Gender, Violence, and Triage: Complainant Identity and Criminal Justice in India

Nirvikar Jassal*

October 1 2021

Please Do Not Cite/Circulate Without Permission

Abstract

Are women hindered vis-à-vis accessing justice? I provide evidence of *institutional triage* in which particular complaints are disadvantaged when passing through nodes of a justice system in which multiple administrators utilize discretion to discriminate. Using an original dataset of roughly half a million Indian crime reports, merged with court files, I find that women's complaints are significantly more likely to be delayed and dismissed at the police station *and* courthouse compared to men. Suspects that female complainants accuse of crime are less likely to be convicted and more likely to be acquitted, an imbalance that persists even when accounting for cases of violence against women (VAW). The application of machine learning to cases reveals—contrary to intuitions of policymakers or judges—that VAW, including the extortive practice of dowry, are not "petty quarrels," but may involve starvation, poisoning, and marital rape. To make a *causal* claim about the impact of complainant identity on outcomes, I utilize a matching technique that uses high-dimensional text data; it underscores why those who suffer from cumulative disadvantage in society may be likely to face challenges whilst seeking punitive justice via formal state institutions.

Keywords: Gender, Crime, Policing, Violence Against Women, Sexual Assault, India

*Postdoctoral Fellow, Stanford University (njassal@stanford.edu). I thank Aprajit Mahajan, Paul Novosad, Sam Asher, Irfan Nooruddin, Alison Post, Abhijit Banerjee, and Margit Tavits, as well as Elliott Ash and Christoph Goessmann for supplementing the data. I am grateful for excellent research assistance at Stanford by Emily Wu and Shirley Cheng.

Introduction

Are minorities disadvantaged in accessing justice, and if so how? These are questions of theoretical and policy relevance, without clear answers. In the largest democracy of India, journalists regularly report that women and minorities are discriminated against when seeking help from the state. Yet, aside from challenges in accessing data that can tackle these puzzles, it remains ambiguous as to whether any disparities are attributable to the *types* of cases registered by such groups or their *identity*. If women are discriminated against, is it because of their gender or the content of their complaints, e.g. harder-to-prove cases of violence against women (VAW)[1]

Not only is there limited research on crime and policing in political science, but also few discussions about inequities in state responses to violence (Htun and Weldon 2012). Investigations into VAW in economics (Jayachandran 2015), sociology (Armstrong, Gleckman-Krut, and Johnson 2018), or criminology (Khan et al. 2020), are typically carried out through the prism of sexual assault (McDougal et al. 2018). In political science, scholarship on VAW has exclusively focused on rape in conflict or post-conflict settings (Karim 2020; Cohen 2013; Agerberg and Kreft 2020), rather than gradations of *everyday* abuse (Khan et al. 2020). And, while an emerging body of work has sought to re-prioritize attention toward criminal justice, most studies experimentally test the impact of police interventions,[2] rather than paint a portrait of the broader system.

I ask whether women in India are less likely than men to access justice when turning to the state, i.e. police *and* judiciary. I advance a theory of 'institutional triage' to explain how officials use discretion to filter cases as complaints funnel through nodes of the justice system. This triage, deployed at specific junctures, marginalizes those who may already suffer from cumulative disadvantage, compounding existing inequalities, including those rooted in gender. To illustrate, I create an original micro-level dataset of the universe of crime from Haryana, part of the Hindi-speaking heartland, and merge them with court files, thereby tracing cases from the second a victim enters a police station until (potentially years) later when a verdict is issued.

The article combines several research questions—e.g., on police accountability toward minorities and/or judicial bias against women—into one holistic study. By linking all arms of the system for the first time, I establish a series of facts, e.g. cases of VAW are likely to be delayed in terms of police registration and court verdict compared to non-gendered crime. Unlike the average 18% conviction for non-gendered cases, VAW results in only 7-10% conviction for suspects. Strikingly, even accounting for VAW, female complainants are significantly more likely to have their cases dismissed, delayed, or result in a suspect's acquittal compared to male complainants. I attempt to provide credible evidence that this is causally identifiable.

The paper aims to make additional contributions. Scholarship has pointed to social impediments hindering women from coming forward to authorities (Iyer et al. 2012; Green, Wilke, and Cooper 2020; Jassal and Barnhardt 2020), with an implicit assumption that if only they can be encouraged to *report* crime, the state may be accommodating. The findings herein not only hint at why VAW carries on with impunity, but also suggest that hesitancy in reporting could be grounded in calculations about the low probability of punitive justice at the conclusion of an arduous process. "Gatekeeping" decisions by police in terms of case registration, while

1. India has been dubbed the most unsafe country for women (Goldsmith and Beresford 2018); 28%, 6.6%, and 78.4% of women report physical violence, sexual assault, and fear of their spouse, respectively (DHS 2017). The UN definition of VAW is, "any act of gender-based violence that results in, or is likely to result in, physical, sexual or mental harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or in private life" (WHO, n.d.). Sexual assault is *one* component of VAW.

2. E.g. community policing, representation, and training (Blair, Karim, and Morse 2019; EGAP 2019).

important, may ultimately have little to do with punishment for crime (Spohn and Tellis 2019).

The study supplements work on bureaucratic discrimination, much of which has focused on ethnicity or involved audit experiments (Butler and Broockman 2011; White, Nathan, and Faller 2015), rather than administrative data (Emeriau 2021). I use the universe of registrations to depict the true "ground reality" for women facing challenges as complaints are being processed, simultaneously quantifying the duration of police investigations, court hearings, and other outcomes, i.e. granular points of interest to scholars of state capacity and South Asia. The work also expands research on gender disparities—which in India have focused on education, income (Calvi 2020), health (Dupas and Jain 2021), and property (Brulé 2020)—to justice delivery.

Another novelty of the study is that it applies unsupervised machine learning to police reports, each of which contain ≈500-word first-person testimonies (Roberts, Stewart, and Airoldi 2016; Roberts, Stewart, and Nielsen 2020). While such methods have been used to probe the content of Arabic *fatwas* (Lucas et al. 2015), Indian rural deliberation (Parthasarathy, Rao, and Palaniswamy 2019), or UK parliamentary debate (Sanders, Lisi, and Schonhardt-Bailey 2017), they have not been applied to the study of crime. The benefits of a text-as-data approach are three-fold. First, it amplifies *victims*' voices, minimizing the researcher's involvement. Second, topic modeling disentangles VAW carried out *in* and *out* of the household, summarizing actual triaged cases, e.g. marital rape or abuse related to women's extortion for dowry. Third, topic-matching diminishes confounding to attempt causal inference using text (Feder et al. 2021).

The study is structured as follows: I outline the theory, contextualize the Indian criminal justice system, and explain the merging process of two distinct records. I present quantitative tests of the argument, utilizing descriptive and OLS analyses, topic modeling, and matching. I discuss the insights, as well as the research agenda that the findings illuminate.

Institutional Triage in Criminal Justice

In a review essay, Kurlychek and Johnson (2019) note that existing studies on U.S. criminal justice tend to examine isolated stages or "episodic disparities" rather than the reproduction of inequality from one body to the next. U.S. studies—which look at *either* the police *or* judiciary—show that African Americans are disadvantaged with regard to bail, sentencing, and incarceration (Arnold, Dobbie, and Yang 2018; Alesina and La Ferrara 2014; Abrams, Bertrand, and Mullainathan 2012; Knox, Lowe, and Mummolo 2020). One reason for the imbalances is what legal scholars call "triage," i.e. lawyers' de-prioritization of minorities' cases (Brown 2004; Richardson 2016). Because public defenders are overworked, their implicit biases produce shortcuts in allocating time or resources, e.g. delaying interviews of witnesses or carrying out shoddy investigations for cases seemingly predisposed to an outcome (Richardson and Goff 2012).

I define institutional triage as a form of *system*-wide discrimination wherein administrators—e.g. from the constable to the judge—leverage the discretion at their disposal to filter or deprioritize specific complaints as they move through nodes in the chain. In criminal justice, these nodes might include (a) *police registration*, e.g. citizens may be turned away or dissuaded from case filing; (b) *police investigation*, e.g. officers may delay inquiries or persuade the complainant to withdraw the report; (c) *preliminary hearing*, e.g. judges may stall arbitration or postpone trial dates; and (d) *court decision*, e.g. judges may acquit rather than convict suspects. Broadly, triage manifests in non-episodic unequal outcomes (*exclusion*), or a disproportionately trying process (*burdens*) across stages (Olsen, Kyhse-Andersen, and Moynihan 2020).

I aim to make a distinction between mere discrimination and triage. First, unlike discrimination that may occur as citizens *avoid* the authorities (e.g. traffic stops or arbitrary arrest), triage exhibits when individuals actively *turn to* the state for grievance redressal (Kruks-Wisner 2021). Second, discrimination may describe single-stages (e.g. stop-and-frisk), whereas triage encapsulates the "squeezing" of requests through multi-nodal agencies (Figure 1). Unlike, say, obtaining a driver's license wherein one agency provides all services, criminal justice is a paradigmatic setting in which triage might manifest because at least two linked bureaucracies are involved in providing services for the *same* complaint.

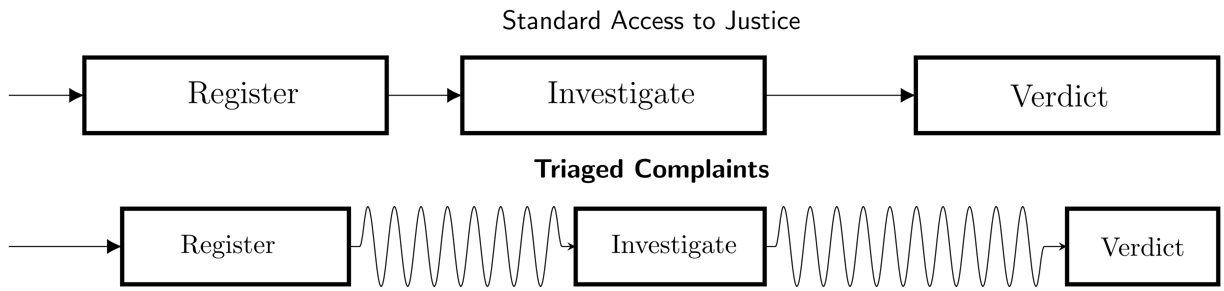


Figure 1: Standard access to justice versus "triage" wherein requests spend longer in-between nodes and have a lower probability of transitioning (as seen in the progressively smaller size of boxes).

However, the theory does not speak to administrator motivation. While triage might certainly be rooted in taste-based discrimination, officials may also be embedded within a milieu (e.g. where domestic violence is seen as a "family matter"), or constrained by resource scarcity (Dasgupta and Kapur 2020). Indeed, low levels of development and layered bureaucracies can result in misgovernance without actors behaving with repressive intent (Banerjee 1997; Slough and Fariss 2021). Officials may even display preference-based discrimination (paternalism), "protecting" victims from the complex (and public) process of accessing formal justice (Bindler and Hjalmarsson 2020). Regardless of motives, a testable implication of triage is that economically or socially disadvantaged groups in society will see a diminished speed and likelihood of their cases crossing the desks of disparate officials, each of whom retain varying levels of discretion.

Charting cases in this way may lead to greater precision. For instance, if police mishandle investigations, judges may have limited evidence; consequently, looking *only* at a single-stage dataset of judicial verdicts may lead to a misleading conclusion that judges are to blame (Lang and Spitzer 2020).[3] Yet, because triage can only be probed by tracing complaints across time and space, it has been challenging to show because of the inability to link multiple nodes.[4] For the first time, I follow administrator decisions sequentially across bureaucracies, which Holland (2016) refers to as "enforcement process tracing." The approach determines, "the number of and type of cases that feed up to the next step of the process until ultimately resulting in a sanction" (Bozçağa and Holland 2018, 303), thereby highlighting bottlenecks and sources of "leakage."

I look at Haryana, a patriarchal region of north India (Jassal 2021). Here, women may be less likely to have organizational support such as access to lawyers (Tellez, Wibbels, and Krishna 2020; Roychowdhury 2021), and cases of VAW may be perceived as difficult to prove and a strain on bureaucratic resources. Culturally, administrators may see women's cases, including VAW that takes place inside the home such as dowry,[5] as a threat to marriage and male dominance.

3. Spohn and Tellis (2019) show how numerous sexual assault cases for which the LAPD have probable cause never yield arrest but are rejected by the District Attorney *prior* to felony charges.

4. See Rehavi and Starr (2014) for a notable exception of multi-nodal data linkage in the United States.

5. Unlike bride-price, dowry involves a wife being coerced, often violently, into providing resources to her spouse

The framework would thus predict that women's cases and VAW will face obstacles vis-à-vis the *process* and *outcomes* associated with formal justice delivery from the stage of entry (police registration) to exit (judicial verdict). I test two sets of hypotheses:

> 1a: *At the stage of entry, women's cases and gendered crime will be more likely to have been delayed vis-à-vis police registration than men's cases and non-gendered crime.*
>
> 1b: *Conditional on police registration, women's cases and gendered crime will be less likely to be sent to court than men's cases and non-gendered crime.*
>
> 2a: *Conditional on entering court, women's cases and gendered crime will be more likely to be delayed vis-à-vis resolution than men's cases and non-gendered crime.*
>
> 2b: *At the stage of exit, women's cases and gendered crime will be less likely to result in a suspect's judicial conviction than men's cases and non-gendered crime.*

Gender and the Indian Criminal Justice System

Crime registration is a citizen's primary step toward formal justice. Registration occurs at police stations run by a head station officer, who is supported by staff (e.g. sub-inspectors). The police are supposed to file all complaints whether they believe them to be valid or not, but in practice have leeway as to which cases are registered. When filed, a case is assigned to a deputy, and, depending on the crime-type, investigations have to be completed within a time-window (e.g. 90 days). If the case is not dropped, or withdrawn, it is sent to the next wing.

The judiciary is related to other former British colonies wherein the Supreme Court sits at the apex of a hierarchy that includes roughly two dozen High Courts, and 7000 district/subordinate courts. Every police station is located within a jurisdiction of a district court; crime reports and any evidence collected during police investigations are assigned to a jurisdictional judge (Ash et al. 2021). These judges may be of the rank District and Sessions Judge down to a Civil Judge–Junior Division. On appeal, a case may travel to a High Court or the Supreme Court.

Figure 2 presents a stylized illustration. Level A represents the abstract concept of all crime, which can never be precisely measured. Level B signifies those who came forward to report (e.g. at a station or help-desk). Within Level 1—when reported crime transition to registered cases—there are two sub-categories: women's complaints and gendered crime (or VAW).[6] (This is illustrated in a Venn diagram because not all VAW is reported by women.[7]) Cases in Level 2 represent those that, after a preliminary investigation, survive police cancellation. The remaining cases, once investigated, enter the judiciary in Level 3. There, unless stalled or dismissed, a verdict may be issued after trials that (dis)favors the complainant in the original crime report.

Judges have greater discretion as to how cases are handled compared to law enforcement. For the police, there are explicit rules that mandate registration of "cognizable" or serious crimes,[8] some introduced after an infamous 2012 gang-rape of a Delhi college student. Police are required to register all gendered complaints—including acid attacks, sexual harassment, trafficking, and

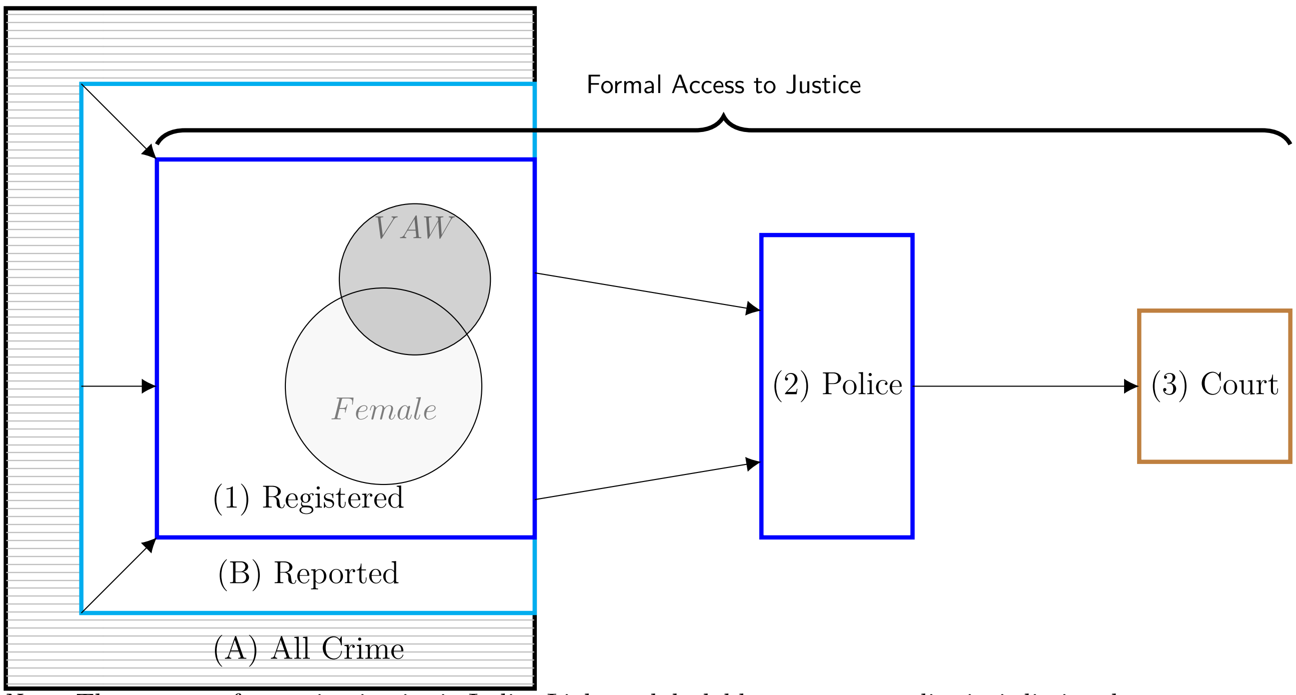
(Anderson 2007; Rao 1993, 1997; Srinivasan and Bedi 2007). Historically associated with small tokens or gifts and originally a practice among the upper caste (Srinivas 1956), it is among the most common gendered crimes in India today (Jassal and Barnhardt 2020). The practice has been linked to wife-beating, murder, and "missing girls" (Rao 1997; Srinivasan and Bedi 2007; Rose 1999; Bhalotra, Chakravarty, and Gulesci 2020).

6. In criminology, the gap between Levels 1-A is called "the dark figure of crime" (Biderman and Reiss 1967).

7. VAW can be further subdivided: abuse inside the household involves the spouse, family, or in-laws.

8. Section 154 of Code of Criminal Procedure.

Figure 2



Note: The process of accessing justice in India. Light and dark blue represent police jurisdiction; brown represents the judiciary. Arrows signify nodes that connect the system. The analyses focus on *all* steps from Levels 1-3.

rape—with the threat of one-year jail time and fine for the officer.[9] Manuals mandate that rape investigations be completed within two-months of filing.[10] Aside from being pressured "from above" via such guidelines, the police are also constrained "from below" where, for example, activists and NGOs assist victims in filing cases, especially VAW (Roychowdhury 2021). The judiciary is exempt from such pressures[11] or from juries, which were formally abolished in 1973.[12]

During registration, police officers stamp Penal Codes to case registrations in order to signal what laws are alleged to have been broken. Gendered Penal Codes (and related "acts") include Section 326-A (acid throwing), Section 376 (rape),[13] Protection of Women from Domestic Violence Act, and others.[14] An important law is Section 498-A. In 1983, a new provision made "cruelty" by a husband (or in-laws) against a wife a crime (Oldenburg 2002).[15] While intended for dowry harassment, the law was applicable to domestic violence.[16] Some politicians argue that

9. Section 166A of the Penal Code.

10. Section 173 of Code of Criminal Procedure.

11. Law enforcement is also constrained and subservient to the bureaucracy (or Administrative Service) and, in practice, answerable to local politicians who hold sway over promotions and transfers (Iyer and Mani 2012).

12. Jury trials had been in operation since British India to 1959. See 1973 Code of Criminal Procedure.

13. See Table A1 for full list. While Section 497 (adultery) might not be considered VAW, I classify all gendered sections as VAW from official lists. This clause was ruled unconstitutional in 2018 (Jassal and Chhibber 2019).

14. There are implicit distinctions between 'heinous' and 'non-heinous' violations. Non-heinous cases include 'compoundable' sections where police are not forced to take action if the victim settles. Gendered cases such as Section 497 (adultery) or Section 312 (causing miscarriage) are compoundable. Bailable, compoundable, and non-cognizable laws are considered the least serious. Section 320 of the Code of Criminal Procedure.

15. Some feminists criticized the clause because it was restricted to married women, and retained a vague definition (Kothari 2005). 'Cruelty' is defined as conduct that drives a woman to suicide, causes grave injury, or endangers life. Section 498-A was followed with Section 304-B or "dowry death," wherein violence related to extortion for dowry culminates in the victim's suicide or murder.

16. The law enabled "dowry" to become a metaphor for all violence in the marital home. In 2005, the Pro-

women exaggerate when registering such cases, even noting, "Many families are destroyed or ruined under such [gendered] provisions, and the legal proceedings go on for years. Men's rights organizations are working to raise awareness...in opposition to women...men should be arrested after proper inquiries rather than on the basis of the woman's complaint" (Verma 2017).

These sentiments are not restricted to politicians. (All-male) benches of the Supreme Court have ruled that domestic violence provisions are, "a license for unscrupulous persons to wreck personal vendetta or unleash harassment [against men]," and a form of "legal terrorism [by women]."[17] The Court has noted, "...complaints under Section 498-A are filed in the heat of the moment over trivial issues without proper deliberations. The learned members of the Bar have enormous social responsibility and obligation to ensure that the social fiber of family life is not ruined or demolished,"[18] and that women should not file cases to, "satisfy the ego and anger of the complainant."[19] These pronouncements imply that women's cases are (a) frivolous, (b) reported in the heat of the moment, (c) submitted by those with an agenda, or (d) best resolved through reconciliation (Basu 2012). I scrutinize these assumptions using two sources of data.

The First-Information-Report Dataset + Judicial Records

In a push for transparency, India made crime or First-Information-Reports (FIRs) accessible (Court 2016). Over several years, I harvested and parsed millions of records; the present study utilizes all 418,190 registrations in Haryana from January 2015-November 2018.[20] I focus on this state for which I translated reports into English, and worked with the local police to collect information about officers and previously inaccessible cases.[21] Aside from particulars about victims, suspects, and officers, FIRs contain descriptions of the incident, generally unaffected by social desirability.[22] Because few people in the Subcontinent have meaningful interaction with law enforcement (CSDS and Cause 2018), crime reports, unlike survey measures, enable us to zero in on individuals who interacted with state officials.[23]

I then merged FIRs with judicial records. India has made (semi-) public the universe of judicial files on a platform called E-Courts, similar to a domain established by China (Liebman et al. 2020). Judicial records contain details about the date of filing/first appearance in court for FIRs, judges assigned, and verdict (if any). With support from scholars at ETH Zürich and the Development Data Lab—who compiled the universe of 80 million records from 2010-2018—I merged these files via the particulars of the police station, complainant name, and other identifiers.[24] Out of 418,190 crime reports, I merged precisely 251,804 or 60.2% to court files, a figure that accurately represents registered cases that were sent to court.[25]

tection of Women from Domestic Violence Act expanded the definition of domestic violence, but also prioritized 'counseling' abused women. Agnes and D'Mello (2015, 80) argue, "...counseling is based on a patriarchal premise and is laden with anti-women biases...advised to "save the marriage" even at the cost of danger to her life."

17. *Sushil Kumar Sharma v. Union of India*, No. 141, 2005.

18. *Preeti Gupta & Anr. v. State of Jharkhand*, Appeal No. 1512, Criminal Appellate Jurisdiction, 2010.

19. *Rajesh Sharma v. State of Uttar Pradesh*, Appeal No. 1265, Criminal Appellate Jurisdiction, 2017.

20. I anonymize the dataset in replication files.

21. The police are exempted from releasing details on 'sensitive' cases involving sexual assault or insurgency.

22. Citizens would have had to provide as much detail to officers to initiate investigation.

23. Victims of VAW, for instance, do not turn to the police as one of the top five sources for help (DHS 2017).

24. Documents produced by each wing are formatted differently, requiring manual re-coding. As a check, Penal Code violations in FIRs were fuzzy matched with those in the court files to ensure cases were correctly merged.

25. As a validation exercise, I show that a third of cases of VAW could *not* be matched to court, reinforcing research based on internal police memos demonstrating ≈30% of crime as cancelled (Jassal 2020).

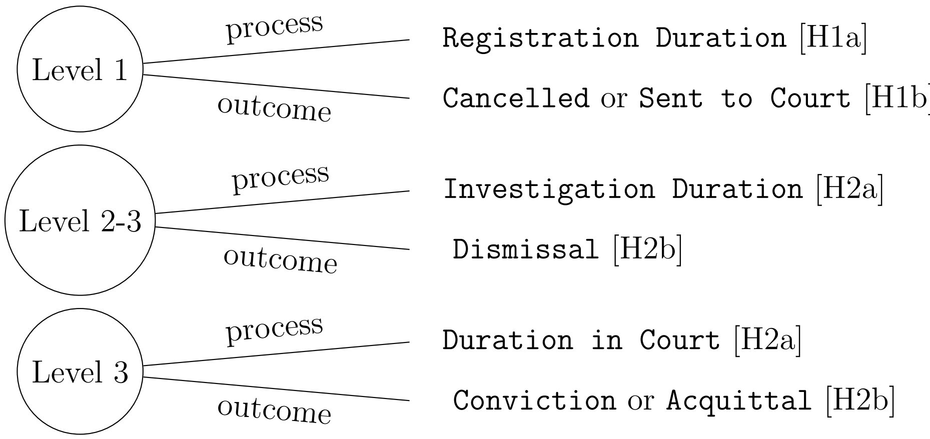


Figure 3: Measures of Institutional Triage and Corresponding Hypotheses

Research Design: OLS, STM & Topical Inverse Regression Matching

To evaluate H1a, I examine the duration of time it took to file an FIR. Each report has dates of case registration, as well as when the complainant told an officer the crime *began* or *ended*. `Registration Duration` reflects the difference between registration date and incident, thus providing an estimate vis-à-vis delays in police filings. To test H1b, I examine the likelihood of a registered case being sent to court. Specifically, non-merged cases are categorized as `Cancelled`, illustrating that law enforcement did not send them to the next branch.

For H2a, I create two measures. First, `Investigation Duration`—the difference (in days) between FIR registration and preliminary hearing in court—estimates the time of police investigation. Second, I create a numeric variable corresponding to the number of days from the preliminary to latest court hearing on file (`Duration in Court`). To evaluate H2b, I create *three* indicator variables of judicial review, i.e. whether the case was ejected by a judge at an initial (bail) hearing (`Dismissal`); or whether, after subsequent trials, the outcome resulted in a suspect's `Conviction` or `Acquittal`. I utilize variations of the following OLS model:

$$Y_i = \alpha + \beta_1 Female_i + \beta_2 VAW_i + \beta_3 (Female \cdot VAW)_i + \vec{\gamma} S_s + \vec{\eta} C_c + \epsilon_i \tag{1}$$

Y is a binary or numeric outcome for crime report i. `Female` is an indicator representing whether the case involved a woman as the primary complainant, while `VAW` signifies whether a gendered Penal Code was affixed to the FIR. S_s and C_c are a set of station- and court-level covariates, e.g. dummies for police station, district, month-year (of registration), rank of investigator, rank of presiding judge, and whether the area in which the case was tackled is urban. When excluding `VAW`, I include fixed effects for the primary[26] Penal Code violation, enabling me to compare differences between complainants *within* categories of crime (e.g. theft). The interaction allows us to observe the difference between men and women for gendered and non-gendered crime. In the Appendix, I breakdown the results for four common types of VAW: female kidnapping, rape, dowry harassment, and criminal force. The standard errors for all models are clustered at the district level. Figure 3 provides a breakdown of the measures.

I also estimate structural topic models (STM) that, in a regression-type framework, can predict whether cases devoted to a topic (e.g. rape) are functions of covariates, e.g. the probability

26. As seen in Appendix Figure A3, most FIRs are combinations of multiple Penal Code clauses, with the first listed generally indicating the case type. There are approximately 1000 unique Penal Codes and special acts.

of being dismissed (Roberts, Stewart, and Tingley 2019; Roberts et al. 2014; Roberts, Stewart, and Airoldi 2016).[27] Unsupervised machine learning de-emphasizes categorizations of crime based on coarse Penal Codes and disaggregates crime, e.g. domestic violence from attempted murder. To do this, I compiled and parsed text from each FIR into an R-readable format, and then translated the (primarily) Hindi text for 418,190 reports (200 million words or ≈450,000 A4-size single pages) using Google Translate.[28]

It is possible that fixed effects OLS models and STM might still lead to imprecise estimates about the impact of complainant gender on, say, conviction. There may be concerns about omitted variable bias or inframarginality (Arnold, Dobbie, and Yang 2018), i.e. even within crime type (e.g. theft), women may report distinct sub-types of cases (e.g. chain-snatching) compared to men (e.g. motorcycle robbery). Consequently, I utilize a *third* method: topical inverse regression matching (TIRM), introduced by Roberts, Stewart, and Nielsen (2020), that allows one to condition on the *content* within FIRs, thereby diminishing confounding.

To implement TIRM, I estimate a STM with a "treatment" (a woman's crime report) as a content covariate. This estimates the relationship between having a female complainant and words in the corpus, as well as how FIRs registered by women discuss topics differently (Roberts, Stewart, and Airoldi 2016). Following Roberts, Stewart, and Nielsen (2020), I extract topic proportions for control FIRs as though they were treated,[29] attaching an estimated propensity score to the topic-proportion vector for every FIR, and then performing coarsened exact matching (Iacus, King, and Porro 2012), in order to fit models predicting conviction or acquittal.

Descriptive Statistics

Figure 4 displays the top Penal Codes appearing in cases registered by female complainants as well as in the category of VAW.[30] Women registered 38,828 or 9% of all FIRs. Descriptively, there are differences in the types of cases registered by women and men (Appendix Figure A1). For instance, for men, the top substantive[31] Penal Codes relate to theft, rash driving, burglary, and public intoxication/bootlegging. The top substantive Penal Code for women is Section 498-A; domestic violence/dowry-related abuse perpetrated by a spouse (or in-laws) was present in 15% of their registrations.[32] Other common gendered Penal Codes include abduction (e.g. kidnapping a woman "to compel her into marriage"/"procuring a minor girl"),[33] "obscene acts/songs,"[34] "criminal force against a woman,"[35] rape, "insulting the modesty of a woman,"[36] stalking, "intent

27. For most analyses, I specify 35-40 topics. As seen in Figure 4 and Appendix Figure A1, most crimes can be slotted into roughly two-dozen Penal Code classifications. I see more repeat topics for values greater than 40.

28. I analyze translations because (a) machine learning, including the STM, were designed primarily for English, and (b) to ease pre-processing, i.e. stemming, lemmatization, and ejection of stop- or common words.

29. The content covariate in the STM knows the weight of each word and topic-word combination. The projection for an FIR would then be the sum of its weighted word counts normalized by FIR length.

30. Appendix Figure A2 presents a heat map illustrating the locations of registrations.

31. Most sections relate to concrete violations, e.g. theft and murder. There are additional clauses that are non-substantive and attached as supplements, e.g. Section 323 (causing hurt), invoked for rash driving to extortion.

32. Many Penal Codes are registered in conjunction with Section 498-A, e.g. "unnatural"/anal sex (for marital rape), or dowry death (when domestic violence culminates in suicide or murder). See Appendix Figure A3.

33. Invoked from cases ranging abductions to young women eloping or running away with boyfriends.

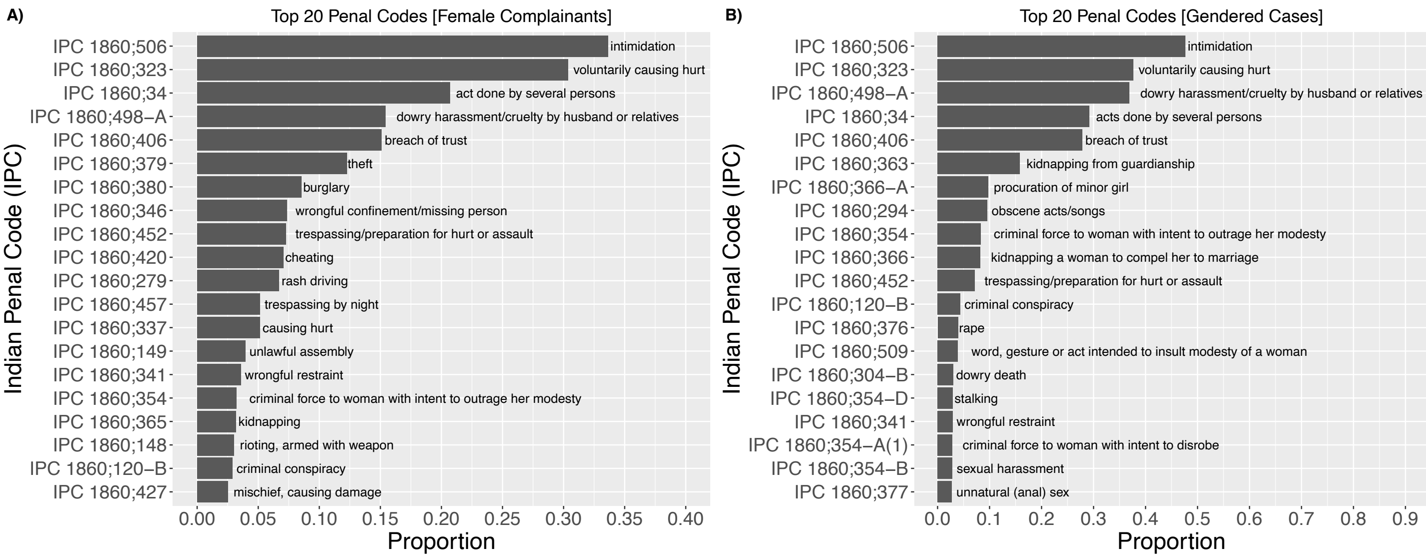
34. Invoked in cases that may include lewd behavior in front of, or towards a woman, as well as 'obscenity.'

35. Invoked in cases ranging from acting aggressively to attempted rape.

36. Invoked in a range of cases, including exhibitionism and invasion of privacy.

to disrobe," sexual harassment, and "unnatural" (anal) sex.[37]

Figure 4: Top Indian Penal Code Sections Listed [Female Complainants and Gendered Crime]



Note: Top twenty Penal Codes attached to women's cases (N=38,828) and gendered crime or VAW (N=20,869). See Appendix Figure A1 for male complainants and non-gendered crime.

Table 1 presents descriptive statistics for the FIR dataset. `Distance` reveals that crime takes place, on average, 6 kilometers from a station. Cases likely have 2 suspects, with crimes registered by women, and VAW, more likely to have a female suspect (`Female Suspects`). (As Jassal and Barnhardt (2020) show, cases of dowry-related oppression may involve the complainant's mother-in-law.) While officers do not always note the ages of victims, non-missing data suggest that complainants are, on average, in their 30s. VAW is likely to have more Penal Codes appended (`No of Sections`), and complainants wait longer at the station in anticipation of registration (9.3 hours). The variables prefixed with 'R:' represent investigator ranks; women's cases are less likely to be assigned to constables (who cannot charge-sheet cases).

Unlike `Pre-Registration Duration`, which reflects the difference between registration date and when a crime first *began*,[38] `Registration Duration` can be seen a measure of police hesitancy in registration. The median days between crime occurrence and registration is 1, with a mean of 28. However, women's cases, as well as VAW, have means of 69 and 113, respectively. In other words, a complainant may have visited a police station to register an FIR but asked to drop the case, or be forced to return at a later date.[39] Prima facie, `Pre-Registration Duration` and `Registration Duration` challenge the assumption that gendered cases are filed, "in the heat of the moment."[40] `No Record` shows 32% of VAW is cancelled at the police-level.[41]

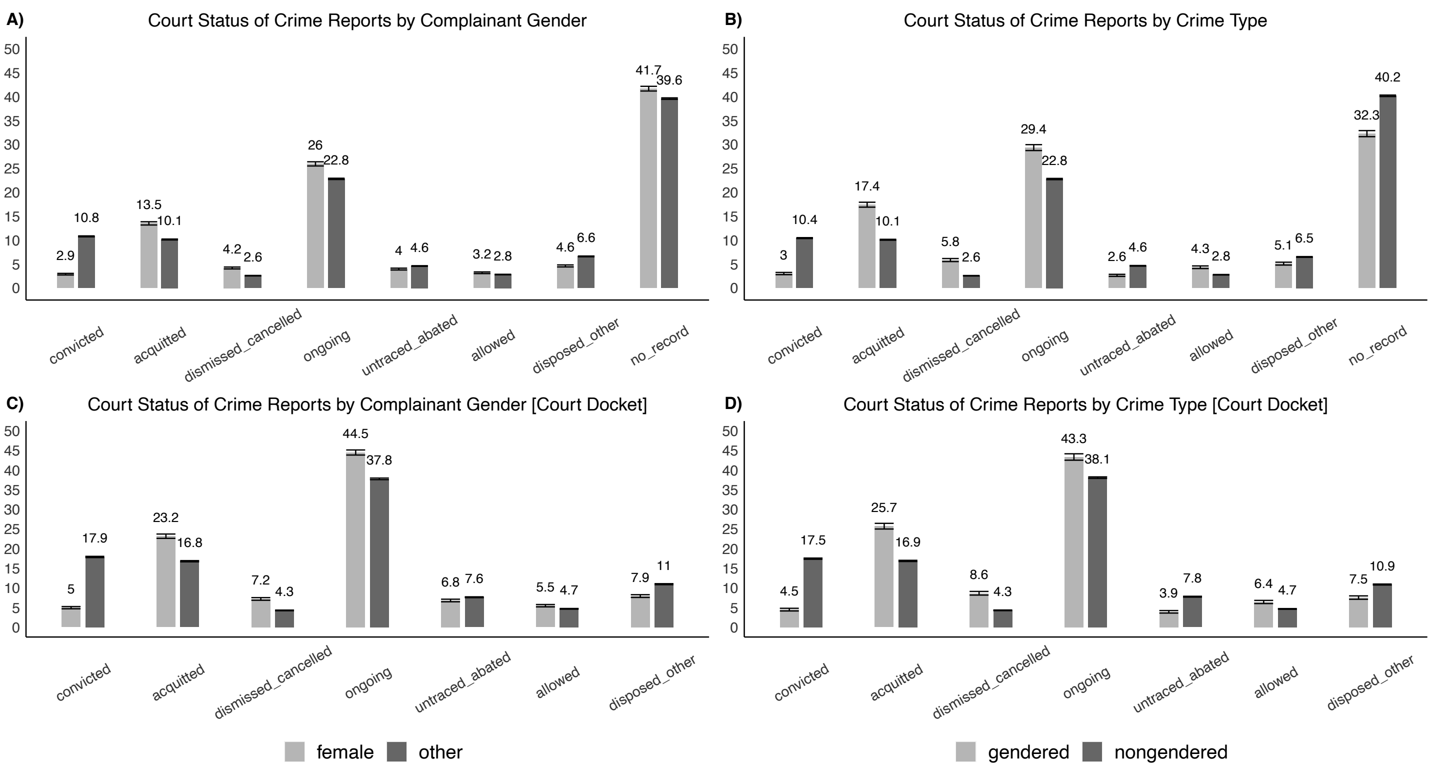
Table 2 highlights variables created post-merging. `Investigation Duration` reflects days between registration and preliminary hearing. The mean number of days spent in the judiciary (`Duration in Court`) is just under a year (336 days), with women's cases, and VAW, spending

37. Invoked in cases of sodomy; this clause was repealed from the statutes in 2018 (Jassal and Chhibber 2019).
38. Therefore potentially illustrative of how long a *complainant* waited to file a case and/or duration of abuse.
39. See Appendix Figure A4 and A5 for a graphical illustration of the inter-quartile range.
40. *Preeti Gupta & Anr. v. State of Jharkhand*, Appeal No. 1512 (Criminal Appellate Jurisdiction, 2010.
41. While it is possible certain cases have transitioned to the judiciary, the FIRs cover 2015-2018. Investigations are supposed to be carried out within 90-days, and the E-Courts database was downloaded in mid-2020. Consequently, the analyses in this study 'allow' a *two-year* window, i.e. far longer than time allotted for investigation.

longer.[42] While most cases are assigned to Judicial Magistrate 1st Class, women's cases, and VAW, are more likely to be assigned to senior judges, e.g. Addl. District Sessions Judge.

Figure 5 illustrates judicial outcomes, which fall into roughly seven categories. `Acquitted` refers to whether the suspect is absolved; `Allowed` denotes if the case entered the judiciary but a trial has not been set; `Convicted` denotes that a suspect was convicted, while `Dismissed` underscores if the case was ejected at a preliminary (or bail) hearing. `Untraced` represents whether the suspect could not be found or brought to court. The remaining outcomes are classified as `Disposed`, indicating that a decision was taken (e.g. fine issued) but further details are unavailable. The cross-tabulations in Figure 5 show that—whether as a function of all registrations (Panels A and B) or simply those in the court docket (Panels C and D)—women's complaints (as well as VAW) are more likely to be listed as on-going (stalled), dismissed, or result in a suspect's acquittal, and less likely to see a suspect sent to prison.

Figure 5: Crime Reports Statuses [Split by Complainant Gender and Crime Type]



Note: Judicial outcomes for cases (% on Y axis). Panels A and B reflect outcomes conditional on police registration. Panel A is separated by female (N=38,828) and male/other complainants (N=379,362). Panel B reflects gendered (N=20,869) and non-gendered crime (N=397,321). Panels C and D reflect outcomes conditional on entering the court docket. Panel C is separated by female (N=22,648) and male/other complainants (N=229,156), and Panel D gendered (N=14,134) and non-gendered crime (N=237,670). 95% confidence intervals included.

42. See Figure A16 for a graphical display.

Table 1: Descriptive Statistics on Select Variables: First-Information-Report (FIR) Dataset

									ALL CRIME			
	Complainant	N	Mean	SD	**Crime Type**	N	Mean	SD	N	Mean	SD	Median
Pre-Registration Duration	Female	33738	181.78	580.08	Gendered	17254	346.80	773.18	381668	49.38	310.38	1.00
	Other	347930	36.54	266.81	Nongendered	364414	35.30	261.17				
Registration Duration	Female	33766	68.94	341.50	Gendered	17269	112.88	440.69	381836	27.78	225.87	1.00
	Other	348070	23.79	210.88	Nongendered	364567	23.75	209.46				
Word Count	Female	38828	577.41	421.49	Gendered	20869	722.44	526.30	418189	452.30	257.86	381.00
	Other	379361	439.49	230.98	Nongendered	397320	438.11	226.72				
Distance	Female	36868	5.96	12.45	Gendered	19585	6.96	14.01	400345	5.52	13.83	3.00
	Other	363477	5.48	13.96	Nongendered	380760	5.45	13.82				
Female Suspects	Female	22022	0.70	1.07	Gendered	17676	0.75	1.08	220943	0.18	0.72	0.00
	Other	198921	0.12	0.65	Nongendered	203267	0.13	0.66				
Total Suspects	Female	22022	2.74	2.35	Gendered	17676	2.78	2.26	220943	2.00	2.42	1.00
	Other	198921	1.92	2.41	Nongendered	203267	1.93	2.42				
Victim Age	Female	17953	35.93	10.39	Gendered	9131	34.35	10.29	192939	38.40	9.07	38.00
	Other	174986	38.65	8.88	Nongendered	183808	38.60	8.96				
No. of Sections	Female	38828	2.57	1.59	Gendered	20869	3.31	1.65	418190	2.11	1.42	2.00
	Other	379362	2.07	1.39	Nongendered	397321	2.05	1.37				
Urban	Female	38141	0.60	0.49	Gendered	20028	0.53	0.50	417322	0.59	0.49	1.00
	Other	379181	0.59	0.49	Nongendered	397294	0.59	0.49				
Hours Waited at PS	Female	38690	7.51	62.08	Gendered	20775	9.32	79.25	416045	7.06	52.72	0.68
	Other	377355	7.01	51.67	Nongendered	395270	6.94	50.95				
Hour Registered	Female	38828	17.37	4.80	Gendered	20869	17.00	5.07	418190	17.20	5.32	19.00
	Other	379362	17.19	5.38	Nongendered	397321	17.21	5.34				
Hour Arrived	Female	38828	16.50	4.83	Gendered	20869	16.14	5.11	418190	16.35	5.43	18.00
	Other	379362	16.34	5.48	Nongendered	397321	16.36	5.44				
R:Head Constable	Female	36959	0.29	0.46	Gendered	19621	0.16	0.36	400086	0.43	0.49	0.00
	Other	363127	0.44	0.50	Nongendered	380465	0.44	0.50				
R:Ass. Sub-Inspector	Female	36959	0.52	0.50	Gendered	19621	0.58	0.49	400086	0.44	0.50	0.00
	Other	363127	0.43	0.50	Nongendered	380465	0.44	0.50				
R:Sub-Inspector	Female	36959	0.16	0.36	Gendered	19621	0.22	0.41	400086	0.10	0.30	0.00
	Other	363127	0.10	0.30	Nongendered	380465	0.10	0.30				
R:Inspector	Female	36959	0.03	0.17	Gendered	19621	0.04	0.21	400086	0.02	0.16	0.00
	Other	363127	0.02	0.15	Nongendered	380465	0.02	0.15				
No Record/Not Sent to Court	Female	38828	0.42	0.49	Gendered	20869	0.32	0.47	418190	0.40	0.49	0.00
	Other	379362	0.40	0.49	Nongendered	397321	0.40	0.49				

Note: Descriptive statistics for variables in the FIR dataset, split by female/other complainants, as well as gendered/nongendered crime. The term 'Other' is used because a small fraction of cases may be brought forward by organizations or institutions rather than individuals. Gendered crime may be brought forward by male or female complainants.

Table 2: Descriptive Statistics: First-Information-Report Dataset Merged With Court Records

									ALL CRIME			
	Complainant	N	Mean	SD	**Crime Type**	N	Mean	SD	N	Mean	SD	Median
Investigation Duration	Female	22471	133.77	206.57	Gendered	14007	113.66	185.91	248920	127.95	204.38	54.71
	Other	226449	127.38	204.15	Nongendered	234913	128.81	205.40				
Dismissed	Female	22648	0.07	0.26	Gendered	14134	0.09	0.28	251804	0.05	0.21	0.00
	Other	229156	0.04	0.20	Nongendered	237670	0.04	0.20				
Ongoing	Female	22648	0.44	0.50	Gendered	14134	0.43	0.50	251804	0.38	0.49	0.00
	Other	229156	0.38	0.48	Nongendered	237670	0.38	0.49				
Acquitted	Female	22648	0.23	0.42	Gendered	14134	0.26	0.44	251804	0.17	0.38	0.00
	Other	229156	0.17	0.37	Nongendered	237670	0.17	0.37				
Convicted	Female	22648	0.05	0.22	Gendered	14134	0.04	0.21	251804	0.17	0.37	0.00
	Other	229156	0.18	0.38	Nongendered	237670	0.17	0.38				
Duration in Court	Female	22522	377.37	368.07	Gendered	14120	378.43	362.50	250287	336.18	365.50	205.00
	Other	227765	332.10	364.99	Nongendered	236167	333.65	365.52				
No. of Hearings	Female	20077	9.82	9.04	Gendered	12852	10.41	9.49	195480	9.84	9.15	7.00
	Other	175403	9.84	9.17	Nongendered	182628	9.80	9.13				
R:Civil Judge Junior Division	Female	22634	0.06	0.25	Gendered	14124	0.06	0.24	251629	0.07	0.25	0.00
	Other	228995	0.07	0.25	Nongendered	237505	0.07	0.25				
R:Judicial Magistrate 1st Class	Female	22634	0.43	0.50	Gendered	14124	0.39	0.49	251629	0.46	0.50	0.00
	Other	228995	0.47	0.50	Nongendered	237505	0.47	0.50				
R:Sub-Divis. Judicial Magistrate	Female	22634	0.08	0.27	Gendered	14124	0.07	0.26	251629	0.09	0.29	0.00
	Other	228995	0.09	0.29	Nongendered	237505	0.09	0.29				
R:Addl. Chief Judicial Magistrate	Female	22634	0.09	0.29	Gendered	14124	0.08	0.26	251629	0.11	0.31	0.00
	Other	228995	0.11	0.31	Nongendered	237505	0.11	0.31				
R:Chief Judicial Magistrate	Female	22634	0.13	0.33	Gendered	14124	0.09	0.29	251629	0.14	0.35	0.00
	Other	228995	0.14	0.35	Nongendered	237505	0.14	0.35				
R:Addl. District Sessions Judge	Female	22634	0.17	0.37	Gendered	14124	0.29	0.45	251629	0.11	0.31	0.00
	Other	228995	0.10	0.30	Nongendered	237505	0.10	0.30				
R:District Sessions Judge	Female	22634	0.03	0.16	Gendered	14124	0.01	0.10	251629	0.02	0.14	0.00
	Other	228995	0.02	0.13	Nongendered	237505	0.02	0.14				
Duration in CJ System	Female	22492	573.19	383.60	Gendered	14110	568.78	381.88	249462	508.71	392.22	435.71
	Other	226970	502.32	392.49	Nongendered	235352	505.11	392.54				

Note: Descriptives statistics for select variables in merged dataset of crime and judicial records, split by female and other complainants, as well as gendered and non-gendered crime. The term 'Other' is used because a small fraction of cases may be brought forward by organizations or institutions rather than individuals. Gendered crime may be brought forward by male or female complainants.

OLS Results

Female Complainants and VAW

Table 3 tests hypotheses outlined in Level 1 (Figure 3). Columns 1-2 show that women's cases have a lag of over a month between incident and registration (significantly longer than the baseline of 24 days). In columns 5-6, when interacting `Female` with an indicator for a case invoking a gendered Penal Code, the gap increases. Put differently, in non-gendered contexts, the gap between crime occurrence and registration is a week longer for women; this gap exceeds 100 days when complaints involve VAW. While this may be reflective of hesitancy in reporting, at the node when cases have not formally entered the books, the police has discretion in forwarding complainants to counseling centers or asking citizens to return later to avoid registration.

Columns 7-8 of Table 3 reveal that women's cases are significantly less likely than men's to be sent to court. However, this does *not* apply to VAW. Conditional on registration, cases of VAW are 7-8% *more* likely to be sent to the judiciary than non-gendered crime. Police officers are bound by rules to ensure (registered) cases of VAW transition or are investigated quickly. For instance, in columns 3-4 of Table 4, cases of VAW are investigated, on average, roughly two-weeks sooner than non-gendered crime (compared to a baseline of 128 days). Columns 6-7 reveal that it is women's *non-gendered* complaints for which investigations are ≈20-days slower.

Figure 6 presents average marginal effects in an easy-to-interpret plot. Panel A suggests that cases of VAW (brought forward by female complainants) have the longest lag between incident and registration. Nevertheless, cases of VAW are, conditional on registration, allowed to pass through the early stages (Panel B and C). At the police-level, gender imbalances for registered cases largely hold in non-gendered contexts, settings where officers are bound by fewer rules.

Table 3: Process and Outcomes: Level 1

	Registration Duration						Cancelled After Registration					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female	45.148***	40.778***			9.100***	6.858***	0.021***	0.025***			0.068***	0.063***
	(6.353)	(7.248)			(2.263)	(2.494)	(0.008)	(0.008)			(0.010)	(0.009)
VAW			89.135***	85.401***	27.459***	22.906**			-0.079***	-0.065***	-0.046***	-0.034**
			(13.777)	(16.049)	(7.740)	(8.986)			(0.014)	(0.011)	(0.017)	(0.014)
Female:VAW					111.601***	117.971***					-0.120***	-0.114***
					(17.815)	(18.085)					(0.018)	(0.019)
Constant	23.788***	8.933***	23.749***	6.915**	23.128***	5.133*	0.396***	0.338***	0.402***	0.397***	0.397***	0.392***
	(2.385)	(2.901)	(2.344)	(2.994)	(2.350)	(2.801)	(0.018)	(0.013)	(0.018)	0.013)	(0.018)	(0.013)
Obs.	381,836	360,022	381,836	360,022	381,836	360,022	418,190	382,265	418,190	382,265	418,190	382,265
R^2	0.003	0.015	0.007	0.019	0.010	0.022	0.0002	0.111	0.001	0.112	0.003	0.113
Controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
PS FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Month Yr FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y

Note: Controls include a numeric variable for distance of crime from station, investigator rank, and urban. Standard errors clustered by district. *p<0.1; **p<0.05; ***p<0.01

Nonetheless, this dynamic changes by the time of the first hearing in court. At this node, having entered the purview of judges where there are few constraints on administrators, triage becomes even more apparent. Columns 7-12 in Table 4 suggest that women's cases—whether in non-gendered *or* gendered contexts—begin to yield negative outcomes for the complainant.

Figure 6 expresses this in Panel D. Specifically, even though women's cases in non-gendered contexts are 1-2% more likely to be dismissed than related cases brought forward by men (compared to a baseline of 4%), this gap *persists* for VAW.

Table 4: Process and Outcomes: Level 2

	Investigation Duration						Court Dismissal					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female	6.390	9.015***			18.875***	19.507***	0.029***	0.014***			0.024***	0.013***
	(4.841)	(3.278)			(4.159)	(2.550)	(0.003)	(0.004)			(0.003)	(0.003)
VAW			−15.146**	−11.009*	−9.306	−4.214			0.043***	0.010	0.049***	0.004
			(6.946)	(5.880)	(7.090)	(6.004)			(0.006)	(0.006)	(0.007)	(0.007)
Female:VAW					−27.209***	−29.353***					−0.032***	−0.002
					(6.155)	(6.700)					(0.006)	(0.007)
Constant	127.378***	116.458***	128.807***	118.368***	127.631***	117.171***	0.043***	0.007	0.043***	0.008	0.041***	0.007
	(5.926)	(16.774)	(5.924)	(16.800)	(6.010)	(16.736)	(0.003)	(0.008)	(0.003)	(0.008)	(0.003)	(0.008)
Obs.	248,920	227,315	248,920	227,315	248,920	227,315	251,804	229,954	251,804	229,954	251,804	229,954
R^2	0.0001	0.069	0.0003	0.069	0.001	0.070	0.002	0.084	0.002	0.083	0.003	0.084
Controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
PS FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Month-Yr FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y

Note: Controls include a numeric variable for distance of crime from station, investigator rank, judge rank, and urban. Standard errors clustered by district. *p<0.1; **p<0.05; ***p<0.01

Columns 1-2 in Table 5 reveal that—for complaints that survive the node of preliminary court dismissal—women's cases spend longer in the judiciary by over a month (compared to a baseline of just under a year). Graphically, Panel E in Figure 6 shows that cases of VAW brought forward by women spend the *longest* time stalled (≈390 days), regardless of whether a verdict was issued.

To investigate whether punitive justice was ultimately meted out, I pay attention to conviction and acquittal. Columns 5-6 in Table 6 demonstrate that cases brought forward by women in non-gendered contexts are 5-6% more likely to result in suspect acquittal (from a baseline of 17%), a figure which is pulled higher if it involves VAW. Women's cases are associated with 10-13% fewer convictions of suspects compared to a baseline of 18% (columns 7-8). Figure 6 summarizes the findings where, in Panel F, we see conviction rates for men who register VAW (e.g. for family or friends) drop from their non-gendered base, but *not* to the same level as women who have only a 7-10% chance of a suspect being convicted in *either* category. The results largely hold when including dummies for over a *thousand* primary Penal Codes (Appendix Table A2).

Heterogeneous Effects Across Gendered Crime

VAW is a broad category. It is plausible that violence perpetrated by a spouse, family, or in-laws would be most likely to be triaged. Consequently, I disaggregate VAW into the four common case types: (a) dowry harassment, (b) female kidnapping, (c) criminal force, and (d) rape.[43] Appendix Table A3 suggests that cases of female kidnapping and "criminal force" are registered

43. These Penal Codes have the least overlap between them, providing variation in gendered crime registered. Dowry (Section 498-A) always involves the spouse or extended family, but this does not apply to rape (Section 376) which is stamped when a non-spouse commits assault. Female kidnappings (Section 366) are usually registered by family/relatives of the complainant rather than the primary victim.

Table 5: Process: Level 3

	Duration in Court					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	45.264***	40.953***			27.042**	33.423***
	(11.042)	(8.827)			(12.088)	(10.229)
VAW			44.781***	33.761***	4.677	9.866
			(10.738)	(10.885)	(10.194)	(11.084)
Female:VAW					47.649***	14.743
					(8.335)	(10.695)
Constant	332.103***	548.265***	333.650***	550.261***	331.976***	547.495***
	(12.660)	(32.900)	(12.347)	(32.835)	(12.755)	(32.800)
Obs.	250,287	228,542	250,287	228,542	250,287	228,542
R^2	0.001	0.200	0.001	0.200	0.002	0.200
Controls	N	Y	N	Y	N	Y
PS FE	N	Y	N	Y	N	Y
Month-Yr FE	N	Y	N	Y	N	Y

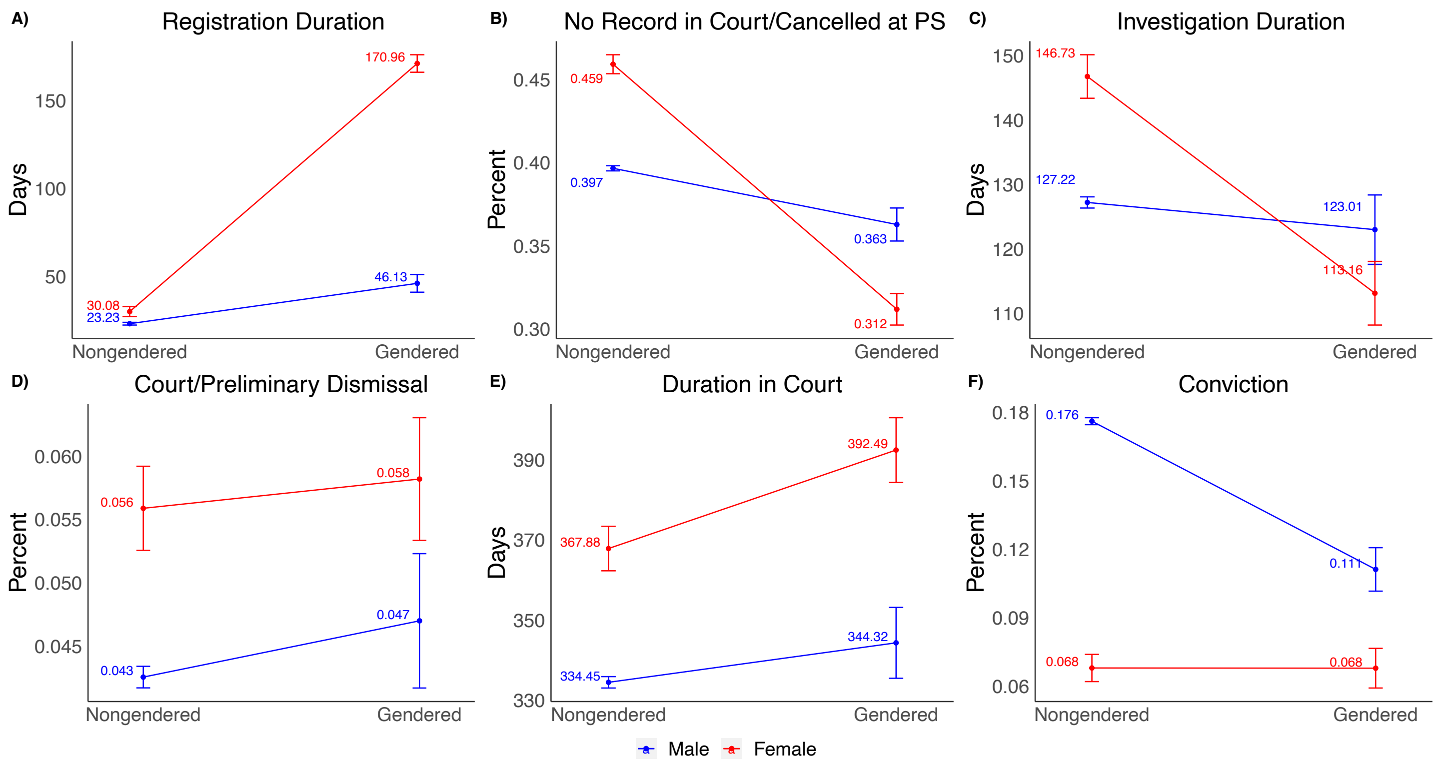
Note: Controls include a numeric variable for distance of crime from station, investigator rank, judge rank, and urban. Standard errors clustered by district. *p<0.1; **p<0.05; ***p<0.01

Table 6: Outcomes: Level 3

	Acquittal						Conviction					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female	0.064***	0.055***			0.056***	0.054***	−0.129***	−0.106***			−0.123***	−0.108***
	(0.009)	(0.007)			(0.008)	(0.006)	(0.012)	(0.010)			(0.011)	(0.010)
VAW			0.088***	0.068***	0.100***	0.080***			−0.130***	−0.081***	−0.122***	−0.065***
			(0.012)	(0.008)	(0.010)	(0.008)			(0.014)	(0.012)	(0.014)	(0.012)
Female:VAW					−0.071***	−0.067***					0.095***	0.065***
					(0.010)	(0.009)					(0.011)	(0.011)
Constant	0.168***	0.393***	0.169***	0.394***	0.165***	0.391***	0.179***	0.267***	0.175***	0.261***	0.182***	0.269***
	(0.017)	(0.018)	(0.017)	(0.018)	(0.017)	(0.018)	(0.014)	(0.028)	(0.014)	(0.028)	(0.015)	(0.028)
Obs.	251,804	229,954	251,804	229,954	251,804	229,954	251,804	229,954	251,804	229,954	251,804	229,954
R^2	0.002	0.124	0.003	0.124	0.004	0.125	0.010	0.101	0.006	0.097	0.012	0.102
Controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
PS FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Month Yr FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y

Note: Controls include a numeric variable for distance of crime from station, investigator rank, judge rank, and urban. Standard errors clustered by district. *p<0.1; **p<0.05; ***p<0.01

Figure 6: Average Marginal Effects for Interactions



Note: Marginal effects based on regressions in columns 7 or 14 in Tables 3-6. All models include controls, month-year, and police station fixed effects. Standard errors clustered by district.

sooner than the baseline, with rape registered around the same time as the average non-gendered case. Dowry/domestic violence is the exception: the lag between the incident and registration can exceed 270 days, and almost a year if the complainant is a woman (Appendix Figure A28), providing suggestive evidence that law enforcement may have initially delayed or diverted complainants.[44] Appendix Table A4 (Figure A29) show that, conditional on registration, VAW is more likely to appear in court records than non-gendered cases, while Table A5 (and Figure A30) illustrate that VAW—except female kidnapping[45]—are investigated (relatively) quickly.

Nevertheless, by the preliminary hearing, VAW, especially dowry harassment, begin to be dismissed at high rates. If complaints happen to cross this node, all four types of VAW spend significantly longer stalled (Table A6). Dowry/domestic violence is among the least likely case to result in conviction (0.7%), comparable to culpable homicide (e.g. rash driving) and real estate

44. This validates the use of `Registration Duration` as a measure of police reluctance in registration; if it only reflected women's anxiety in coming forward, we should also have seen similar lags for rape or "criminal force."

45. Interviews with Haryana police suggest that a large proportion of cases involving Section 366 involve girls, 14-18, who allegedly ran away with partners. Officers believe these cases are not bona fide kidnapping but instead teenagers "rebelling" in conservative settings where there are restrictions on women's mobility. These cases are registered by family members of the victim. One policewoman explained, "Parents refuse to accept that [a woman fell in love] and get an FIR against the boy... As per law, a minor's consent is not consent even if given voluntarily, and thus once police trace the couple or they come back on their own, we get the girl's statement recorded. Many a times, they allege forceful abduction and rape under the coercion of family members...police remain less interested in such crimes. However, they're more responsive if, say, a girl below 10-12 years is missing... According to the *Khap* [village council] rules, girls are forbidden to marry in same *gotras* and nearby villages; apparently all are considered brothers and sisters in a village. Hence, young girls feel compelled to break free, desires which have only been amplified with technology and internet." Personal interview, Crimes Against Women Desk, Haryana.

disputes (Appendix Figure A24 and A21). As Table A6 and Figure A34 demonstrate, the initial variation in how VAW is accommodated at the police-level dissipates such that the sub-types begin yielding higher acquittal rates,[46] and lower convictions (with rape as an exception).[47]

The coefficient on `Female` remains significant in every single model. Further, triage appears most extreme in the mid- to late-stages of justice delivery, "the last mile" at which complaints (considered serious to have been registered/investigated) have spent effort to reach later stages.

Text-As-Data: Structural Topic Modeling

Aside from the usual caveats associated with OLS, there are two challenges. First, categorizations of crime have hitherto relied on Penal Codes. Second, even if we accept that there is a striking gender imbalance, perhaps female complainants *are* more likely to register cases without merit, which the criminal justice system happens to be efficiently weeding out.

To investigate, I apply unsupervised machine learning on victims' testimonies. The technique precludes myself or the administrator (e.g., the officer who stamped Penal Codes) from inserting themselves into the research. Topic modeling estimates relationships between meta-data and topics from the corpus (Roberts, Stewart, and Tingley 2019),[48] thereby facilitating hypothesis testing. Are there, for instance, particular *topics* within the testimonies—including those generally associated with female complainants—that yield lower conviction rates?[49]

As highlighted in Table 1, complaints brought forward by women are longer (VAW has a mean word count of 722).[50] Appendix Figures A37 highlights the kinds of topics that emerge from the *entire* corpus. For women, Figure 7 presents the highest probability as well as FREX (frequent and exclusive) words. Among the top topics that emerge from women's cases involve "fighting" (Topic 14), usually domestic violence. The word clouds for this topic in Appendix Figure A46-A48 underscore terms such as: *wife, hospital, kill, beaten, domest, husband, hurt, blunt.* The kind of theft that female complainants often register is distinct from those associated with men; for women, the most common form of theft is "chain-snatching" (Topic 15), as opposed to auto-theft for men (Topic 22 in Appendix Figure A37).

Figure 8 presents two visualizations. Panel A is a STM of women's complaints with an indicator for conviction as a predictor. Non-gendered cases such as "cheating," "chain-snatching," or "public intoxication" yield better outcomes. Panel B shows correlations (when topics are likely to co-occur within an FIR). Cases involving dowry are clustered at the bottom, with other forms of gendered crime (e.g. rape, domestic violence, and "criminal force") immediately above, suggesting overlap in the kinds of abuse perpetrated in and out of the household.[51]

46. In Appendix Figure A7, *five* of the top ten Penal Codes that have the longest gap between incident and registration are gendered, with dowry being the *most* delayed case (Appendix Figure A6, Figure A14 and A15).

47. Appendix Figure A21 highlights that, while cases of child sexual assault and dowry death have higher conviction percentages (10-17%), cases where a female victim is not alleged to have been raped (by a non-spouse), or not perceived to be grievously injured, have lower conviction rates (e.g. "word or acts intended to insult the modesty of women" (1.3%), and sexual harassment (3.4%). Also see Appendix Figure A21-A22).

48. The method uses the 'bag of words' assumption where each document is a vector containing the count of a word type without reference to order. The resulting Document Term Matrix (DTM) is one where a row represents a document, and a column represents a word (Lucas et al. 2015; Grimmer and Stewart 2013).

49. I utilize the universe of FIRs, and create indicators for whether they eventually resulted in conviction or acquittal (as opposed to analyzing only those in the court docket).

50. See Appendix Figure A36 for a graphical visualization of the spread.

51. Other clusters include cases involving finances, e.g. phishing, real estate and development disputes.

Figure 9 breaks down VAW. Topics range from the extortion of women with compromising photographs/videos (Topic 18) to "trafficking" or being sold into prostitution (Topic 12). While topics involving abuse inside the household appear to be unlikely to result in formal punishment (e.g. dowry), cases involving child abuse and rape have better outcomes (vis-à-vis conviction) (Figure 10). Still, both forms of VAW—*in* or *out* of the household—are likely to yield high rates of acquittal and dismissal (Appendix Figure A45), supporting the OLS analyses.

A theme that emerges from the STM exercise is the prioritization of sons over daughters. Specifically, Topic 7 refers to abandoning or killing babies ("killing the girl child"), Topic 14 refers to (illegal) sex selective diagnostic technologies, and Topic 5 includes unlicensed doctors performing abortions. As highlighted in the word clouds of the Appendix (Figures A51-A53), common words in these categories include: *children, child, medic, drug, abort, kill, patient, ultrasound, pregnant.* A number of inter-correlated topics involve dowry (Topics 6, 23, 13, 9, and 5) in Figure 10. Appendix Figures A46-A48 shows that common words include: *dowry, tortur, parent, cash, daughter, greed, kill, demand, cruelti, in-law, assault.* "Mother-in-law" appears repeatedly, indicating that abuse perpetrated against the victim invariably involves the in-laws as opposed to just an intimate partner.

When disaggregating dowry, the machine is able to separate abuse relating to mental and physical abuse (Topics 1 and 2) from others involving, for instance, violence perpetrated when a victim is pregnant (Topic 3). Topics 6 involves harassment in conjunction with spousal rape; this can be seen in the FREX words of Panel B of Figure 9 that accentuate terms such as *unnatur* (or anal) and *sexual.* Topic 16 is illustrative of FIRs in which complainants explain that they tried to register a case before but were instead asked to reconcile (Jassal 2020). Topics 19 and 20 refer to abusers either deserting their wives or absconding (so as to extract dowry from another victim), and Topic 20 represents cases where suspects starve their wives for extortion. While cases related to rape (by a non-spouse) have a better likelihood of being disciplined (Topic 10), when similar acts are perpetrated by family (Topic 6), triage by the criminal justice system becomes more apparent. The only type of dowry-related abuse that *is* associated with higher levels of conviction is Topic 9, i.e. when harassment has culminated in either a victim's suicide or killing (equivalent to murder).

Figure 7: Top Topics (Female Complainants, N=38,828)

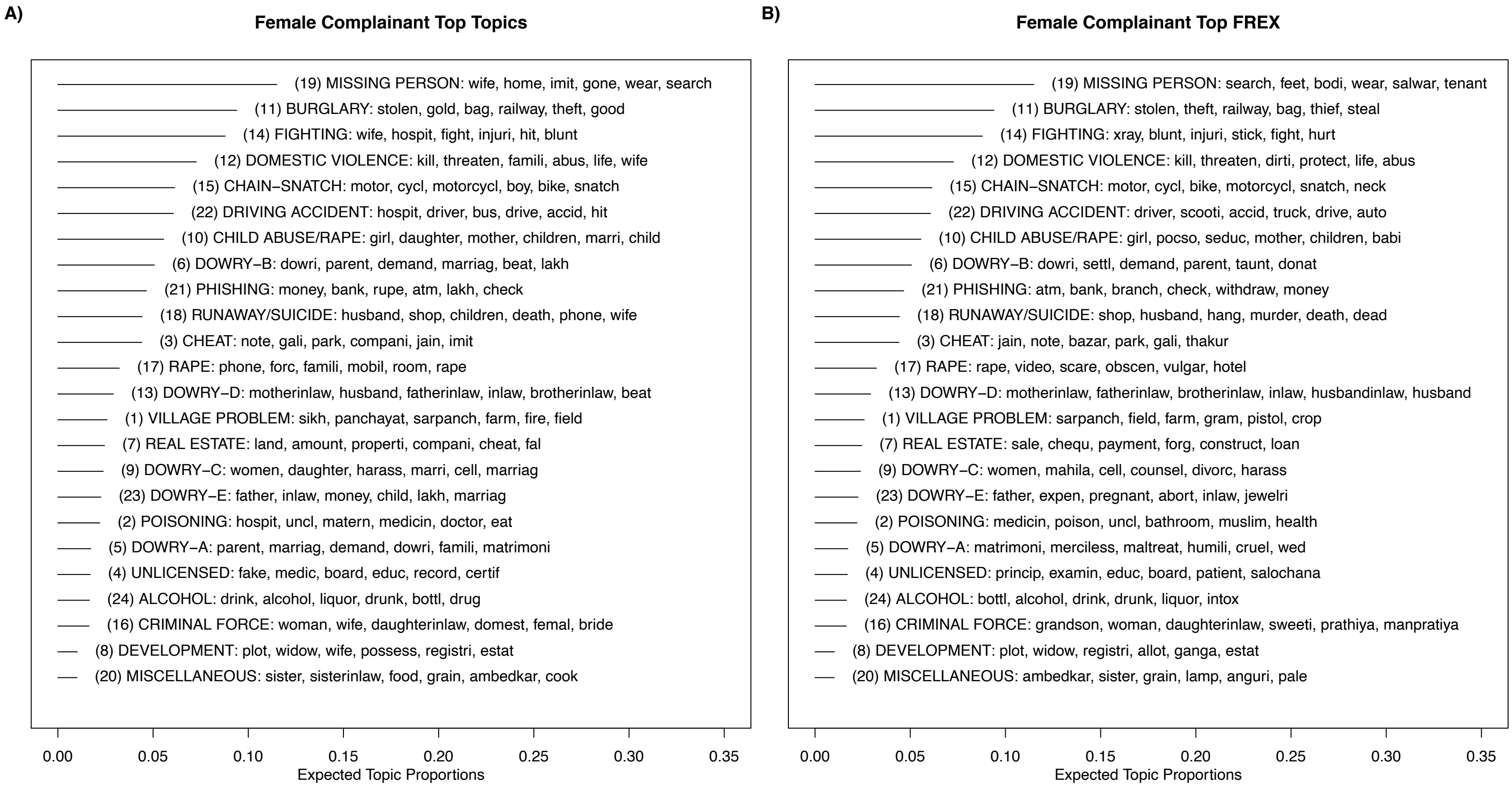
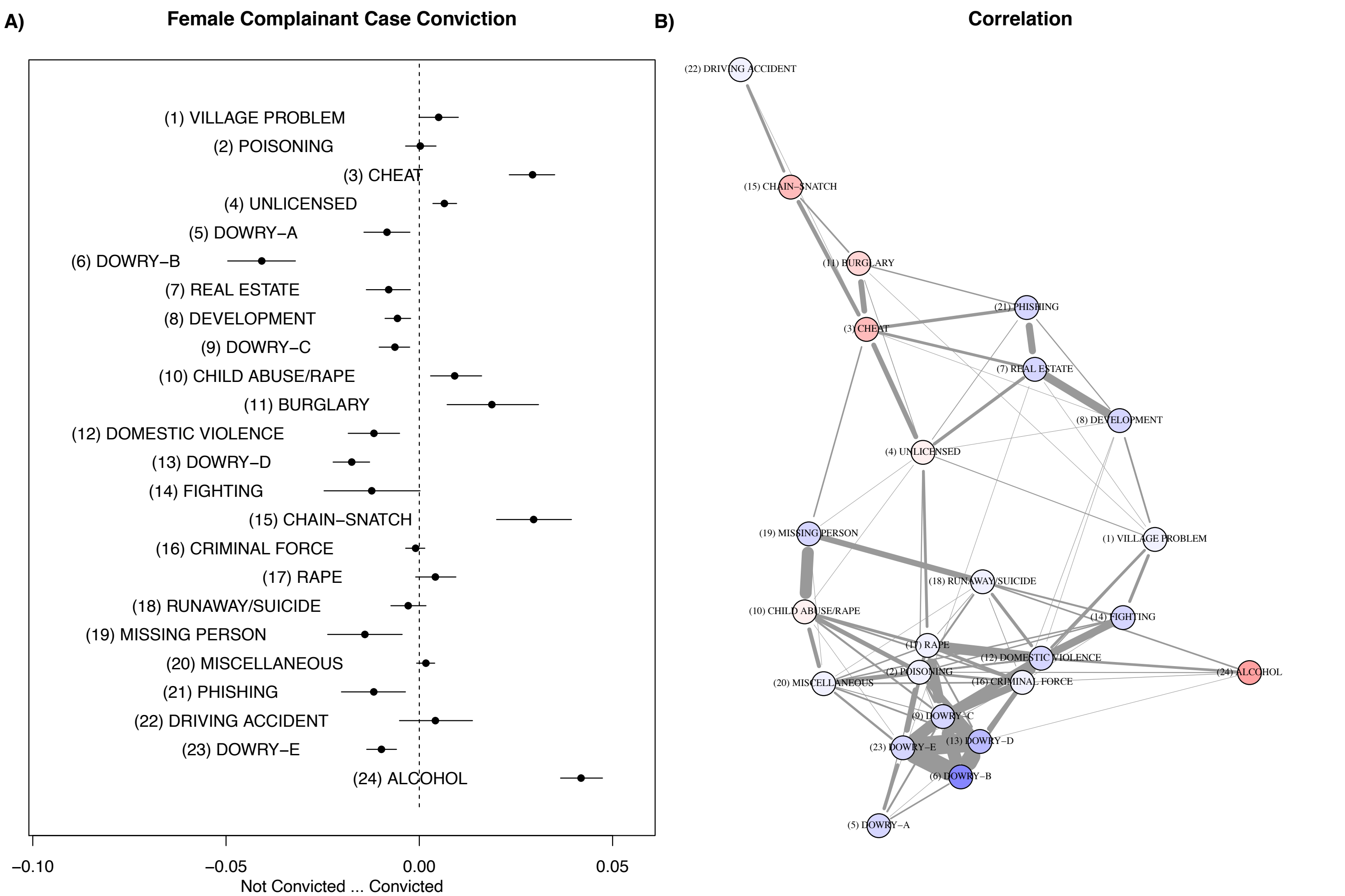


Figure 8: Conviction Rate and Correlation of Topics Associated with Women's Cases



A) STM with binary indicator for conviction. *B)* Topic correlations and magnitude of regression coefficients.

Figure 9: Top Topics (Gendered Crime, N=20,869)

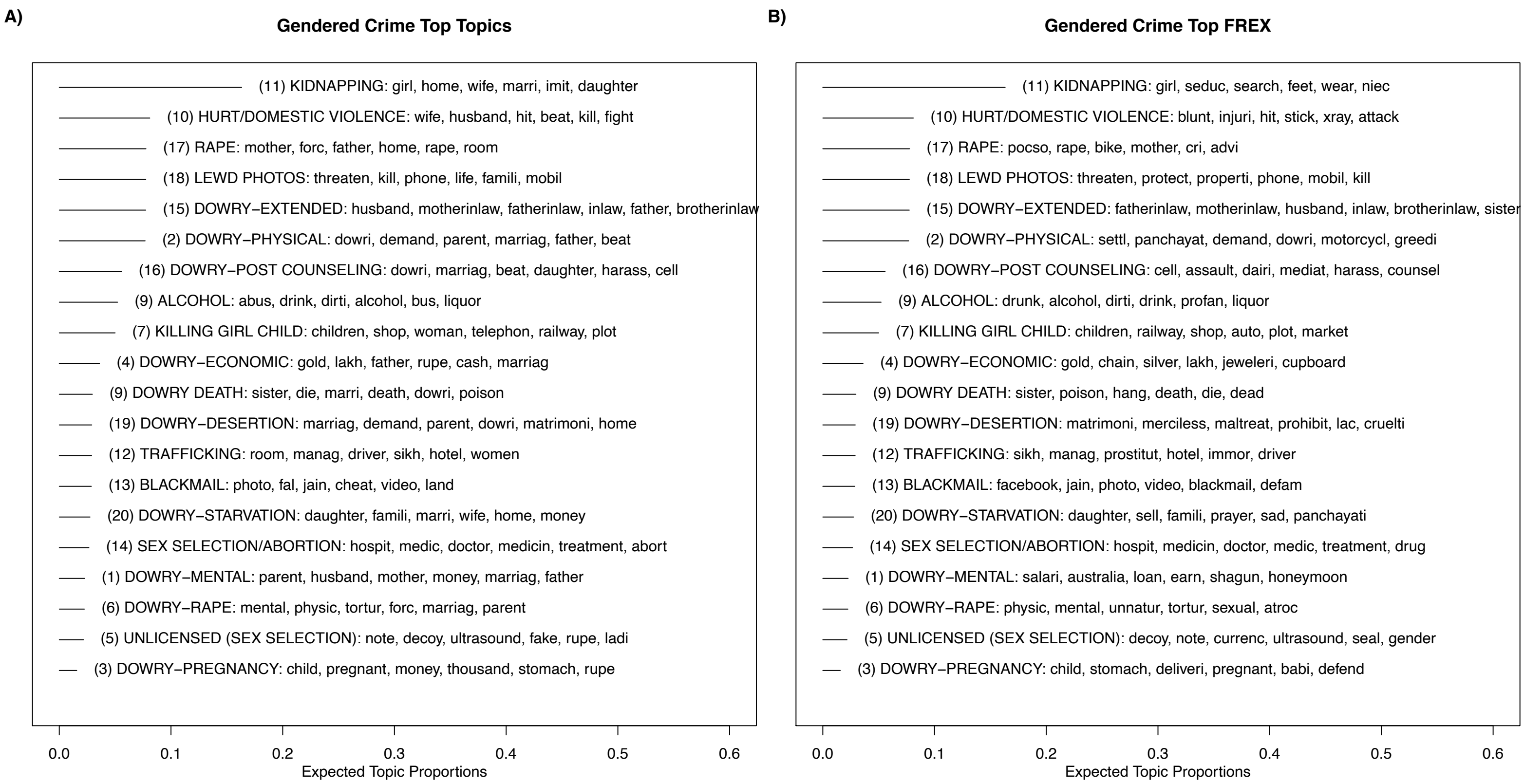
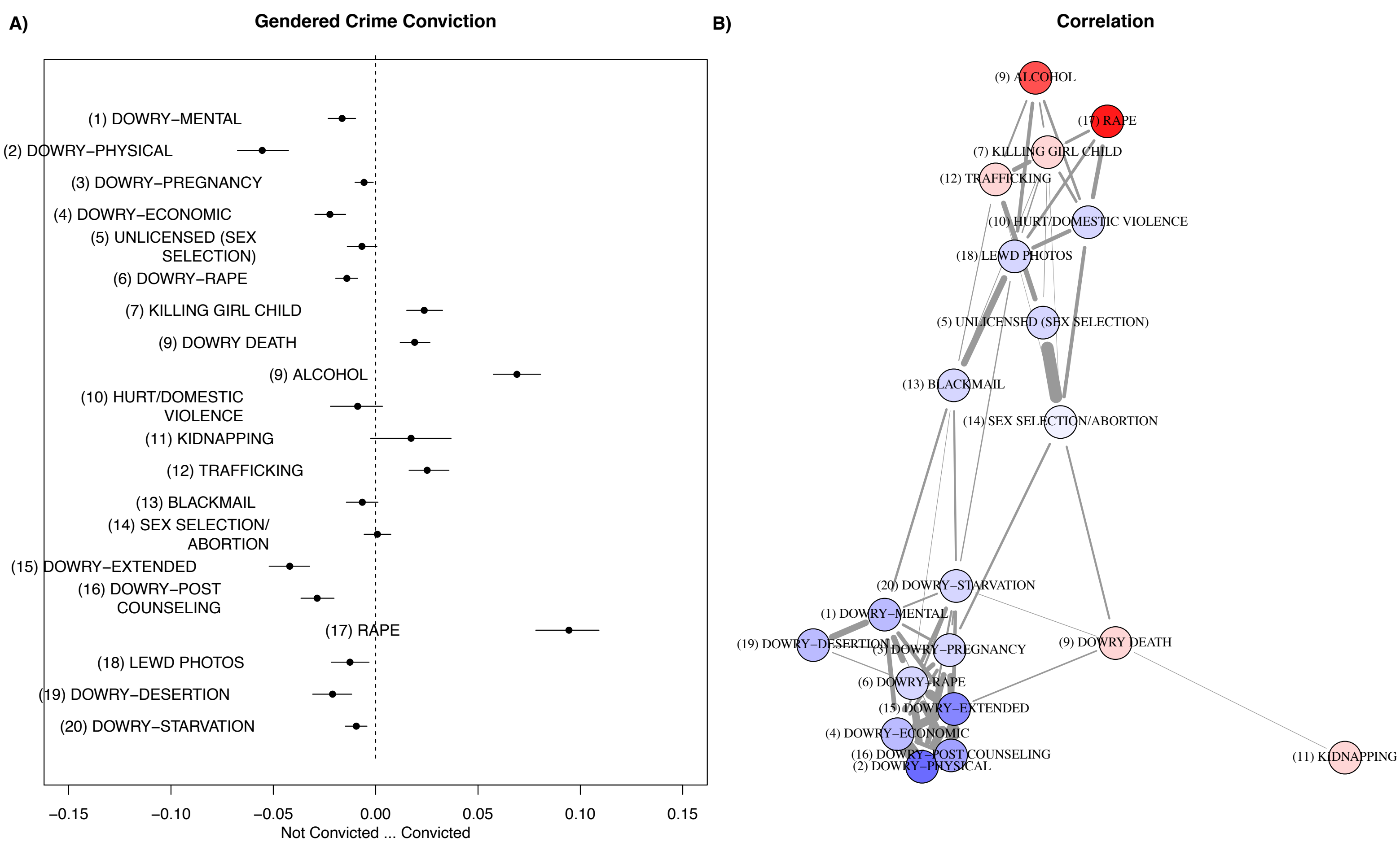


Figure 10: Conviction Rate and Correlation of Topics Associated with Gendered Crime



A) STM with binary indicator for conviction. *B)* Topic correlations and magnitude of regression coefficients.

Text-as-Data: Topical Inverse Regression Matching

The methods used thus far cannot fully account inframarginality, i.e. case-types between complainant gender may be distinct. While an imagined experiment would be to randomly assign individuals to crime, a realistic approach is to leverage the data to match cases on textual (and non-textual) dimensions.[52] Then, after qualitatively ensuring that the technique correctly matched cases (Grimmer and Stewart 2013), compare outcomes.

I use the entire corpus of registered FIRs for topical inverse regression matching (TIRM) introduced by Roberts, Stewart, and Nielsen (2020).[53] Figure 11 is the first balance-test. The grey bars—which highlight the difference between female minus male complainants in the unmatched data—reveal stark differences. Women are more likely to discuss dowry violence (Topic 24), whereas men cases of bootlegging or drunkenness (Topic 4). There are topics that affect both equally, e.g. Topic 13 ("cheating"). Figure 11 shows that while projection matching somewhat improves balance, TIRM is more successful in minimizing differences, similar to topic matching only (despite also balancing on propensity scores).

As a second test, I randomly select and present 12 matched testimonies in Table 7. This is a hard test for balance, and adds a qualitative component to the study. It is a hard test because the machine matched cases *without any* reference to Penal Codes; and still, after TIRM, we see similarities in the Codes simply based on content. In fact, the machine is more successful at categorizations than police officers.[54] [[An outgrowth of this research is that administrators may now be able to use machine algorithms to ensure correct Penal Codes are being utilized, instead of relying on officers' discretion, who may use memory or manuals to classify crimes, potentially "under-weighting" the seriousness of cases or making mistakes. An online tool, called the `Indian-Penal-Code Classifier` under development at Stanford University may (a) ensure accurate charging decisions are applied, and (b) reduce the cognitive load for officers.]]

In rows 2, 3 and 6 of Table 7, we see generic cases registered by either a male or female complainant [identifying information censored]. Row 2 depicts scooter theft, and row 3 a hit-and-run. In the cases of hit-and-run, the machine correctly matched cases not only based on the fact that a crash occurred, but also that the complainants recognize the suspect. Still, despite being topically similar, there remain dissimilarities that the machine cannot (and should not) perfectly match on; for instance, in row 7, the treated and control group involve confidence-tricksters, but the type of con is distinct. The treatment group in the dowry murder case involves the killing of a wife, but in the control condition a wife *and* her child have been found dead.

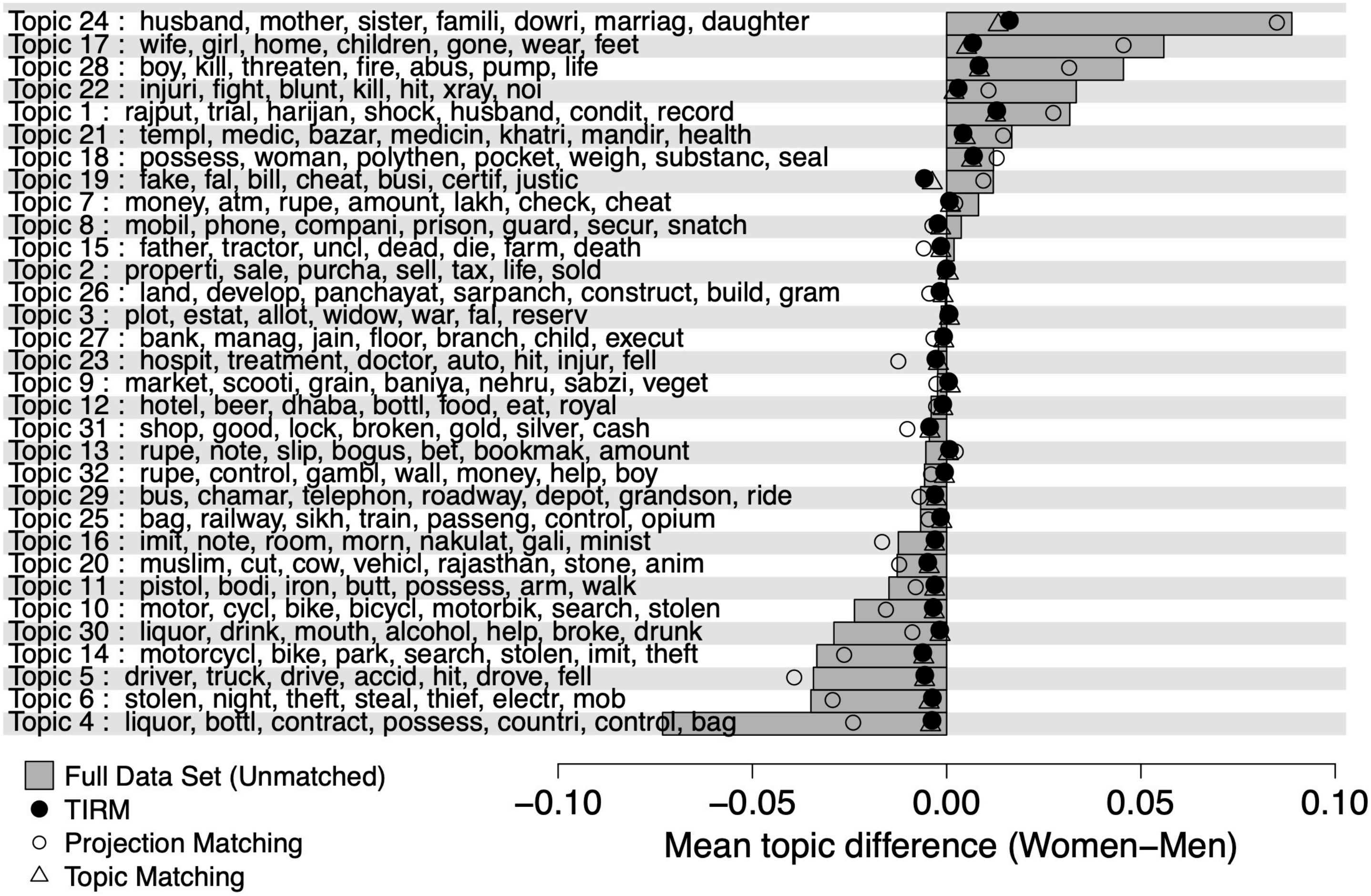
The language in rows 1,4, and 5 is rich, and allows for a brief interpretative exercise. In row 1, we see (relatively less violent) dowry cases wherein victims have been extorted and beaten. Consider the way in which class is foregrounded. In the control group of row 1, the father—who is registering a case on behalf of his child—notes that his daughter is well-educated. The complainant in the treatment group is registering a case against a lawyer and judge, which suggests not only that the perpetrators have influence, but also that they are well-educated; and yet, the suspects allegedly believe they are owed luxury vehicles in view of their "status." Similarly, in row 4, the complainant in the treatment group notes that the in-laws (in likely an

52. I view matching as an additional test rather than a preferred analysis, since it rests on certain assumptions (Sekhon 2009). One also has to consider immpanipulable categories like gender as a "treatment" (Neil and Winship 2019), and potentially minimize the greater hurdles for disadvantaged groups for having come forward.

53. If matched only on propensity scores, treated/non-treated cases may not be topically similar, e.g. extortion might be lumped with bag-snatching because, say, they have equal probability of being registered by women.

54. E.g., Table 7 (row 3), the officer did not attach Section 338 as may have been warranted based on testimony.

Figure 11: Balance Check 1



Note: Balancing estimated topics, and comparison of TIRM with full data set and other matching techniques.

arranged marriage) had been given material goods in accordance "with their status."[55] A puzzle arises as to how justice would vary across these contexts; would the system provide re-distributive justice (financial compensation), especially for losses in the dowry and cheating cases?

Particularly striking in the treatment group of row 5 is that the perpetrators previously went to prison. This raises concerns about the type of punishment that led to the predictable killing of a woman despite the glaring warnings. The reports shed light on criminal impunity, where individuals may be abducted from families in broad daylight, or killed in defiance of the authorities. Many victims are threatened with further violence if they dare to reveal their oppression (e.g. row 5). Clearly, victims in these reports face challenges for breaking their silence, thereby not only hinting at the courage required to register, but also the number of likely unreported cases. The example dowry murders (a type of offense that happens to have the *highest* probability of suspect acquittal, Appendix Table A22), add depth to preceding analyses by illustrating how real human beings are impacted.

In Table 8, `Female` remains significant. Columns 1, 4 show results with only TIRM matching, while Columns 2-3, 5-6 add controls. The results add confidence to the notion that complainant identity specifically yields dissimilar responses to requests for help from the state.

55. More well-to-do individuals might demand luxury vehicles as dowry—which for a less upwardly mobile group could involve a motorcycle instead of car—in addition to the mandatory jewelry and household effects.

Table 7: Balance Check 2 (Hard Test): Matched Cases and Penal Codes [Identifying Information Censored]

Treated First-Information-Report	Matched **Control** First-Information-Report	
...I, Anuja [redacted], daughter of late [redacted]...cruelty and violence which has completely left me traumatized and I am constantly living in fear for my own life...went to my parental house for Pag Phera and returned back at night to my matrimonial house, in Ambala. In the evening all the leftover jewelry (which I was wearing) was taken by my sister-in-law on pretext that it is better to be kept safe with in-laws... After marriage I realised that my husband and my in-laws were downright greedy as they started making more illegal demands for dowry... They used to persistently taunt and harass me for not bringing sufficient amount of cash and gifts. My husband and his father [redacted] also demanded that they have not been given a car according to their 'status,' and should be given a Mercedes or Pajero in dowry. Father-in-law [redacted]...is one of the leading lawyers in the town...his elder son is judge posted as Civil Judge Cum JMIC. My husband...taunting that my parents had not spent money...Since then health has started deteriorating, my mother-in-law and father-in-law became angry and beat me...IPC 323/406/498-A/506	Mr. Sir...[redacted] is my daughter who has studied up to M.Sc., B.Ed. and whose marriage we had with [redacted] Maqsood from Delhi on [redacted]. We had an engagement ceremony which cost Rs.3,00,000 / and gave the boy a gold chain, a gold ring and Rs. 1,51,000 / cash. They then demanded a Scorpio. When we expressed our inability to deliver the Scorpio vehicle, he asked to meet after two days, and I met him on [redacted], he said that we also want Rs.5,00,000 / - cash with the Scorpio. On our refusal, he refused to bring a procession. But we had completed the wedding preparations. Some relatives had arrived. We had booked confectioners, tents, banquet hall... we already spent Rs.10,00,000 /. Then I, and my boy [redacted], my brother-in-law [redacted], our neighbor met them. Sitting and talking, they refused to marry without Rs. 5,00,000 /...The culprits refused to marry my girl after being engaged in the greed of dowry, and I was humiliated and my Rs.20,00,000/ has been lost. Therefore, I pray that legal action should be taken against him and FIR should be lodged...my goods, cash should be returned...Dowry Prohibition Act, 1961;4/3.	Dowry Harassment
I am Ankita [redacted], daughter of Ashok [redacted] [redacted] Colony, [redacted], Punjab. I live in [redacted] Gurgaon. I work in [redacted] company sector [redacted]. On date [redacted] at 10 am I came to company for duty on my scooty. I parked my scooty in the parking lot, and I went to office. When I came back at around 6:00 pm, my scooty could not be found. My scooty color was Gray Model 2014, License [redacted] Engine No [redacted]. I do not know who took it. Please register an FIR for my stolen scooty. IPC 379.	I am Kapil [redacted], son of [redacted] from Nagina. I have a scooty number [redacted] in white. I left my scooty on [redacted] in a plot near [redacted] University. I was giving exam from 2-5 o'clock when I came back, Scooty was not standing there... After that, I had gone to my hometown for some urgent work, and now I am submitting to police. I do not remember the Scooty's engine or chassis number, all papers were in Scooty itself. Please register an FIR for my theft. Phone No. [redacted]. IPC 379.	Scooter Theft
I am Vandana, wife of [redacted] Caste Kamboj, resident of Village [redacted]. I am 30. Yesterday, my boy had gone to [redacted] for tutoring. I was going to pick him up at 6.00 pm on my Activa, License No. [redacted]. While taking U-turn in front of Gupta Petrol Pump, a motorcycle driver from Yamunanagar crashed into me. I fell on the road, and my left leg was seriously injured... My brother noted the License number [redacted]...got admitted to Rama Krishna Hospital Jagadhri for treatment. I am in full consciousness now. The motorcyclist ran away, but I can recognize him if he comes in front of me... IPC 279/337/338.	I am Harsha [redacted], son of Pradeep, Caste [redacted]...I study in B.T.Class. On date [redacted] at around 9:40 PM, I was riding my cycle (License [redacted]) from Sector 13 to Mohan Nagar. Behind me my friend Jagjit [redacted], son of [redacted], caste Jat, was sitting and I was driving. When we reached the telephone exchange, a car came from behind with great speed and carelessness, and hit me, from which I bounced off bike. My head went into the electric pole, and my friend fell on the road. The car no. was [redacted], a Honda I10...the driver's name is Kartik...Strictest legal action should be taken against him. IPC 279/337.	Hit-and-Run
I am [redacted]. Late Shri [redacted] married his girl Puja to [redacted], resident of [redacted] on 21.04.2009. According to his status, everything was given, but after a few months, the accused started harassing the family and demanded a motorcycle. Her family members started beating her. In 2010, he tried to kill her by pouring kerosene on her, but she escaped. For this, [redacted] and his father [redacted] were caught and sent to jail, but later they started living together again. [redacted] and [redacted] brought Puja to Delhi and started harassing her again, saying they want Rs. 1 lakh from her family to start business. The father and mother-in-law [redacted] Devi...started to behave more wrongly till Puja was hanged. Shrimanji is requested to investigate this and please get justice...information was received from Safdarjung Hospital that Puja has died...IPC 304-B.	I have come to complain that my sister Shilpa [redacted] was wife of Sahil [redacted], resident of [redacted] Ground. She was married to Sahil 3 years ago at age 24. Today at 4 o'clock in the evening, we got the news that she and her son Rihansh, aged 2 years, have both been found dead in the bathroom. We got a call from the hospital...Go to [redacted] as soon as possible - we are sure that the death has been caused by dowry demands. We got a call from Shilpa on date [redacted] from Poonam, a resident of Delhi. Shilpa told her that she was being bullied for dowry - Rs. 10 lakh and a vehicle was being demanded...she was being beaten...Please fully investigate that Shilpa's husband Sahil [redacted] has definitely killed Shilpa and her son Rihansh. We hope to take immediate action from you. IPC 304-B.	Dowry/Murder
...Mr. Sir...I am Bimala, wife of [redacted] from Sonipat. This morning my daughter, whose name is [redacted], was abducted by Sagar aka [redacted] and family. She's been taken away. I am getting phone from No. [redacted]. Sagar has threatened to kill her, and said that give 5 lakh rupees or else she will die. We do not know where she is, but the number is telling location Chandigarh. I pray to you that the police administration is involved and it is registered, please do not delay it. Phone no. [redacted]. IPC 365.	Mr. Sir...I am a resident of [redacted] Road Punhana, Mewat, [redacted] Khan. I am a man of peace who abides by the law. On the date [redacted] at around 1 o'clock at night, Hakku son of [redacted] of [redacted]... asked me to open the door...there were two or three others. The men came in and put a *katta* [knife] on my neck and started saying that "if you make noise, we will kill you and your family." They took my girl Shabnam by force and cash of Rs. 32,000 / - and put my girl in a Scorpio. They said they will kill her if we go to police...when we went to Hakku in the morning, he told us that he will not give her at any price...I request, Janab, to take legal action against the people and return my girl to a poor man. IPC 363/366-A.	Abduction
I am a Indira [redacted] wife of Mr. [redacted] from [redacted] Colony, Hisar. I work as an assistant in [redacted]. In January 2016, I got a call from Sachin [redacted], JGS India Trading and Marketing PVT Ltd...a good scheme...where government employees have a big advantage... deposit two lakh twenty thousand rupees in the account of this company, you will get 8000 rupees per month for 12 months...He said that we have benefited thousands of people...Account [redacted]...IFSC Code [redacted]...Sachin threatened me...stole Rs 2,20,000... IPC 406/420	Mr. Sir...I am Gulzar [redacted] son Mr. Sadhu [redacted], resident of [redacted], Ambala city. I have known the suspects for 15-20 years. They said they would help me file to go to Canada in 2015......they told me that they work to send poor people abroad, and with down-payment of Rs 1,50,000 - to 2,00,000. / - one can easily earn more abroad...told me that you should give me all the documents...My shop is located in Grain Mandi...They took my money and now saying they will kill me...retrieve my money which is Rs.6,50,000/... IPC 406/420.	Cheating

Table 8: Impact of Complainant Gender on Conviction/Acquittal After Text-Matching

	Convicted			Acquitted		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.008^{***}	-0.008^{***}	-0.012^{***}	0.010^{***}	0.014^{***}	0.021^{***}
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Constant	0.036^{***}	0.037^{***}	0.052^{***}	0.124^{***}	0.466^{***}	0.682^{***}
	(0.0003)	(0.007)	(0.012)	(0.001)	(0.012)	(0.020)
Observations	337,056	309,008	179,335	337,056	309,008	179,335
R^2	0.0002	0.037	0.066	0.0001	0.093	0.135
Controls	N	Y	Y	N	Y	Y
PS FE	N	Y	Y	N	Y	Y
Month-Yr FE	N	Y	Y	N	Y	Y
Judge Rank	N	N	Y	N	N	Y

Note: Controls include a numeric variable for a crime's distance from a station, investigator rank, as well as whether the registering station is urban. *p<0.1; **p<0.05; ***p<0.01

Discussion

Political science has had limited purchase, even basic descriptive evidence, as to whether the state treats minorities seeking justice differently, especially in the Global South. This paper charts the full trajectory of complaints from the very second that citizens enter a police station until a verdict is issued by court. Having created an original dataset of the universe of crime records from a major Indian state, and then combining it with judicial files, I show that women face a more onerous *process* and unequal *outcomes*.

Unlike medicine, where individual doctors may prioritize patients that have the highest chance of survival, triage in criminal justice I argue reflects a group or complaint-type's relationship with multiple administrators such that episodic discriminations cumulate. Specifically, I find that women may be disadvantaged in terms of (1) delays in registering cases, (2) lower likelihood of cases being sent to court, (3) delays in police investigations, (4) higher levels of case dismissals, (5) delays in court hearings and verdict issuance, (6) higher levels of acquittals and lower convictions for suspects. While VAW is less likely to be cancelled by law enforcement, both categories of crime registered by women are likely to be triaged in the judiciary. Text-matching provides additional evidence on the impact of complainant identity on sanctions for suspects.

I contend that triage may occur when marginalized groups approach formal institutions for grievance redressal; discrimination may *not* be restricted to a single-stage, but might exhibit as complaints *transition* or "squeeze" through the discretionary purview of connected officials, who may utilize tactics at their disposal to (dis)favor complaints. In South Asia, these strategies could include deflecting cases of sexual assault to counseling centers,[56] while in the United States they may comprise securing plea deals to lesser charges (Ransom 2021).

The findings illustrate the importance of being attentive to the workings of criminal justice

56. Mueller-Smith and T. Schnepel (2021) note that Texas may "divert" perpetrators of low-level (drug and property) offenses to community service instead of prison. In India, however, diversion is more often applied to *complainants* rather than perpetrators, and for gender-based violence (Jassal 2020).

institutions when complaints are being processed, long *after* initial registration. In post-colonial contexts, for instance, the state may retain patchworks of red-tape through which triage can sustain. In India, I demonstrate that triage is most extreme at the judicial level where there are few pressures from either "above" or "below."[57] And so, interventions at mitigating discrimination in any one agency may be ineffective unless the manner in which *other* administrators can influence the same case's trajectory is accounted for.

Furthermore, the study expands discussions of VAW—which largely focus on sexual assault in (or after) conflict—by highlighting gradations of *daily* abuse. Dowry, for instance, is a case likely to be triaged; yet, topic modeling reveals that such crimes are not "petty quarrels," but may involve heinous acts including marital rape. This dynamic is evocative of a double-bind: on the one hand, women may be faced with marital violence, and even (dowry) death, in an effort to extract resources from their natal homes; yet, delaying or avoiding marriage comes with its own costs (Carpena and Jensensius 2020; Corno, Hildebrandt, and Voena 2020). While studies on VAW in India have focused on property rights (Panda and Agarwal 2005; Chin 2012), alcohol consumption (Luca, Owens, and Sharma 2015), and culture (Fernandez 1997), a question emerges as to whether perpetrators are *aware* of the inability (or unwillingness) of the state to provide punitive justice, and if this knowledge predisposes them to act.

Subsequent scholarship might systematically probe the motivations of administrators too. Are officials repressive, e.g. triaging cases because of supposed privilege that women exude by coming forward (e.g. without male support)? Or, are they constrained by resource scarcity in an overburdened system? Can cultural forces be at play, e.g. formal justice for women as a threat to male dominance? Do structural barriers have a bearing, e.g. limited access to lawyers, lack of autonomy to follow-up at station- and court-houses, and/or inability to pay bribes?

Aside from opening a research agenda, the data consist of a modern archive that may be useful not only in the present, but also to social scientists and historians a century from now. The cases capture—often in deeply poignant terms—the helplessness of victims, who invariably express that they have turned to formal institutions as a last resort, despite uncertainty in a system's ability to help when much seems lost or destroyed. Other questions worth exploring include: How does gender interact with caste or ethnicity? Is north India representative of other parts of the Subcontinent? Can state policies that make the criminal justice system more demographically representative (for women and minorities) affect the base-line statistics outlined herein? Can finer-grained measures of justice delivery (e.g. monetary compensation) be generated through surveys, especially since many of the complaints remain active?

While the notion that women face hardship in India may be unsurprising to some, others, including judges and policymakers, have vociferously argued that female complainants send men to prison for "petty" offenses, that the Penal Code is stacked in their favor, and that a burgeoning "men's rights movement" should be supported in deterring women's "legal terrorism" (Lodhia 2014; Naishadham 2018). The findings cast doubt on many of these assumptions. Furthermore, the study aims to make a theoretical case for exploring the junctures at which linked institutions are connected, and the varying discretionary authority of bureaucrats across those bodies, in order to understand deeper, multi-layered patterns of discrimination. Exploring what criminal justice triage entails, and where it manifests across institutional designs, may promote theory-building and target reform[58] aimed at improving justice delivery and the quality of democracy.

57. A nationally representative Indian survey shows respondents blaming the judiciary (Appendix Figure A35).

58. While 30% of gendered cases are dismissed by law enforcement in Haryana, newspapers report prosecutors dropped 49% of sexual assault cases in New York City in 2019 (Ransom 2021).

References

Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan. 2012. "Do Judges Vary in Their Treatment of Race?" *The Journal of Legal Studies* 41 (2): 347–383.

Agerberg, Mattias, and Anne-Kathrin Kreft. 2020. "Gendered Conflict, Gendered Outcomes: The Politicization of Sexual Violence and Quota Adoption." *Journal of Conflict Resolution* 64 (2): 290–317.

Agnes, Flavia, and Audrey D'Mello. 2015. "Protection of Women from Domestic Violence." *Economic and Political Weekly,* 76–84.

Alesina, Alberto, and Eliana La Ferrara. 2014. "A Test of Racial Bias in Capital Sentencing." *American Economic Review* 104 (11): 3397–3433.

Anderson, Siwan. 2007. "The Economics of Dowry and Brideprice." *Journal of Economic Perspectives* 21 (4): 151–174.

Armstrong, Elizabeth A., Miriam Gleckman-Krut, and Lanora Johnson. 2018. "Silence, Power, and Inequality: An Intersectional Approach to Sexual Violence." *Annual Review of Sociology* 44 (1): 99–122.

Arnold, David, Will Dobbie, and Crystal S Yang. 2018. "Racial Bias in Bail Decisions*." *The Quarterly Journal of Economics* 133 (4): 1885–1932.

Ash, Elliott, Sam Asher, Aditi Bhowmick, Daniel Chen, Tanaya Devi, Christoph Goessmann, Paul Novosad, and Bilal Siddiqi. 2021. "Measuring Gender and Religious Bias in the Indian Judiciary," 53.

Banerjee, Abhijit V. 1997. "A Theory of Misgovernance." *The Quarterly Journal of Economics* 112 (4): 1289–1332.

Basu, Srimati. 2012. "Judges of Normality: Mediating Marriage in the Family Courts of Kolkata, India." *Signs: Journal of Women in Culture and Society* 37 (2): 469–492.

Bhalotra, Sonia, Abhishek Chakravarty, and Selim Gulesci. 2020. "The Price of Gold: Dowry and Death in India." *Journal of Development Economics* 143:102413.

Biderman, Albert D., and Albert J. Reiss. 1967. "On Exploring the "Dark Figure" of Crime." *The Annals of the American Academy of Political and Social Science* 374:1–15.

Bindler, Anna, and Randi Hjalmarsson. 2020. "The Persistence of the Criminal Justice Gender Gap: Evidence from 200 Years of Judicial Decisions." *The Journal of Law and Economics* 63 (2): 297–339.

Blair, Robert A., Sabrina M. Karim, and Benjamin S. Morse. 2019. "Establishing the Rule of Law in Weak and War-Torn States: Evidence from a Field Experiment with the Liberian National Police." *American Political Science Review* 113 (3): 641–657.

Bozçağa, Tuğba, and Alisha C. Holland. 2018. "Enforcement Process Tracing: Forbearance and Dilution in Urban Colombia and Turkey." *Studies in Comparative International Development* 53 (3): 300–323.

Brown, Darryl K. 2004. "Rationing Criminal Defense Entitlements: An Argument from Institutional Design." *Columbia Law Review* 104:801.

Brulé, Rachel E. 2020. "Reform, Representation, and Resistance: The Politics of Property Rights' Enforcement." *The Journal of Politics* 82 (4): 1390–1405.

Butler, Daniel M., and David E. Broockman. 2011. "Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators." *American Journal of Political Science* 55 (3): 463–477.

Calvi, Rossella. 2020. "Why are older women missing in India? The age profile of bargaining power and poverty." Publisher: The University of Chicago Press Chicago, IL, *Journal of Political Economy* 128 (7): 2453–2501.

Carpena, Fenella, and Francesca R. Jensenius. 2020. "Age of Marriage and Women's Political Engagement: Evidence from India." *The Journal of Politics.*

Chin, Yoo-Mi. 2012. "Male Backlash, Bargaining, or Exposure Reduction?: Women's Working Status and Physical Spousal Violence in India." *Journal of Population Economics* 25 (1): 175–200.

Cohen, Dara Kay. 2013. "Explaining Rape During Civil War: Cross-National Evidence (1980–2009)." *American Political Science Review* 107 (3): 461–477.

Corno, Lucia, Nicole Hildebrandt, and Alessandra Voena. 2020. "Age of Marriage, Weather Shocks, and the Direction of Marriage Payments." *Econometrica* 88 (3): 879–915.

Court, Supreme. 2016. "Youth Bar Association Vs. Union of India." Parliament of India. Accessed August 7, 2017. https://drive.google.com/file/d/0BzXilfcxe7yucEgwdHF2LWgzbTg/view.

CSDS and Common Cause. 2018. "Status of Policing in India Report (SPIR)." Accessed May 12, 2018. http://www.commoncause.in/page.php?id=85.

Dasgupta, Aditya, and Devesh Kapur. 2020. "The Political Economy of Bureaucratic Overload: Evidence from Rural Development Officials in India." *American Political Science Review* 114 (4): 1316–1334.

DHS. 2017. "India National Family Health Survey NFHS-4 2015-16." *Demographic and Health Survey.*

Dupas, Pascaline, and Radhika Jain. 2021. *Women Left Behind: Gender Disparities in Utilization of Government Health Insurance in India.* w28972. Cambridge, MA: National Bureau of Economic Research.

EGAP. 2019. "Metaketa IV: Community Policing." Accessed January 28, 2019. http://egap.org/metaketa/metaketa-iv-community-policing.

Emeriau, Mathilde. 2021. "Learning to be Unbiased: Evidence from the French Asylum Office." *American Journal of Political Science,* 62.

Feder, Amir, Katherine A. Keith, Emaad Manzoor, Reid Pryzant, Dhanya Sridhar, Zach Wood-Doughty, Jacob Eisenstein, et al. 2021. "Causal Inference in Natural Language Processing: Estimation, Prediction, Interpretation and Beyond." *arXiv:2109.00725 [cs].*

Fernandez, Marilyn. 1997. "Domestic Violence by Extended Family Members in India: Interplay of Gender and Generation." *Journal of Interpersonal Violence* 12 (3).

Goldsmith, Belinda, and Meka Beresford. 2018. "Poll Ranks India the World's Most Dangerous Country for Women." *The Guardian* (June 28, 2018).

Green, Donald P., Anna M. Wilke, and Jasper Cooper. 2020. "Countering Violence Against Women by Encouraging Disclosure: A Mass Media Experiment in Rural Uganda." *Comparative Political Studies* 53 (14): 2283–2320.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–297.

Holland, Alisha C. 2016. "Forbearance." *American Political Science Review* 110 (2): 232–246.

Htun, Mala, and S. Laurel Weldon. 2012. "The Civic Origins of Progressive Policy Change: Combating Violence against Women in Global Perspective, 1975–2005." *American Political Science Review* 106 (3): 548–569.

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2012. "Causal Inference without Balance Checking: Coarsened Exact Matching." *Political Analysis* 20 (1): 1–24.

Iyer, Lakshmi, and Anandi Mani. 2012. "Traveling Agents: Political Change and Bureaucratic Turnover in India." *Review of Economics and Statistics* 94 (3): 723–739.

Iyer, Lakshmi, Anandi Mani, Prachi Mishra, and Petia Topalova. 2012. "The Power of Political Voice: Women's Political Representation and Crime in India." *American Economic Journal: Applied Economics* 4 (4): 165–193.

Jassal, Nirvikar. 2020. "Gender, Law Enforcement, and Access to Justice: Evidence from All-Women Police Stations in India." *American Political Science Review* 114 (4): 1035–1054.

———. 2021. "Segregation as Efficiency? Group-Specific Institutions in North India." *The Journal of Asian Studies* 80 (3): 631–661.

Jassal, Nirvikar, and Sharon Barnhardt. 2020. "Do Female Bureaucrats Generate Institutional Legitimacy? Experimental Evidence from India." Working paper, University of California, Berkeley.

Jassal, Nirvikar, and Pradeep Chhibber. 2019. "India in 2018: Crises of Institutional Legitimacy and Gender." *Asian Survey* 59 (1): 85–97.

Jayachandran, Seema. 2015. "The Roots of Gender Inequality in Developing Countries." *Annual Review of Economics* 7 (1): 63–88.

Karim, Sabrina. 2020. "Relational State Building in Areas of Limited Statehood: Experimental Evidence on the Attitudes of the Police." *American Political Science Review,* 1–16.

Khan, Shamus, Joss Greene, Claude Ann Mellins, and Jennifer S. Hirsch. 2020. "The Social Organization of Sexual Assault." *Annual Review of Criminology* 3 (1): 139–163.

Knox, Dean, Will Lowe, and Jonathan Mummolo. 2020. "Administrative Records Mask Racially Biased Policing." *American Political Science Review* 114 (3): 619–637.

Kothari, Jayna. 2005. "Criminal Law on Domestic Violence: Promises and Limits." *Economic and Political Weekly,* 4843–4849.

Kruks-Wisner, Gabrielle. 2021. "Great Expectations, Great Grievances: The Politics of Citizens' Complaints in India." *Comparative Politics.*

Kurlychek, Megan C., and Brian D. Johnson. 2019. "Cumulative Disadvantage in the American Criminal Justice System." *Annual Review of Criminology* 2 (1): 291–319.

Lang, Kevin, and Ariella Kahn-Lang Spitzer. 2020. "Race Discrimination: An Economic Perspective." *Journal of Economic Perspectives* 34 (2): 68–89.

Liebman, Benjamin L., Margaret E. Roberts, Rachel E. Stern, and Alice Z. Wang. 2020. "Mass Digitization of Chinese Court Decisions: How to Use Text as Data in the Field of Chinese Law." *Journal of Law and Courts* 8 (2): 177–201.

Lodhia, Sharmila. 2014. ""Stop Importing Weapons of Family Destruction!": Cyberdiscourses, Patriarchal Anxieties, and the Men's Backlash Movement in India." *Violence Against Women* 20 (8): 905–936.

Luca, Dara Lee, Emily Owens, and Gunjan Sharma. 2015. "Can Alcohol Prohibition Reduce Violence Against Women?" *American Economic Review* 105 (5): 625–629.

Lucas, Christopher, Richard A. Nielsen, Margaret E. Roberts, Brandon M. Stewart, Alex Storer, and Dustin Tingley. 2015. "Computer-Assisted Text Analysis for Comparative Politics." *Political Analysis* 23 (2): 254–277.

McDougal, Lotus, Samuel Krumholz, Nandita Bhan, Prashant Bharadwaj, and Anita Raj. 2018. "Releasing the Tide: How Has a Shock to the Acceptability of Gender-Based Sexual Violence Affected Rape Reporting to Police in India?" *Journal of Interpersonal Violence.*

Mueller-Smith, Michael, and Kevin T. Schnepel. 2021. "Diversion in the Criminal Justice System." *The Review of Economic Studies* 88 (2): 883–936.

Naishadham, Suman. 2018. "Why India's Men's Rights Movement Is Thriving." Vice, April 13, 2018. Accessed April 10, 2020. https://www.vice.com/en_in/article/9b8akp/why-indias-mens-rights-movement-is-thriving.

Neil, Roland, and Christopher Winship. 2019. "Methodological Challenges and Opportunities in Testing for Racial Discrimination in Policing." *Annual Review of Criminology* 2 (1): 73–98.

Oldenburg, Veena Talwar. 2002. *Dowry Murder: The Imperial Origins of a Cultural Crime.* Oxford: Oxford University Press.

Olsen, Asmus Leth, Jonas Høgh Kyhse-Andersen, and Donald Moynihan. 2020. "The Unequal Distribution of Opportunity: A National Audit Study of Bureaucratic Discrimination in Primary School Access." *American Journal of Political Science.*

Panda, Pradeep, and Bina Agarwal. 2005. "Marital Violence, Human Development and Women's Property Status in India." *World Development* 33 (5): 823–850.

Parthasarathy, Ramya, Vijayendra Rao, and Nethra Palaniswamy. 2019. "Deliberative Democracy in an Unequal World: A Text-As-Data Study of South India's Village Assemblies." *American Political Science Review* 113 (3): 623–640.

Ransom, Jan. 2021. "'Nobody Believed Me': How Rape Cases Get Dropped." *The New York Times.*

Rao, Vijayendra. 1993. "The Rising Price of Husbands: A Hedonic Analysis of Dowry Increases in Rural India." *Journal of Political Economy* 101 (4): 666–677.

———. 1997. "Wife-Beating in Rural South India: A Qualitative and Econometric Analysis." *Social Science & Medicine* 44 (8): 1169–1180.

Rehavi, M. Marit, and Sonja B. Starr. 2014. "Racial Disparity in Federal Criminal Sentences." *Journal of Political Economy* 122 (6): 1320–1354.

Richardson, L Song, and Phillip Atiba Goff. 2012. "Implicit Racial Bias in Public Defender Triage." *The Yale Law Journal* 122:24.

Richardson, L. Song. 2016. "Systemic Triage: Review of 'Crook County: Racism and Injustice in America's Largest Criminal Court' by Nicole Van Cleve." *Yale Law Journal* 126:862.

Roberts, Margaret E., Brandon M. Stewart, and Edoardo M. Airoldi. 2016. "A Model of Text for Experimentation in the Social Sciences." *Journal of the American Statistical Association* 111 (515): 988–1003.

Roberts, Margaret E., Brandon M. Stewart, and Richard A. Nielsen. 2020. "Adjusting for Confounding with Text Matching." *American Journal of Political Science* 64 (4): 887–903.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2019. "Stm: An R Package for Structural Topic Models." *Journal of Statistical Software* 91 (1): 1–40.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4): 1064–1082.

Rose, Elaina. 1999. "Consumption Smoothing and Excess Female Mortality in Rural India." *The Review of Economics and Statistics* 81 (1): 41–49.

Roychowdhury, Poulami. 2021. *Capable Women, Incapable States: Negotiating Violence and Rights in India.* Modern South Asia series. New York, NY: Oxford University Press.

Sanders, James, Giulio Lisi, and Cheryl Schonhardt-Bailey. 2017. "Themes and Topics in Parliamentary Oversight Hearings: A New Direction in Textual Data Analysis." *Statistics, Politics and Policy* 8 (2): 153–194.

Sekhon, Jasjeet S. 2009. "Opiates for the Matches: Matching Methods for Causal Inference." *Annual Review of Political Science* 12:487–508.

Slough, Tara, and Christopher Fariss. 2021. "Misgovernance and Human Rights: The Case of Illegal Detention without Intent." *American Journal of Political Science* 65 (1): 148–165.

Spohn, Cassia, and Katharine Tellis. 2019. "Sexual Assault Case Outcomes: Disentangling the Overlapping Decisions of Police and Prosecutors." *Justice Quarterly* 36 (3): 383–411.

Srinivas, M. N. 1956. "A Note on Sanskritization and Westernization." *The Far Eastern Quarterly* 15 (4): 481–496.

Srinivasan, Sharada, and Arjun S. Bedi. 2007. "Domestic Violence and Dowry: Evidence from a South Indian Village." *World Development* 35 (5): 857–880.

Tellez, Juan Fernando, Erik Wibbels, and Anirudh Krishna. 2020. "Local Order, Policing, and Bribes: Evidence from India." *World Politics* 72 (3): 377–410.

Verma, Anshul. 2017. "Need to Amend Section 498A of Indian Penal Code and Section 41A of Code of Criminal Procedure." Lok Sabha Debates, August 3, 2017. Accessed April 10, 2020. http://eparlib.nic.in/handle/123456789/781303.

White, Ariel R., Noah L. Nathan, and Julie K. Faller. 2015. "What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials." *American Political Science Review* 109 (1): 129–142.

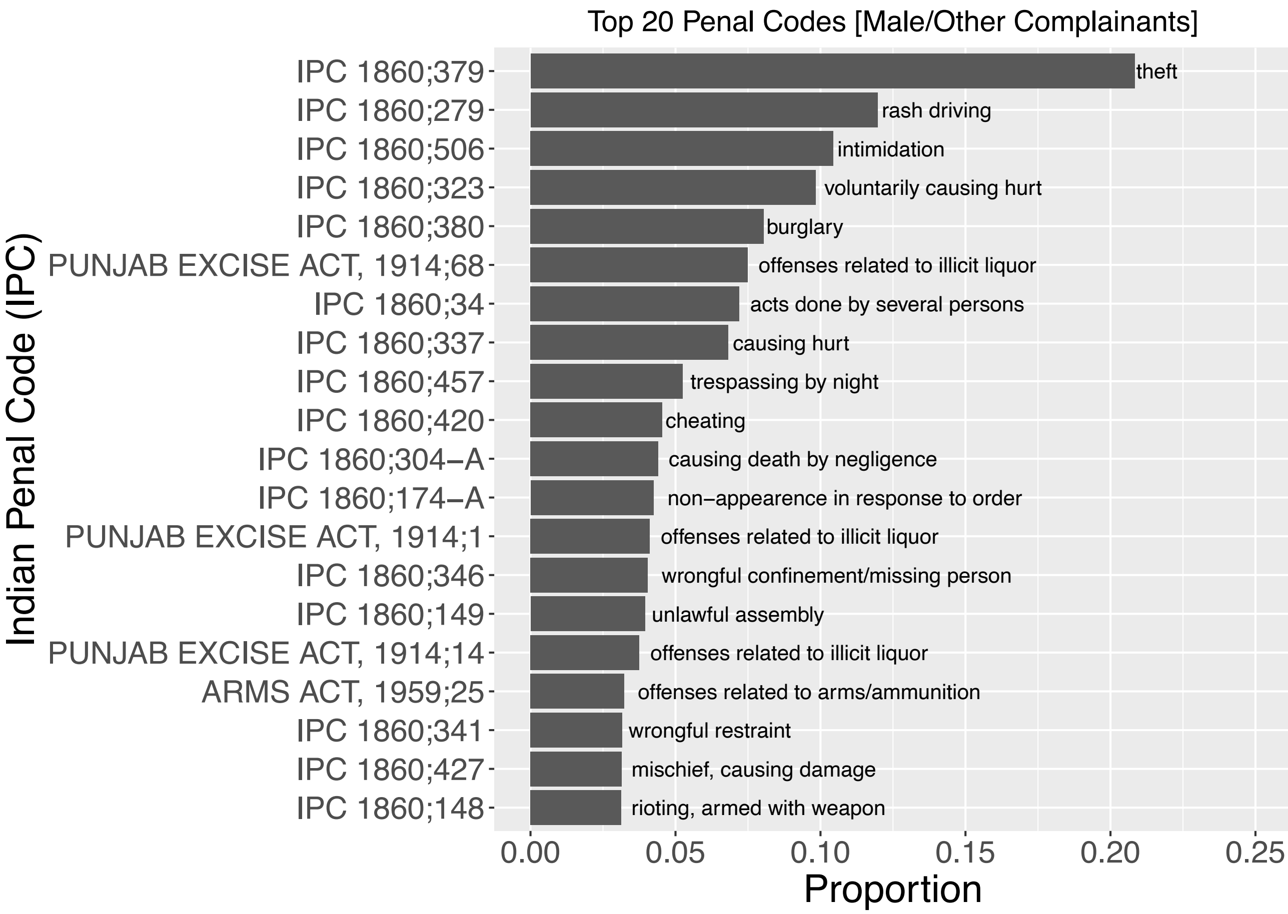
WHO. n.d. "Violence Against Women." World Health Organization. Accessed April 11, 2020. http://www.who.int/reproductivehealth/topics/violence/en/.

Online Appendix

Table of Contents

1 Additional Data on Police Files

Figure A1: Top Indian Penal Code Sections [Male Complainants]



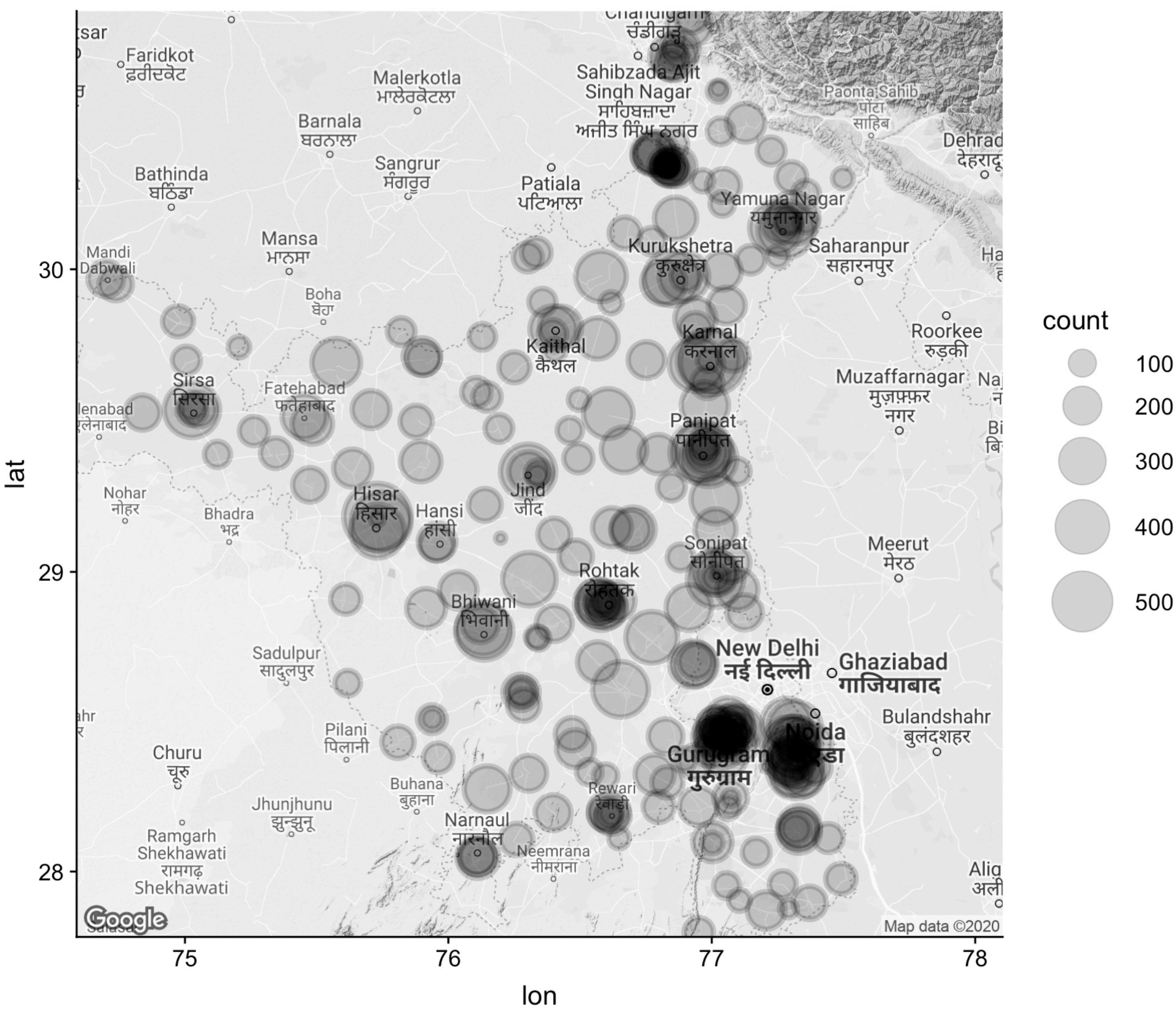
Note: Top twenty Indian Penal Code sections attached to cases brought forward by men/other (N=379,362). The top substantive sections include theft, rash driving, burglary, and illicit liquor/bootlegging.

Table A1: Description of Sections & Special Acts Considered Gendered or 'Crimes Against Women'

Section	Description
IPC 1860;294	obscene acts or songs
IPC 1860;304-B	dowry death
IPC 1860;313	causing miscarriage without woman's consent
IPC 1860;314	death caused by act done with intent to cause miscarriage
IPC 1860;315	act done to prevent child from being born alive
IPC 1860;316	death of unborn child
IPC 1860;318	concealment of birth by secret disposal of dead body
IPC 1860;354	sexual harassment
IPC 1860;366	kidnapping, abducting a woman to compel her to marriage
IPC 1860;366-A	procuration of minor girl
IPC 1860;366-B	importation of girl from foreign country
IPC 1860;376	rape
IPC 1860;376-B	intercourse by husband upon his wife during separation
IPC 1860;376-C	intercourse by person in authority
IPC 1860;376-D	gang rape
IPC 1860;376-E	punishment for repeat offenders
IPC 1860;497	adultery
IPC 1860;498	enticing or taking away a married woman
IPC 1860;498-A	husband or relative subjecting woman to cruelty
IPC 1860;509	word, gesture or act intended to insult modesty of a woman
IPC 1860;306	abetment of suicide
IPC 1860;317	exposure or abandonment of child
IPC 1860;326-A	acid throwing
IPC 1860;326-B	attempted acid throwing
IPC 1860;363	kidnapping from guardianship
IPC 1860;377	"unnatural" sex (anal sex/sodomy)
IPC 1860;494	marrying again during lifetime of husband or wife
IPC 1860;495	concealment of marriage
IPC 1860;496	ceremony gone through without lawful marriage
The Child Marriage Restraint Act, 1929	
The Immoral Traffic (Prevention) Act, 1956	
The Dowry Prohibition Act, 1961	
The Commission of Sati (Prevention) Act, 1987	
Protection of Women Against Domestic Violence Act, 2005	
The Information Technology Act, 2000	
The Indecent Representation of Women (Prohibition) Act, 1986	
Protection of Children from Sexual Offenses Act, 2012	

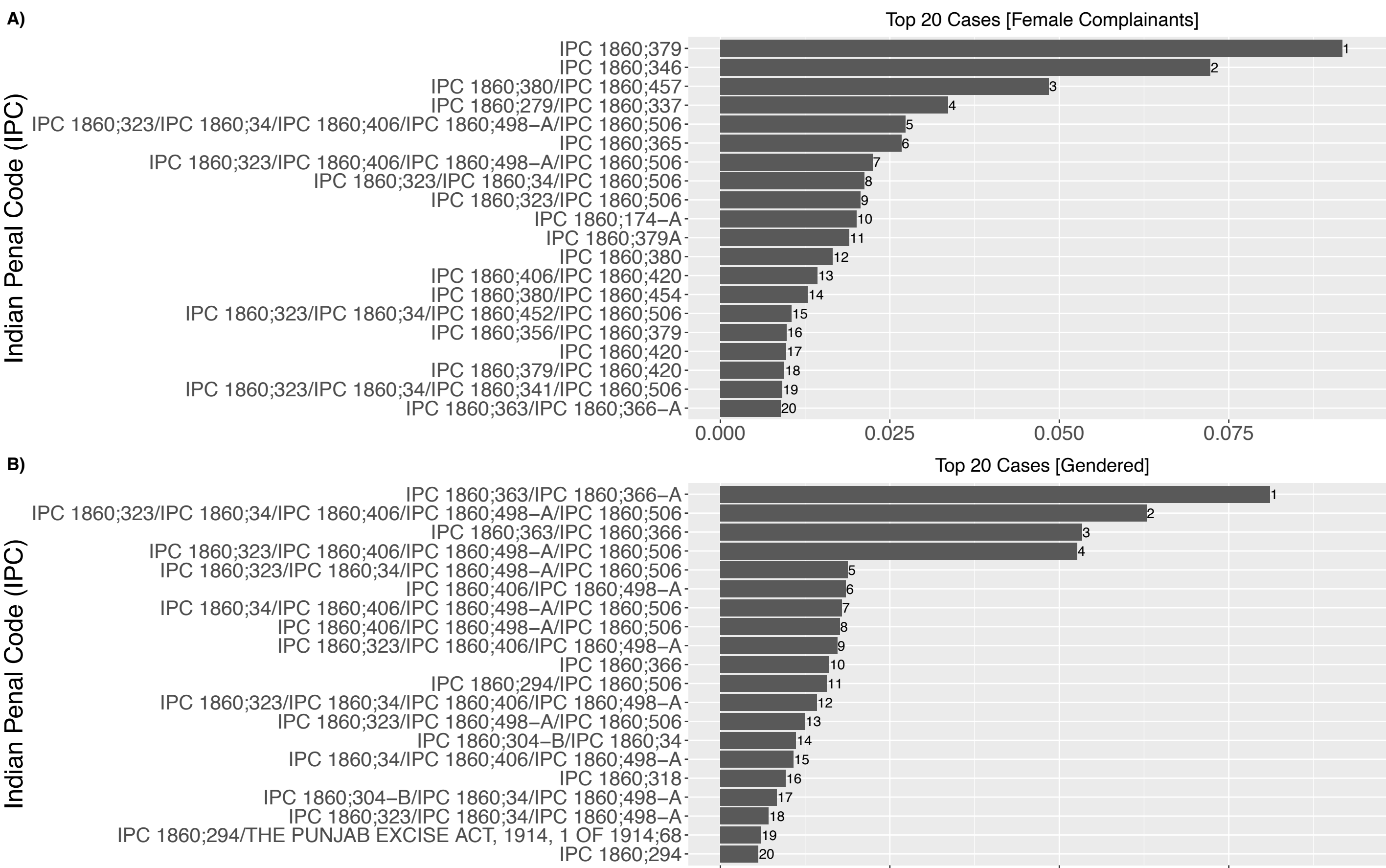
Note: Gendered crimes or 'crimes against women' listed in official government documents. IPC refers to Indian Penal Code. All cases that have one or more of the foregoing Penal Codes appended are categorized as VAW or gendered crime.

Figure A2: Crimes With Female Complainants



Note: Map depicting locations of all Haryana police stations in which female complainants have had cases registered. Dots vary in intensity depending on the total crimes registered for female complainants, 2015-2018 (N=38,828).

Figure A3: Top Cases Registered by Female Complainants and 'Gendered' Crime or VAW

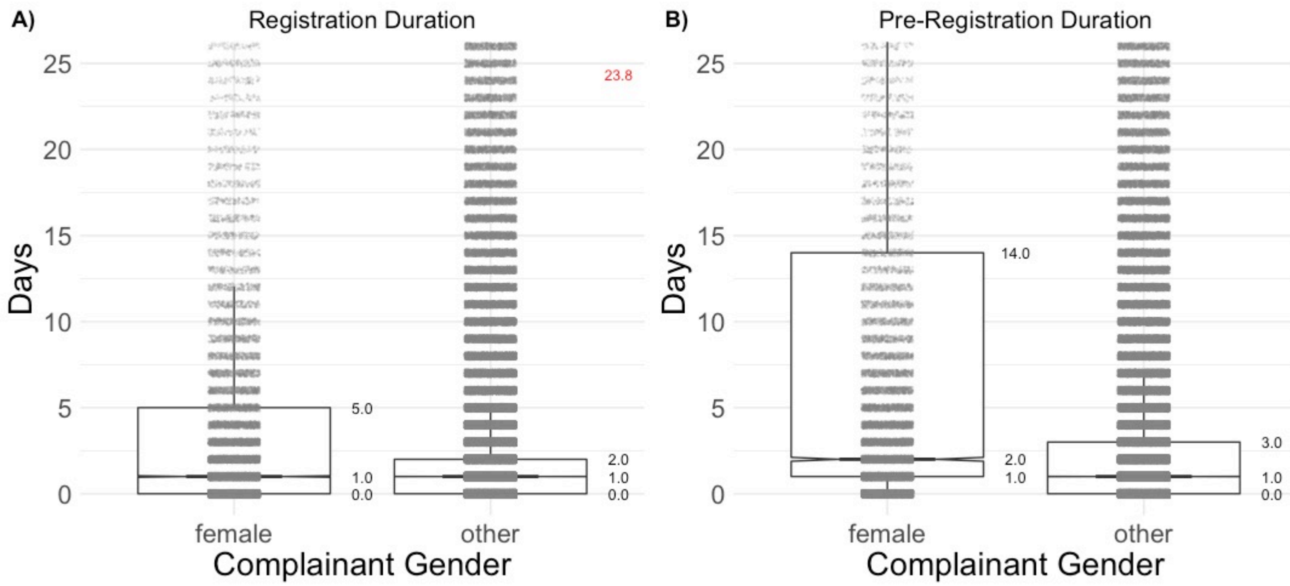


Note: Top twenty *cases* with female complainants (N=38,828) and 'gendered' crime (N=20,869). Most cases are combinations of multiple Penal Code sections. The first Penal Code in the list typically provides an indication of the kind of case, but not always.

2 Triage: PROCESS

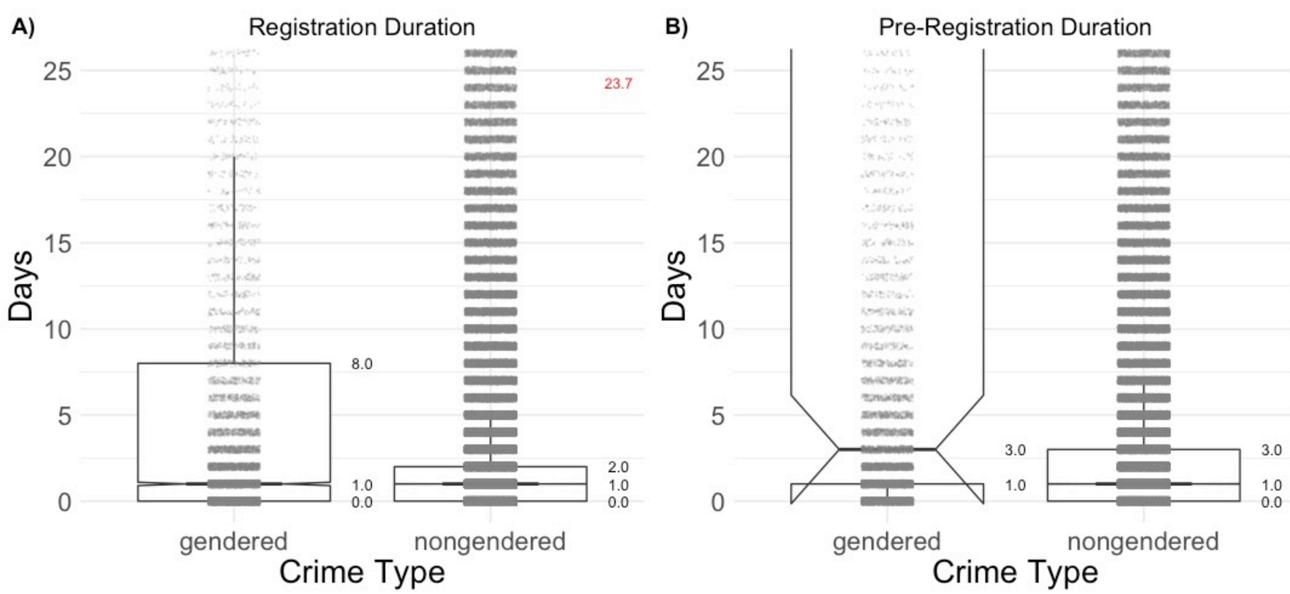
a) Registration Duration

Figure A4: Difference in Days by **Complainant Gender**



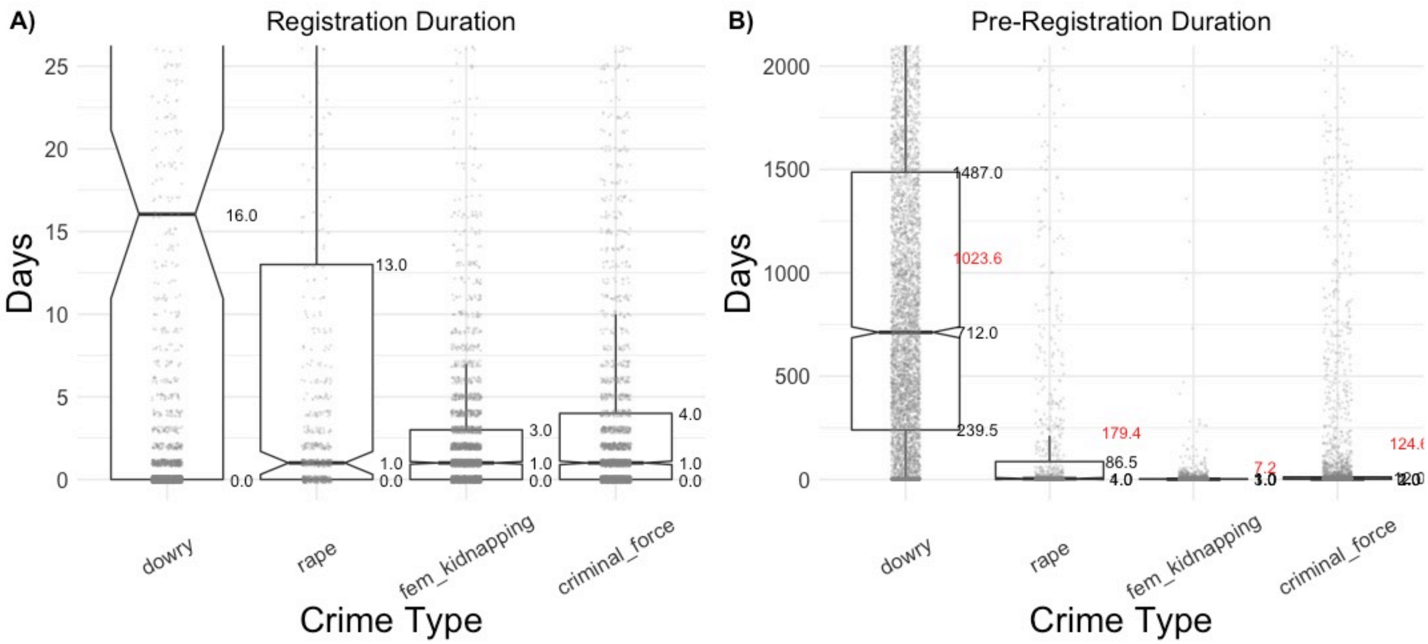
Note: Box plots depicting difference in the date from when the complainant was able to register a case compared to the date the victim told the officer the last incident related to the offense began or ended. Each dot is a registered crime report. Inter-quartile range is depicted, mean cannot be displayed. Women's cases have a longer lag in registration.

Figure A5: Difference in Days by **Crime Type**



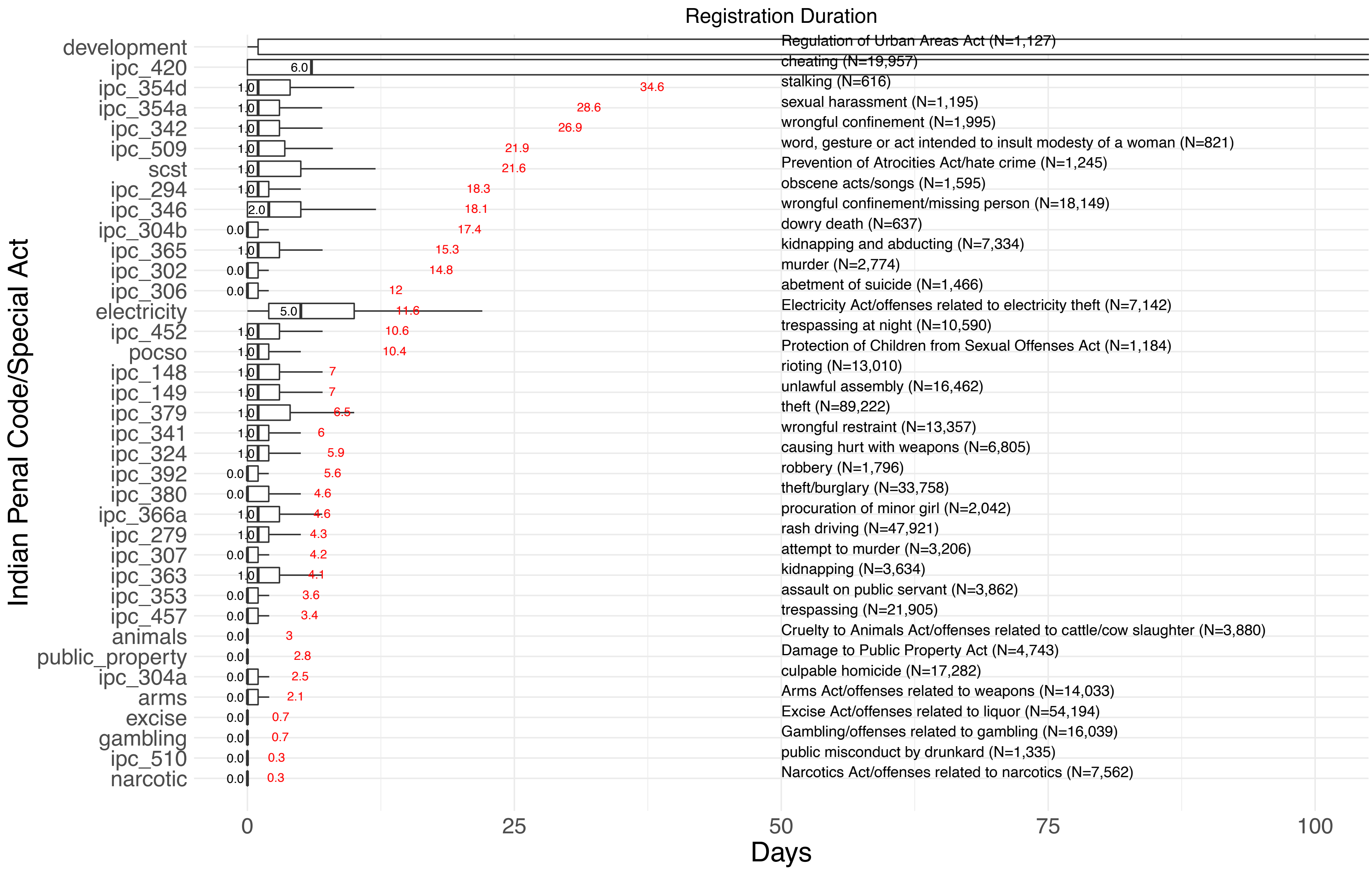
Note: Box plots depicting difference in the date from when the complainant was able to register a case compared to the date the victim told the officer that the offense began or ended. Each dot is a registered crime report. Inter-quartile range depicted, mean cannot be displayed. Gendered cases have a longer lag in registration.

Figure A6: Delays in Case Registration for Particular Gendered Crimes



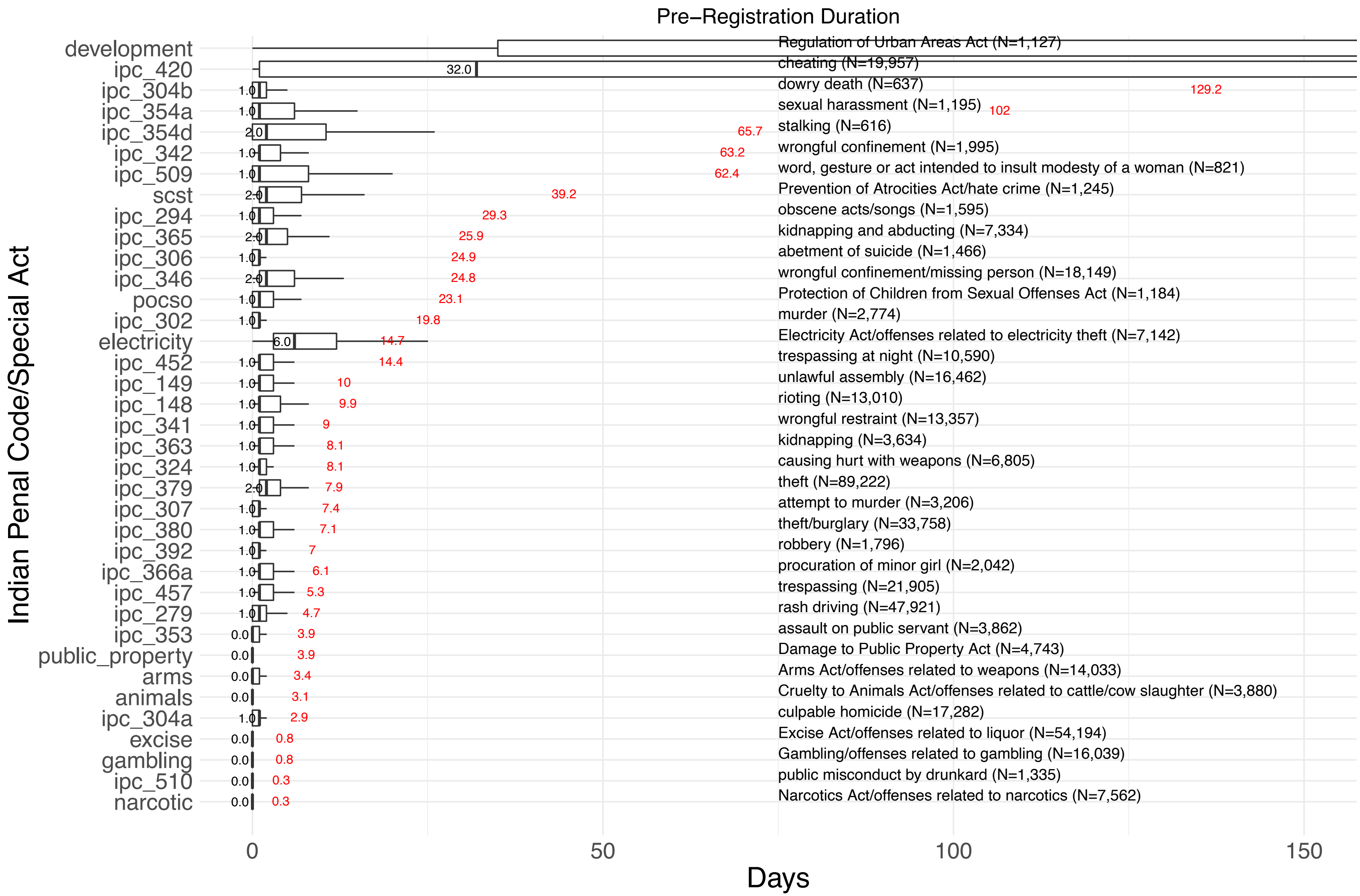
Note: Box plots depicting days waited by specific gendered crime, where each dot is a registered report (FIR). Dowry or Section 498-A (N=7,674); rape or Section 376 (N=1,094); female kidnapping or Section 366 (N=3,754); "criminal force with intent to outrage a woman's modesty" or Section 354 (N=3,804). The difference in days since the last incident related to dowry occurred and when the report was registered is a median of 16 days (mean of 326). Panel B of A6 highlights that the median number of days since the abuse *first began* for dowry harassment/domestic violence is 712 days (mean of 1023.6) or 2.8 years, almost an order of magnitude greater than other gendered crimes.

Figure A7: Difference in Days by Select Penal Code Violations



Note: Box plots depicting difference in the date from when the complainant was able to register a case compared to the date the victim told the officer the last incident related to the offense occurred (split by various violations of the Penal Code). Mean in red. *Five of the top crimes with the longest lag are gendered crimes.*

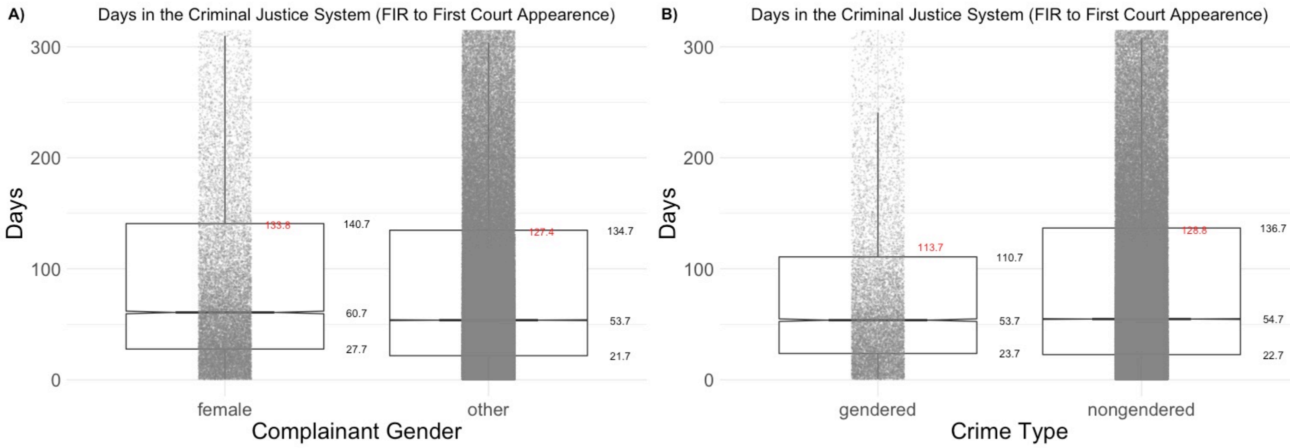
Figure A8: Difference in Days (2) by Select Penal Code Violations



Note: Box plots depicting difference in the date from when the complainant was able to register a case compared to the date the victim told the officer that the first offense related to the crime began to occur (split by various violations of the Penal Code). Mean in red. *Five of the top ten crimes are gendered.*

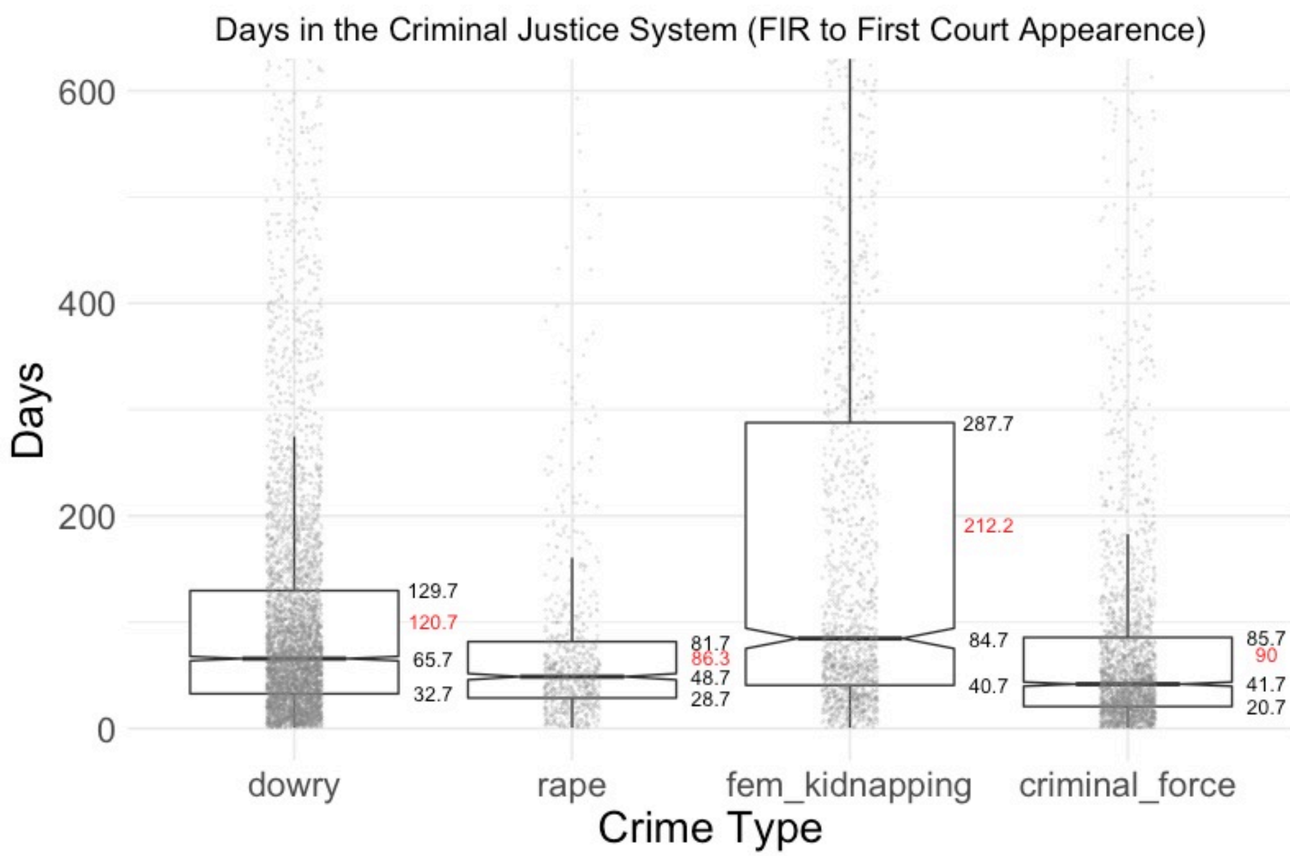
2.1 Investigation Duration

Figure A9: Days Until First Court Appearence



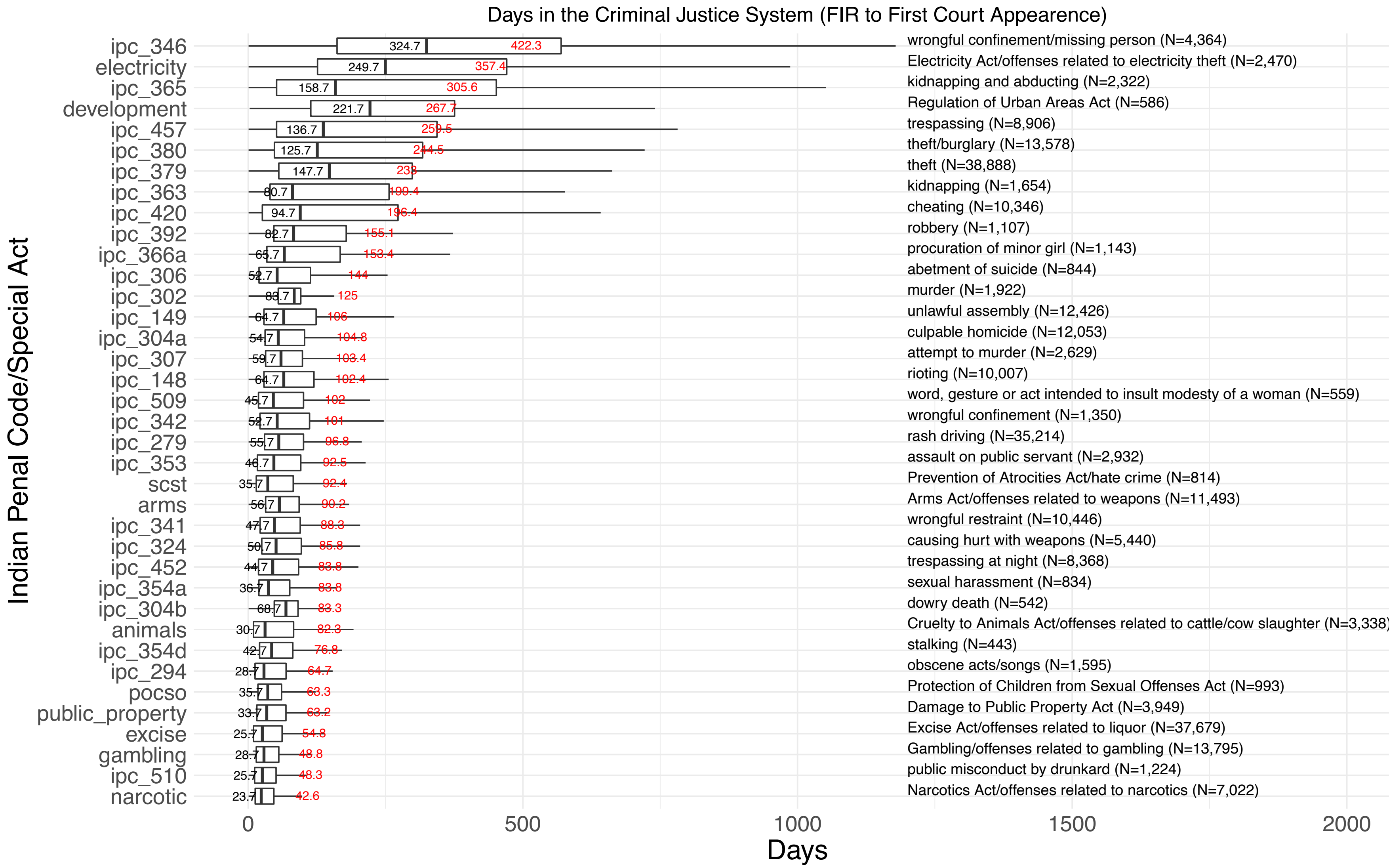
Note: FIRs that could be merged with judicial records. Figures represent the difference in days from the first date that the case appeared in the court files to the date of original crime report registration. Panel A is split by female (N=22,648), and male/other complainants (N=229,156). Panel B is split by gendered (N=14,134), and nongendered crime (N=237,670).

Figure A10: Days Until First Court Appearance for Particular Gendered Crimes



Note: Figure reflects the difference between the first hearing date in the judicial records with date of registration for dowry (N=5,541), rape (N=804), female kidnapping (N=1,685), and "criminal force" (N=2,648). Female kidnapping cases take longer to investigate.

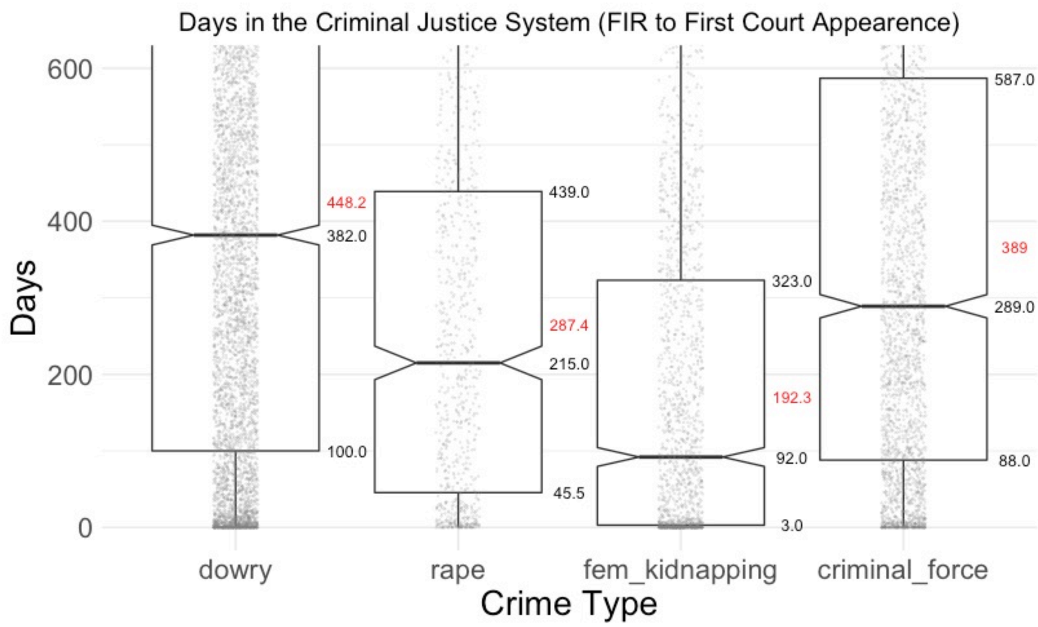
Figure A11: Days Until First Court Appearance for Select Penal Code Violations



Note: Box plots for difference in date from when the complainant was able to register a case compared to when it first entered the court (split by various violations of the Penal Code). Mean in red. Cases such as missing persons and kidnapping take longest to investigate, whereas cases such as public intoxication and drug-use take the shortest.

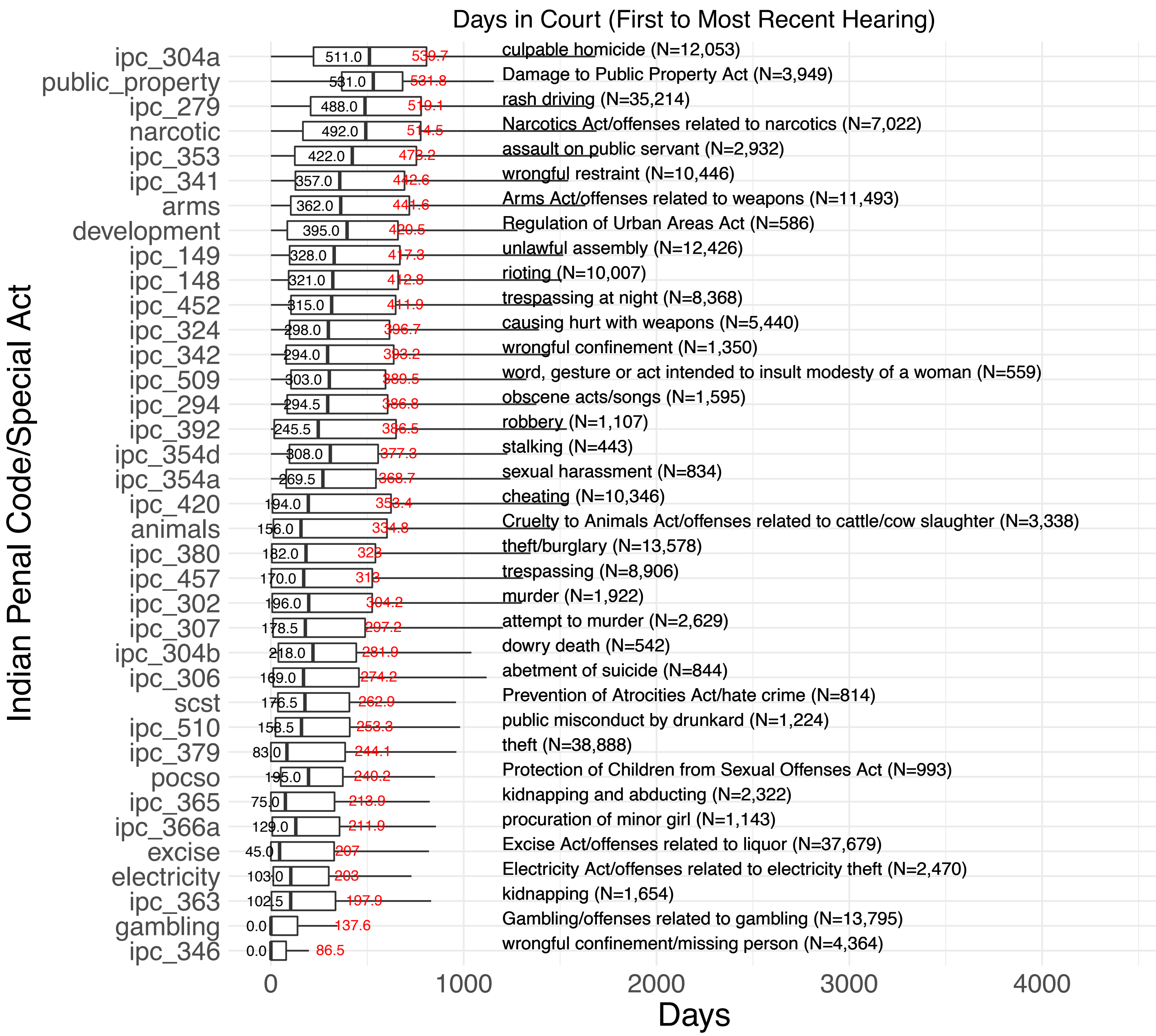
2.2 Duration in Court and Entire Criminal Justice System

Figure A12: Days in Court for Particular Gendered Crimes



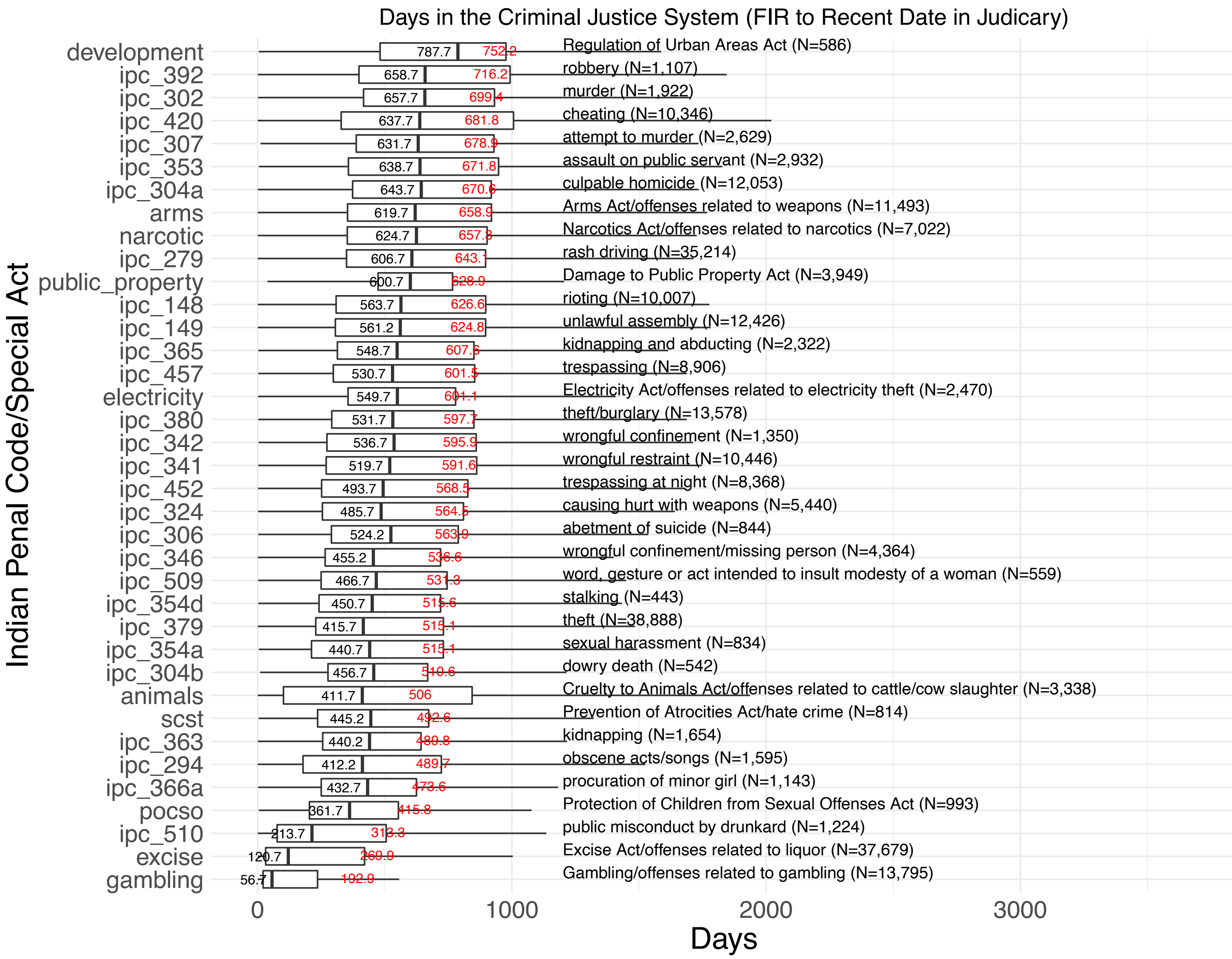
Note: Duration of a case investigation: case registration with police until the date of the first hearing in court. Dowry cases have the longest gap in terms of investigation (even though the suspect—unlike female kidnapping—is generally known).

Figure A13: Days in Court for Select Penal Code Violations



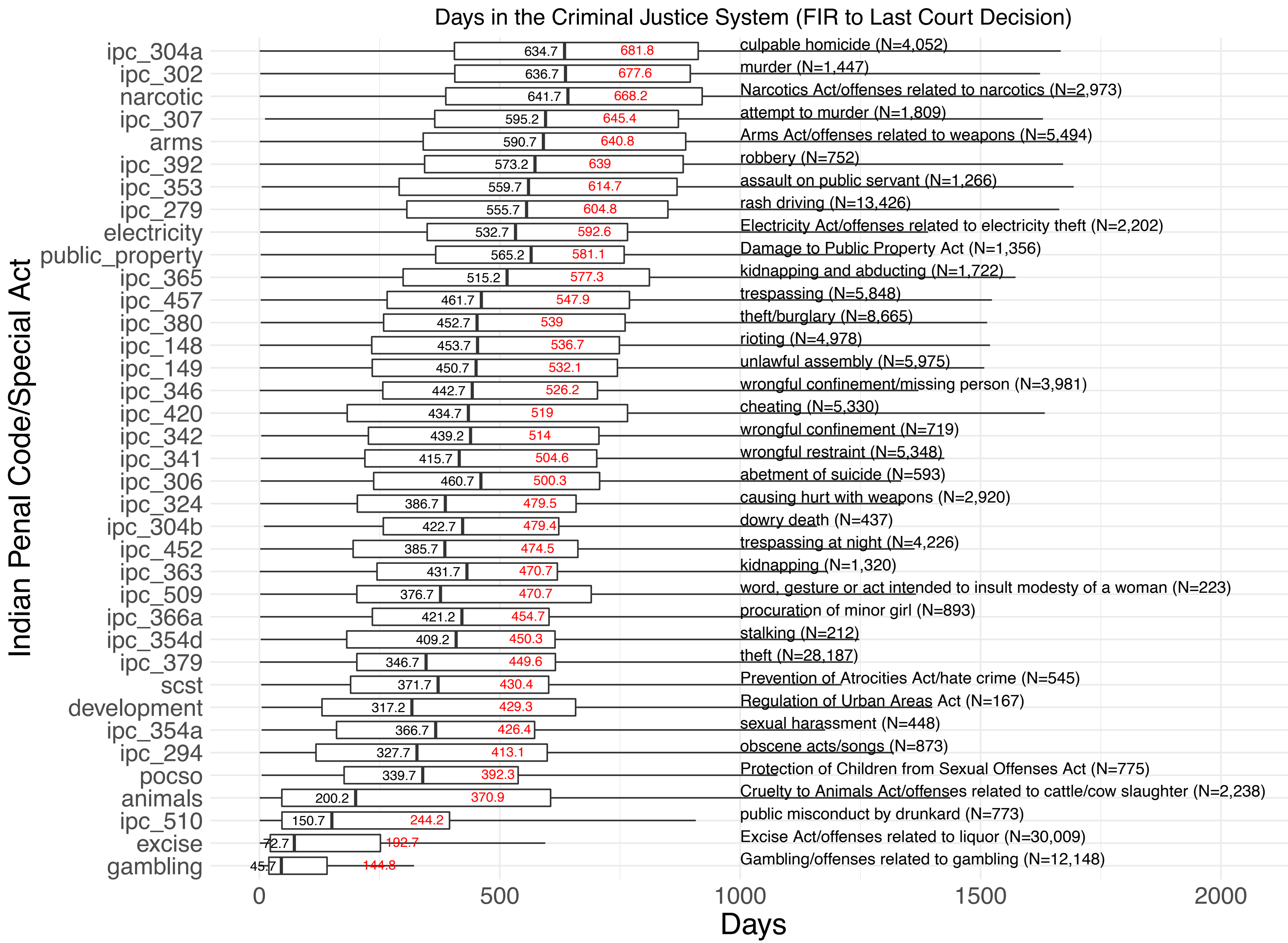
Note: Box plots for difference in the date from when the case entered court and its most recent hearing. Mean in red.

Figure A14: Days in the Entire Criminal Justice System for Select Penal Code Violations



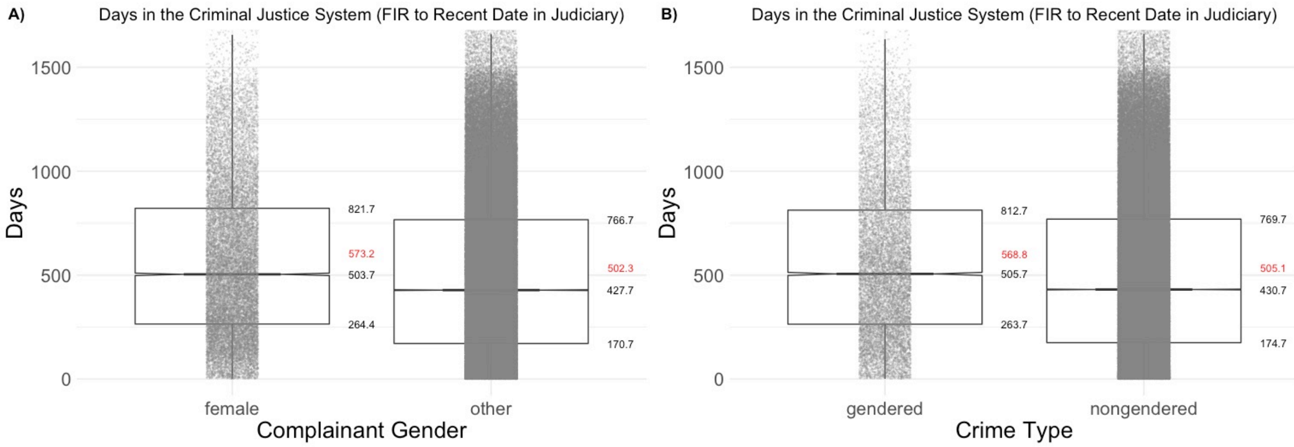
Note: Figure presents box plots for difference in the date from when the complainant was able to register a case compared to most recent hearing date in the judiciary, i.e. including on-going cases (split by various violations of the Penal Code). Mean in red.

Figure A15: Days Until a Final Decision is Reached for Select Penal Code Violations



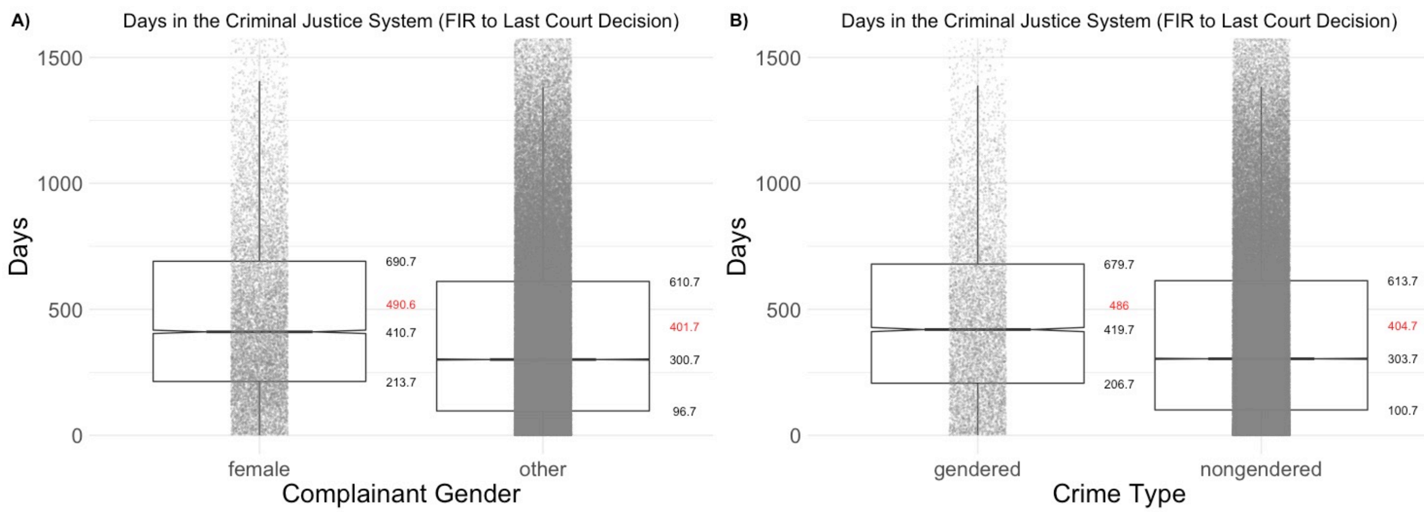
Note: Figure presents box plots for difference in the date from when the complainant was able to register a case compared to the date a decision was made, i.e. excluding on-going cases (split by various violations of the Penal Code). Mean in red.

Figure A16: Days in the Criminal Justice System



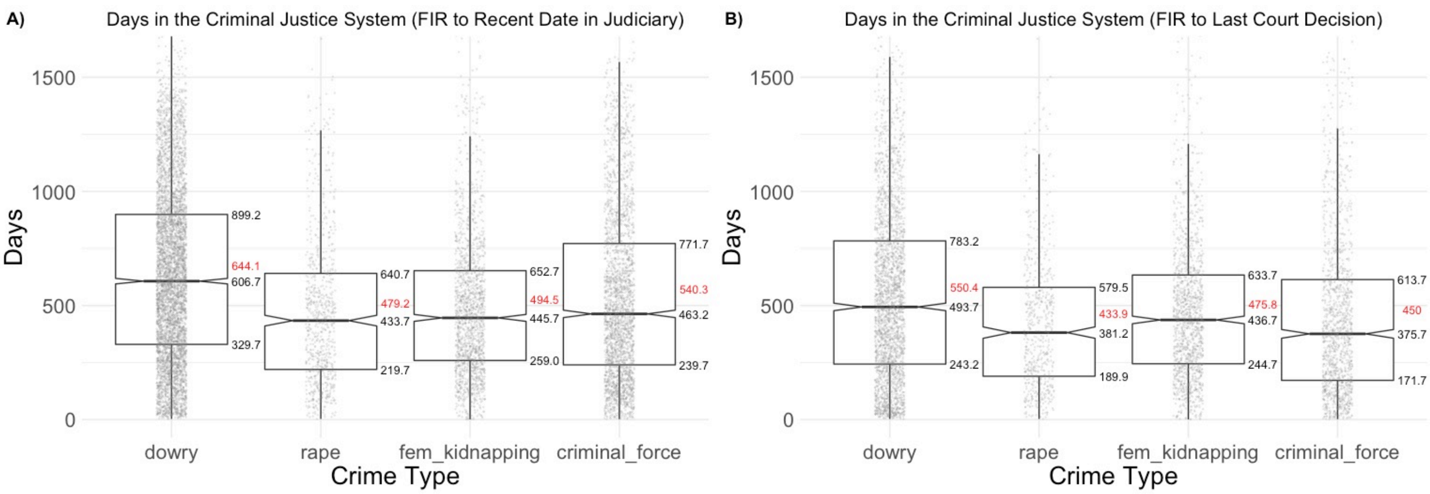
Note: FIRs that could be merged with judicial records. Figures represent the difference in days from the most recent date of the case in the court files from the date of original crime report registration with law enforcement. Panel A is split by female (N=22,648), and male/other complainants (N=229,156). Panel B is split by gendered (N=14,134), and nongendered crime (N=237,670). **Women's cases and gendered crime spend longer in the criminal justice system.**

Figure A17: Days Until a Decision Was Reached by a Judge



Note: FIRs that ultimately had a decision reached by a judge. Figures represent the difference in days from the date a decision was reached from the date of original crime report registration with law enforcement. Panel A is split by female (N=12,572), and male/other complainants (N=142,585). Panel B is split by gendered (N=8,008), and nongendered crime (N=147,149). **Women's cases and gendered crime take longer to reach a verdict.**

Figure A18: Days in the Criminal Justice System for Particular Gendered Crimes

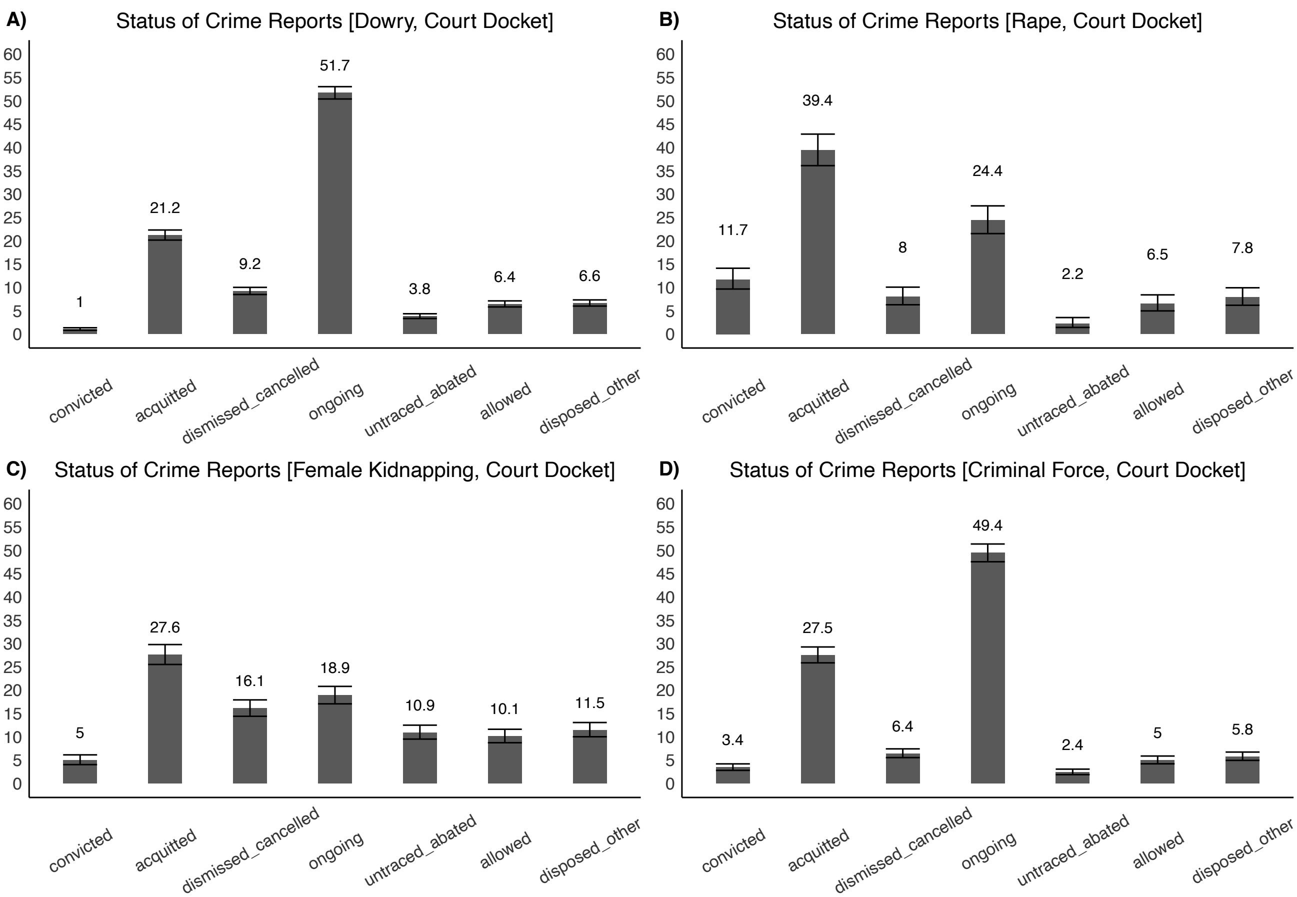


Note: Panel A reflects all cases with court files, and reflects the difference between the most recent hearing date in the judicial records with date of original crime registration for dowry (N=5,541), rape (N=804), female kidnapping (N=1,685), and criminal force (N=2,648). Panel B reflects only those cases that resulted in a decision (excluding on-going cases) for dowry (N=2,680), rape (N=608), female kidnapping (N=1,367), and criminal force (N=1,339). Panel A reveals that gendered cases, especially dowry/domestic violence, are more likely to have a later date associated with the case in the judiciary with a mean of 644 days in the criminal justice system. Of the cases that did in fact reach a decision (including acquittal or dismissal), dowry/domestic violence cases wait, on average, 550 days before a judge issues a final ruling.

3 Triage: OUTCOMES (Function of Court Docket)

3.1 Cross-Tab

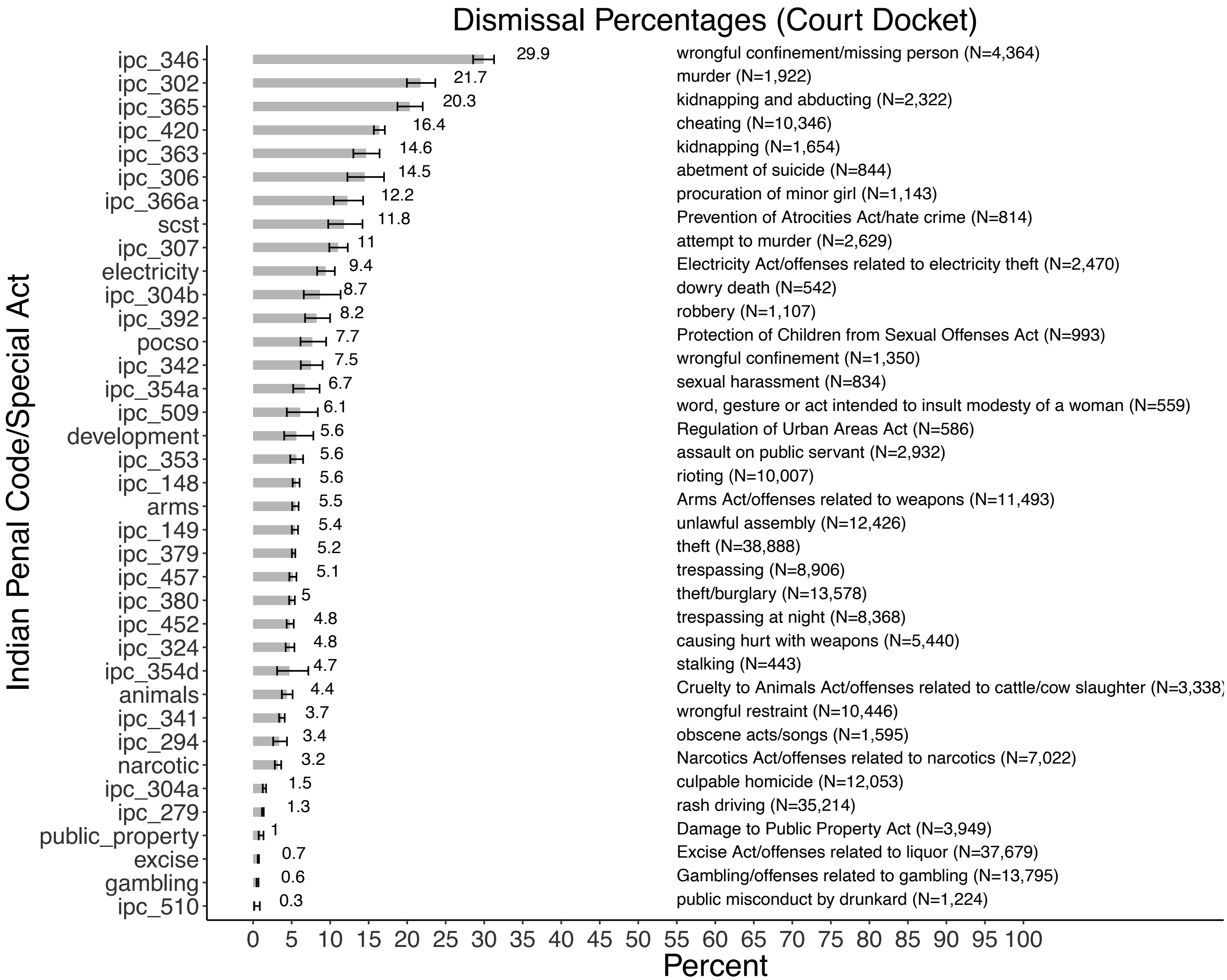
Figure A19: Crime Report Statuses in the Judicial System [Conditional on Having a Court Record]



Note: Breakdown of case statuses for crime reports that have a record in court/could be merged with judicial files, broken down by specific gendered crimes. Panel A reflects dowry cases or those that invoked Section 498-A (N=5,541); Panel B highlights rape cases or those that invoked Section 376 (N=804); Panel C represents female kidnapping or Section 366 (N=1,685); Panel D reflects criminal force with intent to outrage a woman's modesty (N=2,648). Gendered cases have low rates of conviction, with the highest in the category of rape (by a non-spouse).

3.2 Court Dismissal

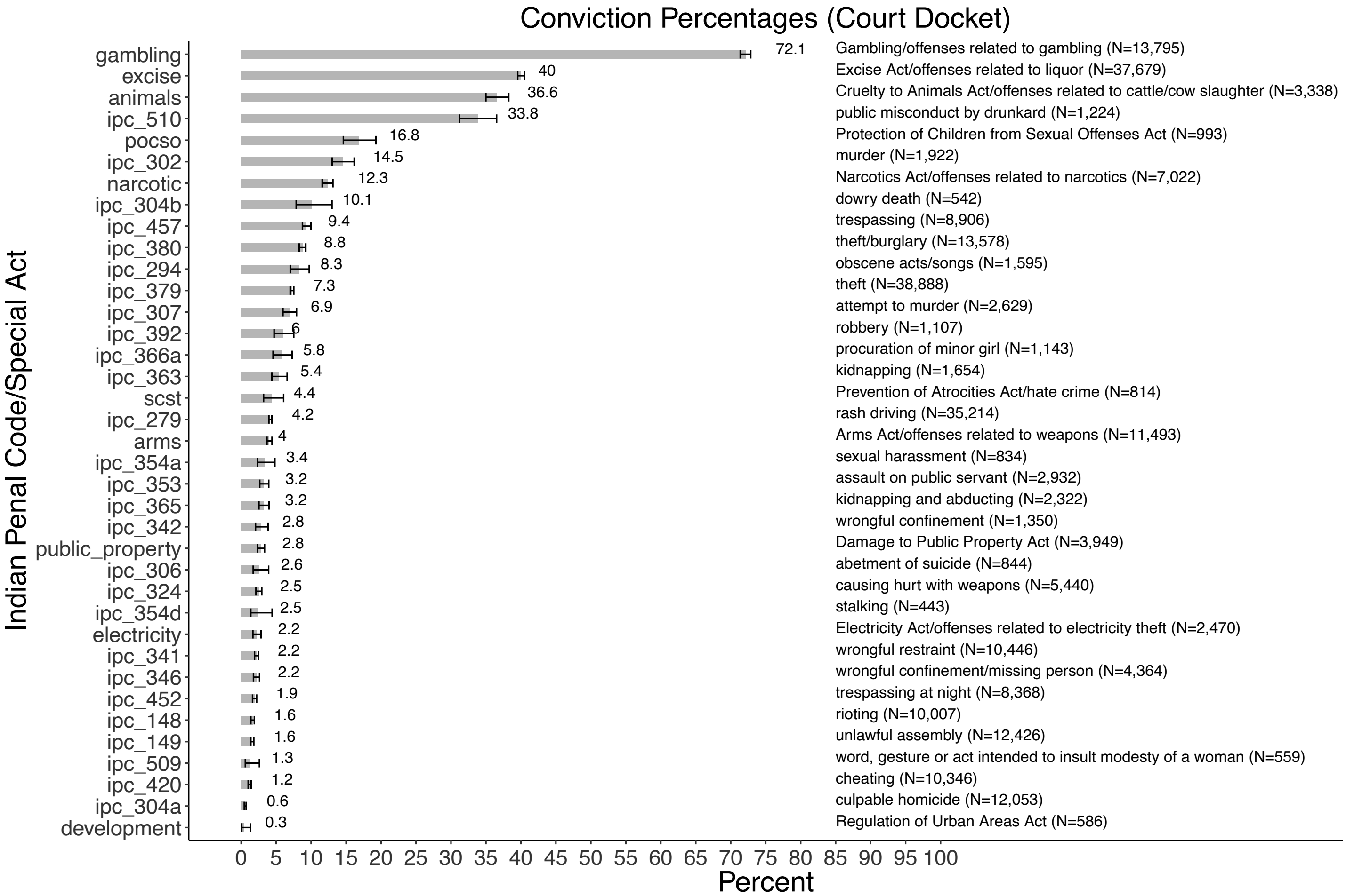
Figure A20: Dismissal Rates of Crime Reports Based on Specific Penal Code Violations [Court Docket]



Note: FIRs that could be merged with judicial records. Figure reveals dismissal rates by cases subset by particular Penal Code violations.

3.3 Conviction

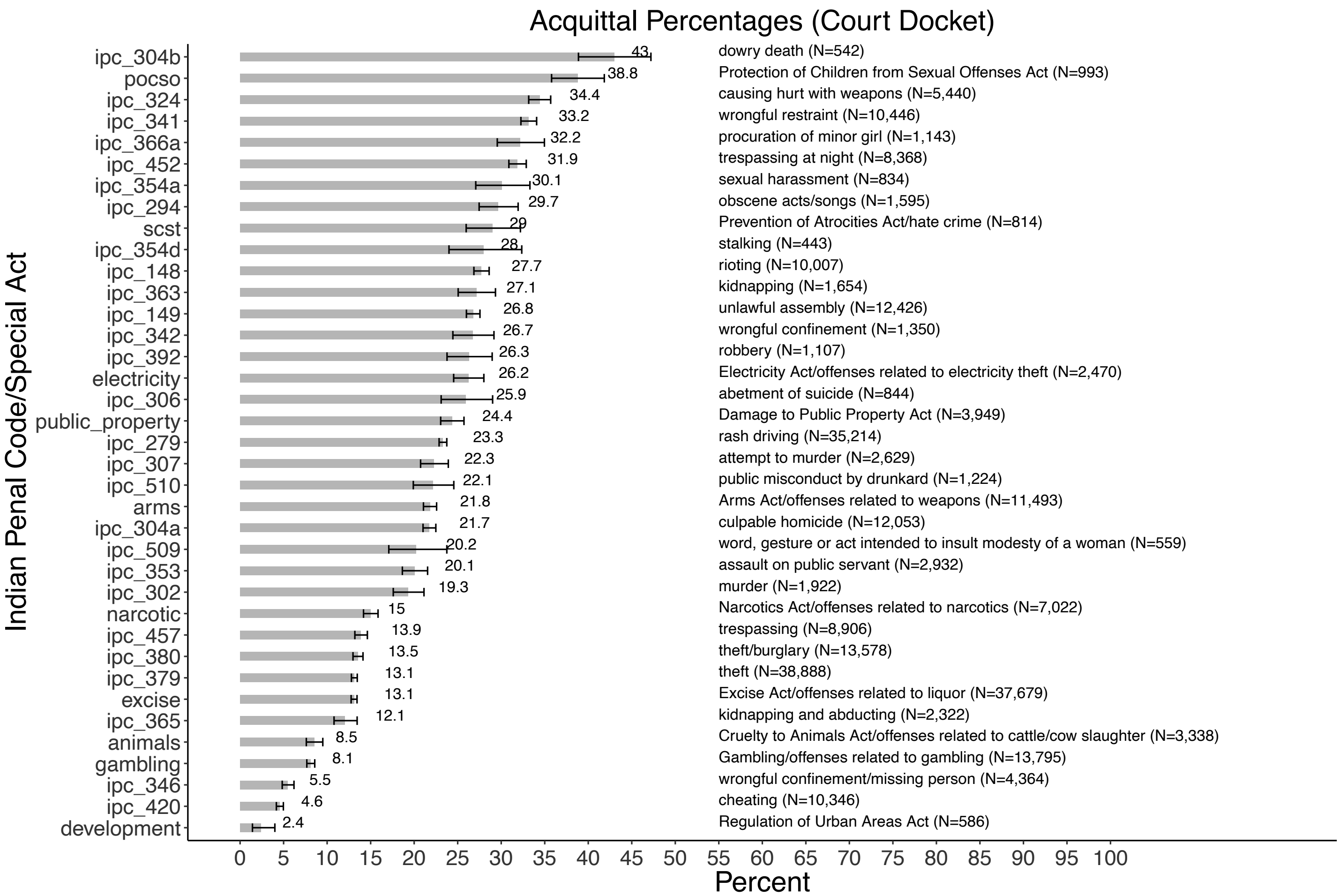
Figure A21: Conviction Rates of Crime Reports Based on Specific Penal Code Violations [Court Docket]



Note: FIRs that could be merged with judicial records. Figure reveals conviction rates by cases subset by particular Penal Code violations. The figure reveals heterogeneity in the types of gendered cases that result in higher rates of conviction. Cases perceived as 'heinous' that involve death (e.g. dowry death) or child rape (Protection of Children from Sexual Offenses Act) have higher convictions than cases seen as 'non-heinous', e.g. sexual harassment or 'insulting the modesty of women.'

3.4 Acquittal

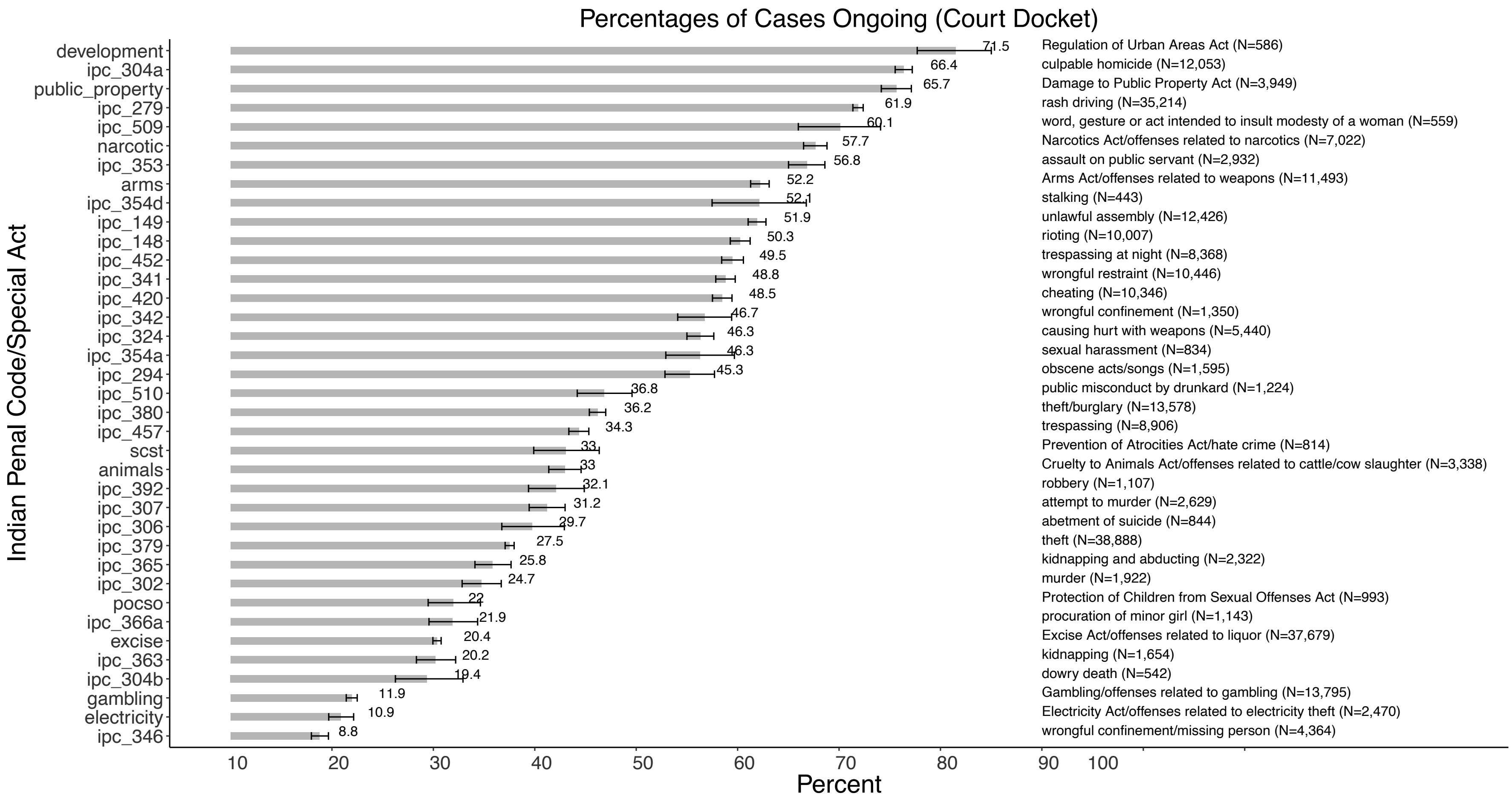
Figure A22: Acquittal Rates of Crime Reports Based on Specific Penal Code Violations [Court Docket]



Note: FIRs that could be merged with judicial records. Figure reveals acquittal rates by cases subset by particular Penal Code violations. Gendered crime have the highest acquittals, whether they are percieved as 'heinous' (e.g. dowry death) or not (sexual harassment).

3.5 Ongoing Cases

Figure A23: On-Going Rates of Crime Reports Based on Specific Penal Code Violations [Court Docket]

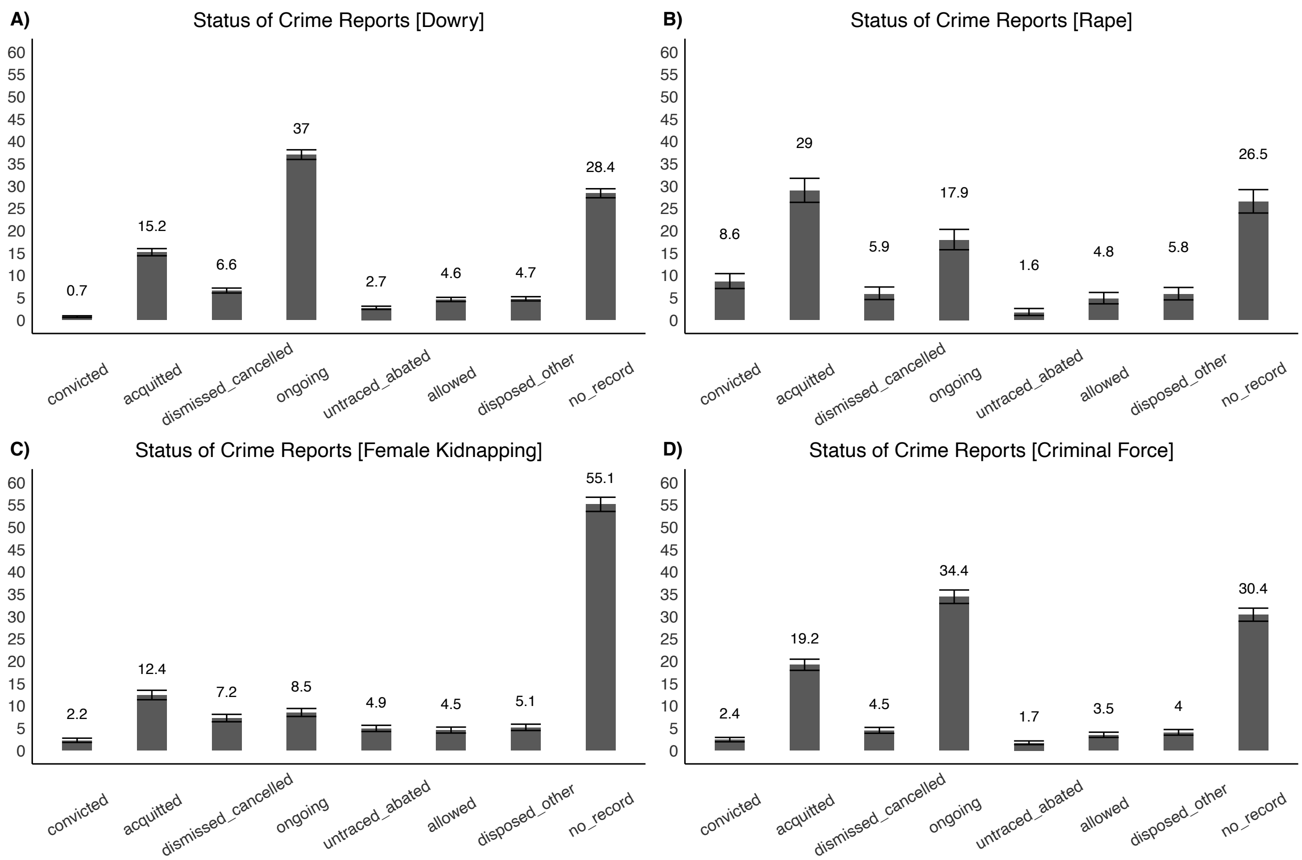


Note: FIRs that could be merged with judicial records. Figure reveals rates of cases ongoing subset by particular Penal Code violations.

4 OUTCOMES (Function of All Registered Crime)

4.1 Cross-Tab

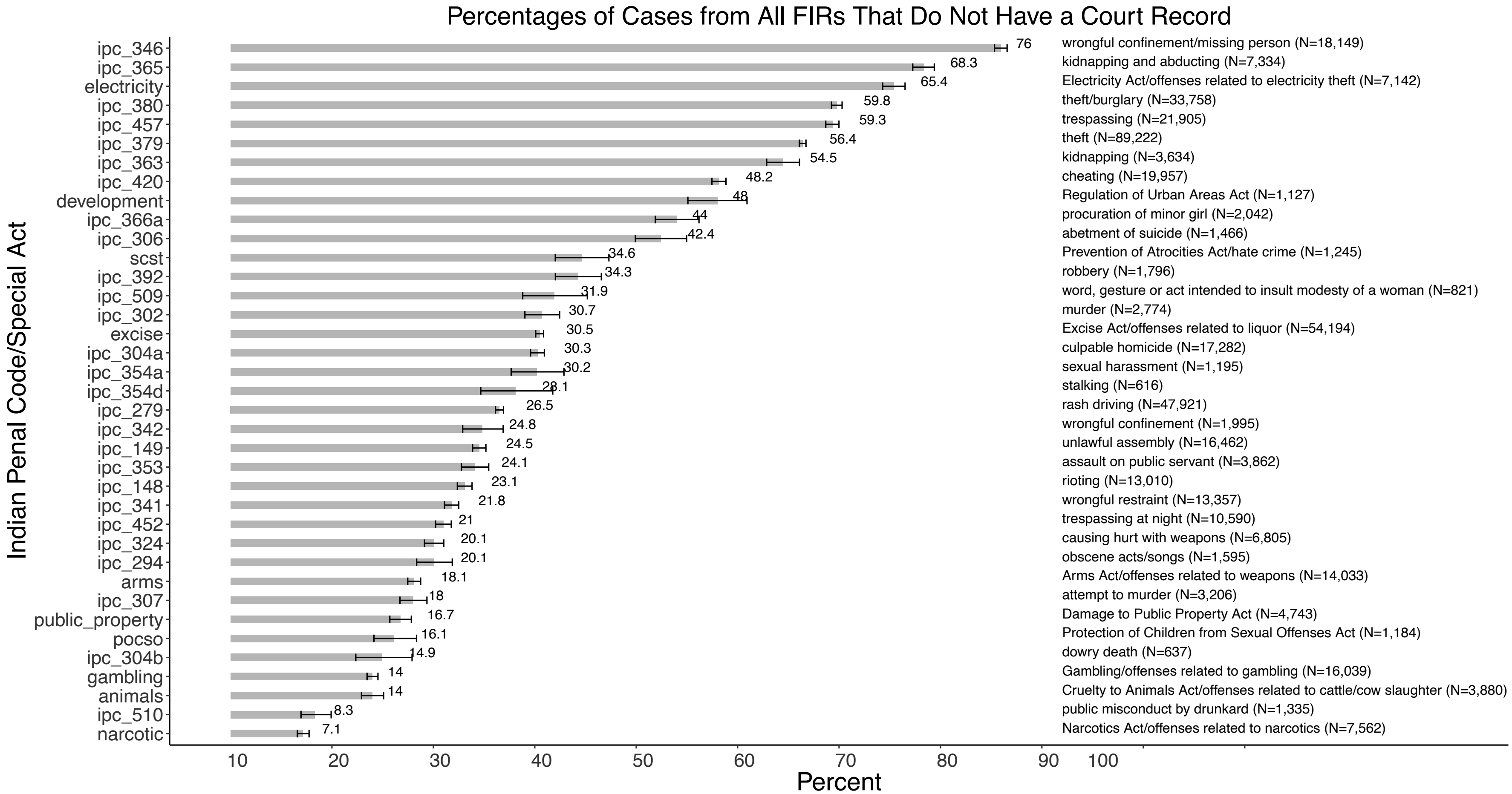
Figure A24: Crime Reports Statuses in the Judicial System [Specific Gendered Crime]



Note: FIRs that could be merged with judicial records. Panel A reflects dowry cases or those that invoked Section 498-A (N=7,674); Panel B highlights rape cases or those that invoked Section 376 (N=1,094); Panel C represents female kidnapping or Section 366 (N=3,754); Panel D reflects criminal force with intent to outrage a woman's modesty or Section 354 (N=3,804). 30% of gendered cases, except for female kidnapping, are cancelled at the stage of law enforcement.

4.2 Cancelled at Station/No Record in Court

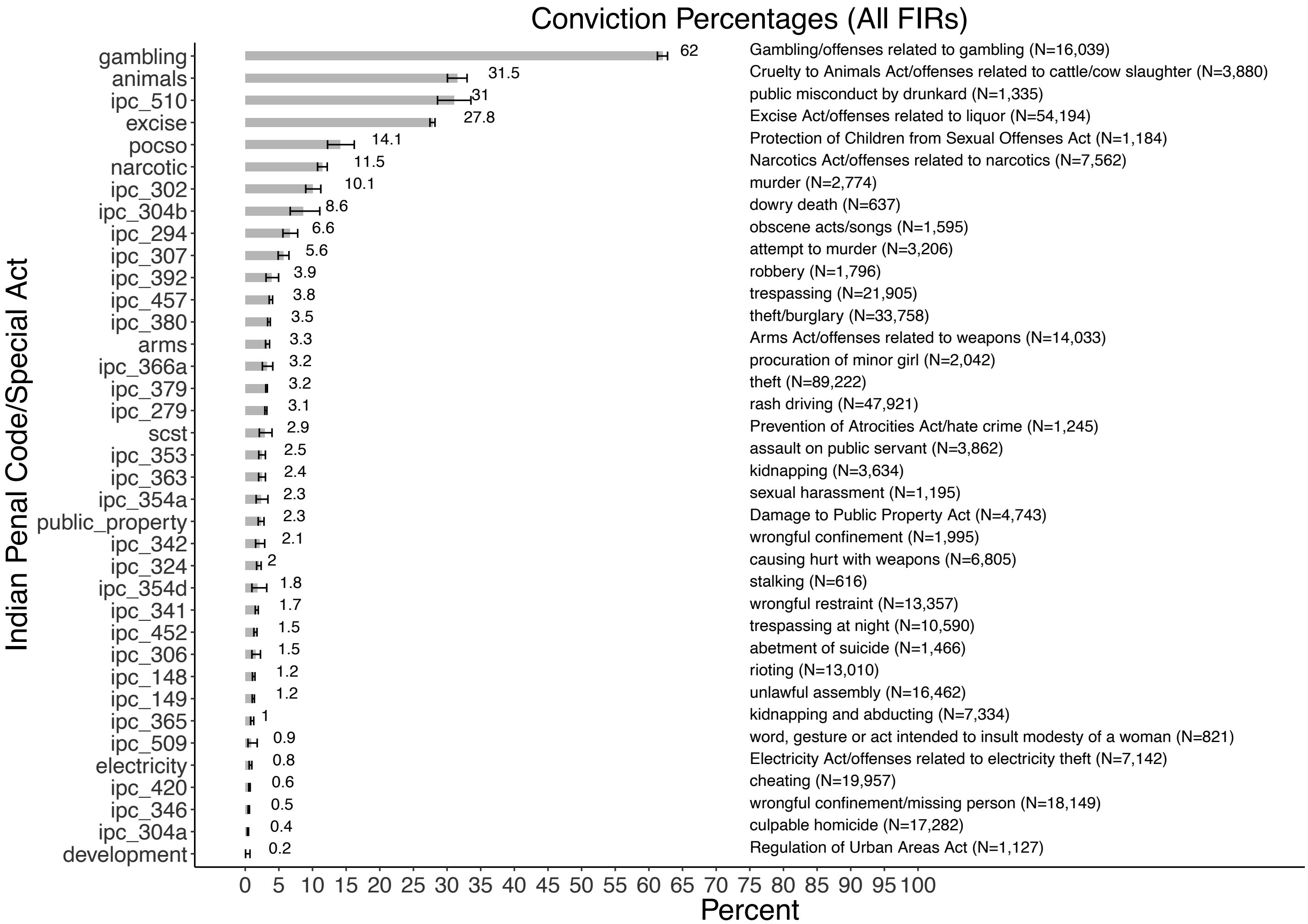
Figure A25: No Record Rates of Crime Reports Based on Specific Penal Code Violations



Note: Figure reveals rates of cases in the FIR dataset that could not be merged with court records/had no record in the judiciary, subset by particular Penal Code violations.

4.3 Conviction

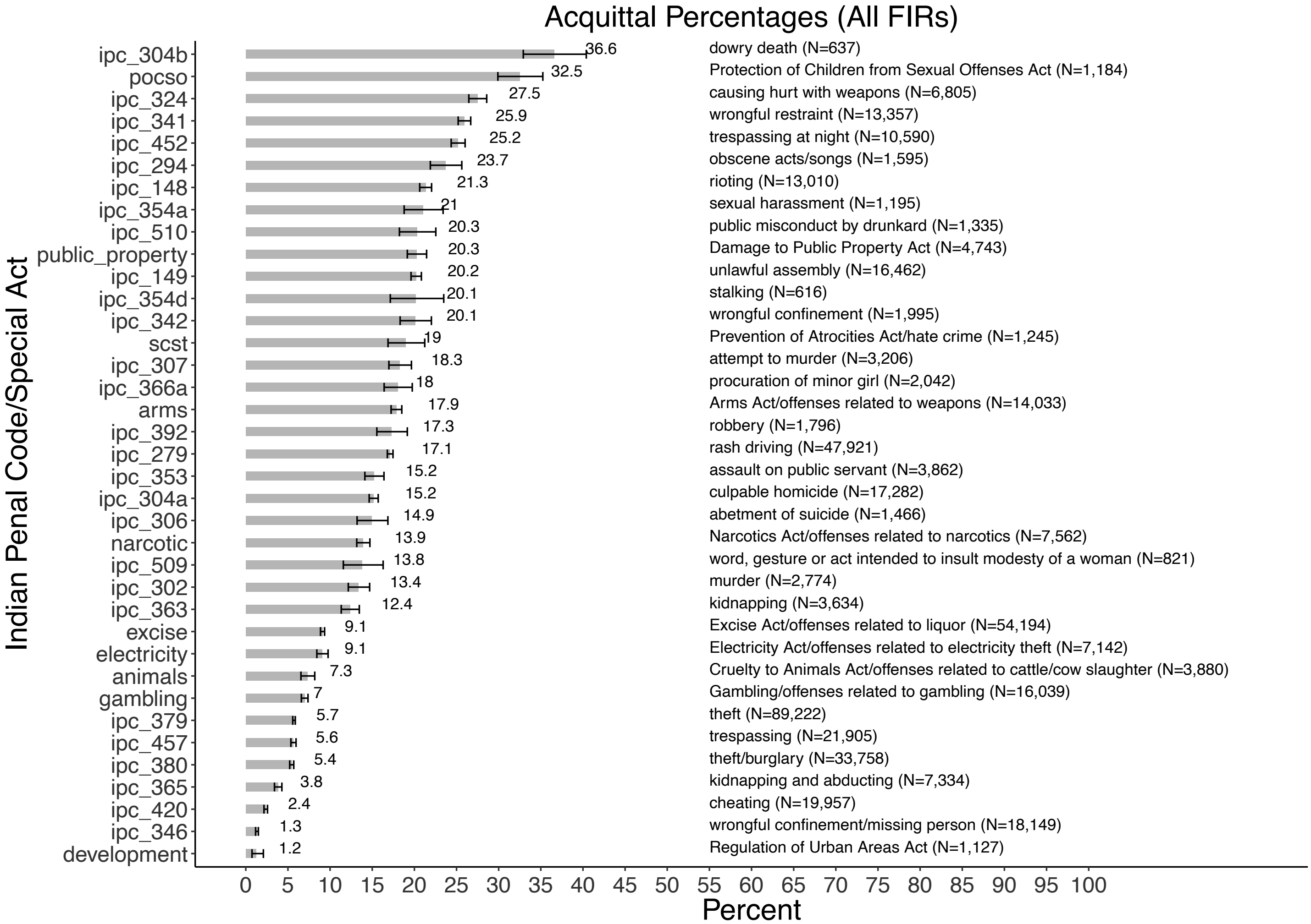
Figure A26: Conviction Rates of Crime Reports Based on Specific Penal Code Violations



Note: Figure reveals conviction rates by cases subset by particular Penal Code violations, as a function of all registered crime.

4.4 Acquittal

Figure A27: Acquittal Rates of Crime Reports Based on Specific Penal Code Violations



Note: Figure reveals acquittal rates by cases subset by particular Penal Code violations, as a function of all registered crime. Dowry death and child sexual assault have the highest rate of acquittals.

5 Additional Tests/Heterogenous Effects

Table A2

	Effects Controlling for Primary Penal Code						
	(1:Registration)	(2:Cancellation)	(3:Investigation)	(4:Court Dismissal)	(5: Court Duration)	(6: Acquittal)	(7: Conviction)
Female	31.529***	0.022***	0.622	0.004*	6.934**	0.001	−0.010***
	(5.666)	(0.005)	(2.438)	(0.002)	(2.999)	(0.005)	(0.002)
Constant	−15.810**	0.450***	135.789	−0.098***	369.940***	0.253**	0.485**
	(7.935)	(0.093)	(87.218)	(0.035)	(130.477)	(0.128)	(0.189)
Controls	Y	Y	Y	Y	Y	Y	Y
PS FE	Y	Y	Y	Y	Y	N	Y
Month-Yr FE	Y	Y	Y	Y	Y	Y	Y
Top/Primary IPC FE	Y	Y	Y	Y	Y	Y	Y

Note: *p<0.1; **p<0.05; ***p<0.01

Table A3

	Registration Duration			
	(1)	(2)	(3)	(4)
Female			8.532***	6.515**
			(2.294)	(2.552)
Dowry	301.380***	311.979***	248.411***	252.896***
	(35.408)	(36.706)	(38.566)	(41.591)
Rape	5.707	−2.533	3.304	−2.934
	(6.477)	(6.253)	(9.196)	(10.010)
Fem Kidnapping	−19.833***	−24.250***	−19.777***	−23.954***
	(2.498)	(2.480)	(2.409)	(2.405)
Criminal Force	−9.600**	−10.981**	−10.051*	−9.656*
	(4.200)	(4.715)	(5.314)	(5.577)
Female:Dowry			61.501**	71.882***
			(24.675)	(23.058)
Female:Rape			−3.205	−4.353
			(10.863)	(11.494)
Female:Fem Kidnapping			−5.088*	−4.745*
			(2.972)	(2.428)
Female:Criminal Force			−6.809	−7.846
			(6.369)	(7.702)
Constant	23.683***	−0.724	23.075***	−1.290
	(2.326)	(3.038)	(2.334)	(3.045)
Observations	381,836	360,022	381,836	360,022
R^2	0.025	0.038	0.026	0.038
Controls	N	Y	N	Y
PS FE	N	Y	N	Y
Month-Yr FE	N	Y	N	Y

Note: Controls include a numeric variable for how far the crime took place from a station, investigating officer rank, as well as whether the registering station is urban. PS stands for police station. Standard errors are clustered by district for all models. *Dowry has longest lag between incident and registration, while female kidnapping is registered sooner.* *p<0.1; **p<0.05; ***p<0.01

Table A4

	No Record in Court			
	(1)	(2)	(3)	(4)
Female			0.055***	0.051***
			(0.010)	(0.009)
Dowry	−0.109***	−0.083***	−0.104***	−0.090***
	(0.017)	(0.014)	(0.022)	(0.021)
Rape	−0.125***	−0.125***	−0.117***	−0.099***
	(0.015)	(0.022)	(0.027)	(0.037)
Fem Kidnapping	0.154***	0.161***	0.163***	0.175***
	(0.028)	(0.026)	(0.028)	(0.028)
Criminal Force	−0.079***	−0.083***	−0.103***	−0.117***
	(0.013)	(0.008)	(0.024)	(0.018)
Female:Dowry			−0.054***	−0.033**
			(0.015)	(0.016)
Female:Rape			−0.047	−0.066
			(0.047)	(0.052)
Female:Fem Kidnapping			−0.074***	−0.098***
			(0.028)	(0.031)
Female:Criminal force			−0.008	0.008
			(0.026)	(0.026)
Constant	0.400***	0.395***	0.395***	0.390***
	(0.018)	(0.013)	(0.018)	(0.013)
Observations	418,190	382,265	418,190	382,265
R^2	0.002	0.113	0.003	0.114
Controls	N	Y	N	Y
PS FE	N	Y	N	Y
Month-Yr FE	N	Y	N	Y

Note: Controls include a numeric variable for how far the crime took place from a station, investigating officer rank, as well as whether the registering station is urban. PS stands for police station. Standard errors are clustered by district for all models. *Most gendered crime types are likely to be sent to court, except female kidnapping which is significantly likely to be cancelled by law enforcement.* *p<0.1; **p<0.05; ***p<0.01

Table A5

	Investigation Duration				Dismissal			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female			14.214***	16.011***			0.024***	0.011***
			(4.195)	(2.780)			(0.003)	(0.003)
Dowry	−4.337	−0.105	−7.392	−2.633	0.047***	0.037***	0.039***	0.007
	(6.999)	(6.940)	(10.275)	(9.465)	(0.009)	(0.010)	(0.010)	(0.011)
Rape	−42.884***	−44.607***	−38.742**	−30.830*	0.022***	−0.083***	0.040**	−0.074***
	(10.722)	(9.212)	(17.220)	(16.439)	(0.007)	(0.011)	(0.018)	(0.020)
Fem Kidnapping	85.726***	85.916***	87.767***	90.886***	0.117***	0.043**	0.121***	0.047**
	(17.214)	(13.304)	(19.419)	(15.965)	(0.013)	(0.018)	(0.014)	(0.020)
Criminal Force	−36.472***	−33.762***	−31.151**	−26.876**	0.013***	0.002	0.023***	0.004
	(7.210)	(7.055)	(13.342)	(13.406)	(0.005)	(0.004)	(0.008)	(0.008)
Female:Dowry			−8.074	−9.884			−0.010	0.030***
			(8.264)	(8.506)			(0.008)	(0.009)
Female:Rape			−14.728	−27.627**			−0.039*	−0.019
			(10.799)	(10.887)			(0.020)	(0.022)
Female:Fem Kidnapping			−18.056	−31.868			−0.036***	−0.024
			(19.970)	(21.005)			(0.013)	(0.018)
Female:Criminal Force			−18.076	−21.552			−0.031***	−0.011
			(12.501)	(13.865)			(0.009)	(0.010)
Constant	127.997***	117.278***	127.036***	116.103***	0.043***	0.007	0.042***	0.006
	(5.830)	(16.823)	(5.946)	(16.729)	(0.003)	(0.008)	(0.003)	(0.008)
Observations	248,920	227,315	248,920	227,315	251,804	229,954	251,804	229,954
R^2	0.002	0.071	0.002	0.071	0.003	0.085	0.004	0.085
Controls	N	Y	N	Y	N	Y	N	Y
PS FE	N	Y	N	Y	N	Y	N	Y
Month-Yr FE	N	Y	N	Y	N	Y	N	Y

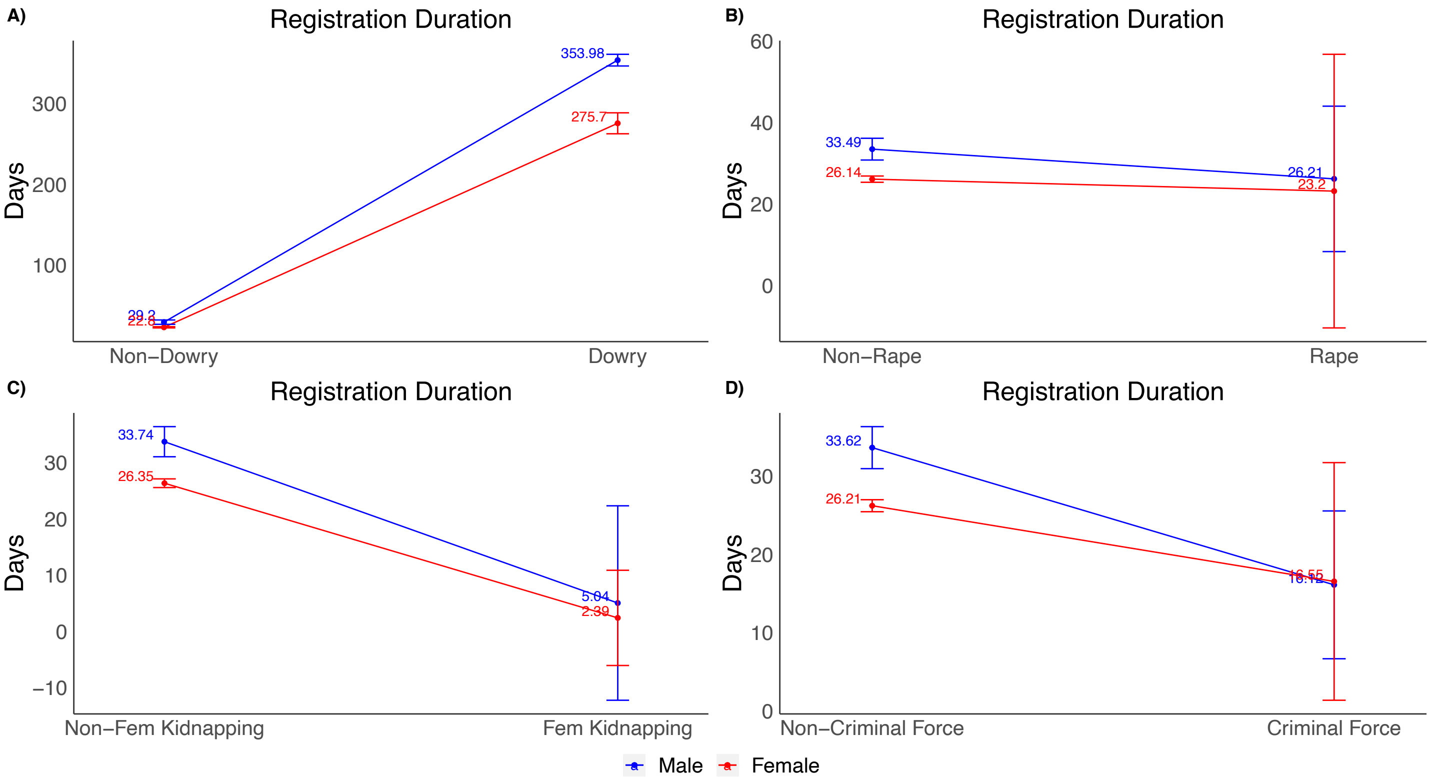
Note: Controls include a numeric variable for how far the crime took place from a station, investigating officer rank, judge rank, as well as whether the registering station is urban. PS stands for police station. Standard errors are clustered by district for all models. *Rape (by a non-spouse) is investigated quickest, while female kidnapping takes longest. There are rules in place that mandate that IPC 376 cases be investigated within 2-3 months.* *p<0.1; **p<0.05; ***p<0.01

Table A6

	Duration in Court				Acquittal				Conviction			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female			29.068**	34.631***			0.061***	0.056***			−0.121***	−0.103***
			(12.129)	(10.227)			(0.008)	(0.006)			(0.011)	(0.010)
Dowry	112.164***	73.441***	83.907***	52.500***	0.032**	0.012	0.072***	0.043***	−0.154***	−0.125***	−0.151***	−0.108***
	(13.098)	(11.907)	(19.044)	(18.829)	(0.016)	(0.011)	(0.017)	(0.014)	(0.013)	(0.016)	(0.015)	(0.022)
Rape	−55.407***	−55.553***	−82.536***	−60.725***	0.204***	0.141***	0.173***	0.112***	−0.019	0.063***	−0.038	0.058
	(15.271)	(16.090)	(18.203)	(22.357)	(0.026)	(0.024)	(0.042)	(0.038)	(0.023)	(0.018)	(0.033)	(0.036)
Fem Kidnapping	−140.816***	−97.035***	−146.490***	−102.858***	0.098***	0.087***	0.095***	0.080***	−0.121***	−0.062***	−0.127***	−0.064***
	(15.037)	(16.850)	(16.174)	(20.384)	(0.023)	(0.017)	(0.024)	(0.019)	(0.013)	(0.015)	(0.014)	(0.016)
Criminal Force	42.188***	33.545***	43.012***	28.422***	0.094***	0.083***	0.089***	0.075***	−0.117***	−0.081***	−0.127***	−0.084***
	(12.649)	(8.678)	(13.319)	(9.590)	(0.023)	(0.018)	(0.023)	(0.017)	(0.014)	(0.011)	(0.019)	(0.017)
Female:Dowry			11.691	−1.691			−0.104***	−0.088***			0.099***	0.064***
			(15.842)	(16.686)			(0.011)	(0.012)			(0.013)	(0.016)
Female:Rape			13.122	−18.552			−0.004	−0.003			0.108***	0.077**
			(17.300)	(19.244)			(0.033)	(0.034)			(0.029)	(0.037)
Female:Fem Kidnapping			10.542	10.175			−0.017	0.006			0.097***	0.062***
			(16.527)	(29.938)			(0.020)	(0.028)			(0.015)	(0.017)
Female:Criminal Force			−23.479*	−18.987*			−0.038*	−0.031*			0.105***	0.081***
			(12.356)	(11.307)			(0.022)	(0.018)			(0.016)	(0.017)
Constant	334.373***	549.745***	332.429***	547.205***	0.171***	0.397***	0.167***	0.392***	0.173***	0.262***	0.181***	0.270***
	(12.250)	(32.884)	(12.730)	(32.827)	(0.017)	(0.018)	(0.017)	(0.018)	(0.014)	(0.028)	(0.015)	(0.028)
Observations	250,287	228,542	250,287	228,542	251,804	229,954	251,804	229,954	251,804	229,954	251,804	229,954
R^2	0.003	0.201	0.004	0.201	0.002	0.123	0.004	0.125	0.006	0.098	0.012	0.102
Controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
PS FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Month-Yr FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y

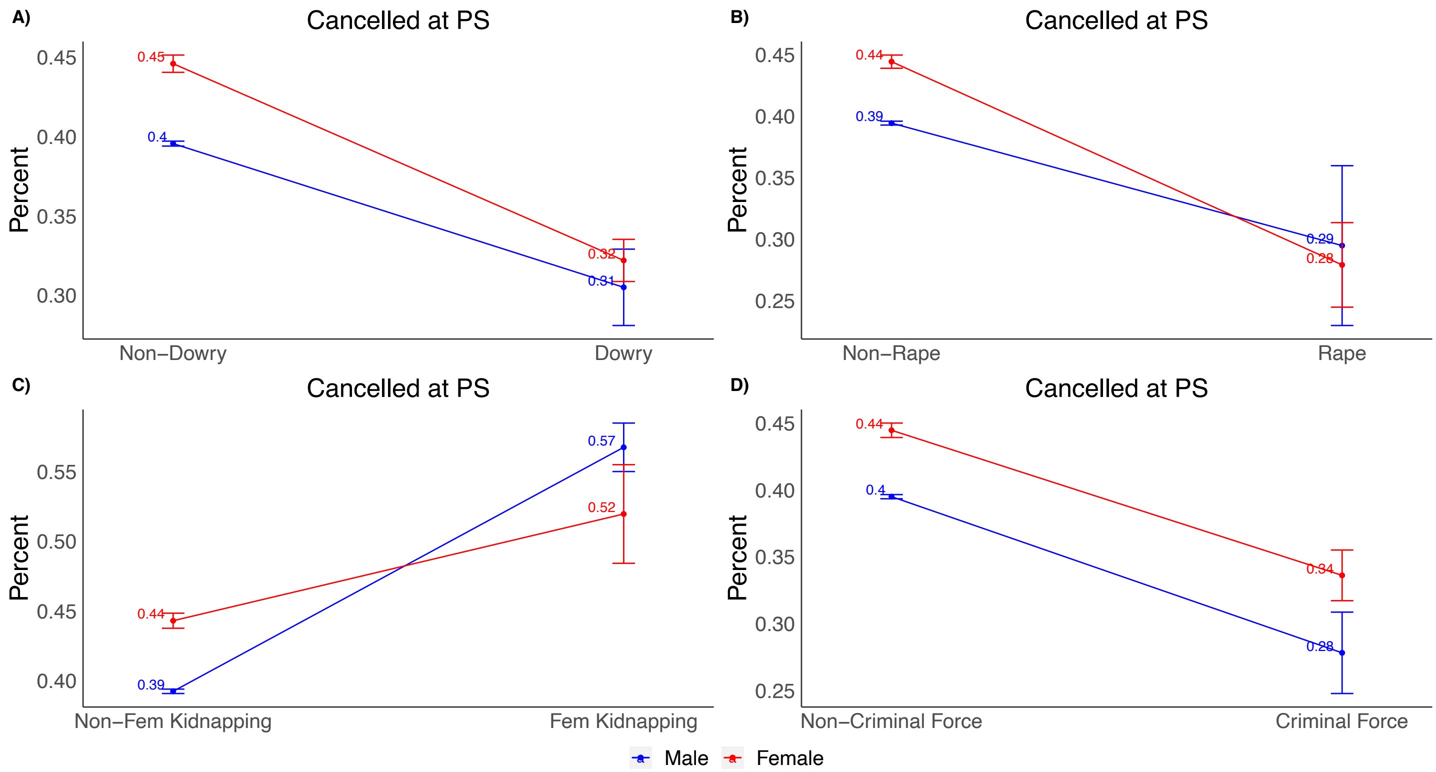
Note: Controls include a numeric variable for how far the crime took place from a station, investigating officer rank, judge rank, as well as whether the registering station is urban. PS stands for police station. Standard errors are clustered by district for all models. Dowry spends longest stalled in court. *Generally, all gendered sub-types are significantly more likely to have a suspect acquitted rather than convicted.* *p<0.1; **p<0.05; ***p<0.01

Figure A28: Average Marginal Effects (Table A3)



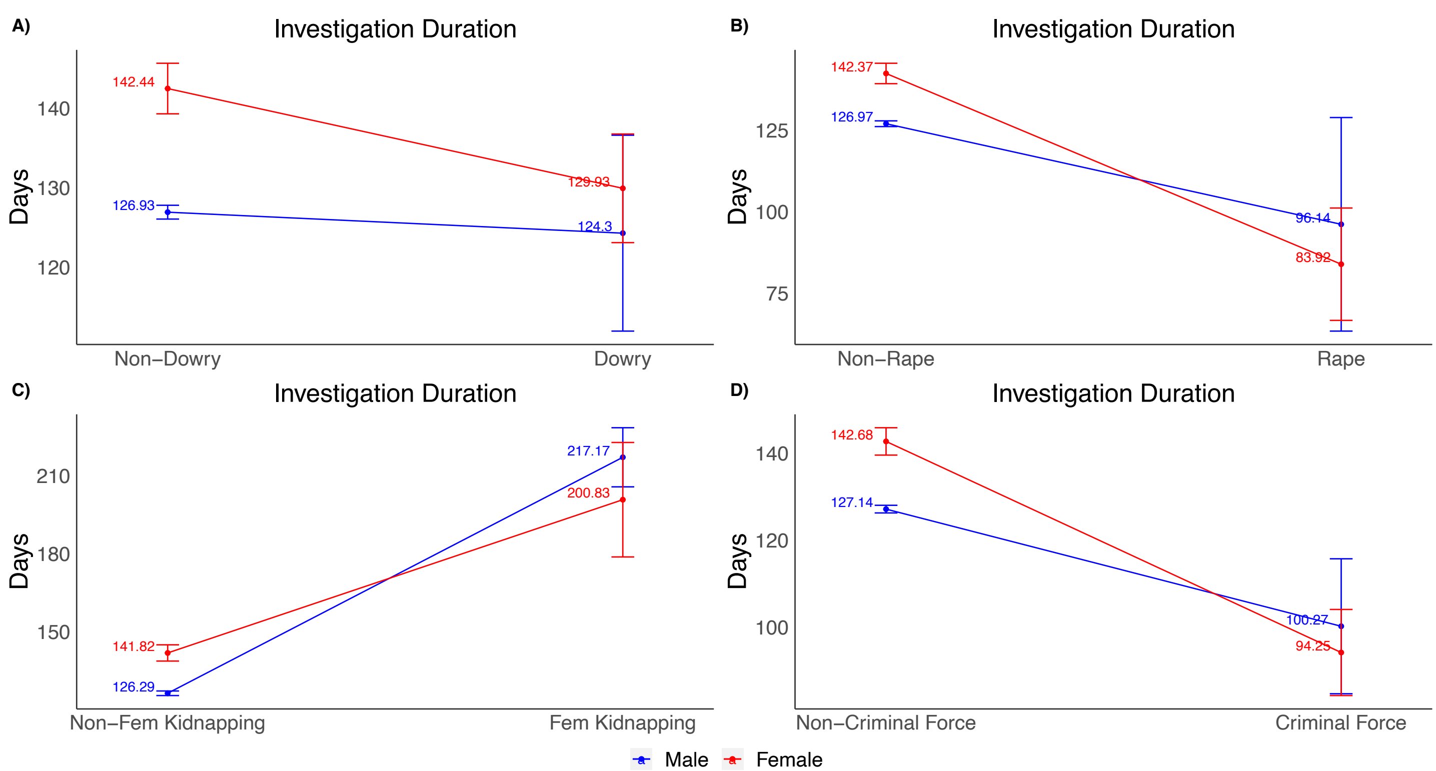
Note: Based on column 4, in Table A3.

Figure A29: Average Marginal Effects (Table A4)



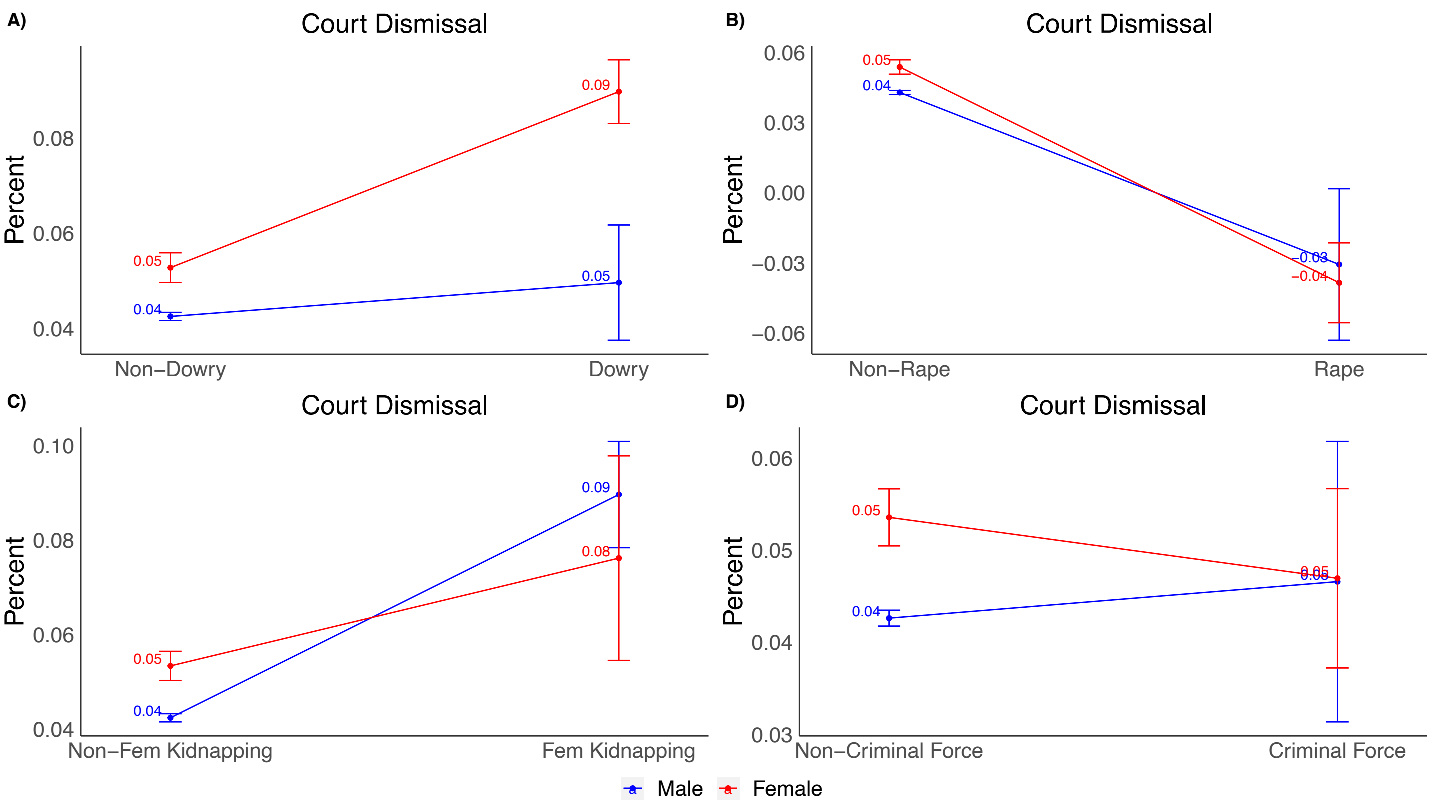
Note: Based on column 4, Table A4 .

Figure A30: Average Marginal Effects (Table A5, Investigation Duration)



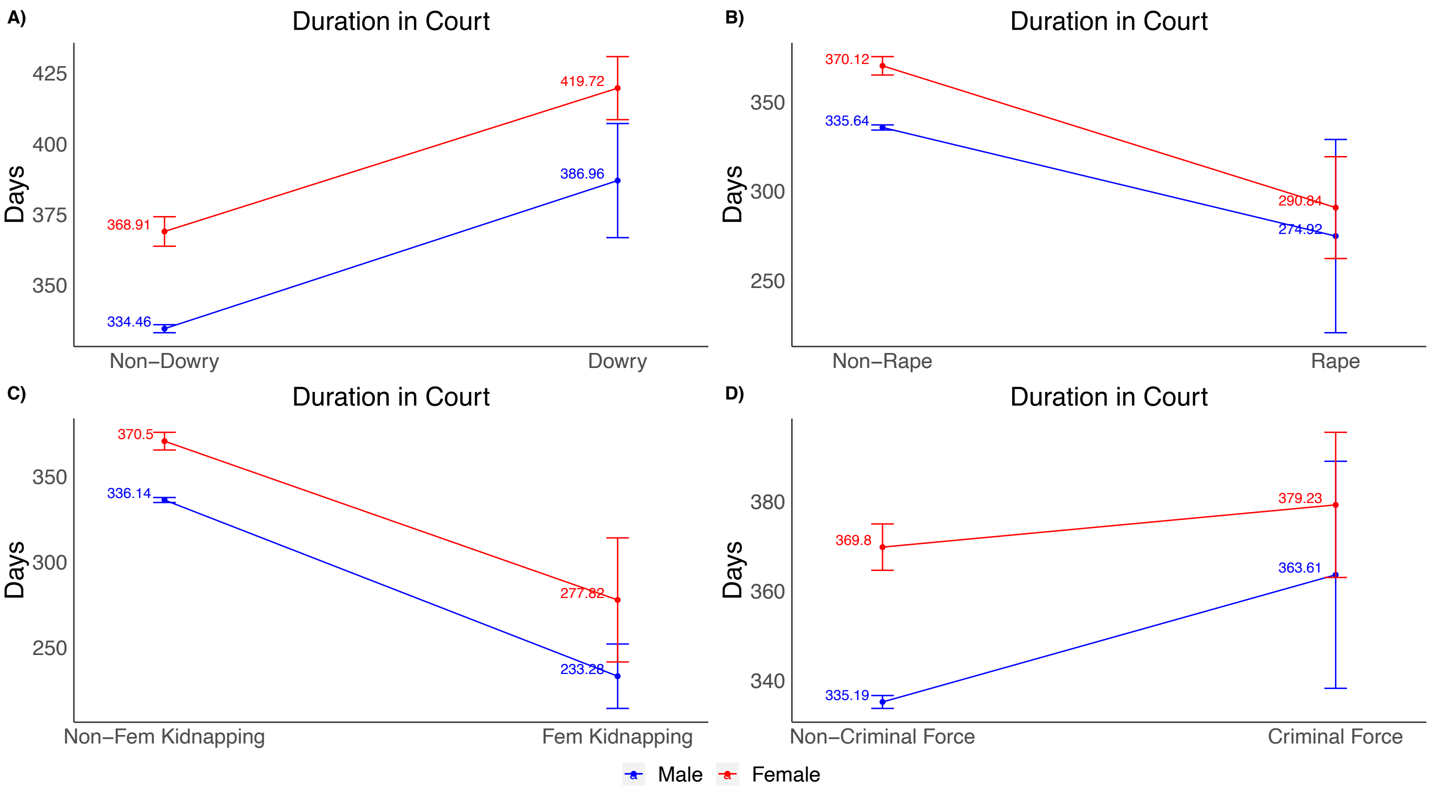
Note: Based on column 4, Table A5 .

Figure A31: Average Marginal Effects (Table A5, Court Dismissal)



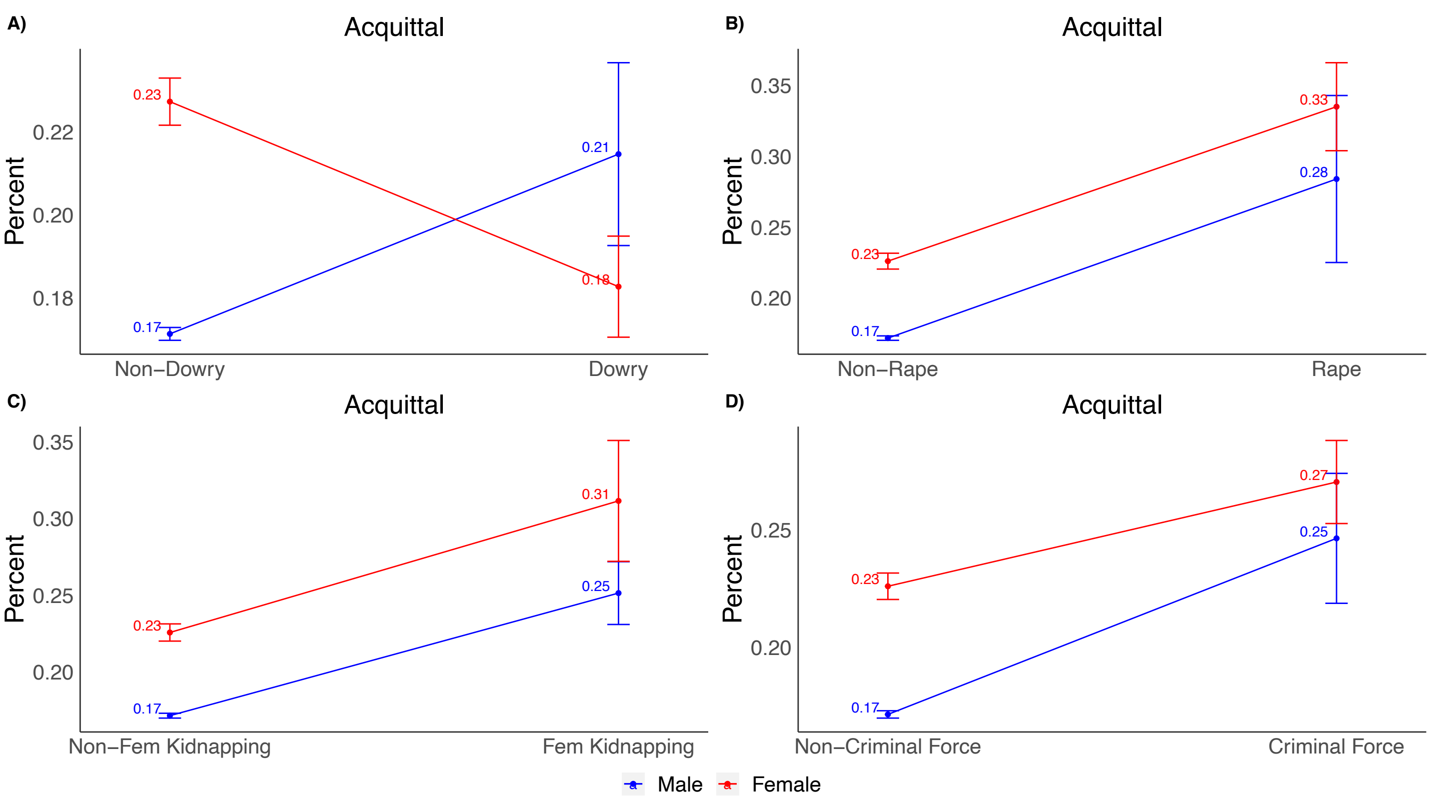
Note: Based on column 8, Table A5.

Figure A32: Average Marginal Effects (Table A6, Duration in Court)



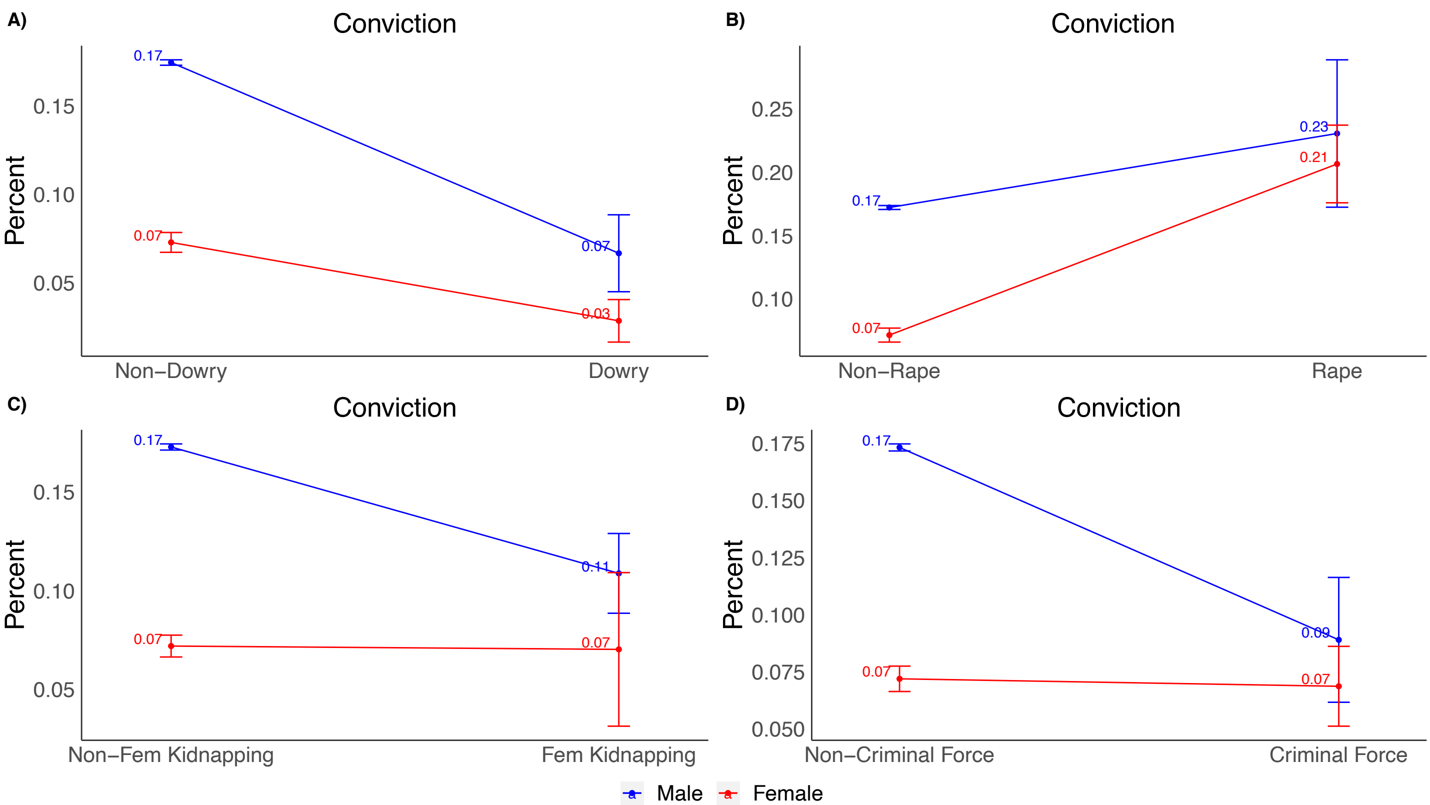
Note: Based on column 4, Table A6.

Figure A33: Average Marginal Effects (Table A6, Acquittal)



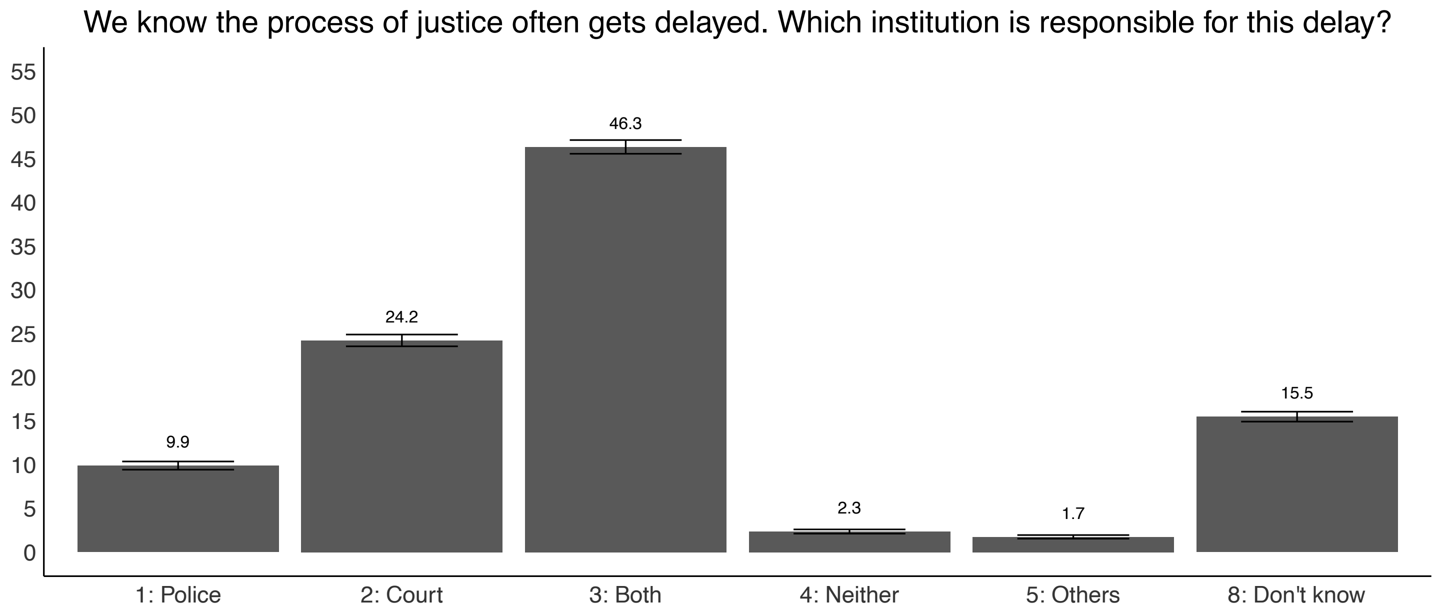
Note: Based on column 8, Table A6.

Figure A34: Average Marginal Effects (Table A6, Conviction)



Note: Based on column 12, Table A6. **Women are unlikely to have a suspect convicted in all categories compared to men.**

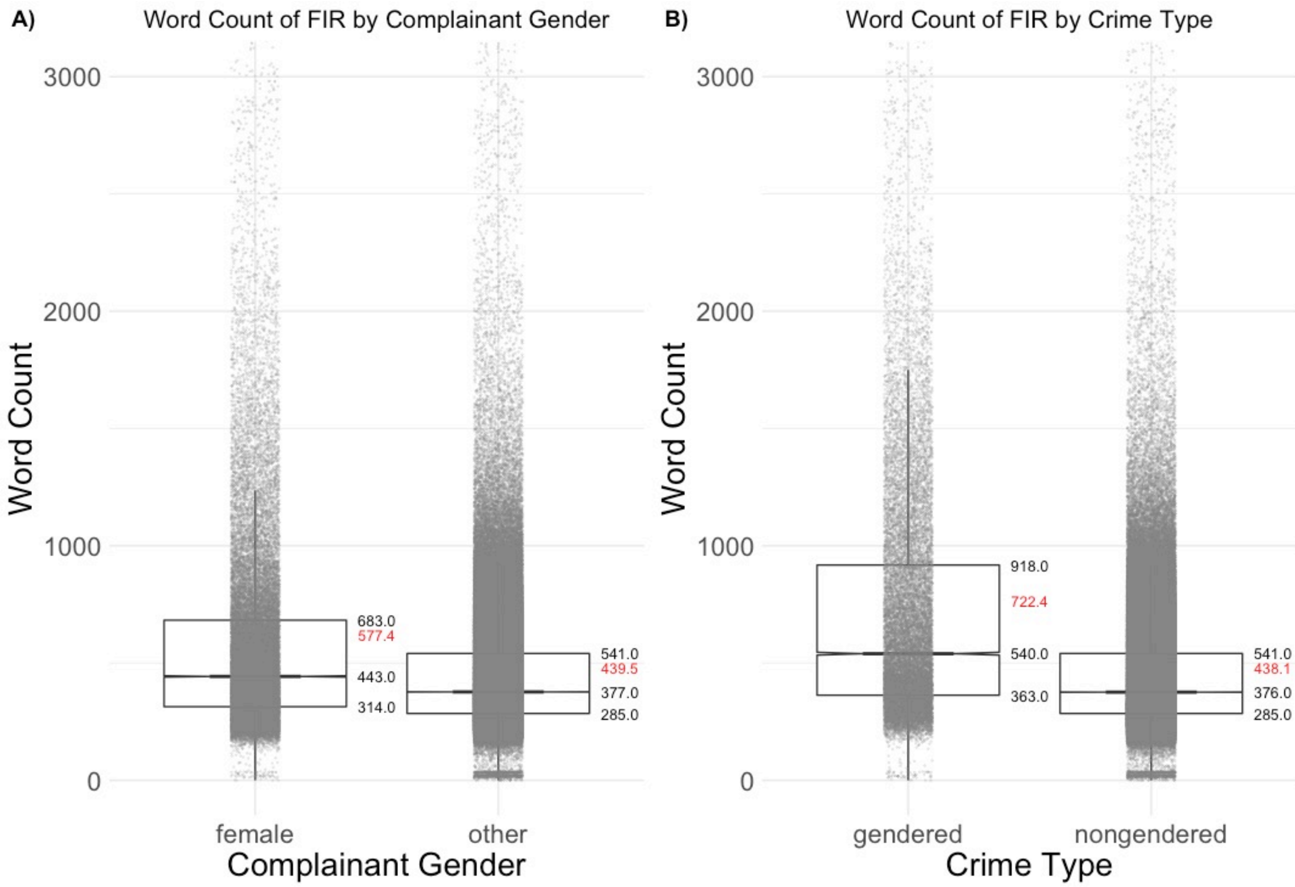
Figure A35: CSDS-Common Cause Survey: Which Institution is to Blame?



Note: Distribution of responses based on the Center for the Study of Developing Societies (CSDS)-Common Cause Survey 2017 (N=15,548).

6 Text-as-Data

Figure A36: Word Count by Complainant Gender and Crime Type



Note: Figure presents box plots for the word count by complainant gender and crime type, where each dot is a registered report (FIR). Y-axis is scaled to a maximum of 3000 words, for ease of visualization. 1st quartile, median, and third quartile included. Mean in red. **Women's cases and gendered crime are significantly longer in terms of the first-person testimonies/contain more detail about the offense.**

6.1 STM on Corpus of Crime

Topics that the machine generated can be identified with highest probability words (Panel A) as well as FREX or frequent and exclusive words to specific topics (Panel B). The top five most common crime types include "public intoxication" and "bootlegging" (Topic 19), "burglary" (Topic 16), "auto theft" (Topic 22/23), and "kidnapping" (Topic 27). As indicated in the FREX words, kidnapping cases usually involve women or girls as victims.[59] Self-explanatory topics include "fighting" (Topic 17),"gambling" (Topic 28), "phone theft" (Topic 26), "driving misdemeanor" (Topic 14), "robbery" (Topic 29), drugs or "narcotics" (Topic 31), and "phishing" (Topic 4).

Machine generated topics that may require additional context include the following: "electricity theft" refers to the illegal connection of wires to power grids (Topic 30). Topic 13 or "injury" can include cases in which a complainant has been hurt from hit-and-runs to construction accidents. Topic 15 refers to absconding from law enforcement or 'jumping bail.' Topic 6 represents cases related to the sand or mining mafia that smuggle or steal natural resources. Topic 24 refers to cases involving fraud and deception, typically financial.[60] Topic 18 or "arms" refer to cases involving unlicensed weapons manufacture and smuggling. The machine coded all cases involving the word 'Muslims' in Topic 25 or "minorities." Topic 1 or "unlicensed" refers to cases involving unlicensed doctors, fraudulent certificates, and fake medical exams. Topic 5 or "cattle" is illustrative of illegal smuggling of cows as well as cattle slaughter. Topic 9 or 'railway' refers to crimes committed in trains or railway platforms, while "accident/attack" or Topic 10 involves someone being attacked, including with a weapon. Topic 20 or "property" and Topic 32 or "real estate" refer to cases involving property and real estate disputes, respectively. Relatedly, Topic 12 or "development" represent illegal land purchases, including by corporations.

Appendix Figure A39 highlight the top topics that are disproportionately associated with female complainants. These include "dowry-A" and "dowry-B" (Topics 3 and 8), as well as "lewd behavior" (Topic 11).[61] 'Lewd behavior' encapsulates cases from blackmailing women in releasing compromising photos[62] to harassing women in public places. The case most likely associated with female complainants are dowry cases, wherein a victim complains to the police about the physical, mental, and emotional abuse her husband and in-laws perpetrate, usually in order to extort money from her natal home. Appendix Figures A41-A44 highlight word clouds associated with each of the topics.

Figure A38 highlight the likelihood of conviction based on the topic metadata, as well as the correlation between topics. In Panel B we see that the machine correctly estimated the relationship between topics where, for instance, Topic 10 and Topic 13 ('accident' and 'injury') are related to each other, as are "cattle" and "minorities," suggesting that Muslims are disproportionately victimized for alleged offenses related to cow slaughter or smuggling. Similarly, "dowry-A," "dowry-B" and "lewd behavior" are all highly correlated in terms of the language used in the crime report. Cases involving public intoxication, fake currency, and gambling have higher rates of conviction. Nevertheless, the plot suggests that topics related to gendered crime, as well as those brought by female complainants, are unlikely to lead to formal punishment.

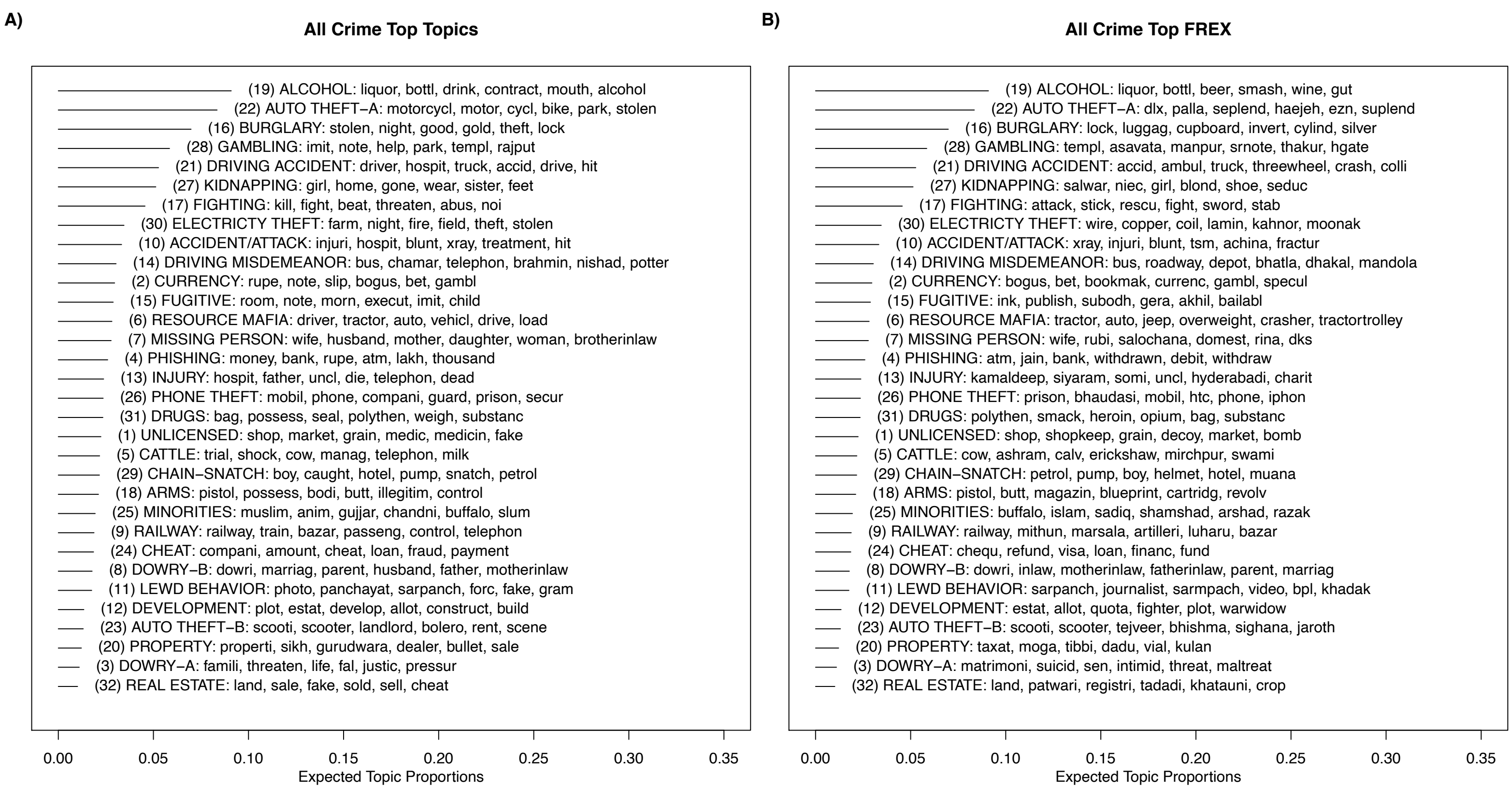
59. Kidnapping may also be closely connected with cases classified as missing persons.

60. These cases generally invoke Indian Penal Code Section 420.

61. Other cases associated with female complainants include phishing (Topic 4), fighting (Topic 17), kidnapping (Topic 27), robbery (Topic 29), and missing persons (Topic 7).

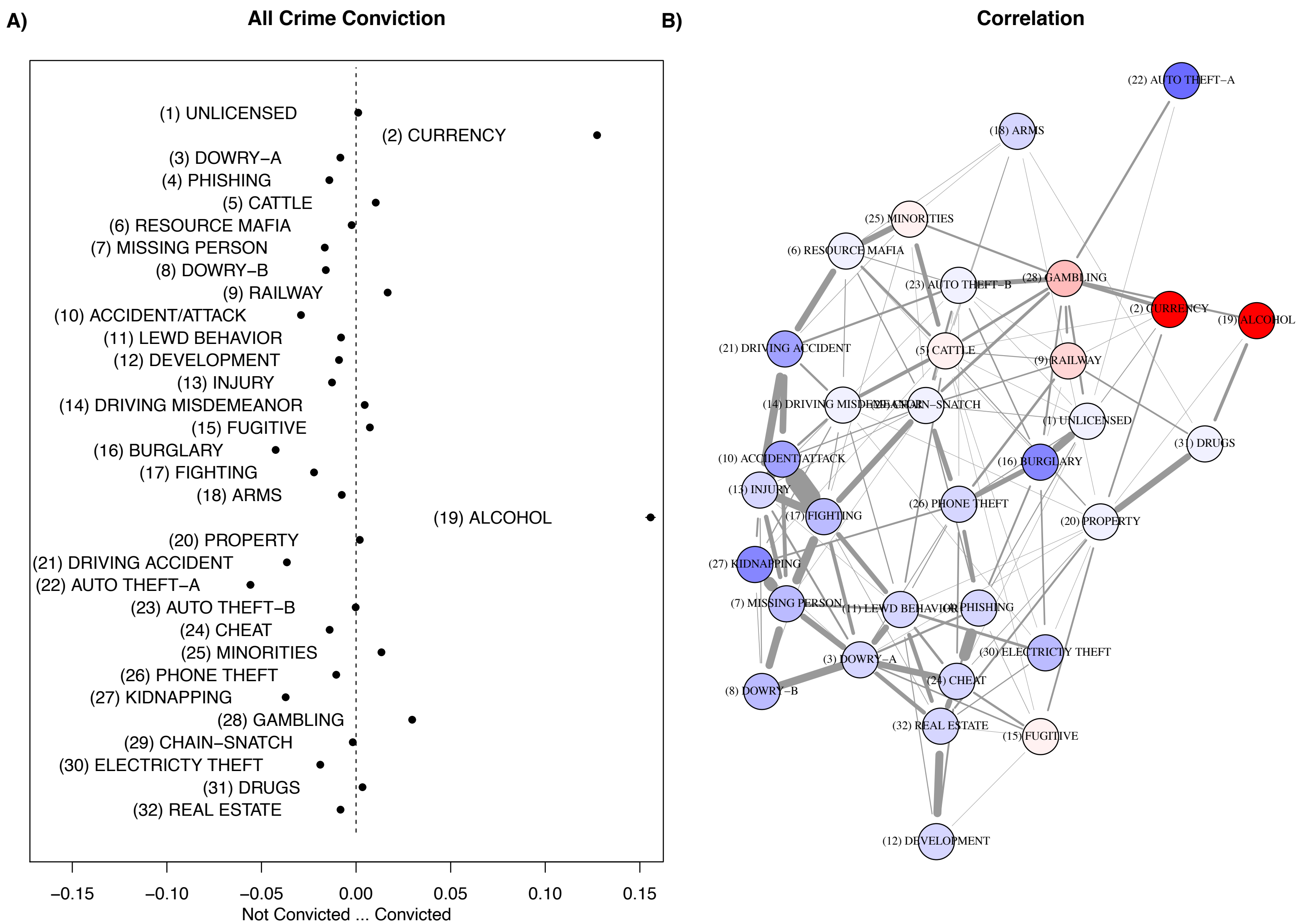
62. Cases associated with the Information Technology Act.

Figure A37: Top Topics (All Crime)



Note: Top topics for entire corpus (N=418,190).

Figure A38: Conviction Rate and Correlation of Topics Across Corpus

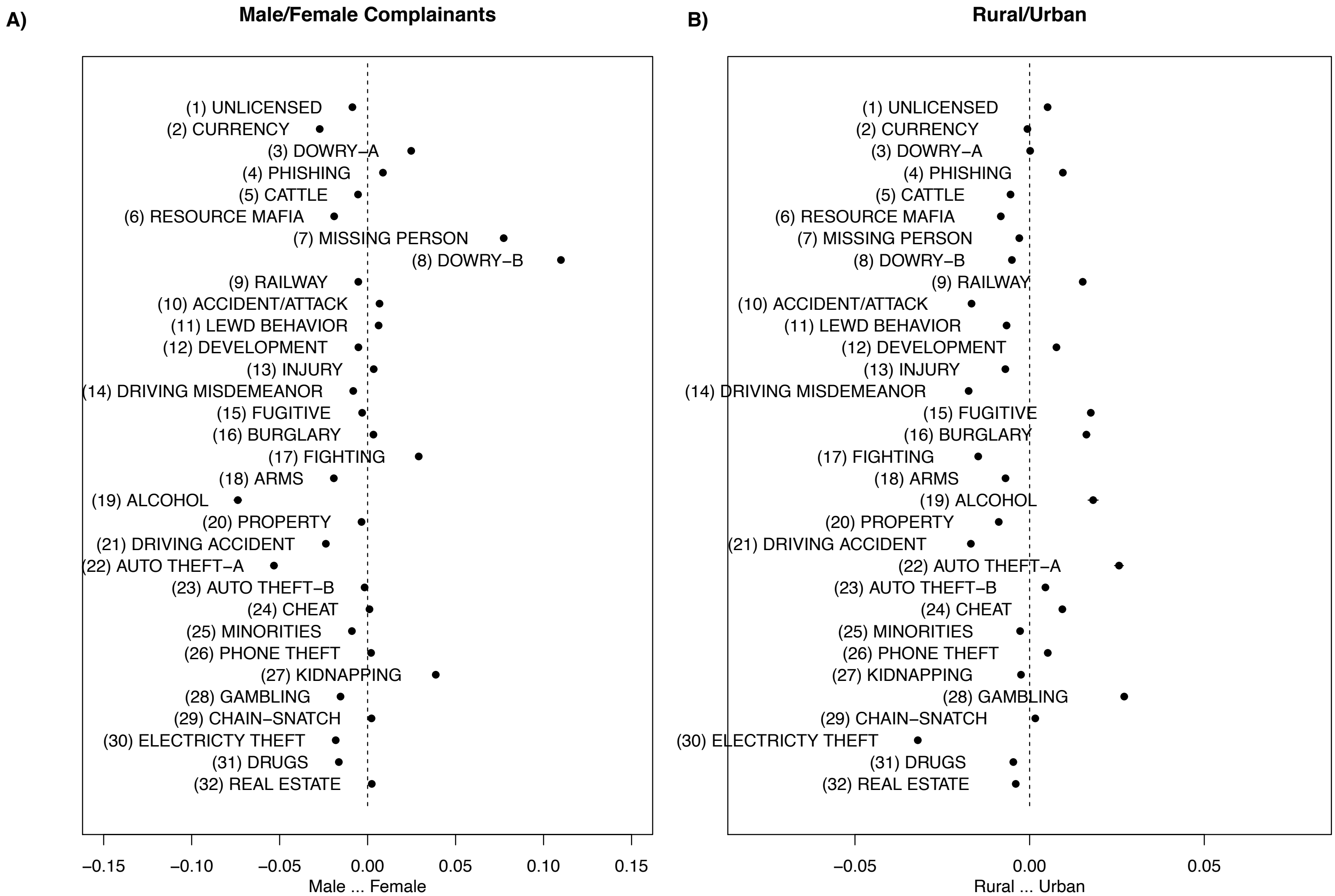


Note: Left: STM estimation of all cases with binary indicator for whether the topic resulted in conviction. *Right:* Network of correlated topics where node color indicates magnitude of regression coefficients (redder nodes indicate positive and bluer negative). Edge width is proportional to the strength of correlation. *Fake currency and public intoxication have better conviction rates, conditional on registration.*

Table A7: Top Word Stems by Topic With FREX (All Crime)

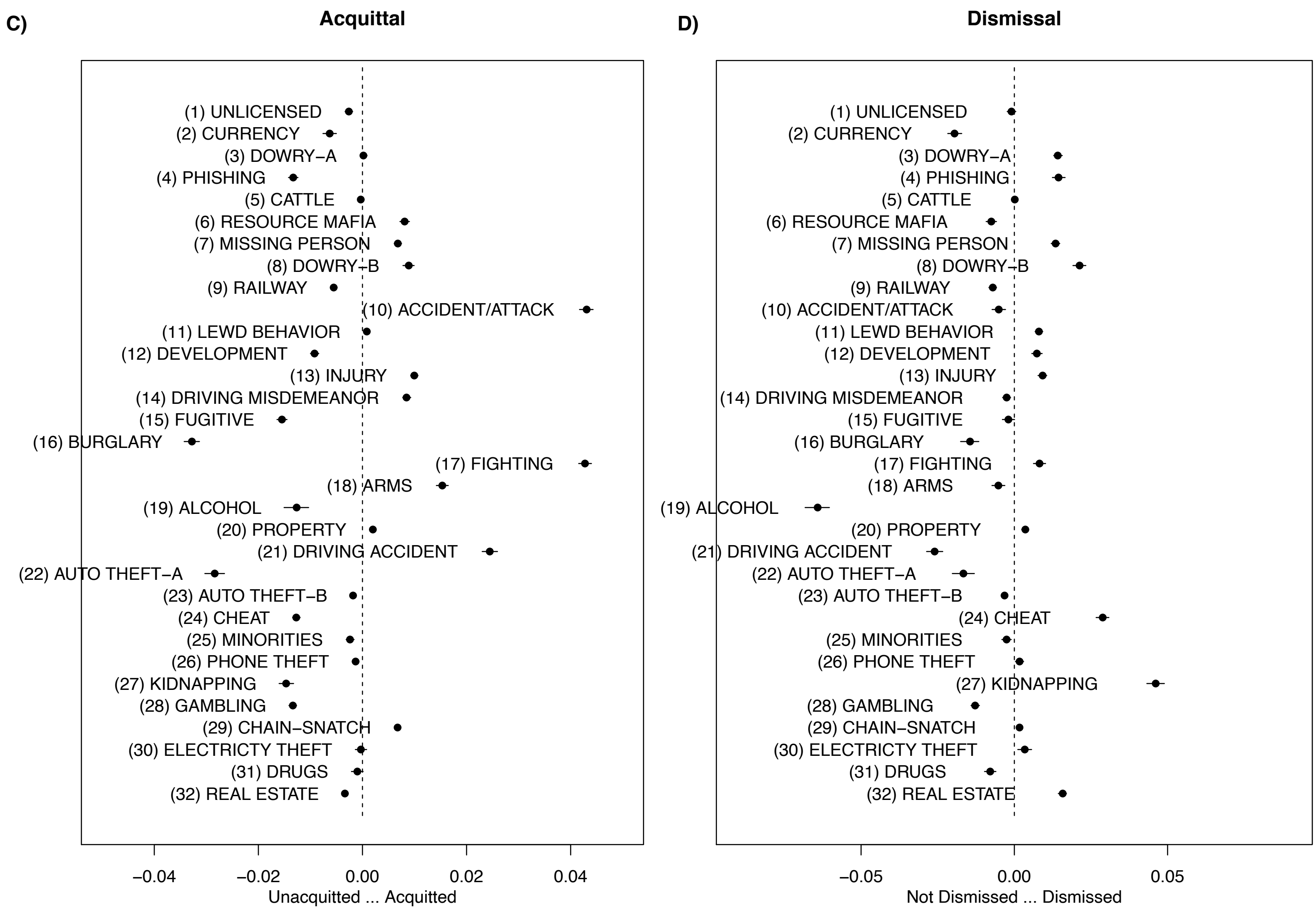
Topic	Top Words
(1) UNLICENSED	Highest Prob: shop, market, grain, medic, medicin, fake, food FREX: shop, shopkeep, grain, decoy, market, bomb, mphw
(2) CURRENCY	Highest Prob: rupe, note, slip, bogus, bet, gambl, amount FREX: bogus, bet, bookmak, currenc, gambl, specul, baj
(3) DOWRY-A	Highest Prob: famili, threaten, life, fal, justic, pressur, suicid FREX: matrimoni, suicid, sen, intimid, threat, maltreat, compromi
(4) PHISHING	Highest Prob: money, bank, rupe, atm, lakh, thousand, check FREX: atm, jain, bank, withdrawn, debit, withdraw, dairi
(5) CATTLE	Highest Prob: trial, shock, cow, manag, telephon, milk, munshi FREX: cow, ashram, calv, erickshaw, mirchpur, swami, morni
(6) RESOURCE MAFIA	Highest Prob: driver, tractor, auto, vehicl, drive, load, rajasthan FREX: tractor, auto, jeep, overweight, crasher, tractortrolley, amw
(7) MISSING PERSON	Highest Prob: wife, husband, mother, daughter, woman, brotherinlaw, children FREX: wife, rubi, salochana, domest, rina, dks, hemlata
(8) DOWRY-B	Highest Prob: dowri, marriag, parent, husband, father, motherinlaw, demand FREX: dowri, inlaw, motherinlaw, fatherinlaw, parent, marriag, taunt
(9) RAILWAY	Highest Prob: railway, train, bazar, passeng, control, telephon, ticket FREX: railway, mithun, marsala, artilleri, luharu, bazar, srm
(10) ACCIDENT/ATTACK	Highest Prob: injuri, hospit, blunt, xray, treatment, hit, hurt FREX: xray, injuri, blunt, tsm, achina, fractur, natija
(11) LEWD BEHAVIOR	Highest Prob: photo, panchayat, sarpanch, forc, fake, gram, villag FREX: sarpanch, journalist, sarmpach, video, bpl, khadak, grafer
(12) DEVELOPMENT	Highest Prob: plot, estat, develop, allot, construct, build, municip FREX: estat, allot, quota, fighter, plot, warwidow, elig
(13) INJURY	Highest Prob: hospit, father, uncl, die, telephon, dead, death FREX: kamaldeep, siyaram, somi, uncl, hyderabadi, charit, poison
(14) DRIVING MISDEMEANOR	Highest Prob: bus, chamar, telephon, brahmin, nishad, potter, mob FREX: bus, roadway, depot, bhatla, dhakal, mandola, surewala
(15) FUGITIVE	Highest Prob: room, note, morn, execut, imit, child, destroy FREX: ink, publish, subodh, gera, akhil, bailabl, bhog
(16) BURGLARY	Highest Prob: stolen, night, good, gold, theft, lock, morn FREX: lock, luggag, cupboard, invert, cylind, silver, laptop
(17) FIGHTING	Highest Prob: kill, fight, beat, threaten, abus, noi, stick FREX: attack, stick, rescu, fight, sword, stab, noi
(18) ARMS	Highest Prob: pistol, possess, bodi, butt, illegitim, control, iron FREX: pistol, butt, magazin, blueprint, cartridg, revolv, wood
(19) ALCOHOL	Highest Prob: liquor, bottl, drink, contract, mouth, alcohol, control FREX: liquor, bottl, beer, smash, wine, gut, patio
(20) PROPERTY	Highest Prob: properti, sikh, gurudwara, dealer, bullet, sale, prevent FREX: taxat, moga, tibbi, dadu, vial, kulan, sukhjit
(21) DRIVING ACCIDENT	Highest Prob: driver, hospit, truck, accid, drive, hit, treatment FREX: accid, ambul, truck, threewheel, crash, colli, oxid
(22) AUTO THEFT-A	Highest Prob: motorcycl, motor, cycl, bike, park, stolen, theft FREX: dlx, palla, seplend, haejeh, ezn, suplend, mblhaameh
(23) AUTO THEFT-B	Highest Prob: scooti, scooter, landlord, bolero, rent, scene, sheep FREX: scooti, scooter, tejveer, bhishma, sighana, jaroth, jupit
(24) CHEAT	Highest Prob: compani, amount, cheat, loan, fraud, payment, paid FREX: chequ, refund, visa, loan, financ, fund, infrastructur
(25) MINORITIES	Highest Prob: muslim, anim, gujjar, chandni, buffalo, slum, cruelti FREX: buffalo, islam, sadiq, shamshad, arshad, razak, sahabuddin
(26) PHONE THEFT	Highest Prob: mobil, phone, compani, guard, prison, secur, manag FREX: prison, bhaudasi, mobil, htc, phone, iphon, emei
(27) KIDNAPPING	Highest Prob: girl, home, gone, wear, sister, feet, children FREX: salwar, niec, girl, blond, shoe, seduc, feet
(28) GAMBLING	Highest Prob: imit, note, help, park, templ, rajput, gali FREX: templ, asavata, manpur, srnote, thakur, hgate, banchari
(29) CHAIN-SNATCH	Highest Prob: boy, caught, hotel, pump, snatch, petrol, forc FREX: petrol, pump, boy, helmet, hotel, muana, mimarpur
(30) ELECTRICTY THEFT	Highest Prob: farm, night, fire, field, theft, stolen, electr FREX: wire, copper, coil, lamin, kahnor, moonak, sirsal
(31) DRUGS	Highest Prob: bag, possess, seal, polythen, weigh, substanc, search FREX: polythen, smack, heroin, opium, bag, substanc, narcot
(32) REAL ESTATE	Highest Prob: land, sale, fake, sold, sell, cheat, registri FREX: land, patwari, registri, tadadi, khatauni, crop, acr

Figure A39: All Crime I



Note: Missing persons, dowry, fighting, kidnapping are likely to have a female complainant. In Panel B, economic offenses (e.g. phishing, development and real estate disputes) are more likely to be urban.

Figure A40: All Crime II



Note:

Figure A41: Word Cloud for 1-8 Top Topics (All Crime)

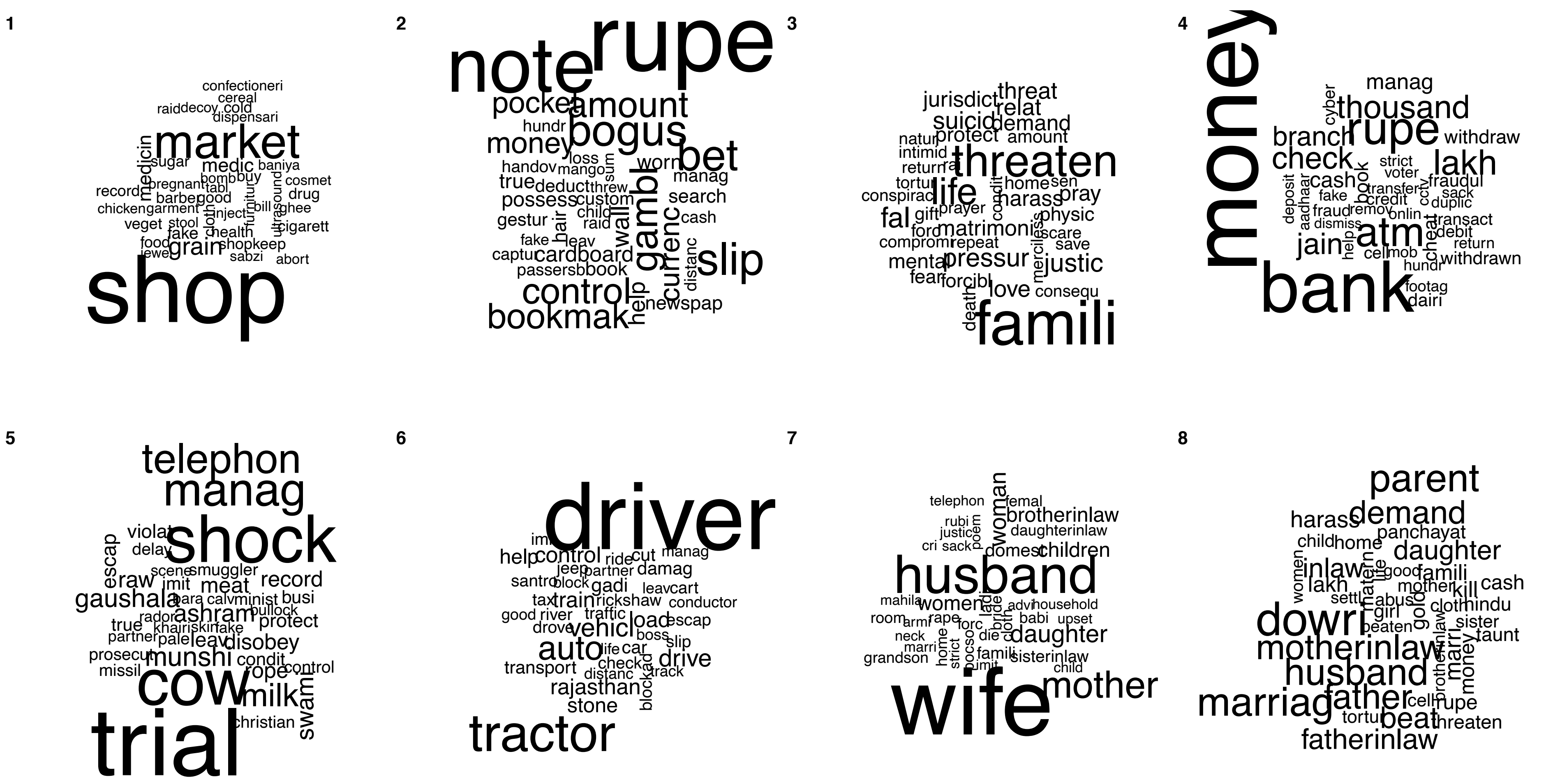


Figure A42: Word Cloud for 8-18 Top Topics (All Crime)

Figure A43: Word Cloud for 19-24 Top Topics (All Crime)

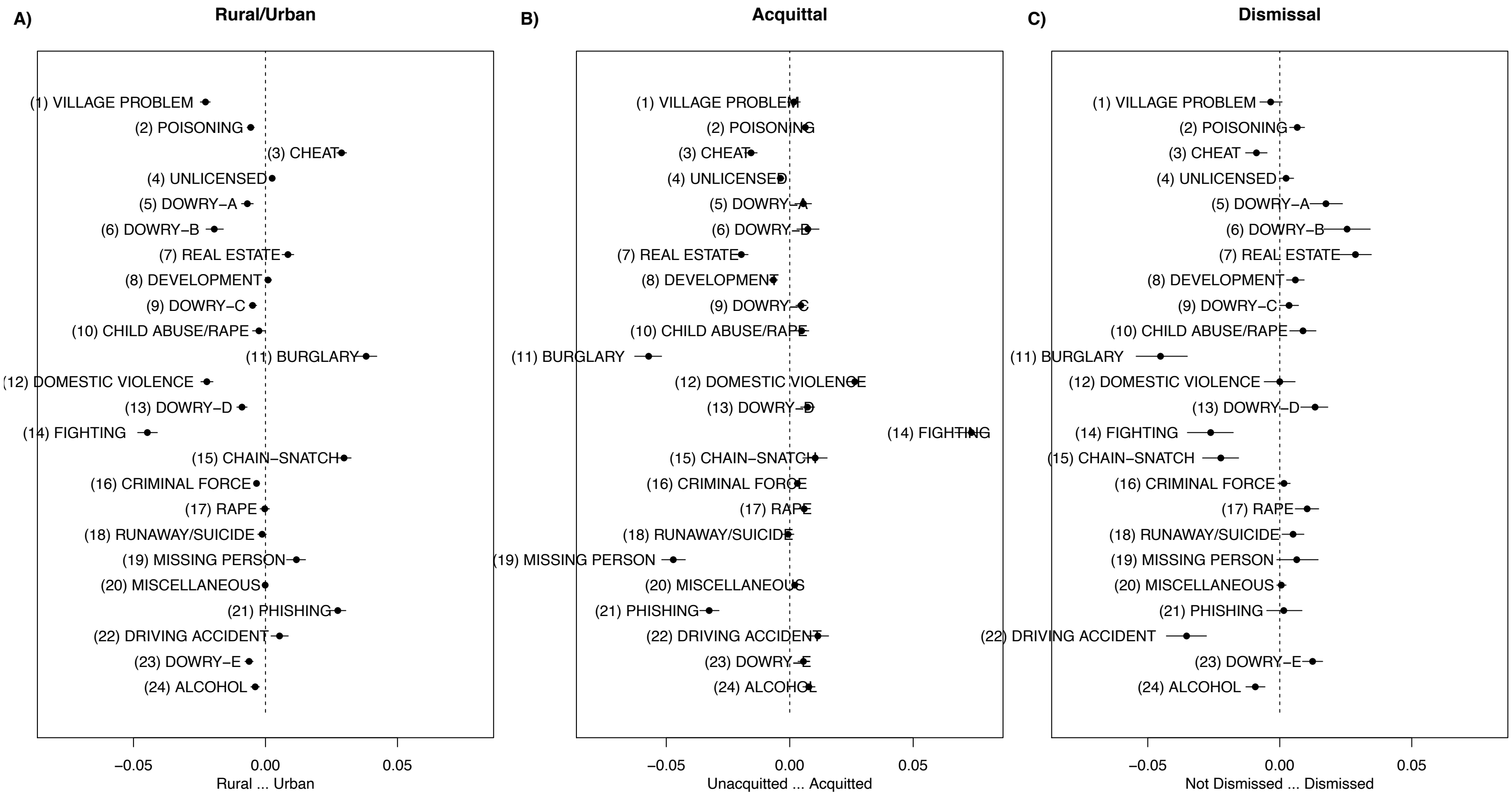
Figure A44: Word Cloud for 25-32 Top Topics (All Crime)

6.2 Female Complainants

Table A8: Top Word Stems by Topic With FREX (Female Complainants)

Topic	Top Words
(1) VILLAGE PROBLEM	Highest Prob: sikh, panchayat, sarpanch, farm, fire, field, land
	FREX: sarpanch, field, farm, gram, pistol, crop, cow
(2) POISONING	Highest Prob: hospit, uncl, matern, medicin, doctor, eat, die
	FREX: medicin, poison, uncl, bathroom, muslim, health, gujjar
(3) CHEAT	Highest Prob: note, gali, park, compani, jain, imit, floor
	FREX: jain, note, bazar, park, gali, thakur, bhagwan
(4) UNLICENSED	Highest Prob: fake, medic, board, educ, record, certif, princip
	FREX: princip, examin, educ, board, patient, salochana, fake
(5) DOWRY-A	Highest Prob: parent, marriag, demand, dowri, famili, matrimoni, home
	FREX: matrimoni, merciless, maltreat, humili, cruel, wed, expect
(6) DOWRY-B	Highest Prob: dowri, parent, demand, marriag, beat, lakh, rupe
	FREX: dowri, settl, demand, parent, taunt, donat, greedi
(7) REAL ESTATE	Highest Prob: land, amount, properti, compani, cheat, fal, loan
	FREX: sale, chequ, payment, forg, construct, loan, forgeri
(8) DEVELOPMENT	Highest Prob: plot, widow, wife, possess, registri, estat, sold
	FREX: plot, widow, registri, allot, ganga, estat, pension
(9) DOWRY-C	Highest Prob: women, daughter, harass, marri, cell, marriag, justic
	FREX: women, mahila, cell, counsel, divorc, harass, mediat
(10) CHILD ABUSE/RAPE	Highest Prob: girl, daughter, mother, children, marri, child, home
	FREX: girl, pocso, seduc, mother, children, babi, posco
(11) BURGLARY	Highest Prob: stolen, gold, bag, railway, theft, good, lock
	FREX: stolen, theft, railway, bag, thief, steal, lock
(12) DOMESTIC VIOLENCE	Highest Prob: kill, threaten, famili, abus, life, wife, beat
	FREX: kill, threaten, dirti, protect, life, abus, save
(13) DOWRY-D	Highest Prob: motherinlaw, husband, fatherinlaw, inlaw, brotherinlaw, beat, home
	FREX: motherinlaw, fatherinlaw, brotherinlaw, inlaw, husbandinlaw, husband, nanand
(14) FIGHTING	Highest Prob: wife, hospit, fight, injuri, hit, blunt, xray
	FREX: xray, blunt, injuri, stick, fight, hurt, noi
(15) CHAIN-SNATCH	Highest Prob: motor, cycl, motorcycl, boy, bike, snatch, chain
	FREX: motor, cycl, bike, motorcycl, snatch, neck, boy
(16) CRIMINAL FORCE	Highest Prob: woman, wife, daughterinlaw, domest, femal, bride, burn
	FREX: grandson, woman, daughterinlaw, sweeti, prathiya, manpratiya, bride
(17) RAPE	Highest Prob: phone, forc, famili, mobil, room, rape, photo
	FREX: rape, video, scare, obscen, vulgar, hotel, facebook
(18) RUNAWAY/SUICIDE	Highest Prob: husband, shop, children, death, phone, wife, die
	FREX: shop, husband, hang, murder, death, dead, hemlata
(19) MISSING PERSON	Highest Prob: wife, home, imit, gone, wear, search, bodi
	FREX: search, feet, bodi, wear, salwar, tenant, rajput
(20) MISCELLANEOUS	Highest Prob: sister, sisterinlaw, food, grain, ambedkar, cook, lamp
	FREX: ambedkar, sister, grain, lamp, anguri, pale, hbc
(21) PHISHING	Highest Prob: money, bank, rupe, atm, lakh, check, thousand
	FREX: atm, bank, branch, check, withdraw, money, withdrawn
(22) DRIVING ACCIDENT	Highest Prob: hospit, driver, bus, drive, accid, hit, treatment
	FREX: driver, scooti, accid, truck, drive, auto, bus
(23) DOWRY-E	Highest Prob: father, inlaw, money, child, lakh, marriag, parent
	FREX: father, expen, pregnant, abort, inlaw, jewelri, womb
(24) ALCOHOL	Highest Prob: drink, alcohol, liquor, drunk, bottl, drug, intox
	FREX: bottl, alcohol, drink, drunk, liquor, intox, drug

Figure A45: Female Complainant Cases



Note: Gendered crime (dowry, fighting, domestic violence) have higher acquittals, unlike cases such as burglary or real estate disputes.

Figure A46: Word Cloud for 1-8 Top Topics (Female Complainants)

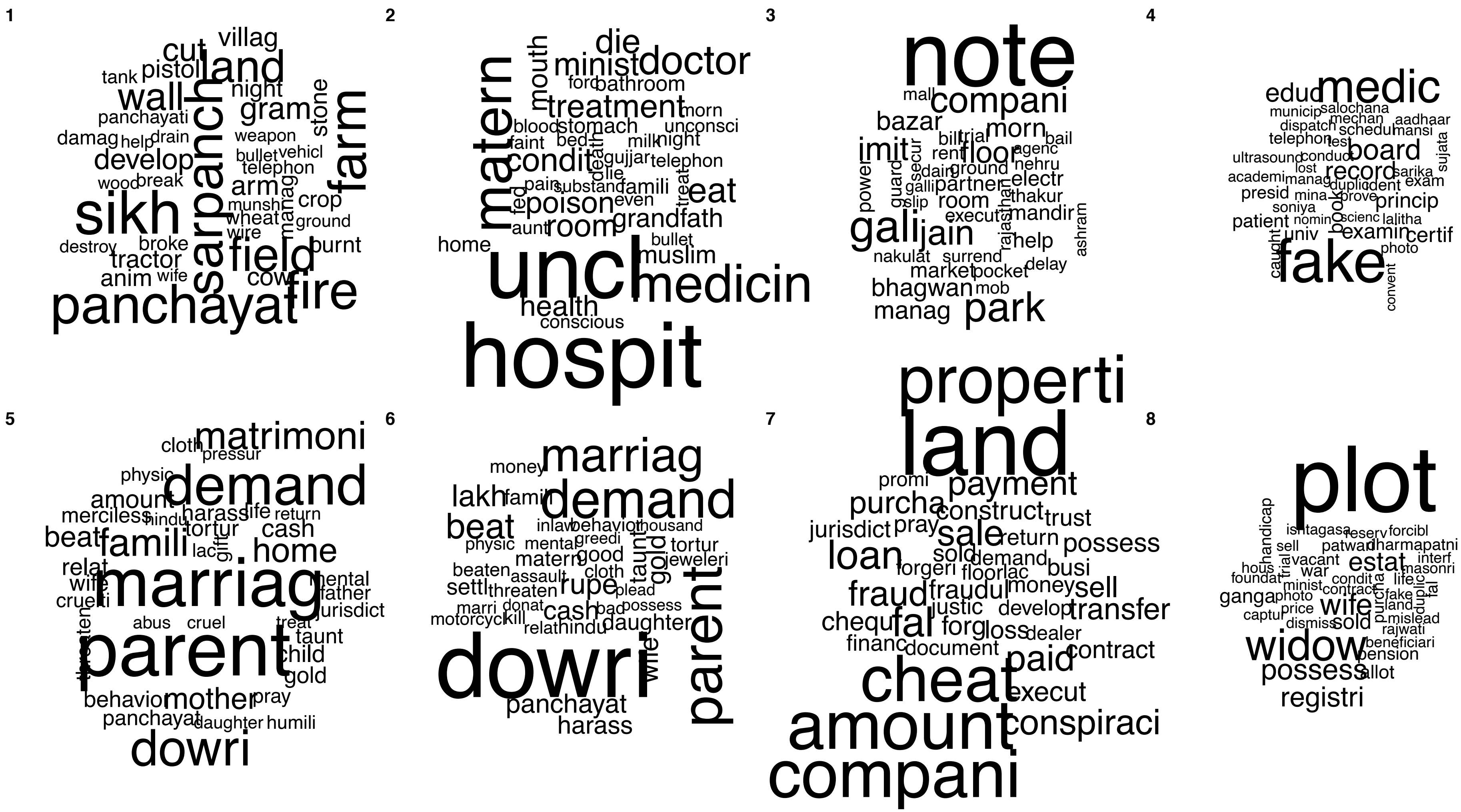


Figure A47: Word Cloud for 9-16 Top Topics (Female Complainants)

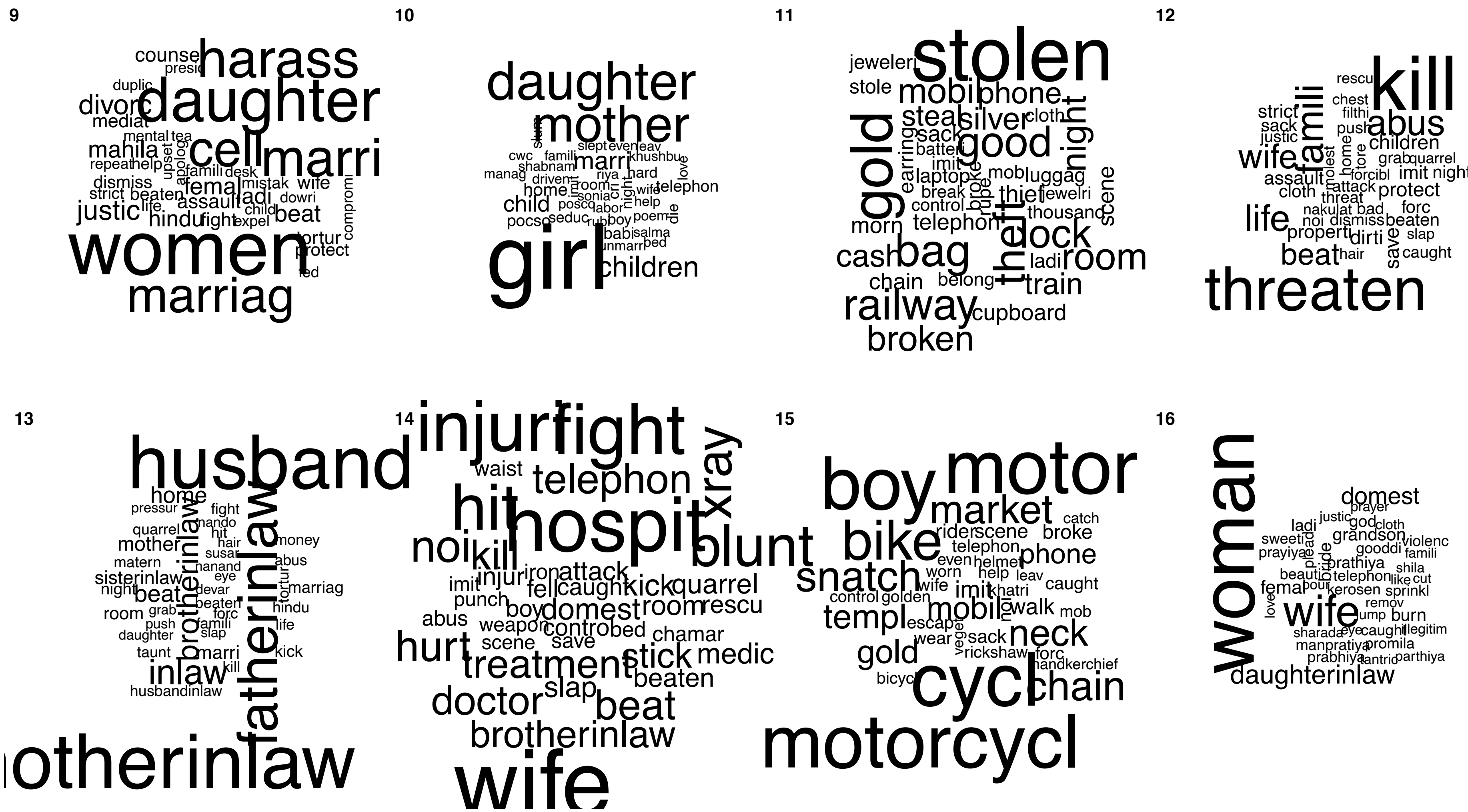


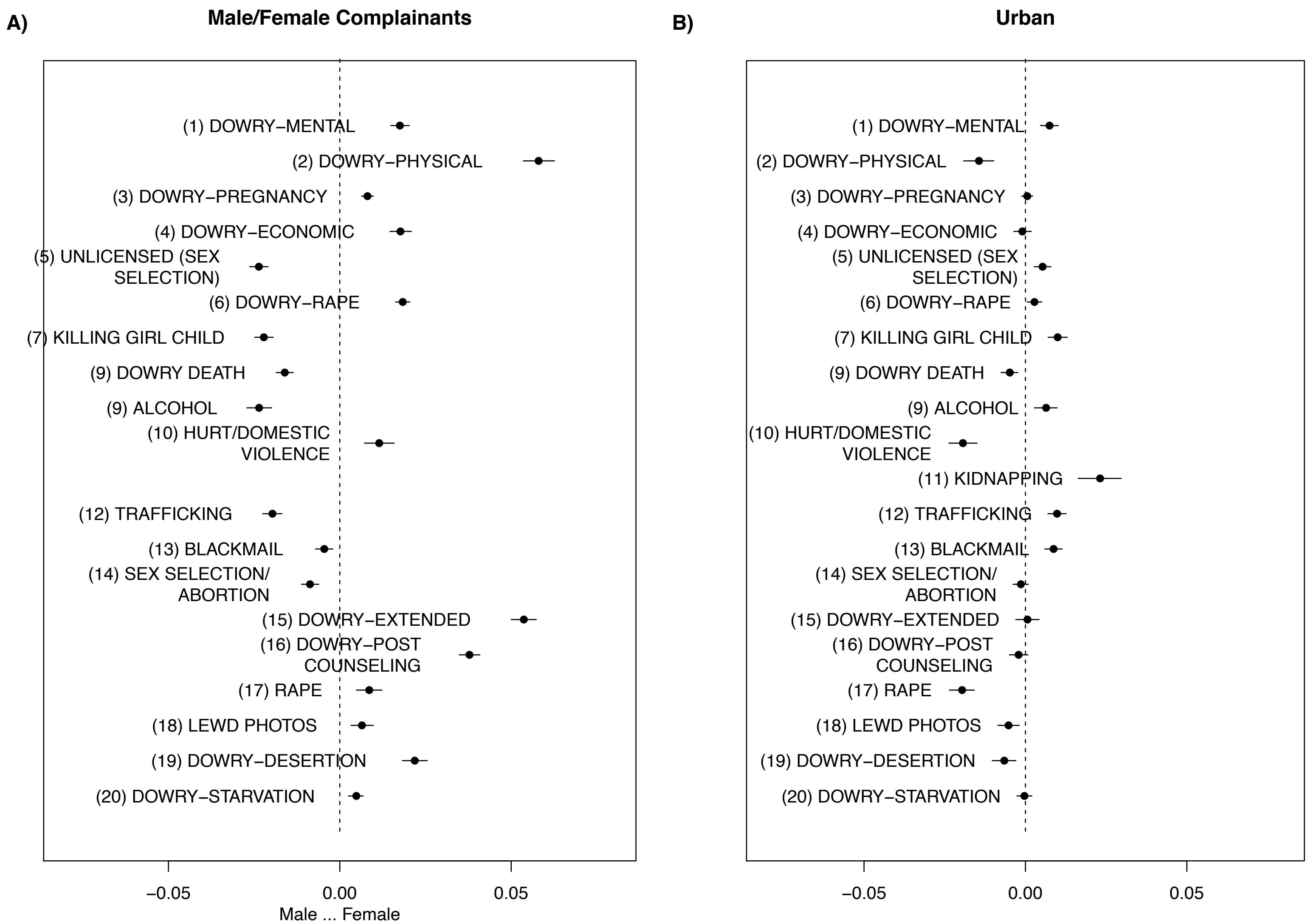
Figure A48: Word Cloud for 16-24 Top Topics (Female Complainants)

6.3 Gendered Crime

Table A9: Top Word Stems by Topic with FREX (Gendered Crime)

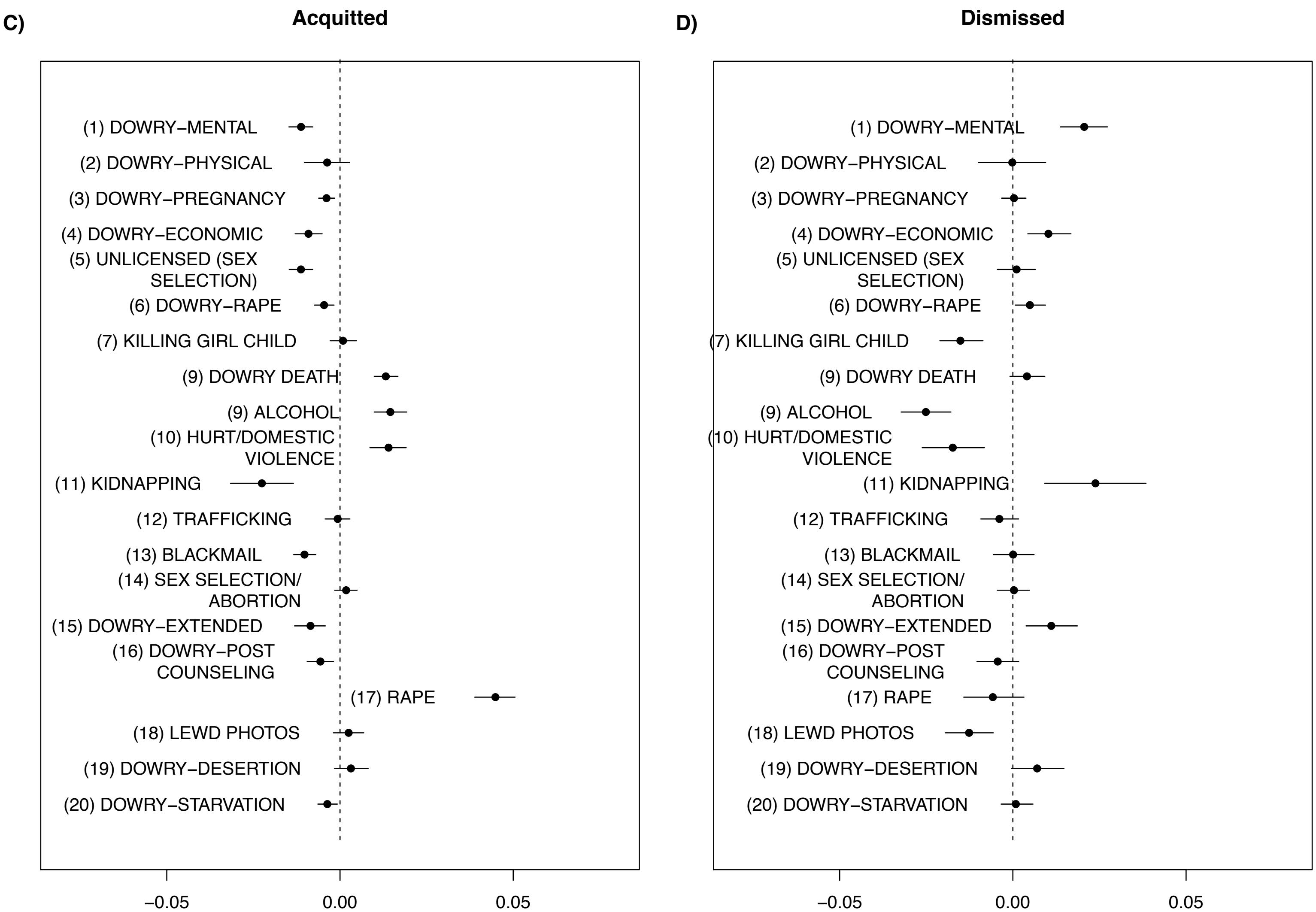
Topic	Top Words
(1) DOWRY-MENTAL	Highest Prob: parent, husband, mother, money, marriag, father, famili
	FREX: salari, australia, loan, earn, shagun, honeymoon, atm
(2) DOWRY-PHYSICAL	Highest Prob: dowri, demand, parent, marriag, father, beat, panchayat
	FREX: settl, panchayat, demand, dowri, motorcycl, greedi, illegitim
(3) DOWRY-PREGNANCY	Highest Prob: child, pregnant, money, thousand, stomach, rupe, babi
	FREX: child, stomach, deliveri, pregnant, babi, defend, thousand
(4) DOWRY-ECONOMIC	Highest Prob: gold, lakh, father, rupe, cash, marriag, money
	FREX: gold, chain, silver, lakh, jeweleri, cupboard, earring
(5) UNLICENSED (SEX SELECTION)	Highest Prob: note, decoy, ultrasound, fake, rupe, ladi, seal
	FREX: decoy, note, currenc, ultrasound, seal, gender, custom
(6) DOWRY-RAPE	Highest Prob: mental, physic, tortur, forc, marriag, parent, pressur
	FREX: physic, mental, unnatur, tortur, sexual, atroc, bad
(7) KILLING GIRL CHILD	Highest Prob: children, shop, woman, telephon, railway, plot, market
	FREX: children, railway, shop, auto, plot, market, sarpanch
(9) DOWRY DEATH	Highest Prob: sister, die, marri, death, dowri, poison, telephon
	FREX: sister, poison, hang, death, die, dead, murder
(9) ALCOHOL	Highest Prob: abus, drink, dirti, alcohol, bus, liquor, drunk
	FREX: drunk, alcohol, dirti, drink, profan, liquor, filthi
(10) HURT/DOMESTIC VIOLENCE	Highest Prob: wife, husband, hit, beat, kill, fight, noi
	FREX: blunt, injuri, hit, stick, xray, attack, rescu
(11) KIDNAPPING	Highest Prob: girl, home, wife, marri, imit, daughter, seduc
	FREX: girl, seduc, search, feet, wear, niec, salwar
(12) TRAFFICKING	Highest Prob: room, manag, driver, sikh, hotel, women, woman
	FREX: sikh, manag, prostitut, hotel, immor, driver, traffic
(13) BLACKMAIL	Highest Prob: photo, fal, jain, cheat, video, land, fake
	FREX: facebook, jain, photo, video, blackmail, defam, fraudul
(14) SEX SELECTION/ABORTION	Highest Prob: hospit, medic, doctor, medicin, treatment, abort, drug
	FREX: hospit, medicin, doctor, medic, treatment, drug, termin
(15) DOWRY-EXTENDED	Highest Prob: husband, motherinlaw, fatherinlaw, inlaw, father, brotherinlaw, bea
	FREX: fatherinlaw, motherinlaw, husband, inlaw, brotherinlaw, sisterinlaw, mater
(16) DOWRY-POST COUNSELING	Highest Prob: dowri, marriag, beat, daughter, harass, cell, women
	FREX: cell, assault, dairi, mediat, harass, counsel, mahila
(17) RAPE	Highest Prob: mother, forc, father, home, rape, room, daughter
	FREX: pocso, rape, bike, mother, cri, advi, posco
(18) LEWD PHOTOS	Highest Prob: threaten, kill, phone, life, famili, mobil, wife
	FREX: threaten, protect, properti, phone, mobil, kill, threat
(19) DOWRY-DESERTION	Highest Prob: marriag, demand, parent, dowri, matrimoni, home, beat
	FREX: matrimoni, merciless, maltreat, prohibit, lac, cruelti, jurisdict
(20) DOWRY-STARVATION	Highest Prob: daughter, famili, marri, wife, home, money, father
	FREX: daughter, sell, famili, prayer, sad, panchayati, adopt

Figure A49: Gendered Crime I



Note:.

Figure A50: Gendered Crime II



Note:.

Figure A51: Word Cloud for 1-8 Top Topics (Gendered Crime)

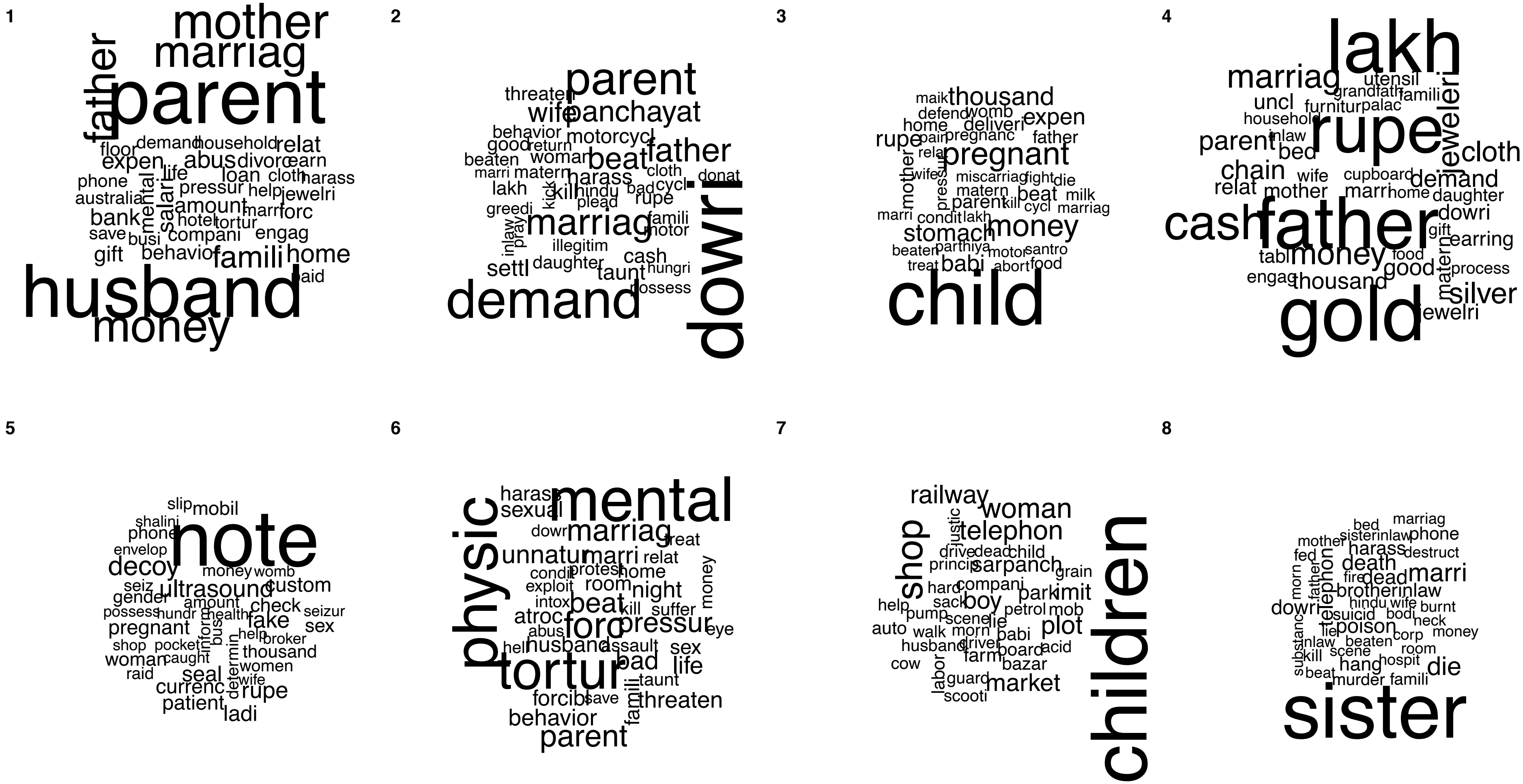


Figure A52: Word Cloud for 9-16 Top Topics (Gendered Crime)

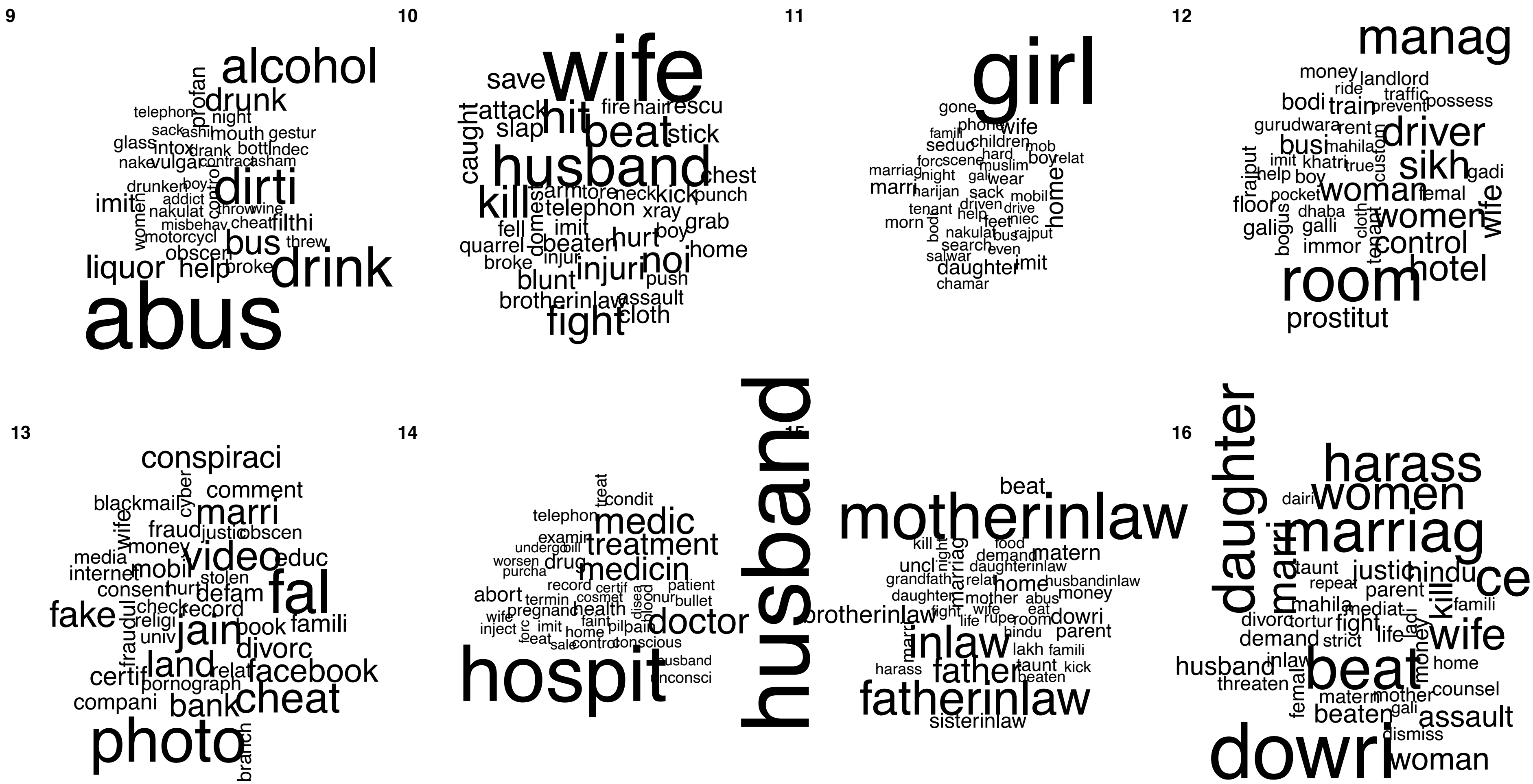


Figure A53: Word Cloud for 17-20 Top Topics (Gendered Crime)