

RESEARCH ARTICLE

GENDER-BASED VIOLENCE

Policing in patriarchy: An experimental evaluation of reforms to improve police responsiveness to women in India

Sandip Sukhtankar^{1,2,3*}, Gabriele Kruks-Wisner⁴, Akshay Mangla⁵

Gender-targeted police reforms are frequently proposed to tackle the global problem of rising yet under-reported gender-based violence (GBV)—but with mixed and often disappointing results. We explore this issue in India, a country with alarming rates of GBV and limited police capacity, by studying the impact of Women's Help Desks (WHDs): dedicated spaces for women in local police stations, staffed by trained officers. Drawing on the largest randomized controlled trial of a police reform to date (180 police stations serving 23.4 million people), we find that officers in stations with WHDs are more likely to register cases of GBV, particularly where female officers run the desks. This suggests that even in resource-constrained and patriarchal environments, police responsiveness can be improved by focusing and mainstreaming attention to women's cases and by greater gender representation within the police.

Countries worldwide grapple with the dual problems of rising gender-based violence (GBV) and weak law enforcement. The ability to register crimes with the police is one essential step in accessing the justice system and promoting women's security. Yet social barriers, including stigma, can deter women from reporting crimes, as can low trust in the police (1, 2). The police, moreover, are often unresponsive to women's concerns, leading to both under-reporting and under-recording of crime and violence (3). The result is a pronounced gap between the incidence of crime and the rates at which such crime is formally recognized.

Gender-targeted police reforms, which seek to increase the accessibility and accountability of the police to women, are regularly proposed to tackle these problems. One set of reforms have worked to increase women's representation within police forces. These initiatives build on theories of representative bureaucracy, which hold that the presence of members of marginalized groups within a public agency improves performance with respect to those groups (4, 5). Evidence on the efficacy of these reforms, however, remains mixed. Some studies find that reporting and arrests for sexual assault increase with the presence of female officers (6, 7). Other research suggests that gender differentials in police behavior are diminished as female

officers operate within male-dominated policing cultures (8, 9), noting a lack of increased gender sensitivity among female officers (10, 11) who, like their male counterparts, often blame victims or dismiss their claims (12, 13).

A second set of reforms involve gender-segregated services such as all-women police stations, established in countries as diverse as Brazil, Ghana, India, Kosovo, and the Philippines. These women-only stations, like other gender-segregated spaces such as women's train carriages or girls' schools (14, 15), operate on the theory that women may be more comfortable in the absence of men. The assumption is that female victims of crime will be more likely to report cases in women-only police stations, and that female officers in those stations will be less constrained by patriarchal policing cultures (16). However, recent research suggests that separating women's cases from other police work makes it less likely that officers in regular, mixed-gender police stations will file cases related to GBV, creating barriers to access that further marginalize women (11, 13, 17).

In this article, we examine a large-scale gender-targeted police intervention that seeks to mainstream rather than segregate attention to women's cases within routine police work, while also testing whether increased representation of female officers better meets women's needs. Our setting is India, a country plagued by some of the world's highest rates of violence against women. Home to entrenched patriarchal norms and chronically weak state capacity (18), India is both a critical case and a representative one, given the prevalence of such conditions in settings throughout the Global South. Theoretical predictions about whether an intervention of this kind ought

to have an impact are ambiguous, given the mixed evidence of existing research, in addition to the difficulty of implementing such reforms. Moreover, efforts to mainstream rather than segregate attention to women's cases have not been rigorously studied.

To fill these gaps, we employ what is to our knowledge the largest randomized controlled trial of a police reform to date, implemented across 180 stations serving a population of 23.4 million. The trial evaluates the impact of introducing Women's Help Desks (WHDs): dedicated spaces mandated to respond to women's cases, located within regular (mixed gender) police stations, staffed by officers who are trained in assisting women, and supported through routine monitoring, coupled with community outreach. We find that officers in stations with WHDs are more likely to register cases of GBV and other complaints filed by women, particularly where female officers are (randomly) assigned to run the help desks. This suggests that, even in resource-constrained and patriarchal environments, police responsiveness can be improved by mainstreaming and focusing attention on women's cases and by greater gender representation within the police. We explore the conditions under which these changes occurred, as well as the limits of the reform—which influenced officer behaviors and interactions with complainants but had little observable effect on the overall rates at which women turn to the police.

Reforming police under patriarchy

India, which was recently labeled the “world's most dangerous country for women” owing to high rates of sexual violence (19), is an instructive place to study gender-targeted police reforms. India ranks 140 of 156 countries on the Global Gender Gap Index—an international measure of gender inequality. This inequality is visible in India's skewed sex ratio at birth (929 girls/1000 boys), low female labor force participation rate (22.3%) (20), and high rates of GBV. An estimated 4 in 10 Indian women report experiencing domestic violence in their lifetimes (21).

GBV is a deeply rooted problem reflecting a nexus of social, economic, and political factors, and state-led interventions—including policing—are only one slice of the necessary responses. Policing is a particularly crucial but fraught arena in which to consider efforts to address GBV. The police are the primary institution for public security, as well as the most visible arm of the state's coercive apparatus. The police are charged with protecting women at the same time that law enforcement itself can play a role in perpetrating violence—including violence against women (22). The police are also critical gatekeepers to the broader justice system, yet where embedded in patriarchal norms can serve to deter or

¹Department of Economics, University of Virginia, Charlottesville, VA 22904, USA. ²Bureau for Research and Economic Analysis of Development (BREAD). ³Abdul Latif Jameel Poverty Action Lab (J-PAL), Cambridge, MA, USA. ⁴Department of Politics, University of Virginia, Charlottesville, VA 22904, USA. ⁵Saïd Business School, University of Oxford, Oxford OX1 1HP, UK.

*Corresponding author. Email: srs8yk@virginia.edu

block women's legal claims. We focus on police registration of cases of GBV as a preliminary step in a long chain to justice, including case investigation, arrests, judicial action, and referrals to social services. Although case registration is just one step, it is a critical one: If police cases are not filed, further legal action cannot occur.

In India, registration takes the form of a First Information Report (FIR). Lodged at a police station, the FIR records information about a "cognizable" offense (classified in the Indian Penal Code as offenses of a serious nature for which a police officer has authority to make an arrest without a warrant and to start investigation without permission from a court) (23). The FIR initiates case investigation and criminal proceedings, as stipulated under the Indian Penal Code. Domestic violence can also be registered in a Domestic Incident Report (DIR), a complaint mechanism created under the Domestic Violence Act of 2005. Filing a DIR initiates civil proceedings and referrals to social services, allows for protection orders and economic support, and may also lead to criminal proceedings. Unlike FIRs, which are filed at police stations, the DIR is filed with the local magistrate (judge). In the latter instance, the police can serve as designated Protection Officers who compile the DIR and bring it before the magistrate. More details on FIRs and DIRs are in Appendix S1.

Women face major obstacles to case registration, reflecting barriers to both demand (that inhibit the reporting of crime) and supply (that inhibit officers from recording cases) (24–26). Even when a woman overcomes social and familial pressures to report a case, officers often resist officially recording it—despite their legal obligation to do so (27). The police's hesitancy to file cases is driven in part by acute resource and capacity constraints, which push officers to lessen their caseloads (28), as well as by political pressure to show lower official crime rates (29). Patriarchal policing norms also push against case registration: Officers are encouraged to "protect families" by promoting reconciliation rather than the legal rights of women (27, 30–32) and often blame victims of sexual assault or question the validity of their claims. A recent report, for example, found that 39% of officers believe that complaints of GBV are unfounded (33). The same report also highlighted patriarchy within India's police stations. Women make up just 7% of the force nationally and face heavy work burdens, as well as workplace discrimination. Female officers, operating in these highly masculine settings, may feel pressure to act as "one of the boys," replicating and expressing patriarchal norms (34). These factors push against case registration: Indeed, there were fewer than four GBV-related FIRs registered per police

station (serving 130,000 people on average) per month at baseline in our data.

Intervention and randomization

With these barriers to case registration in mind, we conducted an experiment to test the impact of a bundle of gender-targeted reforms designed to make the police both more accessible and accountable to female complainants (35). Our study setting is Madhya Pradesh, a large (population 81 million) and ethnically diverse state in north-central India. Madhya Pradesh is representative, in socioeconomic indicators and gender norms, of much of northern India. It is also an illustrative setting in which to explore the problem of under-reported GBV. A report from the state's four largest cities, for example, found that only 1% of women who had experienced violence had reported it to the police (36).

India's federal architecture assigns state governments the responsibility for policing, as well as social programming for women's health and security. The WHD intervention was designed by the Madhya Pradesh Police (MPP) Research and Training Department, in consultation with our research team, lawyers, and GBV experts from civil society, with the goal of overcoming barriers to both the reporting of crimes by women and the recording of such cases by the police.

Undertaking research on GBV and policing is a fraught endeavor, which requires careful ethical consideration. We recognize that our roles as researchers are not and never can be fully neutral and that there is the potential, through research, to introduce risk, activate trauma, and create harm. We discuss our efforts to minimize such risks in Appendix S2. We are guided by recommendations developed specifically for research on GBV (37, 38), as well as by the insights of our local partners, which enabled us to ground global best practices in our study context.

The WHD intervention consists of four components: (i) private spaces (such as a room or cubicle) for female complaints within police stations; (ii) standard operating procedures on how to register cases and assist women visitors, along with officer training on those procedures and routine implementation monitoring; (iii) outreach to local women's and community safety networks; and (iv) assignment of female officers (at the rank of Assistant Sub-Inspector or higher) to run the WHDs.

Police stations—the unit of randomization—assigned to the first treatment arm ("regular" WHDs) received the first three components, whereas those assigned to the second arm (woman-run WHDs) received all four components. WHDs in the second treatment arm were directed to have designated female officers (with 90% compliance). Although there was no prohibition on assigning female officers

to WHDs in the first treatment arm, most (72%) were run by male officers. The control group continued with business as usual.

For the study, the MPP purposively selected 12 districts (out of 51) across the state, automatically including those home to the state's four largest cities and selecting the remainder to be representative of geography, demographics, and socioeconomic conditions across the state (39). We excluded specialized police stations (state or district headquarters, cyber cells, or all-women stations), as well as outposts (smaller than a station). MPP also excluded stations that were remote, servicing entirely rural populations, because they did not consider traffic to those stations high enough to justify a continuously staffed desk. The resulting sample of 180 police stations should thus be viewed as being representative of primarily (although not exclusively) urban and large police stations. Within each district, we stratified these 180 stations by geography (those that were fully urban, and those that also included rural areas) and by the first principal component of a vector of police station size characteristics (35). Within each of these strata, we then randomly assigned the stations into three groups: 61 to the first treatment arm (regular WHDs), 59 to the second (woman-run WHDs), and 60 to the control arm, which received no help desk intervention.

We consider May 2019 as the start date of the full intervention. Training began as early as July 2018, although state elections in December 2018 meant that training did not fully ramp up until March 2019. Our field teams monitored implementation in December 2019, by which point all treatment stations had at least some training, 94% had set aside requisite space, 87% had conducted a community outreach event, and 90% of women-run WHDs had a female officer assigned. However, only 67% had full training at the station level, and community outreach was limited overall given the huge jurisdictions these stations served. Appendix S4 details implementation efforts and timelines.

Data sources

Our aim was to evaluate whether the establishment of the WHDs, as well as assignment of female personnel to the WHDs, improved the responsiveness of police officers to women (40). To that end, we gathered data from five main sources.

1) Administrative data on crimes registered by the police at study police stations from May 2018 through March 2020. Our data are aggregated with no individual or identifiable case details, but FIR categories indicate (i) the number of "crimes against women" (CAW) cases, which include officially designated cases of GBV such as sexual assault, rape, dowry, and other cases (details in Appendix S1);

Table 1. Primary outcomes from administrative data. Each observation represents data at the police station–month level. DIRs are Domestic Incident Reports representing civil complaints of domestic violence (columns 1 and 5). FIRs are First Information Reports, either in cases of Crimes Against Women (CAW) filed by anyone (columns 2 and 6) or in all criminal cases filed by women (columns 3 and 7). Arrests correspond to arrests in CAW cases in a given month (columns 4 and 8). Regular WHD refers to Women's Help Desks without an assigned female officer; woman-run WHDs include an assigned female officer. Strata FE refers to fixed effects for district × urban/rural strata. Standard errors clustered by police station are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	DIRs filed	FIRs filed in CAW cases	FIRs filed by women	Arrests	DIRs filed	FIRs filed in CAW cases	FIRs filed by women	Arrests
Treatment	1.452*** (0.180)	0.542* (0.308)	0.269 (0.168)	0.265 (0.433)				
Regular WHD					1.492*** (0.250)	0.095 (0.339)	0.073 (0.188)	0.126 (0.457)
Woman-run WHD					1.410*** (0.205)	1.014** (0.400)	0.476** (0.207)	0.412 (0.506)
Baseline control	0.192 (0.268)	0.409*** (0.041)	0.344*** (0.036)	0.210*** (0.041)	0.191 (0.265)	0.408*** (0.040)	0.344*** (0.035)	0.209*** (0.040)
Observed	1980	1980	1980	1980	1980	1980	1980	1980
Adjusted R²	0.179	0.270	0.254	0.164	0.179	0.279	0.259	0.164
Control mean	0.047	3.847	2.577	3.433	0.047	3.847	2.577	3.433
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p value: T1 = T2					0.772	0.030	0.058	0.498

and (ii) whether the case was filed by a woman, the latter also incorporating nonviolent crimes.

2) CCTV data from the video feeds of cameras, present in all police stations, focused on the station entrance. MPP provided us with a week's worth of data from each study station, for the hours of 10 a.m. to 10 p.m. each day, at both baseline ($n = 12,537$, January to March 2019) and endline ($n = 9757$, February to March 2020). This footage enables us to measure the number of men and women entering or exiting the police station.

3) A user survey of members of the public who had visited study police stations, asking about their satisfaction with their visit. This was conducted at endline only ($n = 3251$, February to March 2020), drawing from a random selection of visitors within the span of 1 week.

4) A police survey, carried out at baseline ($n = 1950$, September to October 2018) and endline ($n = 1961$, February to March 2020) of personnel in different roles and ranks in study stations. The survey—a representative individual panel with equivalent rank replacements for any transferred officers—captured police perceptions and attitudes on crimes against women.

5) A survey of citizens, carried out at baseline ($n = 5648$ women, 871 men, November 2018 and March to April 2019) and endline ($n = 3376$, July to December 2020, phone survey). The survey—an individual panel sampled from all adult residents living in study police station jurisdictions—asked about perceptions

of safety, opinions of and contact with the police, and experiences of crimes.

Concurrent to this data collection, we also carried out qualitative research—observation and interviews with officers of various ranks—in eight police stations, selected to represent each arm of the study in two purposively selected districts (41).

Statistical methods

We report Intent to Treat estimates comparing average outcomes in treated stations (with either regular or woman-run WHDs) to outcomes in the control stations, as well as outcomes between the two treatment groups, using the most disaggregated measure available. All regressions include district-geographic stratum fixed effects (the level at which treatment probabilities are equal), with a control for the first principal component of variables used for further stratification. Sampled observations are weighted by using inverse sampling probabilities to make the outcomes representative at the police station level. Where available, we include the baseline value of the outcome as a control. We cluster standard errors at the police station, the level of randomization. Our preferred specifications are

$$(1) Y_{ips} = \alpha + \beta \text{treatment}_{ps} + \gamma \text{Y0}_{ips} + \delta_s + \text{pc}_{ps} + \epsilon_{ips}$$

$$(2) Y_{ips} = \alpha + \beta_1 \text{treat1}_{ps} + \beta_2 \text{treat2}_{ps} + \gamma \text{Y0}_{ips} + \delta_s + \text{pc}_{ps} + \epsilon_{ips}$$

where i is the individual, p is the police station, s is the stratum, Y is endline and $Y0$ is the baseline outcome. Using

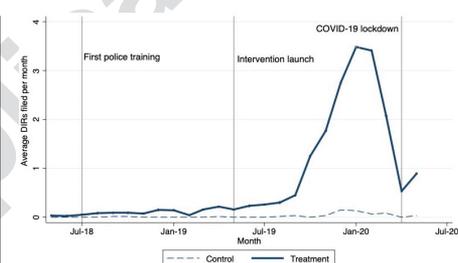


Fig. 1. Domestic incident registration increased sharply. Source: Raw official administrative data.

baseline outcomes, we find no more imbalances across treatment groups than would be expected by chance (tables S1 to S3) (42).

Results

The WHD intervention led to increased registration of women's cases (Table 1). First, we observe an increase in the number of DIRs filed, from practically zero in the control group (where officers had little to no knowledge of the DIR) to 1.5 monthly cases in the treatment group ($p < 0.01$). Figure 1 shows the sharp increase in DIRs after the official launch of the program in May 2019, although some cases were registered soon after training started in treatment stations as early as July 2018. Second, we also observe sizable increases in the number of FIRs filed in CAW cases (14.1%, $p = 0.08$), as well as FIRs filed by women (10.4%, $p = 0.1$). These results reflect an additional 1905 DIRs and 3360 FIRs registered in

treatment stations over the 11 months of the intervention. There were no significant changes in the number of arrests in CAW cases; the coefficient is positive but standard errors are large [95% confidence interval (CI) -0.58, 1.11] (43).

Separating the analysis by treatment group, we see that officers in stations with both regular and woman-run WHDs filed increased numbers of DIRs, with the coefficients statistically indistinguishable from each other. However, the increase in FIRs is entirely driven by the woman-run WHDs (Fig. 2). In regular WHDs, the treatment effects are small and statistically indistinguishable from zero. In the stations with woman-run WHDs, by contrast, FIRs in CAW cases increased by 26.4%, whereas FIRs filed by women rose by 18.5%; these treatment effects are statistically distinguish-

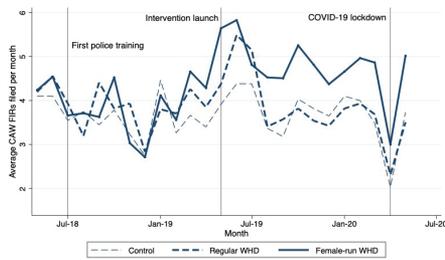


Fig. 2. Registration of crimes against women increased in women-run WHDs. Source: Raw official administrative data.

0.06; Table 1). Meanwhile, implementation quality and officer training both have a strong and significant impact on the number of cases registered for all outcomes (table S4). Finally, increases in registration of CAW cases do not come at the cost of reductions in other cases, with no discernible spillover effect on the overall number of FIRs or on other kinds of police reports (table S5).

These results are driven by changes in police behavior, reflected in a greater likelihood of registering a case once a woman has reported it, but not by any observable changes in the rates at which women report cases or approach the police. There is no impact on the overall CAW rates reported by women in our citizen survey (table S17). Neither is there an impact on the rates at which women visited police stations, based on analysis of CCTV data (Table 2), although these latter results must be viewed with caution given missing data issues (44). These non-impacts are not a result of these variables being uninformative or unpredictable of outcomes in general: Survey CAW numbers are predictive of FIRs registered, baseline CCTV counts are predictive of endline CCTV counts, and the number of female staff (from administrative data) is predictive of the number of female visitors in the CCTV data. Although we do not observe any obvious changes in citizen behavior, it is possible that there were changes in complainant behavior within the police stations, as women decide whether or not to pursue cases once at the station and in interaction with officers. Our research de-

sign and ethical parameters precluded us from observing any such interactions, but we posit that changes in complainant behavior are likely supported by changes in police behavior that we do observe. Among complainants, we do find minor changes in the satisfaction of female visitors to police stations, expressed in exit interviews (Fig. 3). In control stations, respondents generally expressed high levels of satisfaction with their visits (3.16 on a 4-point scale), agreed they were treated respectfully (3.31), and that they felt comfortable discussing their concerns (3.3). The high levels of reported satisfaction may reflect desirability bias among respondents hesitant to critique the police at the station, and given top-coded values, there may be limited scope to observe any improvement due to the WHD intervention. It follows that there are few statistically distinguishable differences in satisfaction between treatment and control stations; only the “comfort” variable is significantly improved given the presence of a WHD ($p = 0.07$, table S7). Notably, however, all three measures of user satisfaction were significantly higher, comparing among treatment stations, in those with better implementation and more training (table S8). In addition, both female and male visitors to treatment stations were significantly more likely to profess satisfaction with the physical infrastructure and conditions of the station, compared to visitors to control stations. Overall, these results suggest modest improvements in citizen perceptions of the police given the presence of a WHD. Meanwhile,

able from those in regular WHDs ($p = 0.03$,

Table 2. Primary outcomes from CCTS data. Observations are at time window–day–police station level, as described in Appendix S2. The number of female visitors is the per-hour count of all women who entered a police station at any time between 10 a.m. and 10 p.m. (columns 1 and 3). The proportion of female visitors is the number of female visitors divided by the number of all visitors captured by CCTV camera in the same hourly durations (columns 2 and 4). Regular WHD refers to Women’s Help Desks without an assigned female officer; woman-run WHDs include an assigned female officer. We use district FE instead of strata FE owing to the unavailability of data from 41 police stations, which results in an unbalanced distribution of treatment and control police stations across some strata. Though not shown in the table, the regressions control for the number of female officers in each police station at endline; fixed effects for time window of day and day of week; the average number of frames per second in the video; and the average number of pixels in the video.

	(1)	(2)	(3)	(4)
	No. of female visitors	Percent female visitors	No. of female visitors	Percent female visitors
Treatment	-0.590 (1.656)	0.005 (0.011)		
Regular WHD			-0.811 (1.894)	0.009 (0.011)
Woman-run WHD			-0.402 (2.097)	0.001 (0.012)
Baseline control	0.226*** (0.052)	0.359*** (0.062)	0.227*** (0.052)	0.359*** (0.061)
Observed	1832	1831	1832	1831
Adjusted R²	0.162	0.168	0.162	0.169
Control mean	13.706	0.092	13.706	0.092
District FE	Yes	Yes	Yes	Yes
p value: T1 = T2			0.857	0.403

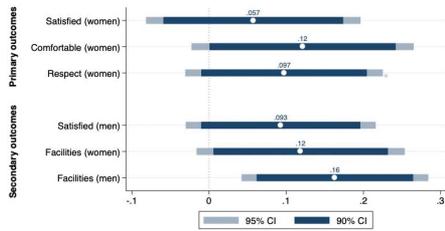


Fig. 3. Women visiting treatment stations were slightly more satisfied. Source: User Survey; coefficients and notes in table S7.

deep-seated police attitudes about gender did not appear to shift overall; there was no significant change in the rates at which officers reported the belief that women file so-called “false cases” against men (an indicator of whether they are inclined to believe women or dismiss their claims) (Fig. 4) (45). Notably, however, female officers in both treatment arms were less likely than those in control stations to ascribe to the narrative of false cases, though there was no such effect on male officers; this suggests that police attitudes shifted more for women. There is some evidence of a shift in officers’ awareness of the general inadequacies of policing with regard to women’s cases (Fig. 4). Both male and female officers in treatment stations were significantly less likely to describe the police as helpful to victims of crimes against women and were also less likely to believe that the police pay sufficient attention to women’s cases relative to other law and order issues (this last result is borderline significant; $p = 0.12$, table S13). We do not interpret these findings as evidence of a shift in gender attitudes, but rather of increased cognizance of the gap between how women’s cases should be handled and how they are handled in practice—a likely consequence of WHD training. Both male and female officers in treatment stations were also more likely to state (unprompted) that cases of crimes against women are among the top priorities within their stations—a likely effect of the attention-focusing presence of the WHDs.

Discussion

Our findings suggest that even in resource-constrained and patriarchal environments, efforts that focus attention on women’s cases can have a substantial impact on police behavior, making officers more responsive to women’s security concerns. This is visible in the higher registration of both FIRs and DIRs (respectively, criminal and civil complaints). The rise in DIRs across both treatment arms is pronounced, because it reflects the adoption of a relatively new practice (since 2005) that remains largely unknown to the police without WHD training, and so is all but absent at baseline and in control stations. Police officers are required to

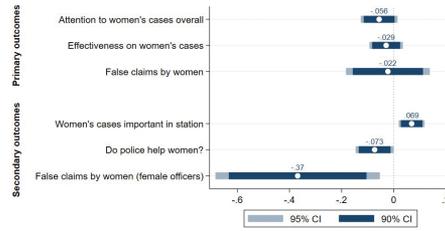


Fig. 4. Police attitudes and awareness. Source: Police survey; coefficients and notes in tables S13 and S14.

assist women in creating the report and must also ensure that the form is lodged with a local magistrate and that social services (such as shelter homes) are accessible—activities that extend beyond conventional police work. Through WHD training, officers gained knowledge of the DIR and learned to coordinate with other state and civil society agencies.

The increase in FIRs for women’s cases is also notable, as is the fact that it is driven almost entirely by woman-run WHDs. This is not simply a function of increased personnel assigned to the woman-run WHDs: Both woman-run WHDs and regular WHDs received additional high-ranked officers to operate the help desk (table S9). It appears, then, that the presence of additional female officers is critical for overcoming barriers to FIR registration. This gender-differentiated effect for FIRs but not DIRs, we suggest, reflects the higher costs to officers of filing an FIR. Unlike the DIR, the FIR automatically initiates a criminal case, requiring substantial investments of police time for investigation and in court proceedings. Moreover, to file an FIR, officers must push against strong norms within the police—articulated in our qualitative research—that prioritize “protecting families” by avoiding legal proceedings, in addition to dismissive narratives about “false cases.” The DIR, though certainly not trivial, requires less from officers, in part because it initiates a civil rather than criminal case, and in part because it passes some of the burden to other government agencies. It is also a relatively new practice that does not have the same norms pushing against it. Filing an FIR, in sum, requires higher levels of officer commitment, which we see primarily within woman-run WHDs.

Questions remain about the sources of commitment and the agency of female officers. More research is required to explore the mechanisms through which female officers affected change through the WHDs, as well as on how the WHD bundle affected female officers. However, our qualitative research suggests two avenues that, together, appear to empower female officers and increase their responsiveness to women. First, the intervention worked to build station-level capacity for action on

women’s cases, channeling resources to treatment stations, while also introducing manuals to clarify complex legal procedures, alongside intensive training. These investments in officer capacity may have been most acutely felt and utilized by female officers, who tend to be more overburdened—given their small numbers and the many tasks assigned to them—relative to male officers (33, 34). Female officers appear to have been particularly responsive to WHD training, which emphasized the legal requirement to file FIRs in the case of cognizable offenses, and which also contained gender sensitization modules that urged officers to listen to women’s claims and not dismiss them out of hand. The impact of this training is visible in the gender-differentiated rates of filing FIRs (more likely in woman-run stations) and shifts in beliefs concerning “false cases” (significant for female but not male officers).

Second, the WHD intervention worked to mainstream gender-responsive policing practices. By locating the help desks in regular police stations, and by training both male and female officers on the mission and operation of the desks, the WHDs helped to give visibility and ascribe value to work on women’s cases, rather than casting such work as peripheral to and therefore of lesser importance than other crime prevention tasks. Further study is required to examine the effects of this mainstreaming, which—in creating an enabling environment in which to focus on women’s cases—may help us understand why female officers were particularly receptive to WHD training and protocols.

These observations carry important insights for debates over representative bureaucracy and the question of whether the presence of members of marginalized groups within a public agency, such as the police, improves performance (4, 5, 30). Our findings suggest that descriptive representation does matter; female officers played a critical role in shaping the impact of the help desks. This, however, was not simply a matter of “feminizing” a police station by adding female officers. Rather, the agency of female officers must be understood as part of the full WHD bundle, including training, infrastructure, and higher-level supports that enabled those officers to work—both as women and for women. Efforts to enhance bureaucratic representation, this suggests, may hinge on institutional supports that activate the agency of underrepresented public personnel. In the case of the WHDs, the mainstreamed nature of the intervention (housed in regular, mixed-gender police stations) may have played a critical role. Although we are not able to directly compare the WHDs to all-women stations, recent research suggests that the act of “segregating” women’s cases may have the perverse effect of marginalizing female complainants (creating barriers and displacing cases) as well as female officers (who are isolated and

F4

sidelined from broader policing structures) (13, 30, 46). Locating women's help desks in regular stations, by contrast, appears to have worked to increase attention to women's security within "normal" police work. Future research is required to systematically explore such mainstreaming attempts, which may prove critical to broader efforts to build and support more representative bureaucracies.

Our study also highlights the limitations of police-centered reforms. First, gender attitudes among the police are difficult to move (although perhaps less so for female officers). Second, even as the WHD program increased the likelihood of the police recording crimes against women, barriers to women reporting such crimes remain. This may reflect limited community outreach implementation, with only 10% of women in our citizen survey aware of WHDs (table S21). Citizens' attitudes toward the police, moreover, did not shift, although this may reflect the short duration of the intervention; beliefs about the police may shift over time if the intervention becomes institutionalized and more visible to citizens. This, however, will depend not only on the ease with which women are able to register cases, but also on the functioning of the broader criminal justice system. Last, the WHD intervention did little to address deep-seated social and economic structures that both drive violence against women and inhibit women's access to justice. Efforts to address GBV necessarily require broader, multipronged approaches that extend beyond police reforms.

Conclusion

The urgent challenge of women's security has prompted police reform proposals worldwide, calling for increased female representation among officers, gender-sensitization training, and specialized services for women. To date, however, the limited evidence available on the efficacy of these reforms has been mixed, if not pessimistic, especially within patriarchal settings. Moreover, efforts aimed at mainstreaming (rather than segregating) these reforms have not been rigorously studied. Our experiment suggests that cautious optimism may be warranted, providing evidence that gender-targeted reforms can, under certain conditions, improve police responsiveness to women.

In addition to furthering our understanding of police reform and representative bureaucracy, our results also carry policy implications as governments in India and elsewhere consider gender-targeted reforms. Indeed, informed in part by these results, MPP has begun scaling up the WHD intervention to 700 police stations across the state. Studying this scale-up may help us understand whether the observed changes in police behavior can be sustained, as well as whether and how attitudinal change among both police and citizens can take hold.

REFERENCES AND NOTES

1. G. Blair et al., *Science* **374**, eabd3446 (2021).
2. Human Rights Watch, "Everyone Blames Me": Barriers to Justice and Support Systems for Sexual Assault Survivors in India (2017).
3. R. Manjoo, "Report of the Special Rapporteur on violence against women, its causes and consequences" (United Nations Office of the High Commissioner for Human Rights, 2012).
4. B. A. Ba, D. Knox, J. Mummolo, R. Rivera, *Science* **371**, 696–702 (2021).
5. L. Keiser, V. M. Wilkins, K. J. Meier, C. A. Holland, *Am. Polit. Sci. Rev.* **96**, 553–564 (2002).
6. K. Meier, J. Nicholson-Crotty, *Public Adm. Rev.* **66**, 850–860 (2006).
7. A. Miller, C. Segal, *Rev. Econ. Stud.* **86**, 2220–2247 (2019).
8. M. Morash, R. Haarr, *Fem. Criminol.* **7**, 3–23 (2012).
9. S. Karim, *Am. Polit. Sci. Rev.* **114**, 536–551 (2020).
10. S. Karim, M. J. Gilligan, R. Blair, K. Beardsley, *Int. Stud. Q.* **62**, 618–631 (2018).
11. N. Wagner, M. Rieger, A. Bedi, W. Hout, *J. Afr. Econ.* **26**, 492–515 (2017).
12. C. Santos, *Women's Police Stations: Gender, Violence, and Justice in São Paulo, Brazil* (Palgrave Macmillan, 2005).
13. N. Jassal, *Am. Polit. Sci. Rev.* **114**, 1035–1054 (2020).
14. C. Jackson, *J. Public Econ.* **96**, 173–187 (2012).
15. V. Ceccato, A. Loukaitou-Sideris, *Transit Crime and Sexual Violence in Cities: International Evidence and Prevention* (Routledge, 2020).
16. E. Perova, S. A. Reynolds, *Soc. Sci. Med.* **174**, 188–196 (2017).
17. S. Amaral, S. Bhalotra, N. Prakash, "Gender, crime and punishment: Evidence from women police stations in India" Working paper, 2021.
18. D. Kapur, *J. Econ. Perspect.* **34**, 31–54 (2020).
19. Thomas Reuters, "The World's Most Dangerous Countries for Women – 2018".
20. *The Global Gender Gap Report 2021* (World Economic Forum, 2021).
21. A. Kalokhe et al., *Glob. Public Health* **12**, 498–513 (2017).
22. M. Berry, M. Lake, *To End Violence Against Women, We Need to Imagine New Futures* (IPI Global Observatory, 2021).
23. Noncognizable offences, by contrast, require police officers to obtain warrants before arrests and court orders before initiating investigation.
24. P. Roychowdhury, *Capable Women, Incapable States: Negotiating Violence and Rights in India* (Oxford Univ. Press, 2020).
25. A. Banerjee, R. Chattopadhyay, E. Duflo, D. Keniston, N. Singh, *Am. Econ. J. Econ. Policy* **13**, 36–66 (2021).
26. L. Iyer, A. Mani, P. Mishra, P. Topalova, *Am. Econ. J. Appl. Econ.* **4**, 165–193 (2012).
27. Commonwealth Human Rights Initiative, *Barriers in Accessing Justice* (CHRI, 2020).
28. Nationally, there are just 1.2 officers per 1000 people in India. The United States, by comparison, has 2.3 police officers per 1000 people, according to the UN Office on Drugs and Crime.
29. On politicization of the police in India, see (47).
30. N. Jassal, *J. Asian Stud.* **80**, 631–661 (2021).
31. CSDS, *Status of Policing in India Report 2018: A Study of Performance and Perceptions* (Centre for the Study of Development Societies, New Delhi, 2018).
32. The practice of reconciliation builds on Indian legislation, including the 1984 Family Courts Act, and various Supreme Court directives, which advise officers to provide "counseling" to preserve families.
33. Common Cause et al., *Status of Policing in India 2019: Police Adequacy and Working Conditions* (Common Cause and Lokniti, 2019).
34. Commonwealth Human Rights Initiative, *Rough Roads to Equality: Women Police in South Asia* (CHRI, 2015).
35. Our Pre-Analysis Plans were registered with the American Economic Association's RCT Registry at <https://www.socialscienceregistry.org/trials/3357>.
36. S. Neville et al., "Evaluation of the Madhya Pradesh Safe Cities Initiative. Baseline Report" (DFID, London, 2015).
37. C. Bradbury-Jones et al., *Best Practice Guidelines in Relation to Gender-Based Violence Research in Low- and Middle-Income Countries* (Univ. of Birmingham/Institute for Global Innovation, 2020).

38. World Health Organization, *Putting Women First: Ethical and Safety Recommendations for Research on Domestic Violence Against Women* (WHO, 2003).
39. Bhopal, Gwalior, Indore, and Jabalpur districts contain the (eponymous) four largest cities in MP; the other districts are Baitul, Balaghat, Morena, Panna, Rattlam, Rewa, Seoni, and Vidisha.
40. The registration of crime must be distinguished from the actual incidence of crime which, by nature, is not observable through official data given the many deterrents to both reporting and recording cases (1).
41. Interviews took place over repeated station visits during a 10-month period and were carried out by two Research Assistants trained in qualitative methods. RAs visited each station approximately twice per month. Each visit lasted for a full day, yielding an average of 21.25 day-long visits per station. RAs interviewed four to six officers of different ranks in each station (including station chiefs, assigned help desk officers, and female officers where present). Interviews lasted for 1 hour on average, and most officers were interviewed multiple times. The qualitative research design is described in our Pre-Analysis Plan, Section 8 (35).
42. DIRs filed are significantly higher if data until April 2019 are considered, which is not surprising because many treatment stations had received training by that time. There is no difference in DIRs when including only data before July 2018 (prior to training). Excluding DIRs, out of 45 coefficients only one is significant at the 5% and one at the 10% level.
43. This lack of movement on arrests is unsurprising given the complex mix of factors that shape whether an arrest is appropriate (given the nature of a case), as well as the fact that some complainants, facing family pressure, may not want perpetrators arrested. Lack of increased arrests, therefore, should not be interpreted solely as an indicator of police inaction.
44. Because of data collection issues (COVID-related and otherwise) with the CCTV data, nearly one-third (55) of our study stations are not included in our online analysis of these outcomes, so these results must be interpreted with caution.
45. The relatively tight standard error bounds visible in the figure suggest that these are precise null effects, which is unsurprising given deep-rooted attitudes.
46. Our aim and empirical strategy did not include a direct comparison of the WHD intervention and all-women stations, which are entirely different models of policing. Moreover, there were only six all-women police stations in our study districts in MP during our study period.
47. S. K. Mitra, "Policing and strategic reform" in *The puzzle of India's Governance: Culture, Context and Comparative Theory* (Routledge, 2007).

ACKNOWLEDGMENTS

This study would not have been possible without the cooperation of the Madhya Pradesh Police department and its officials, in particular V. Kapoor, A. Shankar, R. Kumar Shukla, S. Rana, and R. Tiwari; as well as the enormous efforts of the JPAL-SA research team, in particular S. Bhattacharya, S. Suganya, S. Chandrasekhar, M. Tikku, M. Walia, P. Kumar, P. Baruah, K. Ravishankar, L. Diaz-Martin, S. Ramesh, and A. Bhargava; and the analysis team of J. Le, E. Robertson, and A. Kumar at UVA. We thank J. Doleac, O. Dube, S. Karim, S. Lakhtakia, A. Miller, P. Niehaus, S. Ostermann, N. Prakash, S. Sekhri, M. Williams, four anonymous referees, and participants in various seminars and conferences for comments. **IRB:** The study received IRB approval from the University of Virginia (#2548), the Institute for Financial Management and Research in India (#7107), and Oxford University (SSH_SBS_C1A_18_065). **Registration:** The trial is registered (with Pre-Analysis Plans) in the American Economic Association's randomized control trial registry under ID AEARCTR-0003357. **Funding:** This study was funded primarily by the Crime and Violence Initiative (CVI) of the Jameel Poverty Action Lab (JPAL), as well as the World Bank's Sexual Violence Research Initiative (SVRI), with supplemental funding provided by the University of Virginia's Center for Global Inquiry and Innovation (CGII) and the University of Oxford. **Author contributions:** Authors contributed equally to conceptualization, design, and writing. S.S. was responsible for randomization and data analysis. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** A Pre-Analysis Plan was registered with the American Economic Association's RCT Registry at: <https://www.socialscienceregistry.org/trials/3357>. All code, and survey and administrative data used in the analysis will be

made available (with PII and police station names extracted to preserve anonymity as required by IRB) on the JPAL dataverse hosted by Harvard University (<https://dataverse.harvard.edu/dataverse/jpal>). The raw CCTV videos will not be available given privacy issues, but the final data used for analysis will be made available. All data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials. **License information:** Copyright © 2022 the authors, some rights reserved;

exclusive licensee American Association for the Advancement of Science. No claim to original US government works. <https://www.science.org/about/science-licenses-journal-article-reuse>.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm7387
Appendix S1: Clarifications and further background

Appendix S2: Ethics
Appendix S3: Tables S1 to S24
Appendix S4: Implementation timeline and details, table S25, and figs. S1 to S6
References (48–65)

Submitted 7 October 2021; accepted 4 May 2022
[10.1126/science.abm7387](https://doi.org/10.1126/science.abm7387)

Confidential
Do not duplicate