses for all main hypotheses described in section 4.2.1.

4.2.3.2 Prior Perceptions of Police Oversight

Second, I expect the increase in the perceived amount of police oversight to be largest among those who, at baseline, perceived it unlikely that the police would hear about a mob justice incident in their street and sanction the perpetrators. To assess this idea, I will rely on the item `mob_violence_police_reaction_bl` (see baseline questionnaire and previous PAP) and the outcomes `police_reaction_mob_justice` and `police_intention_mob_justice`. The item `mob_violence_police_reaction_bl` will be treated as a measurement on the household level, assuming that views on the police are correlated within households. I will estimate treatment effects separately among subjects in households that believe it 'Not likely at all' or 'Not very likely' that the police would arrest those involved in mob violence to the conditional treatment effect among subjects in households that believe it 'Somewhat likely' or 'Very likely' that the police would arrest those involved in mob violence.

Specifically, I will test the following hypotheses:

4.2.3.2.1 Conditional effect among those with low priors of police oversight

Estimand: $E[Y(A=1) - Y(A=0)|M=0]$

Hypothesis: The alarm treatment increases perceptions of police oversight among respondents from households that, at baseline, thought it 'Not likely at all' or 'Not very likely' that the police would hear about a mob justice incident and arrest the perpetrators.

Direction: One-tailed (upper)

Outcome: `police_reaction_mob_justice` and `police_intention_mob_justice`

Moderator: `mob_violence_police_reaction_bl`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{a} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{a} * \boldsymbol{m} + \delta \boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.3.2.2 Conditional effect among those with high priors of police oversight

Estimand: $E[Y(A=1) - Y(A=0)|M=1]$

Hypothesis: The alarm treatment increases or decreases perceptions of police oversight among respondents from households that, at baseline, thought it 'Somewhat likely' or 'Very likely' that the police would hear about a mob justice incident and arrest the perpetrators.

Direction: Two-tailed

Outcome: `police_reaction_mob_justice` and `police_intention_mob_justice`

Moderator: `mob_violence_police_reaction_bl`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{a} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{a} * \boldsymbol{m} + \delta \boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.3.2.3 Difference in conditional effects by priors of police oversight

Estimand: $E[Y(A=1) - Y(A=0)|M=1] - E[Y(A=1) - Y(A=0)|M=0]$

Hypothesis: The effect of the alarm treatment on perceptions of police oversight is smaller among respondents from households that, at baseline, thought it 'Somewhat likely' or 'Very likely' that the police would hear about a mob justice incident and arrest the perpetrators.

Direction: one-tailed (lower)

Outcome: `police_reaction_mob_justice` and `police_intention_mob_justice`

Moderator: `mob_violence_police_reaction_bl`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{a} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{a} * \boldsymbol{m} + \delta \boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

If I am able to reject this null hypothesis of no difference between conditional treatment effects on the 10% significance level for one of these two outcomes, I will conduct the same sub-group analyses for all hypotheses described in sections 4.2.2.2 and 4.2.1.

4.2.4 Secondary Outcomes

All hypotheses described so far are closely connected to the theoretical framework. Additionally, I will test two sets of secondary hypotheses that do not follow as directly from the theoretical framework and are more exploratory in nature.

4.2.4.1 Community Relations and Perceptions of Neighbors

The alarm treatment may have effects that go beyond its impact on households who receive an alarm (see also section 5). First, the alarm system may serve as a localized public good: Alarm owners may trigger the alarm on behalf of their neighbors if there is a problem in a neighbor's house. In other words, some of the benefits of owning an alarm in terms of improved service delivery may spill over to neighbors. Moreover, to the extent that neighbors are aware of the increased inclination of alarm owners to rely on the police in emergencies, neighbors may also develop an increased sense of police oversight. These considerations imply that the impact of having a neighbor who was assigned an alarm may be similar to the impact of receiving an alarm. To the extent that alarm owners do still feel slightly better protected than others, the alarm treatment may also create the perception that service delivery by the police is unequal across members of the same community. By reducing the need of alarm owners to rely on help from their community in a one-sided manner, the alarm system may also alter the relationship between alarm owners and their neighbors. To pick up some of these effects, I will test the following hypotheses:

4.2.4.1.1 Increased perception of inequality

Estimand: $E[Y(A=1) - Y(A=0)]$

Hypothesis: The treatment increases the perception of inequality in service provision by the police.

Direction: One-tailed (upper)

Outcome: `perceptions_inequality`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.4.1.2 Increased perception of better protection

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment increases the perception of being better protected by the police than other community members.

Direction: One-tailed (upper)

Outcome: `better_protection`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.4.1.3 Effects on trust in neighbors

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment increases or decreases trust in neighbors.

Direction: Two-tailed

Outcome: `trust_neighbor`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.4.1.4 Effects on perception that neighbors would help with crime

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment increases or decreases the sense that neighbors would help out in a crime situation among both alarm owners and neighbors.

Direction: Two-tailed

Outcome: `neighbors_help`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.4.2 Safety and Victimization

Finally, an important policy-relevant question is whether the alarm system improves the safety of alarm owners. In this regard, I will test the following hypotheses:

4.2.4.2.1 Increased feeling of safety

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment makes subjects feel safer in their homes.

Direction: One-tailed (upper)

Outcome: `feel_safe`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.4.2.2 Reduced risk of victimization

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment reduces the proportion of households in which a crime occurred since last Christmas.

Direction: One-tailed (lower)

Outcome: `crime_victimization`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Main households

Additional remarks: This hypothesis will be tested after collapsing the data to the household level. Households will be coded 1 if any respondent from a given household reports a crime and 0 otherwise.

4.2.4.2.3 Reduced number of crime incidents

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment reduces the number of crime incidents that occurred in a given household since last Christmas.

Direction: One-tailed (lower)

Outcome: `crime_incidents`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Main households

Additional remarks: This hypothesis will be tested after collapsing the data to the household level using means.

4.2.4.2.4 Reduced risk of victimization by violent crime

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment reduces the proportion of households in which a violent crime occurred since last Christmas.

Direction: One-tailed (lower)

Outcome: `violent_crime`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Main households

Additional remarks: This hypothesis will be tested after collapsing the data to the household level. Households will be coded 1 if any respondent from a given household reports a violent crime and 0 otherwise.

4.2.5 Testing for Social Desirability Bias

In order to get a sense of the extent to which treatment-related social desirability bias or experimenter demand effects may drive the results, I will ask respondents about mob justice incidents that happened prior to treatment. Evidence that the treatment affects whether respondents recall any incidents, how many they recall or whether they report having witnessed any of these incidents will be interpreted as evidence in favor of social desirability bias. Specifically, I will test the following hypotheses:

4.2.5.1 Any pre-treatment incidents

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment reduces the proportion of respondents who recall any mob justice incidents that happened in their community in last winter.

Direction: One-tailed (lower)

Outcome: `any_mob_justice_incidents`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.5.2 Number of pre-treatment incidents

Estimand: $E[Y(A = 1) - Y(A = 0)]$

Hypothesis: The treatment reduces the number of mob justice incidents that respondents can recall in their community in last winter.

Direction: One-tailed (lower)

Outcome: `mob_justice_incidents`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.2.5.3 Witnessing pre-treatment incidents

Estimand: $E[Y(A=1) - Y(A=0)]$

Hypothesis: The treatment reduces the share of respondents who can recall any mob justice incidents and report that they have personally witnessed at least on of these incidents in last winter.

Direction: One-tailed (lower)

Outcome: `witness_mob_justice`

Specification: $\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from main households

4.3 Estimation

4.3.1 Main specifications

The unit of observation for almost all analyses will be the respondent. Unless otherwise indicated for specific hypotheses (see previous section), intent-to-treat effects will be estimated using OLS regression and the following specification:

$$\boldsymbol{Y} = \alpha + \tau\boldsymbol{a} + \delta\boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2}$$

where $\boldsymbol{Y}$ is a vector of outcomes; α is an intercept; τ is the intent-to-treat effect (ITT) among the respondents in the subject pool; $\boldsymbol{a}$ is a vector of assignments to the alarm treatment; $\boldsymbol{n}$ is a vector of cluster sizes (number of respondents interviewed in household j) and δ is the associated coefficient; $\boldsymbol{X}$ is a matrix of covariates and $\boldsymbol{\beta}$, again, a vector of associated coefficients; $\boldsymbol{\epsilon}$ is a vector of error terms that allows for clustering at the household level.

This is the same specification as pre-registered in the previous PAP with the exception that I do not condition on block fixed effects. Conditioning on block fixed effects is not required for unbiasedness (treatment assignment probabilities do not vary across blocks) and the large number of small blocks (50 blocks with 5 units each) leads to a substantial increase in the number of parameters to be estimated. It also runs the risk that entire blocks drop out of the analysis, especially in the presence of attrition and when estimating conditional effect.

I condition on cluster size, because differential household sizes induce heterogeneity of cluster sizes in the sample. In all households that have only one member I only interview one respondent. In all other households, two respondents will be interviewed. If potential outcomes are correlated with cluster size, this may bias effect estimates.

The covariates in $\boldsymbol{X}$ will be selected using lasso regression (see details below). For transparency, I will report results with and without the inclusion of the covariates selected through the lasso procedure. The bare-bones version of the specification will still condition on cluster size.

Where the unit of analysis is the household (see hypotheses), the same specification will be used, except for that the error terms will not be adjusted to allow for clustering on the household level.

Unless otherwise indicated, conditional effects and differences in conditional effects will be estimated using the following specification:

$$\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{a} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{a} * \boldsymbol{m} + \delta \boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon, \tag{3}$$

where τ_1 is the ITT among subjects for whom $m = 0$, $\boldsymbol{m}$ is a vector of values of a binary moderator and γ the associated coefficient among respondents who were not assigned to the alarm treatment; τ_2 is the difference between the ITT among subjects for whom $m = 1$ and the ITT among subjects for whom $m = 0$.

4.3.2 Covariate selection

I will use the same lasso procedure described in the previous PAP to select covariates except for that we I omit the block fixed effects from it. The pool of covariates that will be used for the analysis of main households has been described in the previous PAP.

4.3.3 Characterization of uncertainty

In line with what was pre-registered in the previous PAP, randomization inference using the random assignment function described in the previous PAP will be used to calculate p-values on quantities of interest. These p-values will be considered 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.

5 Indirect Effects of Alarm Treatment

The follow-up sample consists of two sets of households: main households who were eligible for direct treatment (receiving an alarm) and neighboring households who were eligible for indirect treatment (having their neighbor receive an alarm). The design thus allows me to analyze not only the direct effect of having been assigned an alarm among those who were eligible to receive an alarm but also the indirect effect of having a neighbor who owns an alarm among the sample of neighbors. Note that these two estimands do not only vary in terms of the intensity of exposure to treatment but also in terms of the sample of individuals that they pertain to. The previous PAP contains more details on how main households have been selected. In this section, I describe hypotheses that pertain to indirect effects of the alarm on neighbors. I use S and s to denote an indicator of assignment to indirect exposure to the alarm among neighbors.

5.1 Effects on Perceptions of Neighbors

One advantage of being able to look at indirect effects on neighbors is that it makes available a separate set of assessments of the willingness of alarm owners to participate in mob justice or rely on the police. One may be worried that direct effects on reports of these behavioral proclivities by alarm owners themselves may be driven by experimenter demand effects. Reports by neighbors about what they believe alarm owners would do may be less susceptible to this problem. Given the close-knit nature of the communities in this project, if those who are protected by an alarm are indeed less willing to participate in mob justice and more willing to collaborate with the police, this change might be perceived by their neighbors. I will therefore test the following hypotheses:

5.1.1 Main Effects

5.1.1.1 Perception that neighbors would rely on the police

Estimand: $E[Y(S=1) - Y(S=0)]$

Hypothesis: Indirect exposure to treatment through a neighbor who owns an alarm increases the perception that one's neighbors would rely on the police.

Direction: One-tailed (upper)

Outcome: Additive index constructed of `alert_police_neighor` and `report_crime_neighbor`

Specification: $\boldsymbol{Y} = \alpha + \tau \boldsymbol{s} + \delta \boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from neighboring households

5.1.1.2 Perception that neighbors would rely on mob justice

Estimand: $E[Y(S=1) - Y(S=0)]$

Hypothesis: Indirect exposure to treatment through a neighbor who owns an alarm decreases the perception that one's neighbors would participate in mob justice.

Direction: One-tailed (lower)

Outcome: Additive index constructed of `neighbor_mob_justice_1`, `neighbor_mob_justice_2`, `neighbor_mob_justice_3`

Specification: $\boldsymbol{Y} = \alpha + \tau \boldsymbol{s} + \delta \boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: Respondents from neighboring households

5.1.2 Sub-group Analyses

I will test the same two hypotheses among neighbors of main households that are most likely to be affected by the two main mechanisms, namely among main households with low priors on the quality of police service and low priors on police oversight. See section 4.2.3 on how these households will be identified.

5.2 Effects on Neighbors' Views and Behaviors

Apart from the fact that neighbors of alarm owners may perceive changes in the views and behavior of alarm owners, it is also possible that neighbors of alarm owners themselves change their views and behavior as a result of indirect exposure to the alarm treatment. On the one hand, those who have an alarm may 'share' the benefits of the alarm with their surroundings by, for example, by triggering the alarm on behalf of their neighbors if a crime occurs in their neighbors' house. In this sense, the alarm system may operate like a localized public good. As a result, the benefits of improved service delivery may spill over to neighbors. Similarly, neighbors may be affected by the increased oversight mechanism. For example, neighbors of alarm owners may refrain from participation in mob justice if they expect alarm owners to alert the police and the police to arrive quickly in response. Finally, to the extent that the alarm system changes community relations (see section 4.2.4.1), these changes may also be sensed by neighbors. I will therefore test the hypotheses specified for the direct effect of the alarm treatment on main households in sections 4.2.1, 4.2.2 and 4.2.4 also in terms of indirect effects of exposure to the alarm treatment among neighboring households.

5.3 Estimation

5.3.1 Main specifications

The specifications that will be used to estimate effects of indirect exposure to treatment among neighbors are identical to those that will be used to estimate effects of direct exposure to treatment among main households (see section 4.3.1), except for that the treatment assignment indicator pertains to assignment to indirect exposure to the alarm treatment. As for effects among main households, standard errors will allow for clustering on the household level.

5.3.2 Covariate selection

For neighbors, covariates will be taken from the follow-up questionnaire, since no baseline or endline data exist for these respondents. The relevant covariates have been labelled as ''Covariate" in the attached questionnaire.[1] See the appendix for a complete list of covariates that will be used for neighbors.

5.3.3 Characterization of uncertainty

I will follow the same strategy outlined for direct effects among main households in section 4.3.3. Randomization inference will proceed in the same way. Note that probabilities of assignment to indirect exposure do not vary across neighboring households, because the sample will contain exactly one neighboring household for every main household in the sample.

[1] Note that the pool of covariates does not include items that are labeled as "Covariate for information treatment". These items are pre-treatment for the information treatments (described below), but not for the main treatment.

6 Information Treatments

6.1 Treatments and Random Assignment

As part of the survey, I will implement two information treatments. Both information treatments consist of extracts from recent news articles that have been published in South African online newspapers (see appendix):

1. *Info Treatment 1 ($Z1$): Police supports harsh sanctions for criminals.* An article that describes the case of two rapists who were sentenced to, in total, 13 life sentences and 240 years in prison. The article underlines the commitment of the police to work hard to ensure that those who commit crimes against women and children receive harsh sentences.
2. *Info Treatment 1 ($Z2$): Police supports harsh sanctions for mob justice.* An article that describes the police's concern with mob justice and and underlines the commitment of the police to sanction those who take the law into their own hand.

The extracts have been translated into the local language. Enumerators will read the news article extracts to the respondent and respondents will be given a copy so that they can follow along. Respondents will not be allowed to keep the copy in order to limit intra-household spillovers in cases where I do not manage to interview all respondents in a household at the same time.

I will implement a full factorial design. The unit of assignment will be the respondent and the assignment procedure will be simple random assignment implemented through surveyCTO, the software used for data collection on tablets. Respondents will be assigned with equal probability to one of four conditions:

- $Z1 = 1$ and $Z2 = 0$
- $Z1 = 0$ and $Z2 = 1$
- $Z1 = 1$ and $Z2 = 1$
- $Z1 = 0$ and $Z2 = 0$

For respondents who receive both information treatments, the order in which the respondent is presented with the two information treatments will be randomized.

6.2 Theoretical Predictions

The idea behind the information treatments is that they manipulate perceptions of, respectively, ω_S and γ, the legal sanction that a criminal receives upon being arrested by the police and the legal sanction for mob justice perpetrators. Perceptions of ω_S and γ should depend on the extent to which individuals believe that the police will make enough investigative effort and gather enough evidence for criminals and mob justice perpetrators to be sentenced effectively. The information treatments are intended to affect these beliefs.

One complication is that mob justice perpetrators are technically a subset of all criminals. The first news article ($Z1$) therefore focuses on rape and, more generally, crimes against women and children. These crimes are quite distinct from mob justice, since mob justice victims are almost always men.

Given this interpretation, the theoretical framework makes several predictions about the effect of the information treatments.

6.2.1 Main Effects on Willingness to Participate in Mob Justice

Both information treatments should reduce the willingness to participate in mob justice, since they, respectively, increase $x_S(c)$ and decrease $x_C(c)$. Denoting the willingness to participate in mob justice by Y, I expect

- $E[Y(Z1 = 1) - Y(Z1 = 0)] < 0$
- $E[Y(Z2 = 1) - Y(Z2 = 0)] < 0$

6.2.2 Interaction with Alarm Treatment

Second, there should be an interaction between the effect of the alarm system (a positive shock to c) on the willingness to participate in mob justice and the two information treatments. To see this, note that the marginal effect of a positive shock to c on $x_S(c)$ is given by $p'(c)\omega_S$ and the marginal effect of a positive shock to c on $x_C(c)$ is given by $-q'(c)\gamma$. Substantively, if people believe that the police have no intention of sanctioning criminals and/or those who participate in mob justice, a closer connection to the state in the form of the alarm may have little effect. If, on the other hand, individuals think that the police is highly motivated to do both, they should change their willingness to participate in mob justice a lot in response to receiving an alarm. In other words, I expect:

- $E\left[Y(\mathrm{A} = 1) - Y(\mathrm{A} = 0) \middle| Z_1 = 1\right] > E\left[Y(\mathrm{A} = 1) - Y(\mathrm{A} = 0) \middle| Z_1 = 0\right]$
- $E\left[Y(\mathrm{A} = 1) - Y(\mathrm{A} = 0) \middle| Z_2 = 1\right] > E\left[Y(\mathrm{A} = 1) - Y(\mathrm{A} = 0) \middle| Z_2 = 0\right]$,

where Y the willingness to participate in mob justice.

Testing these two empirical implications allows me to learn about the extent to which the alarm system does indeed reduce the willingness to participate in mob justice through the two hypothesized mechanisms – better service delivery and increased police oversight.

6.2.3 Effects on Demand For Policing

Finally, apart from the willingness to participate in mob justice, the two information treatments should also have an effect on the extent to which an individual benefits from an increase in c, i.e. on the demand or willingness to pay for a closer connection to the police, be it in the form of an alarm system or other changes that bring the police closer to citizens. To see this, consider an individual who is thinking through how much to pay for a closer connection to the police. In doing so, the individual thinks about a crime that she will be exposed to tomorrow and whether she would report this crime to the police or rely on mob justice to deal with the crime. Given a certain closeness of the police, c, the individual's expected utility from one representative crime is given by the probability that e is low enough for her to rely on the community times the expected utility of community justice, taking into account that e is low, plus the probability that e is too high to rely on the community times the utility of relying on the justice system:

$$Prob\left(x_C(c) - x_S(c) > e\right) * E\left[x_C(c) - e | x_C(c) - x_S(c) > e\right] + Prob\left(x_C(c) - x_S(c) < e\right) * x_S(c)$$

Simplifying, this results in

$$V(c) = \frac{1}{2}\left[x_C(c) - x_S(c)\right]^2 + x_S(c) \tag{4}$$

What matters for the willingness to pay for a closer connection to the state is how this expected value changes as c increases:

$$\begin{aligned} V'(c) &= \left[x_C(c) - x_S(c)\right]\left[x_C^{'}(c) - x_S^{'}(c)\right] + x_S^{'}(c) && (5)\\ &= \left[\omega_C - q(c)\gamma - p(c)\omega_S\right]\left[-q'(c)\gamma - p'(c)\omega_S\right] + p'(c)\omega_S && (6) \end{aligned}$$

As long as this derivative is positive, individuals should be willing to pay for a closer connection to the state and the willingness to pay should be larger the larger this derivative.

How should this derivative change with the information treatments? To get a sense of that, we need to consider the derivative of $V'(c)$ with respect to ω_S and γ:

$$\begin{aligned} \frac{\partial V'(c)}{\partial \omega_S} &= p(c)\left[q'(c)\gamma + p'(c)\omega_S\right] + p'(c)\left[q(c)\gamma + p(c)\omega_S - \omega_C + 1\right] && (7)\\ &= p(c)\left[x'_S(c) - x'_C(c)\right] + p'(c)\left[x_S(c) - x_C(c) + 1\right] \\ \frac{\partial V'(c)}{\partial \gamma} &= q(c)\left[q'(c)\gamma + p'(c)\omega_S\right] + q'(c)\left[q(c)\gamma + p(c)\omega_S - \omega_C\right] \\ &= q(c)\left[x'_S(c) - x'_C(c)\right] + q'(c)\left[x_S(c) - x_C(c)\right] \end{aligned}$$

The first thing to note is that $\frac{\partial V'(c)}{\partial \omega_S} \geq 0$. To see this, note that we have assumed $0 \leq e \leq 1$. As a consequence, for $x_S(c) - x_C(c) < -1$, which is required to make this derivative negative, individuals would always rely on the community and gain utility $\omega_C - q(c)\gamma$, which is independent of ω_S. In other words, $Z1$ should increase the demand for policing, with the exception of corners, where individuals value community justice so much that they would never rely on the police. Intuitively, the marginal value of a connection to the police increases if the police are motivated to ensure that criminals get sanctioned.

The sign of $\frac{\partial V'(c)}{\partial \gamma}$, on the other hand, is ambiguous. It depends, among other things, on ω_C. Intuitively, an increased perception that the police are highly committed to sanction those who engage mob justice may reduce the demand for a closer connection to the police if individuals derive enough utility from community sanctions.

In general, the effect of both treatments should become smaller (in the case of $Z2$ possibly more negative) as the utility of community sanctions, ω_C, increases.

To summarize, denoting the willingness to pay by W for a closer connection to the police, we have

- $E\left[W(Z1 = 1) - W(Z1 = 0) \geq 0\right]$
- $E\left[W(Z2 = 1) - W(Z2 = 0) \lesseqgtr 0\right]$
- $E\left[W(Z1 = 1) - W(Z1 = 0) | \text{high } \omega_C\right] < E\left[W(Z1 = 1) - W(Z1 = 0) | \text{low } \omega_C\right]$

- $E\left[W(Z2=1)-W(Z2=0)|\text{high } \omega_C\right] < E\left[W(Z2=1)-W(Z2=0)|\text{low } \omega_C\right]$

Below, I will consider three proxies for ω_C:a taste for justice to be served immediately, a taste for harsh punishment and a stated preference for mob justice participants to not be punished.

Finally, it is easy to see from the above expressions, that there should be a positive interaction between $Z1$ and $Z2$ in terms of their effect on the demand for policing:

- $E\left[W(Z1=1)-W(Z1=0)|Z2=1\right] > E\left[W(Z1=1)-W(Z1=0)|Z2=0\right]$
- $E\left[W(Z2=1)-W(Z2=0)|Z1=1\right] > E\left[W(Z2=1)-W(Z2=0)|Z1=0\right]$

For example, suppose the effect of $Z2$ on the demand for policing is negative, i.e. knowing that the police is committed to sanctioning mob justice perpetrators reduces the demand for policing. The theory predicts that this negative effect should not be quite as negative among those who also receive the information that the police is highly committed to sanction criminals.

6.3 Outcome-specific Hypotheses

In the following, I will describe the main empirical implications of this theory that I will test using the follow-up data. The risk in terms of non-compliance with assignment to the information treatments seems low. Most hypotheses therefore focus on intent-to-treat effects (ITT) of the information treatments among respondents from both main and neighboring households in the sample. Hypotheses that pertain to the interaction between the alarm treatment and information treatments will be tested among respondents from main households only. Throughout, I use A and a to refer to indicators for assignment to the alarm treatment and $Z1$, $z1$, $Z2$ and $z2$ to refer to indicators for assignment to the two information treatments respectively. I provide details on the empirical specification for each hypothesis. More information on notation is provided in the subsequent section on estimation.

6.3.1 Manipulation Checks

Informally, I will rely on the open-ended responses to the items `think_article`, `learn_article` and `article_comments` to gauge whether respondents understood the articles. More systematically, I will rely on the following hypotheses to understand whether the information treatments induced the desired changes in perceptions of ω_S and γ:

6.3.1.1 Effect of $Z1$ on perceptions of ω_S

Estimand: $E[Y(Z1=1)-Y(Z1=0)]$

Hypothesis: Z1 increases the perception that the police make an effort to ensure that criminals get sanctioned.

Direction: One-tailed (upper)

Specification: $\boldsymbol{Y} = \alpha + \tau \boldsymbol{z_1} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Outcome: `police_punishes_criminals`

Sample: all respondents (main households and neighbors)

6.3.1.2 Effect of $Z2$ on perceptions of γ

Estimand: $E[Y(Z2 = 1) - Y(Z2 = 0)]$

Hypothesis: Z2 increases the perception that the police make an effort to ensure that mob justice participants get sanctioned.

Direction: One-tailed (upper)

Outcome: `police_punishes_mob_justice`

Specification: $\boldsymbol{Y} = \alpha + \tau \boldsymbol{z_2} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: all respondents (main households and neighbors)

If I will not be able to reject the sharp null hypothesis of no treatment effect for any unit for one or both of the two hypotheses above at the 5 percent significance level, I will look at conditional average treatment effects among the sub-groups most likely to be affected by the two information treatments given prior beliefs.

I have two sets of measurements of prior beliefs. First, there are some measures from the baseline survey that capture prior beliefs. These measures have the advantage that they have been obtained prior to both the alarm treatment and the information treatment. That said, these questions were not design with the current theoretical framework in mind and, hence, they are not quite as specific to ω_S and γ as the measurements that are included in the follow-up survey. The latter, however, may be influenced by the alarm treatment. Should the information treatment not have a detectable effect in the aggregate sample, I will therefore conduct two separate sub-group analyses.

The first uses measures of prior beliefs that come from the follow-up survey. Should I be able to reject the sharp null hypothesis for the respective treatment at the 5 percent significance level only in one of those sub-groups and not among the whole sample, I will conclude that the information treatment is only effective in this sub-sample and test all subsequent hypotheses pertaining to this information treatment among this sub-sample only. This statement does not apply to hypotheses which involve interactions with the main alarm treatment (see section 6.3.3), because these measures of prior beliefs are post-treatment from the perspective of the main alarm treatment:

6.3.1.3 Effects of $Z1$ on perceptions of ω_S among those with low prior beliefs (measures from follow-up survey)

Estimand: $E[Y(Z1 = 1) - Y(Z1 = 0)|M = 1]$

Hypothesis: Z1 increases the perception that the police make an effort to ensure that criminals get sanctioned among those who, prior to treatment, did not expect the police to make such an effort.

Direction: One-tailed (upper)

Outcome: `police_punishes_criminals`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{z_1} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{z_1} * \boldsymbol{m} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Moderator: `people_go_free_police`

Sample: all respondents (main households and neighbors)

6.3.1.4 Effects of $Z2$ on perceptions of γ among those with low prior beliefs (measures from follow-up survey)

Estimand: $E[Y(Z1 = 1) - Y(Z1 = 0)|M = 0]$

Hypothesis: Z2 increases the perception that the police make an effort to ensure that criminals get sanctioned among those who, prior to treatment, did not expect the police to make such an effort.

Direction: One-tailed (upper)

Outcome: `police_punishes_mob_justice`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{z_1} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{z_1} * \boldsymbol{m} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Moderator: `police_intention_mob_justice`

Sample: all respondents (main households and neighbors)

The second uses measures of prior beliefs that come from the baseline survey. Should I be able to reject the sharp null hypothesis for the respective treatment at the 5 percent significance level only in one of those sub-groups and not among the whole sample, I will conclude that the information treatment is only effective in this sub-sample and test hypotheses pertaining to interactions between the information treatment and the main treatment in this sub-sample only. This statement does not apply to hypotheses which involve the information treatment only (see section 6.3.3).

6.3.1.5 Effects of $Z1$ on perceptions of ω_S among those with low prior beliefs (measures from baseline survey)

Estimand: $E[Y(Z1 = 1) - Y(Z1 = 0)|M = 1]$

Hypothesis: Z1 increases the perception that the police make an effort to ensure that criminals get sanctioned among those who, prior to treatment, did not expect the police to make such an effort.

Direction: One-tailed (upper)

Outcome: `police_punishes_criminals`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{z_1} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{z_1} * \boldsymbol{m} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Moderator: `people_go_free_bl`

Sample: main households

6.3.1.6 Effects of $Z2$ on perceptions of γ among those with low prior beliefs (measures from baseline survey)

Estimand: $E[Y(Z1 = 1) - Y(Z1 = 0)|M = 0]$

Hypothesis: Z2 increases the perception that the police make an effort to ensure that criminals get sanctioned among those who, prior to treatment, did not expect the police to hear about and sanction mob justice perpetrators.

Direction: One-tailed (upper)

Outcome: `police_punishes_mob_justice`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{z_1} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{z_1} * \boldsymbol{m} + \boldsymbol{X\beta} + \epsilon$

Moderator: `mob_violence_police_reaction_bl` (coded as described in section 4.2.3)

Sample: main households

Should I be unable to reject the null hypothesis of no treatment effect for any unit for one or both of the treatments among the entire sample and among units in the above-described sub-groups, I will investigate the conditional effect of that treatment only among the subjects that have not been assigned to the other treatment. Should I be able to reject the sharp null hypothesis among this subgroup only, then all subsequent hypotheses pertaining to this treatment will only be tested among this sub-group. This applies to all hypotheses, including those that involve the main alarm treatment.

6.3.1.7 Effects of $Z1$ on perceptions of ω_S among those who were not assigned to $Z2$

Estimand: $E[Y(Z1 = 1) - Y(Z1 = 0)|Z2 = 0]$

Hypothesis: Z1 increases the perception that the police make an effort to ensure that criminals get sanctioned among those who, prior to treatment, did not expect the police to make such an effort.

Direction: One-tailed (upper)

Outcome: `police_punishes_criminals`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{z_1} + \tau_2 \boldsymbol{z_2} + \tau_3 \boldsymbol{z_1} * \boldsymbol{z_2} + \boldsymbol{X\beta} + \epsilon$

Sample: all respondents (main households and neighbors)

6.3.1.8 Effects of $Z2$ on perceptions of γ among those who were not assigned to $Z1$

Estimand:$E[Y(Z1 = 1) - Y(Z1 = 0)|Z1 = 0]$

Hypothesis: Z2 increases the perception that the police make an effort to ensure that criminals get sanctioned among those who, prior to treatment, did not expect the police to make such an effort.

Direction: One-tailed (upper)

Outcome: `police_punishes_mob_justice`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{z_1} + \tau_2 \boldsymbol{z_2} + \tau_3 \boldsymbol{z_1} * \boldsymbol{z_2} + \boldsymbol{X\beta} + \epsilon$

Sample: all respondents (main households and neighbors)

6.3.2 Main Effects on Willingness to Participate in Mob Justice

6.3.2.1 Effect of Z1 on willingness to participate in mob justice

Estimand: $E[Y(Z1 = 1) - Y(Z1 = 0)]$

Hypothesis: $Z1$ reduces the willingness to participate in mob justice.

Direction: One-tailed (lower)

Outcome: additive index of `join_beating_3` and `join_beating_4`

Specification: $\boldsymbol{Y} = \alpha + \tau \boldsymbol{z_1} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: all respondents (main households and neighbors)

6.3.2.2 Effect of Z2 on willingness to participate in mob justice

Estimand: $E[Y(Z2 = 1) - Y(Z2 = 0)]$

Hypothesis: $Z2$ reduces the willingness to participate in mob justice.

Direction: One-tailed (lower)

Outcome: additive index of `join_beating_3` and `join_beating_4`

Specification: $\boldsymbol{Y} = \alpha + \tau \boldsymbol{z_2} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: all respondents (main households and neighbors)

6.3.3 Interaction with Alarm Treatment

6.3.3.1 Heterogeneous effects of alarm treatment by $Z1$ on willingness to participate in mob justice

Estimand: $E\left[Y(\mathrm{A} = 1) - Y(\mathrm{A} = 0) | \, Z_1 = 1\right] - E\left[Y(\mathrm{A} = 1) - Y(\mathrm{A} = 0) | \, Z_1 = 0\right]$

Hypothesis: The effect of the alarm treatment on the willingness to participate is smaller (more negative) among those who were assigned to $Z1$.

Direction: One-tailed (lower)

Outcome: additive index of `join_beating_3` and `join_beating_4`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{a} + \tau_2 \boldsymbol{z_1} + \tau_3 * \boldsymbol{z} * \boldsymbol{z_1} + \delta \boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: main households as well as main households and neighbors.

6.3.3.2 Heterogeneous effects of alarm treatment by $Z2$ on willingness to participate in mob justice

Estimand: $E\left[Y(\mathrm{A} = 1) - Y(\mathrm{A} = 0) | \, Z_2 = 1\right] - E\left[Y(\mathrm{A} = 1) - Y(\mathrm{A} = 0) | \, Z_2 = 0\right]$

Hypothesis: The effect of the alarm treatment on the willingness to participate is smaller (more negative) among those who were assigned to $Z2$.

Direction: One-tailed (lower)

Outcome: additive index of `join_beating_3` and `join_beating_4`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{a} + \tau_2 \boldsymbol{z_2} + \tau_3 * \boldsymbol{z} * \boldsymbol{z_2} + \delta \boldsymbol{n} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: main households as well as main households and neighbors.

6.3.4 Effects on Demand for Policing

6.3.4.1 Effect of $Z1$ on demand for policing

Estimand: $E[Y(Z1 = 1) - Y(Z1 = 0)]$

Hypothesis: $Z1$ increases the demand for police.

Direction: One-tailed (upper)

Outcome: additive index of `spend_police_1` and `spend_police_2` and `spend_police_3`

Specification: $\boldsymbol{Y} = \alpha + \tau \boldsymbol{z_1} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: all respondents (main households and neighbors)

6.3.4.2 Effect of $Z2$ on demand for policing

Estimand: $E[Y(Z2 = 1) - Y(Z2 = 0)]$

Hypothesis: $Z2$ increases or decreases the demand for police.

Direction: two-tailed

Outcome: additive index of `spend_police_1` and `spend_police_2` and `spend_police_3`

Specification: $\boldsymbol{Y} = \alpha + \tau \boldsymbol{z_2} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: all respondents (main households and neighbors)

6.3.4.3 Interaction between $Z1$ and $Z2$

Estimand: $E[Y(Z2 = 1) - Y(Z2 = 0)|Z1 = 1] - E[Y(Z2 = 1) - Y(Z2 = 0)|Z1 = 0]$

Hypothesis: The effect of $Z2$ one the demand for police is larger (more positive) among respondents who have also received $Z1$ (and vice versa).

Direction: one-tailed (upper)

Outcome: additive index of `spend_police_1` and `spend_police_2` and `spend_police_3`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{z_1} + \tau_2 \boldsymbol{z_2} + \tau_3 \boldsymbol{z_1} * \boldsymbol{z_2} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Sample: all respondents (main households and neighbors)

6.3.4.4 Heterogeneity in effect of $Z1$ by taste for immediate justice

Estimand: $E[Y(Z1 = 1) - Y(Z1 = 0)|M = 1] - E[Y(Z1 = 1) - Y(Z1 = 0)|M = 0]$

Hypothesis: The effect of $Z1$ one the demand for police is smaller among respondents who have a taste for immediate justice.

Direction: one-tailed (lower)

Outcome: additive index of `spend_police_1` and `spend_police_2` and `spend_police_3`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{z_1} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{z_1} * \boldsymbol{m} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Moderator: `quick_justice`

Sample: all respondents (main households and neighbors)

6.3.4.5 Heterogeneity in effect of $Z2$ by taste for immediate justice

Estimand: $E[Y(Z2 = 1) - Y(Z2 = 0)|M = 1] - E[Y(Z2 = 1) - Y(Z2 = 0)|M = 0]$

Hypothesis: The effect of $Z2$ one the demand for police is smaller among respondents who have a taste for immediate justice.

Direction: one-tailed (lower)

Outcome: additive index of `spend_police_1` and `spend_police_2` and `spend_police_3`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{z_1} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{z_1} * \boldsymbol{m} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Moderator: `quick_justice`

Sample: all respondents (main households and neighbors)

6.3.4.6 Heterogeneity in effect of $Z1$ by taste for punishment

Estimand: $E[Y(Z1 = 1) - Y(Z1 = 0)|M = 1] - E[Y(Z1 = 1) - Y(Z1 = 0)|M = 0]$

Hypothesis: The effect of $Z1$ one the demand for police is smaller among respondents who have a taste for punishment.

Direction: one-tailed (lower)

Outcome: additive index of `spend_police_1` and `spend_police_2` and `spend_police_3`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{z_1} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{z_1} * \boldsymbol{m} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Moderator: `punishment_preferences`

Sample: all respondents (main households and neighbors)

6.3.4.7 Heterogeneity in effect of $Z2$ by taste for punishment

Estimand: $E[Y(Z2 = 1) - Y(Z2 = 0)|M = 1] - E[Y(Z2 = 1) - Y(Z2 = 0)|M = 0]$

Hypothesis: The effect of $Z2$ one the demand for police is smaller among respondents who have a taste for punishment.

Direction: one-tailed (lower)

Outcome: additive index of `spend_police_1` and `spend_police_2` and `spend_police_3`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{z_1} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{z_1} * \boldsymbol{m} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Moderator: `punishment_preferences`

Sample: all respondents (main households and neighbors)

6.3.4.8 Heterogeneity in effect of $Z1$ by taste for imprisonment of mob justice participants

Estimand: $E[Y(Z1 = 1) - Y(Z1 = 0)|M = 0] - E[Y(Z1 = 1) - Y(Z1 = 0)|M = 1]$

Hypothesis: The effect of $Z1$ one the demand for police is smaller among respondents who are opposed to the imprisonment of mob justice perpetrators.

Direction: one-tailed (lower)

Outcome: additive index of `spend_police_1` and `spend_police_2` and `spend_police_3`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{z_1} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{z_1} * \boldsymbol{m} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Moderator: `arrest_mob`

Sample: all respondents (main households and neighbors)

6.3.4.9 Heterogeneity in effect of $Z2$ by taste for imprisonment of mob justice participants

Estimand: $E[Y(Z2 = 1) - Y(Z2 = 0)|M = 0] - E[Y(Z2 = 1) - Y(Z2 = 0)|M = 1]$

Hypothesis: The effect of $Z2$ one the demand for police is smaller among respondents who are opposed to the imprisonment of mob justice perpetrators.

Direction: one-tailed (lower)

Outcome: additive index of `spend_police_1` and `spend_police_2` and `spend_police_3`

Specification: $\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{z_1} + \gamma \boldsymbol{m} + \tau_2 \boldsymbol{z_1} * \boldsymbol{m} + \boldsymbol{X}\boldsymbol{\beta} + \epsilon$

Moderator: `arrest_mob`

Sample: all respondents (main households and neighbors)

6.4 Estimation

6.4.1 Specifications

The unit of analysis will always be the respondent. See individual hypotheses above for the specifications that will be used to test each of them.

All specifications include a matrix $\boldsymbol{X}$ of covariates. The covariates in $\boldsymbol{X}$ will be selected using lasso regression (see details below). With the exception of hypotheses that involve the main alarm treatment (see section 6.3.3), $\boldsymbol{X}$ will always include a de-meaned indicator for whether a respondent's household has been assigned to the alarm treatment as well as interaction terms between the de-meaned indicator and the information treatment indicator(s). For transparency, I will report results with and without the inclusion of the covariates.

With the exception of hypothesis tests that involve the main alarm treatment (see section 6.3.3), standard errors will be adjusted to allow for heteroscedasticity only. For tests of hypotheses that concern interactions with the main alarm treatment described in section 6.3.3, standard errors will be adjusted to allow for clustering at the household level.

6.4.2 Covariate selection

For hypotheses that concern an interaction with the main treatment (section 6.3.3), the pool of covariates and covariate selection procedure will be identical to those described in section 4.3.2 of this pre-analysis plan .

For all other hypotheses which do not involve the alarm treatment, the pool of covariates will consist of survey items that have been asked during the follow-up interview prior to the administration of the information treatments. The pool will include all items that have been flagged as either ''Covariate''or''Covariate for information treatment" in the attached questionnaire. The questionnaire also contains information on how these variables will be coded. A complete list of these covariates is included in the appendix.

6.4.3 Characterization of uncertainty

P-values will be calculated based on randomization inference using the simple random assignment procedure through which information treatments will be assigned. Where hypotheses involve the main alarm treatment, randomization inference will be performed using the assignment function for the alarm treatment described in the previous analysis plan as well as the simple random assignment procedure of the information treatments.

7 Omnibus Tests

In order to deal with problems arising from the large number of comparisons, I group hypotheses mentioned above into three groups of hypotheses, each of which pertains to one of the three main theoretical claims that I seek to test with this round of data collection. For each group, I will rely on non-parametric combination (see Caughey, Dafoe and Seawright, 2017) to test the global sharp null hypothesis that all the component null hypotheses are true. As recommended by Caughey, Dafoe and Seawright (2017), I will rely on permutation inference and use the product function to combine the p-values that pertain to the sub-hypotheses. The ability to reject the respective global sharp null hypothesis will be interpreted as evidence in favor of the respective theoretical claim.

7.1 Theoretical claim 1: The alarm system reduces reliance on mob justice through the ''better service delivery mechanism".

A global p-value will be calculated by combining the p-values from the following sets of hypotheses:

1. All hypotheses described in section 4.2.1 Main Outcomes
2. All hypotheses described in section 5.1.1 Main Effects (Perceptions of Neighbors)
3. All hypotheses described in section 4.2.2.1 Improved Perception of Service Delivery by Police
4. The hypothesis described in section 6.3.3.1 Heterogeneous effects of alarm treatment by $Z1$ on willingness to participate in mob justice

In case that the manipulation checks in section 6.3.1 reveal that the sharp null hypothesis of no treatment effect of $Z1$ on perceptions of the police's intention to sanction criminals can only be rejected among those

with low prior beliefs at baseline, then 4. will be tested only among this subgroup. Should the manipulation check show that this null hypothesis cannot be rejected on the 5 percent significance level among the entire sample or any of the sub-groups, I will also report a global p-value based only on 1., 2. and 3., excluding 4.

Moreover, if the global null hypothesis cannot be rejected on the 5 percent significance level, I will conduct the same test using only the set of respondents for which this mechanism is most likely to apply, namely among those with low prior perceptions of the quality of the police service (see section 4.2.3 for details on how this subgroup is determined).

7.2 Theoretical claim 2: The alarm system reduces the willingness to participate in mob justice through the ''police oversight mechanism".

This theoretical claim will be tested based on the sample of respondents from main households only. A global p-value will be calculated by combining the p-values from the following sets of hypotheses:

1. All hypotheses described in section 4.2.1 Main Outcomes, except for 4.2.1.2.2 and 4.2.1.2.3[2]
2. All hypotheses described in section 5.1.1 Main Effects (Perceptions of Neighbors)
3. All hypotheses described in section 4.2.2.2 Increased Perception of Police Oversight
4. The hypothesis described in section 6.3.3.2 Heterogeneous effects of alarm treatment by $Z2$ on willingness to participate in mob justice

In case that the manipulation checks in section 6.3.1 reveal that the sharp null hypothesis of no treatment effect of $Z2$ on perceptions of the police's intention to sanction criminals can only be rejected among those with low prior beliefs at baseline, then 4. will be tested only among this subgroup. Should the manipulation check show that this null hypothesis cannot be rejected on the 5 percent significance level among the entire sample or any of the sub-groups, I will also report a global p-value based only on 1., 2. and 3., excluding 4.

Finally, if the global p-value indicates that the global null hypothesis cannot be rejected, I will conduct the same test using only the set of respondents for which this mechanism is most likely to apply, namely among those with low prior perceptions of the likelihood that the police would arrest those who participate in mob justice. (see section 4.2.3 for details on how this subgroup is determined).

7.3 Theoretical claim 3: Changes in perceptions of ω_S and γ affect both the willingness to participate in mob justice and the demand for policing in line with the predictions of the theory.

This theoretical claim will be tested based on the sample of respondents from main households and neighbors. A global p-value will be calculated by combining the p-values from the following sets of hypotheses:

1. All hypotheses described in section 6.3.2 Main Effects on Willingness to Participate in Mob Justice
2. All hypotheses described in section 6.3.4 Effects on Demand for Policing

[2]The reason to exclude these two mechanisms is that it is not illegal to support mob justice or reach out to the neighbors for help. As a consequence, I except the oversight mechanism to mainly affect the willingness to participate in mob justice and not the general willingness to rely on neighbors or to express support for mob justice.

Moreover, I will also calculate two separate p-values – one pertaining to hypotheses in the above sets that concern effects of $Z1$ and one pertaining to hypotheses in the above sets that concern effects of $Z2$. In case that the manipulation checks in section 6.3.1 reveal that the sharp null hypothesis of no treatment effect of $Z1$ or $Z2$ on, respectively, the perceptions of the police's intention to sanction criminals or mob justice participants, can be rejected on the 5 percent significance level only among a pre-specified sub-group, the above tests will be conducted among this sub-group.

8 Item-level Missingness

Outcomes will be imputed via chained equations as implemented in the `mice` package. Imputations will take place separately in the categories listed below. These categories respect the fact that certain items are outcomes from the perspective of the main alarm treatment but covariates when it comes to the information treatments. Where a category contains only one variable or where certain missing values cannot be imputed (e.g. because all items in a category are missing), mean imputation is used to eliminate the remaining missing values. To assess the robustness of the results, I will also report results based on listwise deletion.

Neighbor related outcomes

- `trust_neighbors`, `neighbors_help`, `neighbor_mob_justice_1`, `neighbor_mob_justice_2`, `neighbor_mob_justice_3`, `alert_police_neighbor`, `report_crime_neighbor`

Safety and victimization

- `feel_safe`, `victimization`, `crime_incidents`, `violent_crime`

Willingness to rely on police

- `alert_police`, `report_police`, `report_gbv`

Willingness to rely on community

- `alert_community`, `alert_neighbors`, `join_beating`, `join_beating_2`, `beat_known_thief`, `arrest_mob`

Police oversight

- `police_reaction_mob_justice`, `police_intention_mob_justice`, `find_out_stolen_car`, `find_out_illegal_immigrant`, `police_know_name`, `police_know_house`

Police service quality

- `speak_to_police`, `take_problem_seriously`, `response_time`, `lack_of_effort`, `people_go_free_police`, `perceptions_inequality`, `better_protection`

Demand for police

- `spend_police_1`, `spend_police_2`, `spend_police_3`

Willingness to join mob justice (post information treatment)

- `join_beating_3`, `join_beating_4`

Mob justice incidents

- `any_mob_justice_incidents`, `mob_justice_incidents`, `witness_mob_justice`

Manipulation Check 1

- `police_punishes_criminals`

Manipulation Check 2

- `police_punishes_mob_justice`

9 Attrition

I will assess the rate of attrition separately among main households and among neighboring households. Initially, this assessment will be based on the assumption that two respondents should have been sampled in each household, except for households of which I know that there is only a single adult member. Currently, there are 23 such households in the sample of main households. For the latter group of households, I assume that only one respondent should have been interviewed. I will update these assumptions in light of new information about household size that will be acquired during the endline survey.

As a first step, I will test whether there is a relationship between the alarm treatment and whether enumerators report that a household is a single-member household. To assess this, I will create a household-level indicator for whether a household has been reported to be a single-member household and conduct a two-tailed unequal-variances t-test of the hypothesis that treatment does not affect the rate of reported single-member households among main households and among neighbors. I will conduct this test using randomization inference, i.e. I will compare the observed t-statistic to the distribution of t-statistics under random assignment of treatment using the random assignment function described in the previous PAP. If any of these tests produce a p-value smaller than 0.05, I will conclude that information about household size is unreliable and conduct the following two tests under the assumption that two household members should have been sampled in each household. If none of these tests produce a p-value smaller than 0.05, I will proceed based on the assumption that information about household size is accurate.

Next, I will conduct two tests, separately for main households and neighbors, to assess whether attrition is related to the alarm treatment and whether the relationship between baseline covariates and attrition varies across treatment groups. The unit of analysis for both tests will be the respondent and the outcome is an indicator for whether a given respondent attrited.

First, I will conduct a two-tailed unequal-variances t-test of the hypothesis that treatment does not affect the attrition rate among main households and among neighbors. I will conduct this test using randomization inference, i.e. I will compare the observed t-statistic to the distribution of t-statistics under random assignment of treatment using the random assignment function described in the previous PAP.

Second, I will regress an attrition indicator on treatment, a set of baseline covariates, and treatment-covariate interactions. The set covariates that will be used for this test is described in the appendix (see section 11.4). This list contains pre-measurements of all main outcomes and intermediate outcomes as well a measurement of the socio-economic well-being of a household as judged by the baseline enumerators, an item that measures trust in neighbors and an item that measures how safe the baseline respondent felt. Only one respondent per

household was interviewed during the baseline. Baseline measurements will therefore be treated as cluster-level measurements, i.e. all respondents in the same main household will receive the value of the covariate that resulted from the interview with the respondent from that household who was interviewed at baseline. Since neighbors will be interviewed for the first time during the follow-up survey, no baseline measurements exist for them and the test for neighbors will also rely on the measurements from main households. Specifically, all respondents in a neighboring household will receive the value of a given covariate that resulted form the interview with the respondent that was interviewed in the corresponding main household at baseline. The attrition test for neighbors thus focuses on whether neighbors of certain kinds of main households are more likely to attrit in treatment and control. For both main households and neighbors, I will perform an F-test of the hypothesis that all the treatment-by-covariate interaction coefficients are zero. Again, I will rely on randomization inference to conduct this test.

If none of the tests produces a p-value smaller than 0.05, I will report naive estimates among the respondents for whom I have obtained outcome data. Additionally, I will assess the robustness of our results by reporting extreme value bounds.

If one of these tests produces a p-value smaller than 0.05 for main respondents or neighbors, I will rely on a second round of sampling of attrited respondents in that group in combination with an extreme value bounds approach. In this case, I will randomly sample 20 attrited respondents in the respective group, 10 in the control group and 10 in the treatment group.

10 Addendum t-shirt measure

The previous addendum to the first pre-analysis described a household-level t-shirt measure that asks one respondent per household to choose between two t-shirts. At the time of writing, this outcome measure has been collected in all but 12 of the 250 main households. However, there was a slight change in the procedures through which these data have been collected after the first two days of t-shirt distributions. Originally, enumerators did not have instructions pertaining to which person in a given house was supposed to make the t-shirt choice. The idea was that enumerators would ask the person whom they meet first at the household. However, after the first two days of t-shirt distribution (72 households), it seemed that, the share households in which respondent 1 (the respondent who was already interviewed in the baseline) made the t-shirt choice was higher in the treatment group than in the control group. To avoid further imbalances, enumerators were instructed to always ask respondent 1 to make the t-shirt choice in all households.

To date, there are 33 households in which a respondent other than respondent 1 made the t-shirt choice. In these households, respondent 1 will be asked to choose a t-shirt after the follow-up survey interview. Similarly, respondent 1 will be asked to choose a t-shirt after his or her follow-up interview in each of the 12 households for which the t-shirt measure could not yet be collected at all. For households where more than one measurement will be available (those in which a different respondent made the first t-shirt choice), the analysis of the t-shirt choice outcome will be based on the choices made by respondent 1.

Most likely, there will be a small number of remaining households at the end of data collection where respondent 1 never made a choice (e.g. because she could not be found) but where a t-shirt choice was made by a different respondent. To check whether the availability of a choice by respondent 1 remains related to treatment despite the described counter measures, I will regresses an indicator for whether a t-shirt in a given

household was chosen by respondent 1 or by another respondent on a treatment indicator. If I can reject the null hypothesis that whether respondent 1 made the t-shirt choice is affected by treatment on the 5% significance level, I will consider all households for which no choice by respondent 1 is available as attrited for the purposes of the t-shirt measure. If I cannot reject this null hypothesis on the 5% significance level, I will include choices by respondents other than respondent 1 in the analysis. Attrition in the t-shirt measure will be dealt with in the same way as attrition in the survey (see above).

11 Appendix

11.1 Covariates for Analysis of Alarm Treatment among Neighbors

- `female`
- `observed_conditions`
- `floor_material` (indicators)
- `kind_day` (indicators)
- `age`
- `hh_position` (indicators)
- `marital_status` (indicators)
- `education` (indicators)
- `hh_size`
- `n_children`
- `length_stay`
- `traditional_background` (indicators)
- `religious_service`
- `earn_salary`
- `flush_toilet`
- `water_source` (indicators)

11.2 Information Treatments

11.2.1 Info Treatment 1 ($Z1$): Police supports harsh sanctions for criminals.

Rapists sentenced to 13 life sentences and 240 years

Two rapists were combinedly sentenced to 13 life sentences, as well as 240 years imprisonment after a rape and robbery spree in the Brits area in 2016.

Obed Pilusa (31) and Sipho Nampa (31) were found guilty of numerous cases of rape and robbery between January and May 2016 and were sentenced by the Gauteng North High Court.

Pilusa was sentenced to six life sentences for rape and 120 years imprisonment for eight counts of robbery. Nampa was sentenced to seven life sentences for rape and 120 years imprisonment for eight counts of robbery.

The North West Provincial Police Commissioner, Lieutenant General Baile Motswenyane welcomed the hefty sentences.

She congratulated the detectives of the Brits police's Family Violence, Child Protection and Sexual Offences Unit (FCS) for working tirelessly to ensure that the perpetrators were brought to book.

"The sentences will serve as an indication that the police will not hesitate to deal harshly with those who commit crimes against women and children," she said.

Source: https://kormorant.co.za/41975/rapists-sentenced-13-life-240-years/

11.2.2 Info Treatment 1 ($Z2$): Police supports harsh sanctions for mob justice.

Acts of Mob Justice, A Concern to Northwest Police Commissioner

The Provincial Commissioner Lieutenant General Baile Motswenyane is concerned about cases of mob justice that are mushrooming in the province.

According to police spokesperson in the North West, Colonel Sabata Mokwabone, the Provincial Commissioner's concerns stem from several acts of mob justice where even some lives of people who were suspected of having committed crimes were lost.

"Acts of mob justice are condemned in the strongest terms they deserve. On the basis of the Constitution, I therefore make an appeal to communities not to commit acts of mob justice, when you are found, the law will have to deal harshly with you."

There are more than 40 cases of mob justice that have been reported in the province which the police are currently investigating and several suspects have been arrested so far. The Provincial Commissioner has warned that those responsible in perpetuating acts of vigilantism will soon feel the full might of the law.

Source: https://www.northwestnewspapers.co.za/mafikengmail/community/blogs/editor-s-viewpoint/393-acts-of-vigilantism-a-concern-to-nw-police-commissioner

11.3 Covariates for Analysis of Information Treatments

All of the following will be taken from the follow-up questionnaire:

- `female`
- `observed_conditions`
- `floor_material` (indicators)
- `kind_day` (indicators)
- `age`
- `hh_position` (indicators)
- `marital_status` (indicators)
- `education` (indicators)
- `hh_size`
- `n_children`
- `length_stay`
- `traditional_background` (indicators)
- `religious_service`
- `earn_salary`

- flush_toilet
- water_source (indicators)
- trust_neighbor
- neighbors_help
- feel_safe
- victimization
- crime_incidents
- violent_crime
- alert_police
- alert_community
- alert_neighbors
- join_beating
- report_police
- report_gbv
- alert_police_neighor
- neighbor_mob_justice_1
- neighbor_mob_justice_2
- report_crime_neighbor
- punishment_preferences
- quick_justice
- any_mob_justice_incidentss
- mob_justice_incidents
- witness_mob_justice
- neighbor_mob_justice_3
- police_reaction_mob_justice
- police_intention_mob_justice
- beat_known_thief
- join_beating_2
- arrest_mob
- find_out_stolen_car
- find_out_illegal_immigrant
- speak_to_police
- take_problem_seriouslys
- police_know_name
- police_know_house
- response_time
- lack_of_effort
- people_go_free_police
- perceptions_inequality
- better_protection

11.4 Covariates for Attrition Test

- `beat_truck_driver_bl`
- `join_mob_bl`
- `call_police_bl`
- `police_evaluation_bl` (see previous PAP for how this covariate will be constructed)
- `mob_violence_police_reaction_bl`
- `observed_conditions_bl`
- `trust_neighbor_bl`
- `feel_safe_bl`

References

Caughey, Devin, Allan Dafoe and Jason Seawright. 2017. "Nonparametric combination (NPC): A framework for testing elaborate theories." *The Journal of Politics* 79(2):688–701.