

Pre-Analysis Plan for Engendering Policing: Evaluating Reforms to Increase Women's Access to Security and Justice

Gabrielle Kruks-Wisner (UVA), Akshay Mangla (Oxford), and
Sandip Sukhtankar (UVA)

March 5, 2020

1. Introduction

Citizens' ability to access a responsive police force is central to efforts to both prevent and mitigate the effects of crime and violence. And yet it is well known that poor and marginalized populations often lack basic trust in the police (Blair et al. 2019; Ivković 2008). When citizens are hesitant to turn to the police, their access to security and justice is diminished. Moreover, deterrents to future crime are weakened by under-reporting and a resulting culture of impunity. These dynamics are starkly apparent when it comes to women, who are among the least likely to approach the police – as evidenced by globally low levels of reporting of crimes against women (Ellsberg et al. 2015; WHO 2005; Watts & Zimmerman 2002).

These problems are particularly acute in India, which faces rising gender-based violence coupled with under-reporting of these crimes (Thomson Reuters 2018; HRW 2017; CHRI 2015; Simister & Mehta 2010). A recent report from our research site, the state of Madhya Pradesh, finds that in the state's four largest cities, "The overwhelming majority of women who had experienced violence had not reported it to the police" (Neville et al., 2015). Our own baseline surveys confirm the urgency of the situation: 99% of police officers rank crimes against women as a top priority, and yet only 41% of female crime victims reported that they successfully reported their case to the police. Only 18% of police officers surveyed reported having received any training specific to crimes against women or gender sensitization.

The central and state governments in India have pursued a range of high profile initiatives to improve the police's responsiveness to women, including dedicated police vehicles to patrol hot spots for crimes against women, Women's Help Lines, and "Mahila Thanas" (women-only police stations). The effects of these interventions, however, have rarely been rigorously tested. All, moreover, rest on fundamental – but largely unexamined – assumptions related to the role of gender in policing, namely: 1) that women will avail themselves of gender-targeted services; and 2) that women will be more comfortable approaching female officers, who in turn will respond with greater sensitivity to women's requests for assistance.

Scholarship on demographic balance in frontline bureaucracies has grappled with these issues, but has reached ambiguous conclusions. On the one hand, a body of work on local bureaucracies suggests that greater representation of marginalized groups within public agencies can improve outcomes for those very groups (Selden 1997; Meier et al. 1999). For example, studies in education find that having more women and minority teachers improves the performance of girls and minority children (Keiser et al 2002, Meier and O’Toole 2006). Studies of crime data in the U.S. also find that as female representation within the police increases, reporting of crimes against women also increases, and that female officers help prevent future domestic violence (Meier and Nicholson-Crotty 2006, Miller and Segal 2019). On the other hand, another body of work suggests the need for caution in drawing links between bureaucratic representation and performance (Wilkins and Keiser 2006). A recent field experiment on ethnic balance in the Liberian police force, for example, found that minority police officers were more rather than less discriminatory against minority civilians (Blair et al. 2019). Also in Liberia, Karim et al. (2018) find no evidence of increased sensitivity to gender-based violence among female officers. In the United States, studies suggest that gender-differentials in police performance are diminished by the fact that female officers tend to operate within a male policing culture (Collins 2004; Morash & Haarr 2012).

Whether women are more likely to approach the police given specialized gender services, and whether they are more likely to do so when those services are delivered by women, are thus open questions. We study these questions via a randomized control trial (RCT) in India that examines an intervention that aims to increase the accessibility and responsiveness of the police to women and, by extension, increase the rates at which women turn to the police to report crime and seek assistance. The intervention, developed in partnership with the Madhya Pradesh Police (MPP), aims to promote and sustain women-centered policing practices through the establishment of Women’s Help Desks (WHDs¹): dedicated spaces in local police stations staffed by officers trained to assist women in registering cases and pursuing other forms of assistance, coupled with sustained police training and community outreach programs.

2. Experiment design

2.1. Intervention

The Women’s Help Desks represent a treatment bundle that combines four components:

- 1) setting aside private physical spaces (such as a room or cubicle) within police stations for the desk;
- 2) the development of standard operating procedures (SOPs), along with training on these SOPs, to guide officers assigned to the desks as well as others in the police station on how to register cases and assist women visitors;

¹ The desks are referred to locally as “URJA” desks, an acronym for Urgent Relief and Just Action which also means “energy” or “enthusiasm” in Hindi.

- 3) outreach to local women’s networks with the aim of socially embedding the WHDs in their surrounding communities; and
- 4) the allocation of additional female officers to the WHDs.

Our study includes two treatment arms and a control group; the first treatment arm will have the first three components above (but not the allocation of additional female officers), while the second arm includes all four components.

The physical spaces, while simple and cost effective, are designed to ensure basic privacy and protection to women seeking assistance. The SOPs, which were designed in consultation between MPP, legal experts, and civil society organizations with gender expertise, are reinforced through a training program that includes repeated rounds of both district and local, station-level training, as well as regularized contact with a newly established WHD-District Liaison officer charged with supporting the WHDs. The outreach program requires regularized meetings and day-to-day “beat” contact by patrolling officers with local institutions such as day care centers, health centers, and women’s self-help groups, as well as the formation of a WHD community advisory board – members of which act as liaisons between the WHD and the broader community. The allocation of additional female officers pulls from a roster of new trainees, made possible by a MPP personnel change requiring that 30% of all new recruits be women.

2.2. Randomization

For the study, the Madhya Pradesh Police selected twelve districts (out of 51) across the state to be representative of geography, demographics, and socio-economic conditions: Baitul, Balaghat, Bhopal, Gwalior, Indore, Jabalpur, Morena, Panna, Ratlam, Rewa, Seoni, Vidisha. The districts are spread across the state, and are generally diverse and similar to the remaining districts with one important exception: the districts with the four largest cities in the state (Indore, Bhopal, Jabalpur, Gwalior) were purposively chosen, since the police view crimes against women as more prevalent in urban areas.

Our unit of randomization is the police station: treatment stations mandated to start WHDs will be compared to a control group of stations with no WHDs. All treatment stations receive the first three components: physical space for a desk, training and standard operating procedures, and community outreach. Half of the treatment stations are randomly assigned to also receive the fourth component, namely the allocation of female personnel to run the WHD.²

After removing specialized police stations (headquarters, cybercells, aforementioned “Mahila Thanas”, etc.) as well as those the police indicated were “extremely remote” from the list of police stations in our 12 study districts, we have 180 police stations in our sample. Within each district, we stratified police stations by geography (urban and rural) and by the first principal

² Note that while WHDs in the second treatment arm will have designated female officers, there is no prohibition on assigning female officers to WHDs in the first treatment arm.

component of a vector of police station characteristics (number of assigned officers; number of registered cases; population served). We randomized our three study arms (two treatment and one control) within these strata, in the following manner: within each district-geography stratum, we sorted police stations into blocks of three by principal component score, and randomly assigned the police stations within the block so every block had one station in each of the three groups. We assigned the remaining police stations that were not in blocks of three as follows: blocks of a singleton police station were simply randomly assigned into one of the three study arms; blocks of two police stations were sequentially assigned by randomly assigning the first station into one of the three study groups, and the second into one of the remaining two groups.

The result of this process was 61 police stations assigned to the first treatment arm (WHDs with first three components), 59 to the second (WHDs with first three components + female police officers assigned), and 60 to the control arm.

2.3. Timeline

Training for the intervention was officially launched as early as July 2018. However, administrative delays and, importantly, state elections in December 2018 (that resulted in a change in government) meant that training did not fully begin until March 2019. The intervention became fully operational in May 2019. We conducted intervention monitoring surveys in August 2019 and December 2019 to evaluate adherence to the randomization protocols.

3. Data sources

We gather six main sources of data for the study: administrative data on crimes reported; the number of people approaching police stations from CCTV data; a user satisfaction survey of people who visit police stations; a survey of police officers at assigned police stations capturing perceptions and attitudes towards crimes against women; a citizen survey on perceptions of safety, opinions of and contact with the police, and experiences of crimes; and a qualitative study of police stations in each arm of the study.

3.1. Administrative data

The administrative data comprise aggregated records of the number of First Information Reports (FIRs), Non-Cognizable Offence Reports (NCRs), and Domestic Incident Reports (DIRs) filed at study police stations for the period prior to the start of the intervention as well as the period of the intervention. These reports serve as the initial registration of crime cases by the police, and are generally filed by crime victims at police stations. The reports are recorded on official ledgers by the police, and uploaded into the Crime and Criminals Tracking Network System (CCTNS), a police database. We will have access to aggregated data without any personal identifiers, filed by category of the the Indian Penal Code (IPC) that the crime falls under, which allows us to separate those categorized as “crimes against women” (CAW) from other crimes.

3.2. CCTV data

The CCTV data comprise one week of video feeds from the camera focused on the entrance of the police station during the hours of 10 am to 10 pm. Baseline video feeds were collected during January and March 2019, and we plan to collect endline feeds during February and

March 2020. The video data will be analyzed to record the number of men and women entering or exiting the police station, using machine learning and computer vision tools.³

3.3. User satisfaction survey

The user satisfaction survey is a short (5 minute) module asking visitors to police stations about their satisfaction with their visit. The survey will be conducted at endline only (February-March 2020) at study police stations, approaching visitors immediately after their experience with the police. We aim to interview at least 25 people spread over the course of one week at a given police station.

3.4. Police survey

The police survey comprises 1950 police officers posted at study police stations at baseline (September-October 2018). We will attempt to re-interview the baseline sample in February-March 2020, with equivalent rank replacements for those officers transferred out of study police stations.

3.5. Citizen survey

The citizen survey comprises 5648 women and 871 men sampled from the jurisdiction of study police stations at baseline (in two phases, November 2018 and March-April 2019). We will attempt to re-interview the baseline respondents in March-April 2020, in order to obtain an individual-level panel.

3.6. Qualitative study

The qualitative study consists of ethnographic observation and interviews with officers (of various ranks) within 8 police stations, selected to represent each arm of the study in two purposively selected districts. The study includes baseline and endline observational reports on gendered policing practice in the selected stations; weekly interviews and observations within those same stations for the duration of the intervention implementation period; focus group discussions with community organizations associated with the police stations; and interviews with district-level police officials. We describe this in more detail in Section 8.

4. Sampling

4.1. User

We will attempt to conduct user-satisfaction surveys for a period of one week at each police station, interviewing at least 25 visitors, using the following criteria:

1. Any person who voluntarily comes to the police station to seek assistance can be interviewed.
2. Accused or people who have been called in to give their statement regarding a case or have been called for investigation will not be interviewed. A quick question asking whether they have come voluntarily or have been asked to come for investigation will resolve this.
3. Any person who is either in visible physical or emotional distress, in an inebriated state, or otherwise unfit to give answers will not be interviewed.

³ Further details are available in “Leveraging Modern Computer Vision Tools for Social Scientists: A Case Study”, a note produced by our Research Associate Eric Robertson.

4. We will divide the days of the week into morning and afternoon slots, and for each slot surveyors will have a target of surveys to complete that will enable a spread of interviews over the course of the week.
5. Surveyors will continue interviewing people until they reach their target for the slot, ignoring any individuals who leave the police station while an interview is being conducted.

4.2. Police

We obtained a list of officers posted at our study stations from the police administration division, and sampled roughly 1/5th of the 8185 officers posted. We stratified officers by gender and rank (head constable and below were considered low ranks, all others were high rank), and sampled officers in proportion to rank with two important exceptions: the head of the police station - the Station House Officer (SHO) - was included in each station, and we oversampled female officers to ensure at least 2 women were interviewed in every police station. We will reweight our regressions using inverse sampling probabilities in order to ensure representativeness of our surveys at the police station level.

4.3. Citizen

The citizen survey was conducted in two phases: a first phase based on sampling from a citizen database ("Samagra") maintained by the government conducted in November 2018, and a second phase based on sampling of GIS locations conducted in March-April 2019. We used the Samagra database in 4 districts: Balaghat, Panna, Ratlam, and Rewa. In the remaining districts, where we uncovered problems with the Samagra database, we switched to sampling GIS locations. In both phases, the relevant unit of representation is the police station; within a police station jurisdiction, sampling was done using a consistent strategy (drawing from either Samagra or GIS data). An equal number of police stations were sampled using the different strategies across treatment and control groups.

For both strategies, we initially sampled population clusters using Probability Proportional to Size (PPS) sampling. For the Samagra data, we sampled Gram Panchayats (village administrative unit) in rural areas and "colonies" (urban neighborhood unit) in urban areas. For the GIS data, we sampled census villages (after matching them to GIS maps of police jurisdictions) in rural areas, and 100 meter square grids in urban areas (using UN World Population data for population density information).

From the sampled population clusters, we sampled individual names for the Samagra data, and (populated) GPS coordinates for the GIS data.⁴ We

⁴ Survey enumerators were trained to follow a random walk using steps digitized within the survey instrument to the closest surveyable house if the sampled GPS coordinate was of a non-residential nature.

interviewed one adult female from each household thus sampled, and one adult male from every 5th household thus sampled.

5. Primary outcomes of interest

5.1. Administrative data

5.1.1. Number of FIRs filed in CAW cases [FIRs filed in CAW cases controlling for CAW crimes from survey].

Since improving the ease of reporting crimes is one of the goals of the intervention, we will examine the number of FIRs filed at the police station level. An FIR is a report filed by a crime victim of a “cognizable” offense; this allows the police to make arrests without warrants, and both allows and obligates them to investigate the case.⁵ We will have access to FIR data by police station and category of crime (in particular, by the Indian Penal Code number under which the FIR is filed), allowing us to determine which are Crimes Against Women (CAW) cases (as determined by the National Crime Records Bureau, NCRB). If we have access to monthly data, we will report a figure showing monthly trend by treatment and control, in addition to an aggregate of post-intervention data. We will also show a specification that controls for the number of crimes corresponding to the CAW cases as reported in our citizen survey, to account for the possibility that the change in cases filed is due to changes in the actual crime rate rather than the intervention.

5.1.2. Number of FIRs filed by women complainants.

An FIR can be filed by anyone who has direct information about the crime committed. Since the intervention is targeted to women, we will also examine the number of FIRs that are filed by women in any type of case (i.e. not just CAW cases), *conditional on being able to obtain these data*. It is possible that data on who files the FIR will not be available to us, in which case we will not report this outcome.

5.1.3. Number of DIRs filed [DIRs filed controlling for domestic incident crimes from survey].

A DIR is a report of domestic violence as prescribed in The Protection of Women from Domestic Violence Act (2005).⁶ It is similar to an FIR, but prescribes additional duties in such cases for the police as well as other government agencies that support women. We will also show a specification that controls for the number of crimes corresponding to the domestic violence cases (as defined by the Act) reported in our citizen survey, to account for the possibility that the change in cases filed is due to changes in the actual crime rate rather than the intervention.

5.1.4. Action taken on FIRs filed in CAW cases.

⁵ See <https://www.humanrightsinitiative.org/publications/police/fir.pdf> for more details.

⁶ See <https://indiankanoon.org/doc/542601/> for details.

Conditional on obtaining data, we will create an index of action taken by the police on FIRs filed in CAW cases. This will include data such as the number of arrests, the number of cases in which charges were filed, etc. We will update the pre-analysis plan to describe the precise construction of this index once we know what data may be available, prior to data analysis.

5.2. CCTV data

5.2.1. Number of women approaching police stations [Proportion of women approaching police stations (women/total visitors)].

Increasing approachability to the police for women is one of the main outcomes the intervention aims to achieve. We will analyze the CCTV data using computer vision tools to produce a simple number of women who have visited police stations. Our machine learning algorithm is able to distinguish men from women; however, it will separately count individuals as they both enter and leave the police station. While the number of women approaching the police is the primary outcome, we will also check whether the proportion of women visitors has changed. We will control for the number of female officers assigned to the police station to ensure that the additional assigned officers are not driving any changes.⁷

5.3. User satisfaction survey

Improving the experience for women at the police station is the primary goal of the intervention. We will thus examine the following outcomes from the user satisfaction survey

5.3.1. For women: satisfaction with visit

Responses to the question “How satisfied or dissatisfied were you with your visit to the station today?” on a 4 point scale.

5.3.2. For women: comfort during visit

Responses to the question “How comfortable or uncomfortable did you feel discussing your issue with the police?” on a 4 point scale.

5.3.3. For women: did you feel you were treated respectfully

Response to the question “Did you feel you were treated respectfully?” on a 4 point scale.

5.4. Police survey

5.4.1. Perceptions of cases related to women

Changing police perceptions of and sensitivity to cases related to women is one of the main goals of the intervention. Specifically, we will examine

⁷ Since we cannot determine the identity of visitors to the police station, we will not be able to determine simply from the CCTV data whether any changes in the number or proportion of women is due to differences in the number of *individuals* visiting the station versus changes in the *number of times* a given individual visits the police station. However, we can i) use the citizen survey data as an additional check on the number and composition of visitors to the police station and ii) determine whether the treatment had an impact on repeat visits using a question on the user satisfaction survey.

the response to the question “Do you think cases related to women receive too much, too little, or just the right amount of attention by the police, relative to other crime and law and order issues?”

5.4.2. Perceptions of false claims by women

One of the main deterrents to women’s reporting is the fear of not being believed, and this is borne out by preliminary surveys which indicate that a large proportion of police believe that a majority of cases filed by women involve false claims. We will examine to what extent the intervention affects these perceptions through the question “In your thana (station), how common a problem is it for women to report false cases about crimes committed by men?”

5.4.3. Perceptions of police effectiveness in responding to crimes against women

Improving the police’s response to crimes against women is one of the key goals of the intervention. We examine the extent to which the intervention affects police perceptions of effectiveness through the question “How effective do you think the police officers in your thana are in handling cases of crimes against women?”

5.5. Citizen survey

5.5.1. For women who have visited police station since start of intervention: satisfaction with visit

How women are treated at the police station is one of the main outcomes the intervention aims to influence. We will report the response to the question “How satisfied were you with your interaction with the police?” for women respondents who visited police stations since the start of the intervention.⁸

6. Secondary outcomes of interest

6.1. Administrative data

6.1.1. Total number of FIRs filed

In order to check whether the intervention increased the number of FIRs filed overall [as opposed to the proportion of FIRs filed in CAW cases], we will also examine the total number FIRs filed at the station level.

6.1.2. Number of NCRs filed in CAW cases [NCRs filed in CAW cases controlling for CAW crimes from survey]

A Non-Cognizable Offense Report (NCR) is a report filed in cases that are generally less serious. In these cases, the police cannot arrest anyone without prior authorization from courts, and in general are not

⁸ It is possible treatment affects the composition of women that approach the police station, and this affects a number of primary outcomes such as satisfaction with the police in the citizen survey as well as other outcomes in the user satisfaction survey outcomes. To help interpret results, we will test whether treatment impacts composition of women who have visited a police station since the start of the intervention using the citizen survey data.

obligated to investigate. We will examine whether the intervention led to changes in the number of NCRs filed in CAW cases, conditional on getting access to data (it is possible that NCR data is not categorized like the FIR data is). We will also report specifications that control for the number of CAW cases from the citizen survey.

- 6.1.3. Inbound calls to Dial-100 (by women and/or calls dealing with CAW)
Dial-100 is the emergency response system of MP Police, that relies on a professional call center and GIS linked emergency response vehicles staffed by police. In some cases, the intervention drew on the Dial-100 system to carry out community outreach (for example, providing information about the WHDs when a vehicle responds to a call by a woman). It is possible, however, that Dial-100 calls serve as either substitutes or complements to police station visits. We will thus examine whether the number of calls to Dial-100 made by women and/or dealing with crimes against women (depending on data availability) is affected by the intervention. This outcome is conditional on obtaining data. Moreover, it is likely that we only get data at the vehicle level, rather than by police station jurisdiction; we will match vehicles to the police station they are generally attached to.

6.2. CCTV data

- 6.2.1. Total number of people approaching police station
In order to check whether the intervention increased foot traffic overall [as opposed to the proportion of women going to the police station], we will also examine the total number of people approaching the police station.

6.3. User satisfaction survey

- 6.3.1. For men: satisfaction with visit
To check whether having a WHD at the station also affects men (either positively or negatively), we will examine men's satisfaction via the response to the question "How satisfied or dissatisfied were you with your visit to the station today?" on a 4 point scale.
- 6.3.2. Separately for men and women: satisfaction with facilities
We will examine whether satisfaction with facilities increased as a result of the intervention by examining (separately for men and women) the response to the question "How satisfied or dissatisfied were you with the facilities (seating arrangements, toilets, etc) available at this police station?" on a 4 point scale.
- 6.3.3. Separately for men and women: expectation that issue will be resolved
We will examine whether the intervention increased confidence that the concern brought to the police would be resolved by examining (separately for men and women) the response to the question "What expectation do you have that your issue will be resolved?" on a 4 point scale.

6.4. Police survey

- 6.4.1. Perceptions of activities given most importance in thana

One of the aims of the training was to get police to stop thinking of crimes against women as not part of “hard core” policing. In this vein, we examine whether priorities in policing have changed, by measuring whether police bring up “pursuing cases against women” as one of the three main priorities given most importance in the police station.⁹

6.4.2. Perceptions of activities important to officer personally

In this case, we examine the same question as above, except we ask about the priorities given most importance by the officer him/herself.

6.4.3. Perceptions of help provided by police to victims

Attempting to get police to help women more is a goal for police leadership. We examine the extent to which the intervention affects (at least) police perceptions of assistance provided through the question “How helpful do you think the police are in providing support to women who have been victims of violent crime?”

6.4.4. Gender sensitivity in responses to women’s cases

Since gender sensitivity was an important component of training, we examine whether this had an effect on how police report (unprompted) their responses to what they would do when women come into the police station. We add up the number of steps that are consistent with the WHD’s standard operating procedures to be followed when a woman approaches the police station to complain about “eve-teasing” (sexual harassment) or grievous hurt (separate questions) by men, and then regress the total number on treatment indicators.¹⁰

6.5. Citizen survey

Given the short time frame of the intervention (less than a year), it is unlikely that it affects actual crime rates or safety for women; our theory of change would suggest a significant period between an increase in reporting of crimes, more effective tackling of these reports, and an eventual deterrent effect. Moreover, only a small proportion of the population will likely come in contact with the police in the period of the intervention, so it is unlikely to change perceptions of the police or of safety. Nonetheless, since these are eventual goals, and for completeness, we will report the following outcomes from the citizen survey:

6.5.1. Perceptions of how police deal with crimes against women

⁹ The full list is 1. Performing everyday beat; 2. Apprehending suspects; 3. Law and order; 4. Crime investigation; 5. Dial-100 emergency response; 6. Traffic management; 7. Criminal case paperwork; 8. Court duty; 9. Intelligence gathering; 10. Pursuing cases related to women; 11. VIP duty; 12. Victim support.

¹⁰ The full list is 1. Ensure her physical safety; 2. Calm her down and let her tell her story; 3. Record her statement; 4. Refer her for medical care; 5. Refer her for counseling; 6. Talk to her family members; 7. File a FIR; 8. Open an investigation; 9. File a DIR; 10. File an NCR; 11. Refer her to legal aid or other related agencies. Of these, numbers 1-4 are always consistent with the SOPs in the case of “eve-teasing”, while 1-4, 7, 9 are always consistent in the case of grievous hurt. We will simply add up the number of consistent responses that the police respondent supplies unprompted.

This consists of an index of responses to the question “How well do the police in your area deal with the following issues?”, where we list eight issues¹¹ related to crimes against women.

6.5.2. Perceptions of safety for women

This consists of an index of perceptions of safety that women experience in i) their neighborhood, ii) their home, and iii) in large public places.¹²

6.5.3. The rate of crimes against women

The rate of crimes against women (with crimes against women defined above as per NCRB categorizations).

6.5.4. How easy or difficult it is to file FIR

For women who have visited police stations in the past year, the response to the question “How easy or difficult was it to file a FIR?” on a 4 point scale from very easy to very difficult.

7. Statistical methods

7.1. Estimation

We will report Intent to Treat (ITT) estimates comparing average outcomes in treatment and control groups. For all outcomes, we will compare getting any treatment versus control, as well as tests comparing treatment arm 1 versus 2. Our outcome measures are defined either at the individual (citizen or police) or police station level; we will use the most disaggregated measure available. All regressions will include district-geographic stratum fixed effects (level at which treatment probabilities are equal), with a control for the first principal component of “size” variables (pc) used for further stratification. Regressions will be weighted using inverse sampling probabilities to make the outcomes representative at the police station level.

$$(1) \quad Y_{ips} = \alpha + \beta \cdot \text{treatment}_{ps} + \delta_s + pc_{ps} + \epsilon_{ips}$$

$$(2) \quad Y_{ips} = \alpha + \beta_1 \cdot \text{treat1}_{ps} + \beta_2 \cdot \text{treat2}_{ps} + \delta_s + pc_{ps} + \epsilon_{ips}$$

where i is the individual, p is the police station, and s is the stratum.

In addition, where available (for most data sources, other than the user satisfaction survey), we will include the baseline value of the outcome as a control. For outcomes from the police survey, where we expect significant attrition from the baseline sample due to routine transfers, and where we will sample replacement officers at endline, we will use the baseline mean at the police station level as the baseline outcome value for the replacement officers, along with an

¹¹ 1 Eve-Teasing, verbal harassment; 2 Domestic violence, cruelty; 3 Rape, sexual assault, molestation; 4 Dowry; 5 Child marriage; 6 Sex discrimination; 7 Human trafficking; 8 Forced prostitution. The responses can be “good, fair, and poor”, which we will score as 1, 0, and -1; the index will be a simple average of the scores across the 8 categories.

¹² Responses range from “very secure, secure, insecure, very insecure”; we will assign scores from 1-4, with 4=very secure, and the index will be a simple average of scores across the 3 questions.

indicator for replacement status. We will consider the specification including baseline outcomes as our primary specification.¹³

$$(3) \quad Y_{ips} = \alpha + \beta \cdot \text{treatment}_{ps} + \gamma \cdot Y0_{ips} + \delta_s + \rho_{ps} + \epsilon_{ips}$$

$$(4) \quad Y_{ips} = \alpha + \beta_1 \cdot \text{treat1}_{ps} + \beta_2 \cdot \text{treat2}_{ps} + \gamma \cdot Y0_{ips} + \delta_s + \rho_{ps} + \epsilon_{ips}$$

where $Y0$ is the baseline outcome.

7.2. Inference

We will cluster standard errors at the police station level, which is the level of randomization. We will not adjust standard errors for the possibility of mistaken inference due to multiple hypothesis testing; instead, we have pre-specified primary outcomes to report, and wherever there is a family of outcomes we will combine into an index as described in Sections 5 and 6 above.

7.3. Balance

We will test for balance at baseline using specifications 1 and 2 on outcomes used to stratify randomization, as well as our primary outcomes that are collected at baseline.

7.4. Attrition

Attrition is relevant for the citizen survey and the police survey. For the citizen survey, we will report attrition rates for the three treatment groups, and check for differential attrition across these groups. To do so, we will regress an indicator for individuals surveyed at baseline but not at endline on indicators for treatment groups as well as interactions of these indicators with primary outcomes at baseline. If an F-test of these indicators and interactions reveals significant differential attrition, we will present Lee bounds on our estimates.

For the police survey, we expect significant attrition from the baseline sample because of routine transfers. We have thus sampled replacements for the police officers who we will survey at endline, and will consider the officers sampled at endline as our main sample. In addition, treatment arm 2 involves the assignment of a female officer to be a help desk officer; some of these officers may be new and additional to the station. We will thus check for differential attrition following the procedure above only across treatment arm 1 and the control group, and report Lee bounds if we find significant evidence of differential attrition in this case.

7.5. Heterogeneity

We will test for heterogeneity along the following five pre-specified dimensions, for the specific group of outcomes listed below under each particular dimension.

¹³ For primary outcome 5.5.1 (satisfaction with police), it is unlikely that we have corresponding baseline data, since very few women will have visited the police station both in the baseline and endline reference periods. We will thus consider the specification without baseline data to be the main specification for this outcome.

- 7.5.1. Implementation quality. Using data from two rounds of implementation monitoring, we will generate an index of implementation quality.¹⁴ We will examine heterogeneity in our primary outcomes by interacting treatment with this index.
- 7.5.2. Police training. Using data from implementation monitoring and our endline police survey, we will generate an index of strength of training exercises.¹⁵ We will examine heterogeneity in primary and secondary outcomes in our police survey and user satisfaction survey by interacting treatment with this index.
- 7.5.3. Community outreach. Using data from implementation monitoring and our endline surveys, we will generate an index of strength of community outreach.¹⁶ We will examine heterogeneity in primary and secondary outcomes in our citizen surveys and CCTV data by interacting treatment with this index.
- 7.5.4. Gender. We will examine whether there are any differences by gender in the treatment effect on the police survey primary outcomes.
- 7.5.5. Geography. We will examine whether there are any differences in impact on our primary outcomes by police station location - rural or urban. We will interact an indicator for rural location with the treatment indicators, dropping the district-geography indicator and including a simple district fixed effect instead.

8. Qualitative analysis

8.1. Introduction

¹⁴ The index will comprise of a simple summation of the number of “yes” responses to the following questions in the final round of intervention monitoring in treatment stations: i) Is there a help desk for women in the thana? ii) Is the URJA program active in the thana? iii) Is there a separate room allotted for the Help Desk? iv) Does the URJA Desk have a dedicated assigned officer? v) Has there been any URJA Desk related training at the thana level for all the thana staff? vi) Has the Help Desk officer attended any training on operating the URJA desk at PTRI Bhopal, MPPA Bhauri or the district control room? vii) Does the general register contain entries? viii) Is the plan for doing outreach activities filled out by the Help Desk officer? ix) Does the thana have a list of members who belong to the Shakti Samiti?

¹⁵ The index will comprise of a score between 0-5, based on responses across the final round of intervention monitoring and the police endline survey. This includes two components: a 0-1 point range for the proportion of officers who have attended training as reported separately in the monitoring exercise and a summary of responses from the police surveys; and 1 point each for a “yes” response to the following three questions: i) Has there been any URJA Desk related training at the thana level for all the thana staff? ii) Has the Help Desk officer attended any training on operating the URJA desk at PTRI Bhopal, MPPA Bhauri or the district control room? iii) Has the SHO officer any training on operating the URJA desk?

¹⁶ The index will comprise of a score between 0-4, based on responses in the final round of intervention monitoring. This includes two components: a 0-1 point range for the number of community outreach events held divided by the maximum number of community outreach events held; and 1 point each for a “yes” response to the following three questions: (a) Is the plan for doing outreach activities filled out by the Help Desk officer? (b) Does the thana have a list of members who belong to the Shakti Samiti? (c) Do you conduct meetings with Shakti Samiti members?

In order to understand the *mechanisms* by which the WHD treatment bundle may influence police behaviour, we also incorporate a qualitative research component as part of our data collection efforts. Below we describe: (1) the rationale behind our qualitative research, (2) the research design and case selection strategy, (3) the qualitative field methods to used, and lastly, (4) how we intend to integrate qualitative data analysis with our quantitative analysis.

8.2. Rationale for Qualitative Research

Qualitative research is needed to pry open the “black box” of the police station and unpack the organizational factors that enable (and impede) the handling of cases involving women. The qualitative study of police stations allows us to decipher how police officers on the frontlines understand their duties and the actions that are required of them, as well as how they interpret and implement the WHD intervention. Through the collection of qualitative data, we aim to identify *how* the WHD treatment bundle operates inside the police station (our unit of analysis) to produce changes (or not) in police behaviors and associated outcomes of interest.

Our approach follows in the tradition of qualitative and mixed-methods research on police organizational culture and behavior (for example, Van Maanen 1975; Wilson 1978; Bradford et al 2018; Jauregui 2016). These studies reveal important differences between the formal institutions of law enforcement and the informal norms, practices and interpretation of the law by police officers on the frontlines. Our qualitative study of the WHD intervention also aims to contribute to a growing set of studies that combine field experiments with detailed qualitative research (Bamberger, et. al. 2010; Rao et al. 2017). Qualitative field research aids in the discovery of new hypotheses and unanticipated mechanisms that cannot be prespecified. It may also help to cast light on other, unintended or additional effects of the treatments that are not directly captured by other data sources, or that are not precisely quantifiable.

8.3. Research Design and Case Selection

8.3.1. Multi-level Comparative Design

The design of our qualitative research mimics the RCT design by engaging in in-depth study in police stations within each arm of the study, while also engaging in multi-level comparisons across districts. of the WHD intervention. This multi-level comparative framework acknowledges the realities of governance, which involves coordination between administrative tiers and subdivisions. In preliminary rounds of open-ended qualitative fieldwork, the significance of the police hierarchy was a recurring theme. In particular, we identified the crucial role of the district police chief, known as the Superintendent of Police (SP), and his/her team of senior officers, along with the Station House Officer (SHO) who oversees the police state, and finally, the interface between frontline officers and citizen groups. While the help desks operate inside the police stations, activating them fully requires initiative and coordination by these multiple stakeholders. Our qualitative research thus embeds the police station within a more complex web of governance relations that may influence (from above and below) the everyday functioning of the help desks.

8.3.2. District selection

Our multi-level comparative design starts at the district level. Of the 12 RCT study districts, we selected two districts for qualitative research. The two districts were selected to match on key variables associated with governance and women's security. These included the local administrative structure and capacity of the police, rates of literacy for the overall and female population, the sex ratio, and levels of economic development and urbanization. Adopting otherwise similar districts, qualitative research aims to study (potential variation in) the leadership, managerial processes and other informal practices that may influence police behaviour at the station-level and the operation of help desks.

8.3.3. Police station selection: Matched Triads and Rural Stations

Next, we selected eight police stations in total for intensive study. Station selection mimicked the design of the RCT. First, we selected two matched-triads—three stations from each district—to conduct structured comparisons. The three stations in a triad were chosen from each of the three RCT study groups: (1) treatment arm 1, (2) treatment arm 2 and (3) control. Beyond this, we had two further criteria for selecting stations in these triads. First, we chose stations that were “typical” for each district, i.e. those with average levels of resources and having no special amenities or functions. Second, the police stations were chosen to match on observable station-level characteristics (e.g. physical resources, station infrastructure and quantity of personnel) as well as environmental features of the station neighbourhoods (e.g. level of urbanization).

Our station-level case selection generated two matched triads of police stations that were observably similar and representative of typical stations in the two study districts. In addition, we selected two more treatment police stations (in arm 2) that are located in predominantly rural sections of the study districts. These two stations were also matched on observable characteristics. The inclusion of rural police stations for qualitative research was critical to account for the contextual differences in policing and women's security between rural and urban geographies.

8.4. Qualitative Field Methods

In each of the treatment stations, we performed qualitative process tracing of the WHD intervention, allowing us to study implementation closely over time. Parallel research in control stations enabled us to compare how women's cases are handled in the absence of a help desk.

To investigate how the WHD intervention may drive officer behavior and shape the responsiveness of police stations to cases involving women, we use a variety of qualitative field methods. These include officer interviews, ethnographic observations inside police stations, focus group discussions with community groups and observations of the police-citizen interface. For each method, we developed semi-structured instruments that had consistent guidelines to standardize the scope of research across field sites, but also left room for the discovery of

unanticipated themes. We also sought to triangulate across the different methods and data sources to ensure the robustness of qualitative findings.

8.4.1. Qualitative baseline and endline assessments

In all eight of study stations (treatment and control), we first conducted a baseline qualitative assessment. The baseline assessment sought to capture the pre-existing organizational norms and practices relating to cases involving women. Methods for the baseline include: (1) intensive ethnographic observations within the station and (2) interviews with a wide set of officers. Through ethnographic observation, we sought to unpack the organization within the police station, including the operation of hierarchy and task delegation, what activities occupy officers' time, what types of cases gain priority and what external forces influence police activities. The officer interviews focused on the priorities and perceptions regarding cases involving women, as well as handling those cases. We will carry out the same qualitative assessment exercises at the endline of the study.

8.4.2. Ethnographic observations

In subsequent rounds of fieldwork, bi-weekly visits were made to each study station to perform interviews and ethnographic observations. Ethnographic observations focused on both the routine operations in each station as well as the functioning of the help desks, systematically capturing the interaction between the help desks and the police station. Our emphasis was to understand how, incrementally, the WHD intervention shapes organizational behavior and practices in the station, and, whether and how changes varied between the different study stations. The inductive approach to qualitative research provides opportunities to discover unexpected but important aspects of policing, such as "counseling," an informal system adopted in some police stations for hearing complaints from women and resolving family disputes (which may interact with the WHD intervention). And the long duration of our fieldwork allows us to observe how organizational culture and leadership (e.g. of the SHO) interacts over time with socio-structural factors (e.g. resource constraints and gender norms) to influence the WHD intervention and the handling of cases involving women.

8.4.3. Interviews

For the interviews, we purposively selected a range of officers in each station, based on rank, gender and level of experience. The officer interviews followed semi-structured questionnaires that included modules on officer background and motivation, perspectives on crimes against women, the prioritization of women's cases and officer perceptions of the wider community. We also had dedicated interview modules for the station's SHO and other senior staff, female officers and the dedicated help desk officer. We paid special attention to the perspectives and experiences of female officers in interviews to help decipher the potential impact of having women stationed at the help desk, thereby comparing the mechanisms across the two treatment arms. We interviewed officers stationed at the help desk, including female officers, who we

theorize to be potential agents of change. We will aim to assess how female officers are utilized within the thana, and how they understand and take up help desk responsibilities. We will also interview officers (male and female) who are not stationed at the help desk to assess whether and how their priorities and behavior may have changed.

8.4.4. Police-community Interface

Community outreach is a key component of the WHD intervention and a plausible mechanism for pulling in women's cases to the help desk. We developed three approaches to qualitatively assess outreach and the broader police-community interface in our study stations. First, interview questionnaires contained modules pertaining to police perceptions of the community and the priority attached to outreach activities. Second, our ethnographic observation guide included a module on police-community interactions. These included observations of routine interactions between citizens and officers in the police station campus and out on the beat, dedicated outreach activities conducted by police officers (e.g. at schools and community centers), and finally, meetings held by citizen committees created by the police. Third, we will develop a questionnaire for conducting focus group discussions of community groups created under the WHD intervention (in control stations, these groups include established community policing citizen committees). The focus group discussions will allow for a guided conversation among citizens on key themes, including perceptions of women's safety, accessing the police, responsiveness of the police, and what role the desks may (or may not) have in shifting citizen perceptions of the police.

8.4.5. Senior Officers

In order to embed the study police stations within a multi-level analysis, we also engaged in qualitative research at the district level, where the office of the SP is located. The SP office oversees all aspects of policing in the district, helps set priorities for police stations and evaluates station performance. The SP office also handles the monitoring of the WHD intervention and thus plays a key role in the oversight of cases involving women. Our qualitative research involves interviews with the SP and other senior officers and ethnographic observation of the SP's office. Interviews follow a semi-structured protocol, with modules on the priorities of the SP, the SP's approach to leadership and frontline officer motivation, perceptions of and involvement with the WHD intervention, monitoring of cases involving women, and the SP-community interface and outreach. The ethnographic observation of the SP office will cover these themes but also examine communication between the SP's office and police stations in the study district.

References

- Bamberger, M., Rao, V., & M. Woolcock. (2010). "Using mixed methods in monitoring and evaluation: experiences from international development." *World Bank Policy Research Working Paper No. 5245*.
- Blair, R. Karim, S., Gilligan, M., Beardsley, K. (2019). Policing Ethnicity: Lab-in-the-Field Evidence on Discrimination, Cooperation and Ethnic Balancing in the Liberian National Police (Jan 9, 2019). [Available at SSRN: <https://ssrn.com/abstract=2772634>].
- Bradford, B., Loader, I., Jauregui, B., & J. Steinberg, eds. (2016). *The Sage handbook of global policing*. London: Sage.
- CHRI (2015). "Crime Victimization and Safety Perception: A Public Survey of Delhi and Mumbai." New Delhi: Commonwealth Human Rights Initiative. [Available at: http://www.humanrightsinitiative.org/download/1461662128Crime%20Victimization_Soft%20File_distribution.pdf].
- Collins, S. (2004). Sexual Harassment and Police Discipline: Who's Policing the Police? *Policing: An International Journal of Police Strategies & Management*, 27(4), 512-538.
- Ellsberg, M., Arango, D. J., Morton, M., Gennari, F., Kiplesund, S., Contreras, M., & Watts, C. (2015). Prevention of Violence Against Women and Girls: What Does the Evidence Say? *The Lancet*, 385(9977), 1555-1566.
- HRW. (2017) "'Everyone Blames Me': Barriers to Justice and Support Systems for Sexual Assault Survivors in India." Washington, DC: Human Rights Watch. [Available at: https://www.hrw.org/sites/default/files/report_pdf/india1117_web.pdf].
- Ivković, S.K. (2008). A Comparative Study of Public Support for the Police. *International Criminal Justice Review*, 18(4), 406-434.
- Jauregui, B. (2016). *Provisional authority: Police, order, and security in India*. Chicago, IL: University of Chicago Press.
- Karim, S., Gilligan, M. J., Blair, R., & Beardsley, K. (2018). International gender balancing reforms in postconflict countries: Lab-in-the-field evidence from the Liberian national police. *International Studies Quarterly*, 62(3), 618-631.
- Keiser, L. R., Wilkins, V. M., Meier, K. J., & Holland, C. A. (2002). Lipstick and Logarithms: Gender, Institutional Context, and Representative Bureaucracy. *The American Political Science Review*, 96(3), 553-564.
- Meier, K. J., & Nicholson-Crotty, J. (2006). Gender, representative bureaucracy, and law enforcement: The case of sexual assault. *Public Administration Review*, 66(6), 850-860.

- Meier, K. J., & O'Toole, L. J. (2006). Political Control versus Bureaucratic Values: Reframing the Debate. *Public Administration Review*, 66(2), 177-192.
- Meier, K. J., Wrinkle, R. D., & Polinard, J. L. (1999). Representative Bureaucracy and Distributional Equity: Addressing the Hard Question. *The Journal of Politics*, 61(4), 1025-1039.
- Miller, A. R., & Segal, C. (2019). Do female officers improve law enforcement quality? Effects on crime reporting and domestic violence. *The Review of Economic Studies*, 86(5), 2220-2247.
- Morash, M., & Haarr, R. N. (2012). Doing, redoing, and undoing gender: Variation in gender identities of women working as police officers. *Feminist Criminology*, 7(1), 3-23.
- Neville, S; Hilker, M; Humphreys, M; Husain, S; Khan, S; Lindsey, S. (2015). "Evaluation of the Madhya Pradesh Safe Cities Initiative." Baseline Report. London: DFID.
- Rao, V., Ananthpur, K., & Malik, K. (2017). The anatomy of failure: an ethnography of a randomized trial to deepen democracy in rural India. *World Development*, 99, 481-497.
- Selden, S. C. (2015). *The Promise of Representative Bureaucracy: Diversity and Responsiveness in a Government Agency*. Armonk, NY: ME Sharpe.
- Simister, J., & Mehta, P. S. (2010). Gender-based violence in India: Long-term trends. *Journal of Interpersonal violence*, 25(9), 1594-1611.
- Thomas Reuters. (2018) "The World's Most Dangerous Countries for Women – 2018." Thomson Reuters Foundation Annual Poll. <<https://poll2018.trust.org>>.
- Van Maanen, J. (1975). "Police socialization: A longitudinal examination of job attitudes in an urban police department." *Administrative science quarterly*, 20 (2):207-228.
- Watts, C., & Zimmerman, C. (2002). Violence Against Women: Global Scope and Magnitude. *The Lancet*, 359(9313), 1232-1237.
- WHO. (2005). "WHO Multi-Country Study on Women's Health and Domestic Violence Against Women: Summary Report of Initial Results on Prevalence, Health Outcomes and Women's Responses." Geneva: World Health Organization. [Available at: http://apps.who.int/iris/bitstream/10665/43310/1/9241593512_eng.pdf].
- Wilkins, V. M., & Keiser, L. R. (2006). Linking Passive and Active Representation by Gender: The Case of Child Support Agencies. *Journal of Public Administration Research and Theory*, 16(1), 87-102.
- Wilson, J. Q. (1978). *Varieties of Police Behavior: The Management of Law and Order in Eight Communities*. Cambridge, MA: Harvard University Press.