Eyes on the Street: Mass Transit and Women's Safety in Public Spaces

Anisha Garg*

September 14, 2026

Please click here for the latest version.

Abstract

Can mass transit make the places it connects safer for women? I study the staggered arrival of first Delhi Metro stations across police jurisdictions between 2016 and 2019. I construct a high-frequency dataset from more than 300,000 police reports, 88% geocoded, and estimate effects using staggered difference-in-differences. A jurisdiction's first station reduces reported harassment of women by 32% relative to the pre-opening mean among treated jurisdictions, with no comparable change in other crimes against women. Harassment falls within one kilometer of a station without evidence of displacement to the adjacent ring. The decline is concentrated in evening hours and neighborhoods with low baseline activity, while independent safety audits show improved lighting, visibility, and perceived safety. These findings identify a gendered place-based safety return to transit that conventional evaluations may overlook, with implications for where new lines and stations are built.

Keywords: Public Transit, Urban Transformation, Crimes Against Women, Safety.

JEL Codes: J16, K42, R41, O18, R23.

*University of Warwick. Email: anisha.garg@warwick.ac.uk. I am deeply grateful to my supervisors, Robert Akerlof, Sharun Mukand, Debraj Ray and Andreas Stegmann, for their unwavering guidance and support. I also extend my thanks to Wiji Arulampalam, Yatish Arya, Manuel Bagues, Dan Bernhardt, Sonia Bhalotra, Nathan Canen, Nikhil Datta, Ashwini Deshpande, Mirko Draca, Bhaskar Dutta, James Fenske, Ludovicca Gazzè, Selim Gulesci, Bishnupriya Gupta, Clément Imbert, Amrita Kulka, Devesh Rustagi, Sara Spaziani and participants at the SIOE Annual Conference (UNSW), the STEG Annual Conference (CEPR and Oxford), the LEO Economics Conference (Plaksha), the Meeting of Young Minds in Frontier Economics (IIT Bombay), the CPR–CSH Urban Workshop, ACEGD (ISI Delhi), the Delhi Winter School (DSE), the Violence Against Women Workshop (Warwick), and the CAGE-AMES seminar (Warwick) for their invaluable feedback. I am grateful to the Joint Commissioner of Delhi Police for guidance on police jurisdictions and crime data. I also thank Kalpana Viswanath and the SafetiPin team for sharing the SafetiPin app data. All errors are my own.

1 Introduction

Urban transportation infrastructure does more than connect places; it can transform them. The economics literature studies how transit reduces commuting times and costs (Anderson, 2014; Akbar et al., 2023; Tsivanidis, 2026), expands access to jobs and services (Khanna et al., 2022; Zárate, 2022), and reshapes the location and value of economic activity within cities (Gibbons and Machin, 2005; Heblich et al., 2020). Yet we know much less about the externalities generated when transit transforms the public spaces immediately surrounding stations. This paper studies one such place-based externality: safety.

This externality is particularly consequential for women, for whom the risk of harassment can restrict mobility and access to the city (Borker, 2021). In Urban India, 70% of women surveyed say that fear of harassment in public spaces affects whether they work (Beaman et al., 2026).[1] A new transit station may improve safety by attracting shops, vendors, lighting, and pedestrian activity, turning quiet streets into active public spaces. This possibility echoes the seminal argument of Jacobs (1961) that streets are kept safe not only by the police but also by the everyday watchfulness of those who use them—what she called "eyes on the street." Importantly, if transit stations generate such informal surveillance, their safety return may be greatest in low-activity neighborhoods where few eyes were present initially, rather than in established centers that already generate substantial activity. Accounting for this benefit could therefore change both the measured welfare gains from transit expansion and the relative attractiveness of alternative station locations. I ask whether expanding mass transit makes surrounding public spaces safer for women.

I study this question in Delhi, India, where women's safety in public space remains an acute concern. In a survey conducted by UN Women and the International Center for Research on Women, 95% of women and girls reported feeling unsafe in the city's public spaces (UN Women and ICRW, 2013). Delhi also provides a useful empirical setting: the rapid expansion of the Delhi Metro brought new stations to different parts of the city at different times. The net effect of a station opening, however, is ambiguous *a priori*. Greater activity around a station may increase informal surveillance, but the accompanying concentration of people, including more women, may also increase exposure to harassment. Moreover, any decline in the immediate vicinity of a station could reflect the displacement of harassment to nearby streets.

I estimate the net effect of a station opening using the staggered difference-in-differences

[1] Similarly, in a 2021 survey across 140 Indian cities, 52% of respondents reported having turned down an education or work opportunity because they considered the commute unsafe (Ratho and Jain, 2021); in Jordan, nearly half of the women surveyed had turned down a job opportunity because of inadequate public transportation, and 62% had experienced harassment during their commutes (Aloul et al., 2018).

estimator of Callaway and Sant'Anna (2021). I compare changes in police jurisdictions as they receive their first Metro station between 2016 and 2019 with contemporaneous changes in jurisdictions that have not yet received one. The identifying assumption is that, absent a station opening, treated jurisdictions would have followed the same trends in reported harassment as the not-yet-treated comparison jurisdictions. The design does not require the cross-sectional placement of stations to be random. Nevertheless, as a check on selection into treatment during the sample period, I find no systematic differences in 2011 Census demographics between jurisdictions that receive a station and those that remain untreated throughout. More directly relevant to the identifying assumption, I find no evidence of differential pre-treatment trends in reported harassment.

The headline result is that the opening of a police jurisdiction's first Metro station reduces reported harassment of women by approximately **32%** relative to the pre-opening mean among treated jurisdictions. The magnitude is stable across alternative comparison groups and heterogeneity-robust estimators (Sun and Abraham (2021), Borusyak et al. (2024), de Chaisemartin and D'Haultfœuille (2020), and Gardner (2022)), and the conclusion is unchanged under alternative functional forms, inference procedures, and allowances for spatial correlation in the errors.

The police-jurisdiction design, however, cannot reveal how the effect varies with distance from a station or whether some harassment is displaced just beyond its immediate vicinity. Testing for such spatial displacement requires the precise location of each reported crime. To obtain these locations, I construct what is, to my knowledge, the first high-frequency dataset of Delhi First Information Reports with point-level spatial resolution. The dataset contains more than 300,000 reports published by the Delhi Police between 2016 and 2019, 88% of which I geocode to the reported location of the crime. Because these reports are published only as individual, bilingual, scanned PDFs, I convert them into a structured crime-level dataset using optical character recognition and text parsing, classify offenses using the legal sections recorded in each report, and harmonize changes in police boundaries into 147 consistent territorial jurisdictions.

I then use these point locations in a supplementary ring-based analysis that estimates changes in reported harassment within one kilometer of a new station and in the surrounding distance bands. Reported harassment falls within one kilometer of a station, and I find no evidence that it rises in the adjacent ring: the station appears to import safety without exporting harassment next door. It is also outcome-specific: I find no comparable change in domestic violence or in other crimes against women, such as rape and abduction. This outcome specificity weighs against the decline reflecting a general change in recorded crime or police registration practices.

The remaining results provide a set of mechanism tests that point toward informal surveillance. If stations deter harassment by increasing activity and "eyes on the street," the decline should be most evident when station-related activity reaches the surrounding neighborhood, be largest where stations add the most activity, and be accompanied by improvements in the street environment. The temporal pattern fits the first prediction. The decline in reported harassment is concentrated in the evening, when Metro passenger flows disperse from employment centers toward residential and mixed-use neighborhoods (Xavier et al., 2025); I find no comparable decline during the morning commuting hours.

The cross-sectional heterogeneity further helps distinguish informal surveillance from a uniform increase in formal police surveillance. The reduction in harassment is largest in neighborhoods with low baseline nighttime luminosity, where there was less activity—and therefore fewer informal observers—before the station opened. Under a simple explanation in which the effect is driven solely by the formal security or policing introduced with each station, the reduction should not systematically depend on how active the neighborhood was before the opening. Instead, the effect is greatest precisely where the marginal contribution of additional activity to informal surveillance should be largest. The concentration of the effect in low-activity neighborhoods also has a direct implication for transit policy. The safety benefits of transit therefore need not be greatest in the locations that conventional evaluations based on travel demand and accessibility value most highly. Accounting for this place-based externality could alter the spatial allocation of new lines and stations within a city.

The results also weigh against an explanation based solely on changes in crime reporting. One possibility is that commuters become less likely to report harassment because they are in a hurry. Such an explanation, however, should plausibly produce a decline during the morning commute as well, which I do not observe. Moreover, independent safety audits, collected separately from police reports, provide complementary evidence of changes in the surrounding environment: lighting and visibility improve after a station opens, as does women's perceived ability to walk alone at night, while footpath quality—an attribute that the proposed mechanism does not predict should respond—does not change. Although changes in reporting or police behavior cannot be ruled out entirely, the timing of the decline, its concentration in neighborhoods with low baseline activity, and the independently measured improvements in street conditions and perceived safety are difficult to reconcile with either a uniform increase in formal policing or a pure reporting effect. Taken together, the evidence is consistent with Metro stations improving women's safety by activating the surrounding public space and increasing informal surveillance.

Finally, descriptive evidence from hostel listings suggests a gender-differentiated response

in the local accommodation market: accommodation serving women is located closer to Metro stations following their opening than men-only accommodation. Although this pattern may reflect improved transport access as well as safety and is not interpreted causally, it is consistent with safer and better-connected neighborhoods becoming more attractive locations for accommodation serving women.

This paper contributes first to the literature on urban transport and crime. Existing studies emphasize how transit changes the mobility of offenders and victims or the economic opportunities available to potential offenders. Phillips and Sandler (2015) show that transit can spread crime along rail networks by lowering offenders' travel costs; DeAngelo et al. (2019) find that service disruptions increase crime in transit-dependent neighborhoods, consistent with the local isolation of offenders and victims; and Khanna et al. (2022) show that improved transit access can reduce criminal participation by expanding legitimate employment. I identify a complementary, place-based effect and provide evidence consistent with a channel operating through local activity and informal surveillance. Moreover, the geocoded crime data allow me to distinguish local deterrence from spatial displacement: reported harassment falls near new stations without evidence of an offsetting increase in adjacent areas. Thus, transit can change the conditions under which opportunistic public-space offenses like street harassment occur rather than merely redistribute crime through the network.

Second, the paper speaks more broadly to urban and transportation economics. Existing research studies how transportation infrastructure affects commuting, market access, land values, and the spatial distribution of economic activity (Gibbons and Machin, 2005; Heblich et al., 2020; Khanna et al., 2022; Zárate, 2022; Tsivanidis, 2026), while evaluations of rail investment concentrate on its network benefits (Winston and Maheshri, 2007; Murty et al., 2006). I identify an additional place-based safety return: by changing local activity and surveillance conditions, a station can make its immediate surroundings safer. Because this effect is largest in neighborhoods with low baseline activity, the spatial distribution of safety benefits need not coincide with that of conventional benefits based on travel demand and accessibility. Accounting for this omitted benefit could therefore change both the estimated social return to transit expansion and the spatial allocation of new lines and stations within a city.

Third, the paper contributes to the economics of crimes against women. Existing work examines determinants of violence, including behavioral shocks (Card and Dahl, 2011), access to alcohol (Luca et al., 2015), and household sanitation infrastructure (Hossain et al., 2022). A related body of work studies how women's political representation and gender-focused police reforms shape reporting, enforcement, and access to justice (Iyer et al., 2012; Miller and Segal, 2019; Jassal, 2020; Amaral et al., 2022; Sukhtankar et al., 2022). I add evidence

on a distinct margin: general-purpose urban infrastructure can reduce harassment in public space by changing the surrounding environment. Unlike interventions centered on formal law enforcement or services targeted specifically at women, Metro expansion can generate safety as a by-product of local activity and informal guardianship. To my knowledge, this is the first causal evidence that transit expansion makes surrounding public spaces safer for women. More broadly, the results show how urban infrastructure that is ostensibly gender-neutral can generate gender-specific welfare gains by changing women's ability to use public space.

The remainder of the paper proceeds as follows. Section 2 describes the setting and the construction of the data. Section 3 presents the research design, its identifying assumptions, and the main results. Section 4 presents evidence consistent with informal surveillance and examines alternative explanations. Section 5 concludes.

2 Background and Data

This section describes the institutional setting and the data used in the analysis. I first trace the phased expansion of the Delhi Metro and explain how station openings are assigned to police jurisdictions and distance rings. I then discuss women's safety and the institutional process through which crimes are recorded in Delhi, before detailing the construction of the daily, point-geocoded FIR panel that forms the paper's central data contribution. Finally, I describe the additional data used to construct population denominators, assess treatment balance, measure changes in the public environment, characterize baseline local activity, and examine responses in the accommodation market.

2.1 The Delhi Metro and the Assignment of Treatment

The Delhi Metro is a rapid-transit rail system built to improve connectivity and relieve road congestion across the National Capital Region. Its first line was inaugurated on December 25, 2002, and the network expanded in phases thereafter. By 2019, three phases had been completed, bringing the system to 12 lines and 285 stations spanning over 389 km – a total that includes the Noida Aqua Line and Gurgaon's Rapid Metro (Delhi Metro Rail Corporation, 2020).[2] This expansion brought stations to different parts of Delhi at different times and provides the spatial and temporal variation used in the paper.[3]

The analysis covers January 2016 through December 2019. The sample begins in 2016

[2]Noida Aqua Line and Gurgaon's Rapid Metro fall outside Delhi's police jurisdictions and therefore outside the paper's sample.

[3]Appendix Table A.1 summarizes the expansion under each phase, and Appendix Figure C.1 maps the Metro network.

because comparable report-level crime data are not publicly available for earlier years and ends in 2019 to avoid the disruptions to mobility, Metro use, policing, and crime reporting associated with the COVID-19 pandemic beginning in early 2020. I obtain each station's latitude, longitude, opening month, and local police station from the Delhi Metro Rail Corporation (DMRC) website, and use the reported police station to assign the station to a police jurisdiction. This administrative assignment is preferable to a spatial overlay because Metro lines and stations often lie on or close to police-jurisdiction boundaries, making a purely coordinate-based assignment sensitive to small differences in boundary placement.

The baseline jurisdiction-level analysis excludes jurisdictions that already contained a Metro station before January 2016. Because these jurisdictions are already treated when the crime panel begins, their first-station event predates the panel and no pre-treatment outcomes are observed for them within the sample. Among jurisdictions untreated at the beginning of 2016, a jurisdiction becomes treated in the month its first Metro station opens and remains treated thereafter; openings of subsequent stations do not change its treatment status. This definition captures the transition from having no station to becoming connected to the Metro network. Of Delhi's 147 police jurisdictions, 68 already contained a station before January 2016 and are excluded from the baseline jurisdiction-level design. Among the remaining 79 jurisdictions, 25 receive their first station between 2016 and 2019, while 54 remain untreated throughout the sample period.[4]

The station coordinates also allow me to examine where within police jurisdictions the estimated decline occurs. For this supplementary spatial analysis, I use 57 Metro stations that opened between 2016 and 2019, including new stations in jurisdictions that were already connected to the network before 2016. These openings can be included because the objective is not to estimate the effect of a jurisdiction's first connection, but to trace changes in reported harassment by proximity to each new station. I construct distance bands around the stations and aggregate point-geocoded crimes within them. I then examine whether the decline is concentrated in the area immediately surrounding a station and whether it is accompanied by an offsetting increase in nearby areas. Thus, the timing of first-station openings across jurisdictions provides the paper's primary identification strategy, while the station coordinates allow me to characterize the spatial distribution of the effect and assess displacement. Section 3 describes the identifying assumptions of the jurisdiction-level design

[4]As an alternative specification, I expand the sample to include jurisdictions that already contained a Metro station before 2016 but received an additional station during the sample period more than one kilometer from any pre-existing station. I use the same one-kilometer threshold as the ring-based design in Section 3.2, which finds that the effect of a station on reported harassment is concentrated within one kilometer and does not extend into the 1–2 km band; a new station sited farther than this from any existing station therefore represents a treatment unaffected by the prior network. For these jurisdictions, treatment is defined by the opening month of the new, spatially distinct station.

and the implementation of the supplementary spatial analysis.

2.2 Women's Safety and Reported Crime in Delhi

Harassment in public space is a widespread feature of women's experiences in Delhi. Among Indian states and union territories, Delhi had the highest rate of reported crimes against women in 2023—133.6 per 100,000 female population, compared with a national average of 66.2 (National Crime Records Bureau, 2023). Reported crimes capture only part of the underlying problem. In a survey conducted by UN Women and the International Center for Research on Women, 95% of women and girls reported feeling unsafe in Delhi's public spaces (UN Women and ICRW, 2013). The conduct relevant to this paper ranges from unwanted sexual comments and gestures to groping, stalking, and other forms of assault, and may occur on streets, in markets, and while accessing public transportation.

In India, an alleged crime formally enters the police system through a First Information Report (FIR), which records the initial information provided to the police about the alleged offense. In the Delhi Police data used here, each FIR is associated with the territorial police station responsible for the location of the incident. Police jurisdiction therefore reflects where the alleged crime occurred rather than where the victim or accused resides. Each FIR records the date on which the report was filed and, when available, the date and time at which the alleged offense occurred. I assign crimes to months using the reporting date, which is observed for every FIR. The date of occurrence is missing for 12.93% of reports filed under the Indian Penal Code (IPC 1860). The hour-of-day analysis instead uses the reported time of occurrence and is necessarily restricted to FIRs for which that information is available.

An FIR also records the sections of criminal law under which the case is registered. I use these sections to construct a transparent and replicable measure of *reported harassment crimes against women*. The primary outcome includes FIRs registered under the women-specific IPC sections covering assault or criminal force with intent to outrage a woman's modesty, sexual harassment, assault with intent to disrobe, voyeurism, stalking, and words or gestures intended to insult a woman's modesty, together with obscene acts or words in a public place (IPC sections 294, 354, 354A–354D, and 509).[5] Each FIR is counted once, irrespective of the number of qualifying sections it contains.

I also construct broader and comparison outcomes, including all crimes against women, domestic violence, and non-gender-specific street crimes such as theft and robbery. These outcomes help distinguish a change specific to harassment from broader changes in crime, policing, or reporting. Defining every outcome from the complete set of legal sections recorded

[5]Appendix Table B.3 reports the complete mapping of IPC sections into crime categories.

in each FIR, rather than classifying the free-text narrative, makes the classification procedure reproducible and avoids assigning offenses on the basis of subjective textual interpretation.

The analysis nevertheless measures *reported* crime rather than the unobserved incidence of crime. A change in FIRs may reflect a change in underlying offenses, victims' willingness or ability to report them, or police registration practices. I therefore do not interpret changes in FIRs mechanically as equivalent changes in victimization. The empirical analysis uses the timing and location of the effects, comparisons across crime categories, crowdsourced safety audits, and reports filed at Metro police stations to examine these alternative explanations.

Constructing a High-Frequency, Geocoded Crime Panel. The central data contribution of the paper is the construction of a daily, geocoded panel from the universe of FIRs published by the Delhi Police between January 2016 and December 2019. Although the police make these reports publicly available on their website, they do not release them as a structured database. Each FIR appears as an individual, bilingual Hindi–English, multi-column scanned PDF produced from a completed official form. Consequently, neither the form fields nor the incident information can be analyzed directly at scale.

I download 357,802 FIR PDFs and convert them into a structured crime-level dataset through a multistage pipeline. I first preprocess the document images and apply optical character recognition to recover their contents. I then parse the recognized text to extract the reporting date, time and place of occurrence, police station, and applicable statutes and legal sections. Finally, I clean and standardize the extracted fields, classify offenses, and geocode the reported place of occurrence. The pipeline successfully parses 356,140 reports, or 99.54% of the downloaded collection. A manual audit covering 10% of the FIRs confirms that the OCR and parsing procedures recover the key fields with high accuracy. Appendix Section B describes the construction and validation procedures in detail.

The empirical analysis classifies crimes using the sections recorded under the IPC 1860, and of these successfully parsed 356,140 reports, 296,869 are registered exclusively under the IPC 1860 and constitute the IPC corpus used to construct the crime outcomes. Reporting dates are available for every report in this corpus, and IPC sections are successfully extracted for all but two reports.

I geocode the address recorded in the *Place of Occurrence* field of each FIR using ArcGIS together with Delhi Police jurisdiction-boundary shapefiles. These entries vary in their level of detail: some identify residential addresses or named shops, while others refer to landmarks such as schools, places of worship, overhead water tanks, or roundabouts. I retain a geocode only when the information supports an unambiguous point coordinate and exclude ambiguous matches. This procedure geocodes 312,836 of the 356,140 parsed reports, or 87.84%, and

261,493 of the 296,869 IPC reports, or 88.08%. I therefore refer to these observations as point-geocoded reports, without implying that every coordinate identifies an exact building address. Appendix Table B.4 shows that coverage is high and stable across years. Coverage for harassment reports ranges from 91.3% to 94.3% across the four years and is also similar across treated, untreated, and already-treated jurisdictions.

Creating a consistent administrative geography requires an additional harmonization step. Delhi contained 180 territorial police stations during the sample period, but their boundaries changed as new stations were carved out of existing jurisdictions. To prevent changes in administrative boundaries from generating artificial changes in measured crime, I combine all police stations involved in a boundary change into a common territorial unit whose composition remains fixed throughout the panel. This pure aggregation retains every FIR filed at the affected stations and yields 147 harmonized territorial jurisdictions. I apply the same police-station-to-jurisdiction crosswalk to the local police-station assignments reported by DMRC, ensuring that Metro treatment and crime outcomes are defined over the same 147 units.

I separately exclude police units that do not represent fixed residential territories. These include crime, special, economic-offences, and vigilance cells, as well as railway and Metro police stations. Railway and Metro police oversee transport networks rather than geographically defined residential catchments; they therefore have neither fixed territorial boundaries nor corresponding resident populations and cannot serve as units in the jurisdiction-level panel. After these exclusions, the territorial IPC sample – covering the 147 harmonized jurisdictions – contains 286,677 reports, or 96.57% of the IPC corpus, and retains approximately 99% of reported harassment crimes against women.

The exclusion of Metro-police FIRs from the territorial panel also defines the spatial scope of the main outcome. The baseline analysis measures reported harassment in the urban areas surrounding stations rather than incidents reported within the Metro system itself. I do not, however, discard reports filed at Metro police stations. I retain them as a separate series and use them to test whether the decline in harassment recorded by territorial police stations is offset by an increase in reports registered with the Metro police. This exercise helps distinguish an improvement in safety around stations from substitution in reporting between territorial and Metro police.

To my knowledge, this is the first dataset to render the full collection of Delhi FIRs at daily frequency and point-level spatial resolution. Its three dimensions of granularity serve different parts of the analysis. The harmonized police jurisdictions permit comparisons around the staggered arrival of first Metro stations. Reporting dates can be aggregated into the monthly panel used in the jurisdiction-level design, while reported occurrence times

permit the mechanism analysis to examine when incidents occur. The point coordinates make it possible to estimate effects within narrow distance bands around stations and test for spatial displacement independently of police-jurisdiction boundaries.

The standard source for crime statistics in India is the National Crime Records Bureau's annual *Crime in India* reports, which I use to describe aggregate crime in Delhi. Those data, however, cannot support the empirical analysis in this paper. NCRB statistics are annual and district-level and therefore lack the temporal and geographic precision needed either to exploit the staggered arrival of first Metro stations across police jurisdictions or to examine how changes in reported harassment vary with distance from a new station. NCRB also applies a "principal offense rule," under which a report is classified according to its most serious offense. This can omit harassment sections when they appear in an FIR alongside a more serious charge. Constructing the data from all IPC sections recorded in each underlying FIR avoids this loss of information. The resulting dataset—with each report dated by its filing date, assigned to a consistent territorial geography, spatially located where possible, and classified using its complete set of legal sections—supports the jurisdiction-level causal design, the supplementary spatial analysis of concentration and displacement, and the mechanism analysis.

2.3 Additional Data

The remaining datasets perform distinct roles in the empirical analysis. Census data provide population denominators and predetermined demographic characteristics; location-specific safety audits measure changes in the surrounding public environment independently of police reports; nighttime lights characterize baseline local activity; and hostel listings provide suggestive evidence on a possible downstream response in the accommodation market.

Demographic data. I construct population and demographic characteristics from the 2011 Census of India. The Census reports these variables for towns and villages rather than for police jurisdictions or distance rings. I therefore delineate the boundaries of Census towns and villages using digital map products from the Survey of India as a guide and follow the unit definitions in the Census Administrative Atlas.[6] I obtain the corresponding demographic counts from the 2011 Census Primary Census Abstract (Office of the Registrar General and Census Commissioner, India, 2011) and spatially aggregate them to police jurisdictions and to the distance rings used in the point-level analysis. Where a Census town or village intersects more than one jurisdiction or ring, I allocate its population in proportion to the share of its

[6]Survey of India: https://onlinemaps.surveyofindia.gov.in/Digital_Product_Show.aspx. Census Administrative Atlas: https://censusindia.gov.in/nada/index.php/catalog/40.

area falling within each unit, following high-resolution spatial work on India (Asher et al., 2021). This procedure imposes uniform population density within each Census unit.

Population provides the denominator used to express reported crimes per 100,000 residents in both the jurisdiction-level and ring-based designs. I additionally construct household size, literacy, employment, and caste composition, including gender-disaggregated measures where available. Because these characteristics are measured in 2011, before the analysis period and the station openings studied here, I use them to assess balance between jurisdictions that receive their first station during the sample and those that remain untreated.

Location-specific safety audits. To measure the surrounding public environment independently of police reports, I use location-specific safety audits collected through the Safetipin mobile application (Viswanath and Basu, 2015).[7] The data contain approximately 15,000 individual, geocoded audits conducted across Delhi between 2013 and 2024. A location may be audited more than once, and each observation records the location and date of the audit.

Each audit evaluates the location on a scale from zero to three along nine dimensions. *Lighting* measures the amount of illumination; *openness* captures the availability of clear sightlines and the absence of visual obstructions; and *visibility* measures whether a person at the location can be seen from nearby buildings, shops, or entrances. *People* records the presence of other people; *security* captures visible police or private guards; and *walkpath* evaluates the availability and quality of pedestrian space. *Public transport* measures the availability and proximity of transport options; *gender usage* records the presence of women and children relative to men; and *feeling* captures the auditor's perceived safety at the location. Safetipin also combines these dimensions into a composite safety score.

I use the nine audit dimensions and the composite safety score as separate outcomes. For comparability across measures, I standardize each outcome by subtracting its pre-opening mean and dividing by its pre-opening standard deviation, calculated using audits conducted within one kilometer of stations before they open. These audits provide evidence on both observable street conditions and perceived safety that does not depend on whether an incident was reported to or registered by the police. In Section 4, I compare changes within 0–1 kilometer of a station with those in the 1–2 kilometer ring to assess whether station openings are accompanied by the environmental changes predicted by informal surveillance. The analysis uses the complete 2013–2024 audit sample, thereby providing a longer post-opening window than the crime panel.

[7] Safetipin: https://safetipin.com.

Baseline nighttime activity. I measure baseline local activity using nighttime luminosity from the VIIRS Day/Night Band, processed as part of NASA's Black Marble VNP46A2 product (Román et al., 2018). Nighttime lights are widely used as a proxy for local economic activity (Henderson et al., 2012; Asher and Novosad, 2020). I extract radiance through Google Earth Engine and calculate mean luminosity in 2016 within one kilometer of each of the 57 stations used in the ring-based design. In doing so, I exclude any area lying within one kilometer of a Metro station that opened before 2016, so that the baseline measure does not capture activity in the immediate surroundings of a pre-existing station.

The first station opening in the analysis occurs in May 2017, making 2016 luminosity predetermined for every station in the sample. I divide the 57 stations into relatively high- and low-baseline-activity groups at the median of the station-level luminosity distribution. The 1–2 kilometer ring provides the comparison area in the corresponding spatial analysis. Fixing the classification using 2016 luminosity prevents subsequent Metro expansion or changes in lighting technology during the sample period from affecting a station's baseline-activity group (Kyba et al., 2017).

Hostel locations. Finally, I use hostel listings from MagicBricks to provide suggestive evidence on a possible downstream response in the accommodation market. I collected 4,374 listings on May 18, 2026, of which 4,265 contain coordinates. After removing exact duplicate records, invalid coordinates, and listings without a reported operating-since year, 2,632 listings remain. The data provide each hostel's address and coordinates, the year in which it reports having begun operations, and a structured classification of whether it serves men only, women only, or both genders.

To reduce contamination from pre-existing Metro access, I exclude any hostel located within one kilometer of a station already operating before 2016, regardless of when the hostel began operating. I then match each remaining hostel to its nearest station among the 57 stations that opened between 2016 and 2019 and calculate the distance between them. I classify a hostel as pre-opening if its reported operating-since year precedes the station's opening year and as during-or-post-opening if it began operating in or after that year. I compare mean distance across these periods separately for men-only hostels and for women-only and mixed-gender hostels.

This comparison is descriptive because the data constitute a cross-section of hostels listed on MagicBricks in May 2026 and do not include establishments that had closed or left the platform before collection. It therefore does not identify a causal effect on hostel entry or resident composition, but provides suggestive evidence on whether surviving accommodation serving women is located differently around new Metro stations.

3 Empirical Strategy and Main Results

I first estimate the effect of a jurisdiction's first Metro station using the staggered timing of station openings across police jurisdictions. I then report the dynamic effects and examine robustness to alternative comparison groups, estimators, functional forms, inference procedures, and samples. Next, I use the point-geocoded crime data to determine where within a jurisdiction the decline occurs and whether it is offset by displacement to nearby streets. I subsequently examine whether the estimated decline is specific to harassment. The jurisdiction-level design provides the primary causal estimate; the spatial analysis establishes how that effect is distributed around a station.

3.1 Jurisdiction-Level Design and Results

3.1.1 Setup and treatment

The unit of observation is police jurisdiction i in month t. The estimation sample contains the 79 jurisdictions that did not contain a Metro station in January 2016, observed through December 2019. Of these, 25 receive their first station during the sample period and 54 remain untreated through December 2019. The other 68 of Delhi's 147 jurisdictions already contained a station when the panel begins. Because their first-station event predates the crime data, I exclude them from the baseline jurisdiction-level analysis.

The outcome Y_{it} is the number of reported harassment crimes against women per 100,000 residents in jurisdiction i and month t. Let G_i denote the month in which jurisdiction i receives its first Metro station, and set $G_i = \infty$ for jurisdictions that remain untreated throughout the sample period. Jurisdictions sharing the same opening month form treatment cohort g. A jurisdiction becomes treated in the month its first station opens and remains treated thereafter. Consequently, jurisdictions treated later can serve as comparisons until their own stations open, while jurisdictions that remain untreated through December 2019 can serve as comparisons throughout the panel.

Estimator and target parameter. I estimate group-time average treatment effects using the staggered difference-in-differences estimator of Callaway and Sant'Anna (2021). Let $Y_{it}(g)$ denote the potential outcome in month t if jurisdiction i receives its first station in month g, and let $Y_{it}(0)$ denote its potential outcome if it remains untreated. For cohort g in a post-treatment month t, the group-time average treatment effect is

$$ATT(g,t) = \mathbb{E}\left[Y_{it}(g) - Y_{it}(0) \mid G_i = g\right], \qquad t \geq g. \tag{1}$$

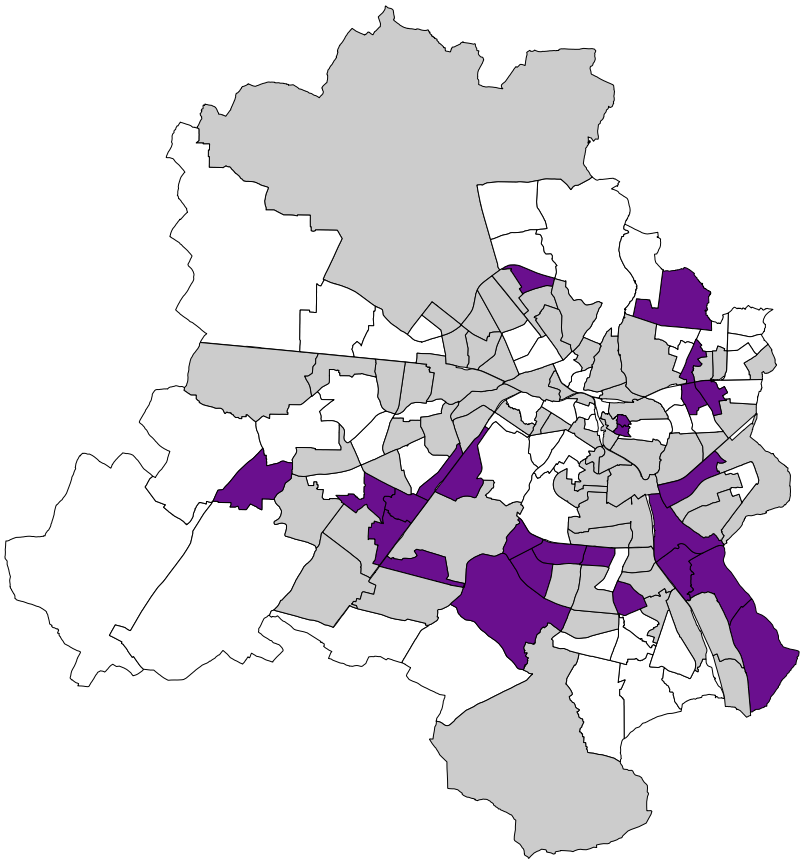


Figure 3.1: Treatment and control police jurisdictions

Notes: Police jurisdictions of Delhi by treatment status: treated (first Metro station opens during the sample period), untreated (no station through December 2019), and already treated (station before January 2016). Already-treated jurisdictions are excluded from the baseline jurisdiction-level analysis.

Under the Assumptions 1 and 2 stated below, and because the baseline specification includes neither covariates nor sampling weights, this parameter is identified by

$$ATT(g,t) = \mathbb{E}\left[Y_{it} - Y_{i,g-1} \mid G_i = g\right] - \mathbb{E}\left[Y_{it} - Y_{i,g-1} \mid G_i > t\right], \qquad t \geq g. \tag{2}$$

The first term is the change in reported harassment for cohort g between the month before its station opens and month t. The second is the contemporaneous change among jurisdictions that remain untreated in month t. Because $G_i > t$ includes both later-treated jurisdictions before their own openings and jurisdictions untreated throughout the panel, the composition of the comparison group can vary across months. Already-treated jurisdictions are never used as controls. This avoids the contaminated comparisons that can arise in conventional two-way fixed-effects estimators when treatment effects vary across cohorts or exposure lengths (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfœuille, 2020).

Aggregation. I aggregate the group-time effects by event time $e = t - g$, the number of months since a station opened. Let T denote the final month of the panel, let $\mathcal{G}_e = \{g : g + e \leq T\}$ denote the set of treatment cohorts observed at event time e, and let N_g be the number of jurisdictions in cohort g. For each post-treatment event time, the dynamic effect is

$$\theta(e) = \sum_{g \in \mathcal{G}_e} \left(\frac{N_g}{\sum_{h \in \mathcal{G}_e} N_h} \right) ATT(g, g+e), \qquad e \geq 0. \tag{3}$$

Thus, $\theta(e)$ averages the effects for cohorts observed at the same length of exposure, weighting each cohort by its number of jurisdictions. At longer exposure lengths, the estimate is necessarily based on the earlier-treated cohorts that remain observable at that event time.

The headline estimand is the event-study aggregation's average across all observed post-treatment exposure lengths. Let $\mathcal{E}_+$ denote the set of nonnegative event times included in the event study. The average post-treatment effect is

$$\overline{\theta}_{\text{post}} = \frac{1}{|\mathcal{E}_+|} \sum_{e \in \mathcal{E}_+} \theta(e). \tag{4}$$

The event-study figure also reports pre-treatment placebo estimates. Because the long-difference specification uses $g - 1$ as the common reference period, these estimates compare changes between earlier months and the month immediately preceding treatment in cohort g with the corresponding changes among eligible comparison jurisdictions. The estimate for $e = -1$ is normalized to zero and omitted. These pre-treatment estimates are not treatment

effects; they provide evidence on whether treated and comparison jurisdictions were following different trends before station openings. Baseline analytical standard errors are clustered at the police-jurisdiction level.

Identifying assumptions. Identification rests on parallel trends and no anticipation.

Assumption 1 (Parallel Trends)**.** For every treatment cohort g and post-opening month $t \geq g$, absent a Metro station opening, jurisdictions first treated in month g would have experienced the same change in reported harassment between month $g-1$ and month t as jurisdictions not yet treated by month t:

$$\mathbb{E}[Y_{it}(0) - Y_{i,g-1}(0) \mid G_i = g] = \mathbb{E}[Y_{it}(0) - Y_{i,g-1}(0) \mid G_i > t]\,.$$

Assumption 2 (No Anticipation)**.** Reported harassment does not respond to a station before it opens:

$$Y_{it}(g) = Y_{it}(0) \qquad \text{for all } t < g.$$

The design does not require the cross-sectional placement of Metro stations to be random. It allows jurisdictions receiving stations to differ systematically from untreated jurisdictions in their levels of reported harassment and observed characteristics. It requires, however, that their reported harassment would have evolved similarly in the absence of a station opening. As a check on selection into treatment during the sample period, Table A.3 compares jurisdictions receiving their first station between 2016 and 2019 with those remaining untreated through December 2019. The two groups exhibit no systematic differences in population, household size, literacy, employment, or caste composition measured in the 2011 Census. Because these characteristics predate the analysis period, this exercise provides evidence that eventual treatment during the sample is not systematically associated with the observed demographic characteristics. It does not directly test parallel trends or rule out selection on unobserved characteristics.

The pre-opening event-study coefficients reported below provide more direct evidence on differential outcome dynamics. They are also relevant to no anticipation because Metro construction may be visible before a station formally opens. I therefore examine whether reported harassment begins to change before the opening date and separately conduct a placebo exercise that advances treatment by three months in Section 3.1.3.

Finally, because Metro stations sometimes lie near jurisdiction boundaries, the jurisdiction-level treatment indicator may classify some locations within the affected area of a new station as untreated. To the extent that an opening also reduces harassment in these neighboring

locations, this contamination of the comparison group would attenuate the jurisdiction-level estimate toward zero. The spatial analysis in Section 3.2 addresses this limitation directly.

3.1.2 Main result

The opening of a jurisdiction's first Metro station substantially reduces reported harassment crimes against women per 100,000 residents. Table 3.1 reports an average across the post-opening event-time estimates of -2.53 reports per 100,000 residents per month, with a standard error of 0.98. Relative to the pre-treatment mean of 7.88 reports per 100,000 among treated jurisdictions, this estimate implies a decline of approximately **32%**. The corresponding 95% confidence interval, $[-4.45, -0.61]$, excludes zero.

Table 3.1: The Effect of the First Metro Station on Reported Harassment Crimes Against Women

	Reported harassment per 100,000
Post-opening ATT	-2.533^{***}
	(0.976)
Pre-opening average	0.659
	(1.844)
Implied change (% of pre-treatment mean)	-32%
Pre-treatment mean (treated)	7.884
Jurisdiction fixed effects	Yes
Month-year fixed effects	Yes
Observations	3,792
Treated / never-treated jurisdictions	25 / 54

Notes: Callaway and Sant'Anna (2021) estimator using doubly-robust inverse-probability weighting with not-yet-treated comparison units. "Post-opening ATT" is the average of the post-opening event-study coefficients (reported harassment per 100,000 residents); "Pre-trend (avg.)" is the average of the pre-opening coefficients, a test of parallel trends. Standard errors, clustered at the police-jurisdiction level, are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

3.1.3 Dynamic effects

Figure 3.2 plots the post-treatment dynamic effects from (3), together with pre-treatment placebo estimates constructed using month $g - 1$ as the common reference period. Before a station opens, the estimates are small, exhibit no systematic trend, and are statistically indistinguishable from zero. The event study therefore provides no evidence that treated and comparison jurisdictions were following different trends in reported harassment before opening. After opening, the effect emerges within the first few months and remains negative throughout the observed exposure window, with estimates generally between -2.5 and -4

reports per 100,000. This pattern indicates a sustained reduction in reported harassment rather than a short-lived change around the opening month.

As an additional check on anticipation, I assign treated jurisdictions a placebo opening date three months before their actual opening and estimate the average effect over the three months immediately preceding the true opening. The resulting estimate is 1.90 reports per 100,000 residents, with a standard error of 1.83. None of the three monthly estimates shows a statistically significant decline. This evidence is consistent with reported harassment not responding before stations begin operating.

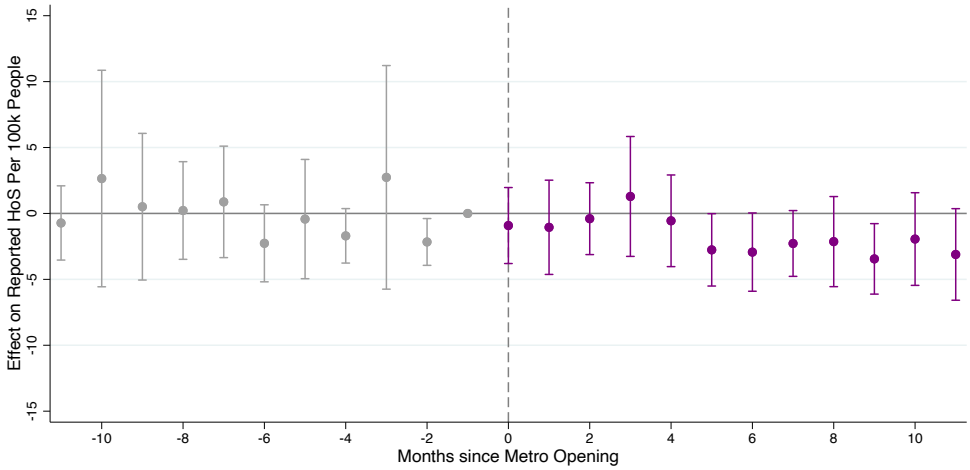


Figure 3.2: Event study: Effect of a jurisdiction's first Metro station on reported harassment crimes against women

Notes: The figure plots post-treatment dynamic effects $\theta(e)$ and pre-treatment placebo estimates constructed using $g-1$ as the common reference period using the Callaway and Sant'Anna (2021) estimator, plotted against months since a jurisdiction's first Metro station opened. The outcome is reported harassment crimes against women per 100,000 residents. Bars denote conventional pointwise 95% confidence intervals. Standard errors are clustered at the police-jurisdiction level. Estimates at longer event times are based on fewer treatment cohorts because only jurisdictions receiving stations earlier in the sample contribute to those event-time effects.

3.1.4 Robustness

I next examine whether the jurisdiction-level result is sensitive to the construction of the comparison group, estimator, outcome transformation, composition of the estimation sample, inference procedure, or spatial correlation in the errors.

Alternative comparison groups. Restricting the comparison group to the 54 jurisdictions that remain untreated throughout the 2016–2019 crime panel yields an almost identical estimate of -2.54 reports per 100,000 residents per month, with a standard error of 0.98. I

then impose a more demanding restriction, retaining as controls only jurisdictions whose first Metro station opened between January 2020 and June 2026. These jurisdictions were untreated throughout the crime panel but were subsequently incorporated into the network, making them potentially more comparable to jurisdictions treated during 2016–2019 on the determinants of station placement. The estimate remains essentially unchanged at -2.51, with a standard error of 1.02. Appendix Table A.4 reports these results.

Alternative estimators. The heterogeneity-robust estimators of Sun and Abraham (2021), Borusyak et al. (2024), de Chaisemartin and D'Haultfœuille (2020), and Gardner (2022) produce estimates of -2.54, -2.47, -1.96, and -2.47, respectively. All four estimates are negative and statistically significant. A Goodman–Bacon decomposition shows that comparisons in which earlier-treated cohorts serve as controls for later-treated cohorts account for only 3.6% of the weight in the static two-way fixed-effects estimate (Goodman-Bacon, 2021). Even so, I continue to rely on the heterogeneity-robust estimators reported above, following the literature's recommendation to explore alternative estimators even when the two-way fixed-effects estimate does not appear biased: the comparisons implicit in the two-way fixed-effects estimator remain opaque, and the weights it places on the underlying cohort-level average treatment effects need not correspond to a policy-relevant parameter (de Chaisemartin and D'Haultfœuille, 2020). Appendix Table A.5 reports the estimates.

Functional form and count models. The result is not an artifact of expressing harassment as a linear rate. A static Poisson pseudo-maximum-likelihood model of harassment counts, including jurisdiction and month-year fixed effects and using jurisdiction population as the exposure, yields a coefficient of -0.30 ($p = 0.024$). This coefficient implies a reduction of approximately 26%. Applying the Callaway–Sant'Anna estimator after transforming the harassment rate using the inverse hyperbolic sine, which accommodates jurisdiction-months with no reported harassment, produces an average post-opening estimate of -0.72, with a standard error of 0.21. The corresponding average pre-treatment estimate is -0.07, with a standard error of 0.23 (Appendix Table A.6).

Influential jurisdictions and treatment cohorts. The result is not driven by a single treated jurisdiction. Re-estimating the model while omitting one treated jurisdiction at a time produces coefficients ranging from -2.78 to -1.79, all of which remain statistically significant. The conclusion is also robust to removing one treatment cohort at a time. The least negative estimate, obtained after excluding the earliest treatment cohort, is -1.67 and remains statistically significant at the 5% level. Although smaller than the baseline

coefficient, it still represents a decline of approximately one-fifth relative to the pre-treatment mean. Finally, excluding jurisdictions above the 90th percentile of the 2016 harassment-rate distribution yields an estimate of −2.51, with a standard error of 0.70. These exercises show that the headline result is not generated by a small number of unusually high-crime jurisdictions or a single opening cohort. Appendix Table A.7 reports these results.

Multiplier-bootstrap inference. Because only 25 jurisdictions are treated during the sample, I supplement the conventional cluster-robust inference with the multiplier bootstrap implemented by Callaway and Sant'Anna (2021), clustering at the police-jurisdiction level. The resulting simultaneous confidence bands include zero throughout the pre-opening period. Post-opening point estimates remain consistently negative, and several continue to exclude zero even under simultaneous inference.

Spatially correlated errors. Because neighboring jurisdictions may experience correlated shocks, I also estimate a static two-way fixed-effects specification using Conley standard errors with spatial cutoffs of 3, 5, and 10 kilometers and a temporal lag of 12 months (Conley, 1999). The point estimate is −2.91 reports per 100,000 residents per month. The standard error is approximately 1.24 under each distance cutoff, leaving the estimate statistically significant at conventional levels. This exercise addresses inference under spatial correlation; its static two-way fixed-effects coefficient is not the same estimand as the average post-treatment effect reported in the baseline specification (Appendix Table A.8).

3.2 Spatial Distribution and Displacement

The jurisdiction-level analysis provides the primary estimate of the effect of gaining a first Metro station, but police jurisdictions are too coarse to reveal the spatial incidence of that effect. The observed decline could be concentrated in the station's immediate surroundings, extend more broadly across the jurisdiction, or be accompanied by an increase in harassment on nearby streets. Moreover, administrative boundaries do not necessarily coincide with physical exposure to a station, particularly when stations lie near jurisdiction borders. I therefore use the point-geocoded crime data to trace changes in reported harassment per 100,000 residents across concentric distance rings around new stations. This supplementary spatial analysis locates the decline identified at the jurisdiction level and examines whether stations make their immediate surroundings safer or merely displace harassment to adjacent areas.

Constructing station catchments. I assign every location within five kilometers of the 57 new Metro stations to its nearest station. This creates mutually exclusive catchments and prevents crimes or population from being counted more than once where station buffers overlap. I construct the catchments using all 57 station locations and hold them fixed throughout the panel, regardless of when each station opens. Within each catchment, I construct 0–1, 1–2, 2–3, 3–4, and 4–5 km distance bands and calculate reported harassment per 100,000 residents in each station-band-month cell. I exclude areas within one kilometer of a station operating before 2016 so that exposure to the pre-existing network is not attributed to a new station opening.

Changes by distance. I first document post-opening changes separately across distance bands. Let $\mathcal{B} = 0, 1, 2, 3, 4$ denote the set of bands, with $b = r$ corresponding to locations between r and $r + 1$ kilometers from station s. Taking the 0–1 km band as the reference category, I estimate

$$Y_{sbt} = \delta \, \text{Post}_{st} + \sum_{r \in \mathcal{B} \setminus \{0\}} \gamma_r \, \mathbf{1}\{b = r\} + \sum_{r \in \mathcal{B} \setminus \{0\}} \beta_r \left[\mathbf{1}\{b = r\} \times \text{Post}_{st} \right] + \mu_s + \lambda_t + \varepsilon_{sbt}, \quad (5)$$

where Y_{sbt} is reported harassment per 100,000 residents in station s, distance band b, and month t, and Post_{st} equals one from the month station s opens onward. Station fixed effects μ_s absorb permanent differences across station catchments, distance-band fixed effects γ_r capture average differences across bands, and calendar-month fixed effects λ_t absorb changes common to Delhi in each month. Standard errors are clustered at the Metro-station level. The coefficient δ gives the post-opening change in the 0–1 km band. Each β_r measures how the post-opening change in outer band r differs from that in the 0–1 km band, so $\delta + \beta_r$ gives the corresponding change in band r.

Table A.9 and Figure 3.4 provide the results. Reported harassment declines by approximately 2.4 reports per 100,000 residents within one kilometer of a new station, with a standard error of 0.975. The estimate is statistically significant at the 5% level. By contrast, the estimated changes in the 1–2, 2–3, 3–4, and 4–5 km bands are small and statistically indistinguishable from zero.[8] Calendar-month fixed effects absorb changes common across Delhi, so the outer-band estimates indicate the absence of an additional station-specific change farther from a new station rather than the absence of a broader citywide trend. Taken together, these descriptive estimates document a sharply localized decline around new stations. I next compare the inner and outer bands directly to assess the robustness of this

[8] Reported harassment declined across the city over this period; the month-year fixed effects absorb this common trend, so the flat outer bands reflect the absence of any station-specific change beyond one kilometer rather than an absence of any change at all.

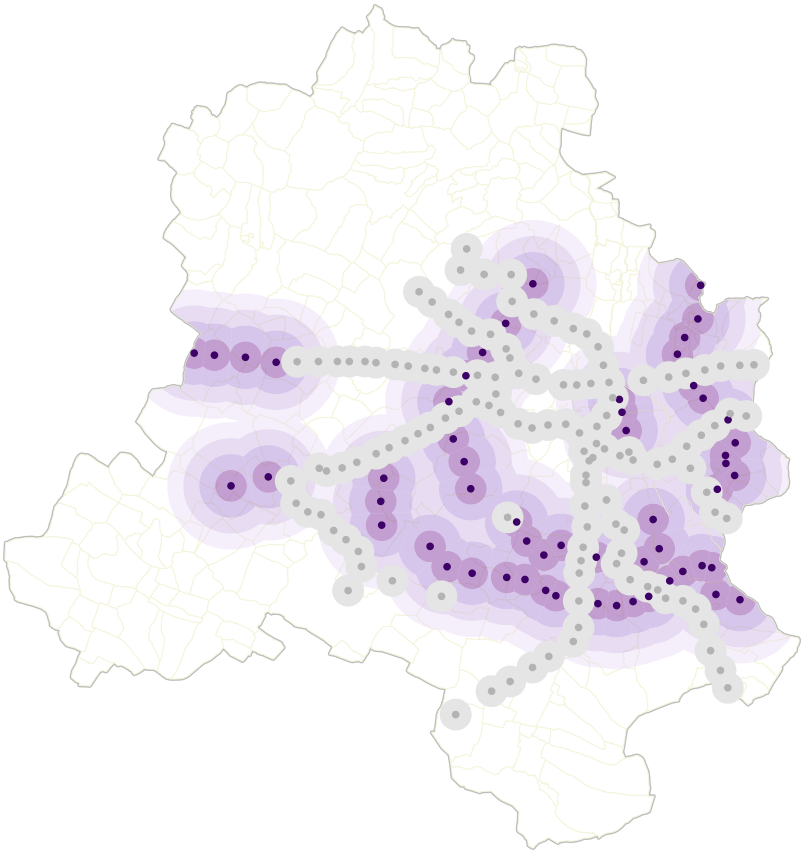


Figure 3.3: Distance bands around newly opened Metro stations

Notes: The figure maps the fixed distance bands around the 57 Metro stations opening during the sample period. Each location is assigned to its nearest new station using Voronoi polygons, so catchments are mutually exclusive (Appendix Section B.5). Areas within one kilometer of a station operating before 2016 are excluded.

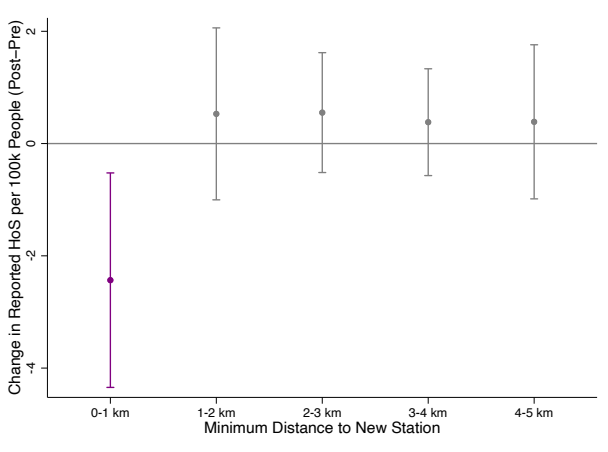


Figure 3.4: Post-opening changes in reported harassment by distance from a new station

Notes: Each point reports the estimated post-opening change in reported harassment crimes against women per 100,000 residents within the indicated distance band, obtained from Equation (5). Bars denote 95% confidence intervals, with standard errors clustered at the Metro-station level.

spatial pattern and test more explicitly for displacement.

Relative and nested-ring estimates. I next estimate the change in the inner ring relative to nearby areas using the Callaway and Sant'Anna (2021) estimator. In the baseline specification, each station contributes two spatial units: an inner 0–1 km cell and an outer 1–2 km cell. The inner cell enters treatment cohort G_s, defined by station s's opening month, while the outer cell is coded as a comparison unit throughout the panel. Before G_s, the inner cell is not yet treated. For each cohort and month, the estimator compares the change in newly treated inner cells with the contemporaneous change among outer cells and inner cells associated with stations that have not yet opened. This comparison is informative about relative spatial incidence if, absent an opening, reported harassment in the treated inner cells would have followed the same trends as these eligible comparison cells. The pre-opening estimates provide no evidence of differential trends between them (Table 3.2). As in the jurisdiction-level analysis, I aggregate group-time effects by event time using cohort-size weights and report the average across the resulting post-opening event-time estimates. Standard errors are clustered at the Metro-station level.

Using the adjacent 1–2 km band as the comparison area, reported harassment within one kilometer declines by 3.227 reports per 100,000 residents, with a standard error of 1.26. I then hold the treated 0–1 km ring fixed while successively expanding the outer comparison area to 1–3, 1–4, and 1–5 km. In each specification, I aggregate crimes and population across the relevant outer bands to form one comparison cell per station. The corresponding estimates are −2.896, −2.826, and −2.876, with standard errors of 1.18, 1.17, and 1.14, respectively.

All four estimates are statistically significant at the 5% level. The stability of the estimates as the comparison area expands indicates that the relative decline is not sensitive to the width of the outer ring (Table 3.2).

Table 3.2: Ring Estimates: Robustness to the Width of the Control Band

	Control band (treated band: 0–1 km)			
	1–2 km (baseline)	1–3 km	1–4 km	1–5 km
Post-opening ATT	-3.227^{**}	-2.896^{**}	-2.826^{**}	-2.876^{**}
	(1.260)	(1.178)	(1.173)	(1.143)
Pre-trend (avg.)	−0.033	0.438	0.404	0.484
	(2.996)	(2.966)	(2.970)	(2.960)
Station × band fixed effects	Yes	Yes	Yes	Yes
Month-year fixed effects	Yes	Yes	Yes	Yes
Observations	5,472	5,472	5,472	5,472

Notes: Each column is a separate Callaway and Sant'Anna (2021) estimate comparing the treated 0–1 km ring to a single control ring spanning the indicated distance band, where the control ring aggregates reported harassment crimes and resident population across its constituent one-kilometer bands. Estimation uses doubly-robust inverse-probability weighting with not-yet-treated rings as the comparison group. "Post-opening ATT" is the average of the post-opening event-study coefficients (reported harassment per 100,000 residents); "Pre-trend (avg.)" is the average of the pre-opening coefficients, a test of parallel trends. Standard errors, clustered at the Metro-station level, are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Together with the band-specific estimates, which show no statistically significant increase in any outer band, these results provide no evidence that the decline within one kilometer is offset by displacement to areas up to five kilometers away.

Outcome specificity. I next examine whether this localized decline is specific to harassment or reflects a broader change in reported crime. I apply the same Callaway–Sant'Anna ring specification separately to each crime category shown in Figure 3.5 (Appendix Table A.10), treating the 0–1 km cell as exposed from the station's opening month and using the corresponding 1–2 km cell as the comparison area. The category of other crimes against women excludes harassment. Harassment is the only category for which I estimate a statistically significant post-opening decline. I find no comparable change in domestic violence, other crimes against women such as rape and abduction, or non-gender-specific street crimes such as theft and robbery. This pattern weighs against explanations based on a general change in crime, police activity, or the propensity to register offenses following a station opening.

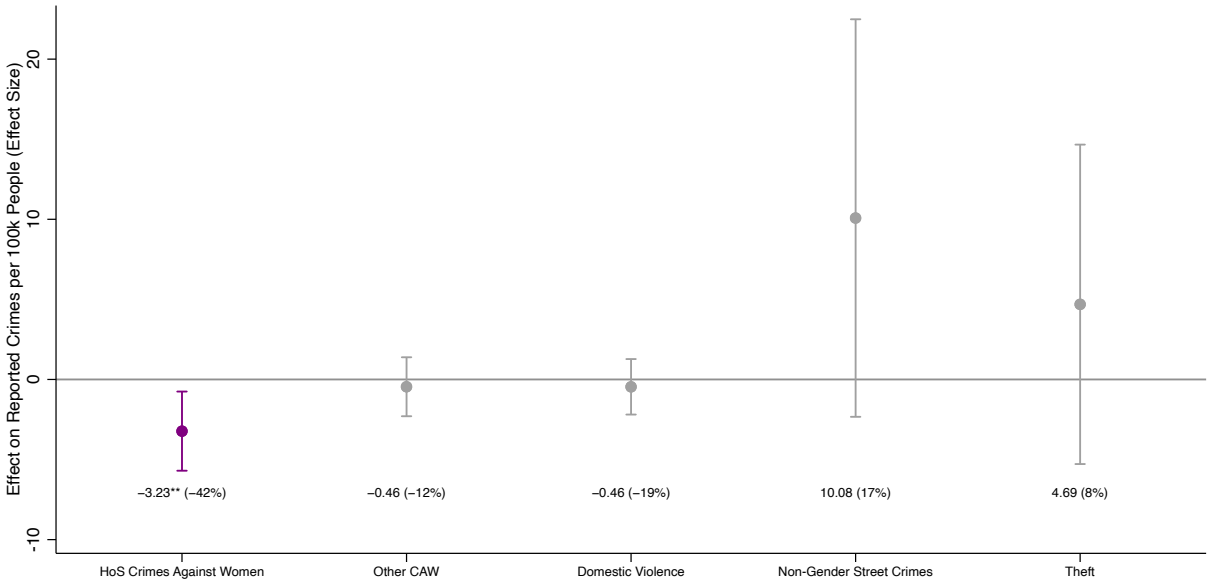


Figure 3.5: Effects of Metro station openings across crime categories

Notes: Each point reports the average across post-opening event-time estimates from the Callaway and Sant'Anna (2021) ring specification for the indicated category of reported crime per 100,000 residents. The 0–1 km cell is treated from the month its associated station opens, while the corresponding 1–2 km cell is coded as a comparison unit throughout the panel. The spatial sample contains 57 Metro stations opening between 2016 and 2019. Bars denote 95% confidence intervals, with standard errors clustered at the Metro-station level. Labels report the point estimate and, in parentheses, the estimate as a percentage of the corresponding pre-opening mean in treated 0–1 km cells. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Spatially distinct openings in already-connected jurisdictions. The spatial analysis shows that the decline in reported harassment is concentrated within one kilometer of a new station. This finding motivates an additional exercise using jurisdictions that already contained a Metro station before 2016 but received another station during the sample period more than one kilometer from any pre-existing Metro station. Although these openings do not represent a jurisdiction's first connection to the network, they newly expose a spatially distinct part of the jurisdiction to the immediate station environment. I expand the baseline sample by adding the 13 qualifying jurisdictions and assign their treatment dates according to the opening of the new, spatially distinct station. The resulting estimate therefore combines first-station openings in the original treated jurisdictions with additional-station openings in already-connected jurisdictions. The average post-opening estimate is -2.00 reports per 100,000 residents per month, with a standard error of 0.706. This composite estimate shows that the main conclusion is robust to expanding the sample to include additional stations that serve previously unexposed parts of already-connected jurisdictions. Appendix Table A.7 reports the estimate.

Displacement into the Metro system. A station might also displace harassment from surrounding streets onto Metro platforms or trains. Such incidents are filed at dedicated Metro police stations and are excluded from the territorial-jurisdiction panel. To assess whether this form of displacement could account for the estimated decline, I conduct a back-of-the-envelope calculation that applies the baseline average rate reduction of 2.533 reports per 100,000 residents uniformly to the population of every treated jurisdiction in each post-opening month. This calculation implies approximately 1,180 fewer harassment reports between January 2016 and December 2019. Over the same period, Metro police stations serving the 57 newly opened stations recorded only 42 reports meeting the same harassment definition, accounting for less than 4% of the implied decline. Reported harassment within the Metro system therefore cannot quantitatively account for the estimated reduction around station openings.

Together, the jurisdiction-level and spatial results show that Metro station openings are followed by a large, outcome-specific decline in reported harassment concentrated within one kilometer of the station, with no evidence of an offsetting increase in adjacent streets or within the Metro system. These findings establish where the decline occurs but do not by themselves determine why stations make surrounding public spaces safer. I next examine whether the timing and heterogeneity of the effect, together with changes in the observed street environment, are consistent with stations increasing local activity and informal surveillance.

4 Mechanisms

Why does a Metro station reduce reported harassment in the surrounding public space? I consider three explanations. First, a station may generate local activity and informal surveillance—the "eyes on the street" of Jacobs (1961)—that increase the likelihood that harassment is observed and potential offenders are deterred. Second, the formal policing or private security associated with a station may increase deterrence. Third, the incidence of harassment may remain unchanged while victims' reporting behavior or police registration practices change. These explanations are not mutually exclusive, and no single test can distinguish them conclusively. They do, however, generate different predictions that can be evaluated jointly.

An informal-surveillance mechanism should be most evident when station-related activity reaches the surrounding neighborhood, should generate larger safety gains where little activity existed beforehand, and should coincide with improvements in the visibility and perceived safety of nearby public space. A uniform increase in formal security may also reduce harassment and should be reflected in visible security around stations, but it does not by itself predict a systematically larger effect in initially inactive neighborhoods. A pure reporting explanation does not predict independently measured improvements in the street environment; moreover, a simple account in which hurried commuters become less likely to report should plausibly apply during both morning and evening travel peaks. I organize the evidence around these contrasting predictions before assessing what the results imply for the competing explanations.

4.1 When: The Effect Is Concentrated in the Evening

The timing of the effect provides the first test. Using more than 80 million fare-collection records from 237 Delhi Metro stations across the National Capital Region (NCR), Xavier et al. (2025) document a morning travel peak between 07:00 and 10:00 and an evening peak between 17:00 and 20:00. During the morning peak, the stations making the largest contributions to link-level crowding are concentrated around major employment destinations, whereas during the evening peak they are more widely distributed across residential and mixed-use districts. This spatial reversal bears the familiar signature of the daily commute: flows concentrated around employment centers in the morning disperse toward residential neighborhoods at the end of the working day.[9]

I assign each reported incident to one of four mutually exclusive six-hour intervals—03:00–

[9] Although the data come from September 2023, after the period studied here, they provide independent descriptive evidence on the daily timing and spatial distribution of Delhi Metro passenger flows.

09:00, 09:00–15:00, 15:00–21:00, and 21:00–03:00—using its recorded time of occurrence, and re-estimate the 0–1 versus 1–2 km specification separately for each interval.[10] Figure 4.1 (Appendix Table A.11) shows that the reduction is concentrated between 15:00 and 21:00, the interval containing the evening travel peak. Reported harassment falls by approximately 2.0 reports per 100,000 residents during this interval, and this is the only six-hour estimate that is statistically distinguishable from zero. I find no comparable decline in either of the intervals overlapping the morning peak or during the remainder of the day.

This concentration is consistent with informal surveillance being especially valuable when station activity reaches surrounding neighborhoods during the evening, as natural visibility declines. Taken alone, however, the timing does not exclude a role for formal security, which may also vary over the day. It does weigh against the simple reporting explanation that commuters forgo reporting because they are in a hurry: morning travel also has a pronounced peak, but reported harassment does not decline in either six-hour interval that overlaps it.

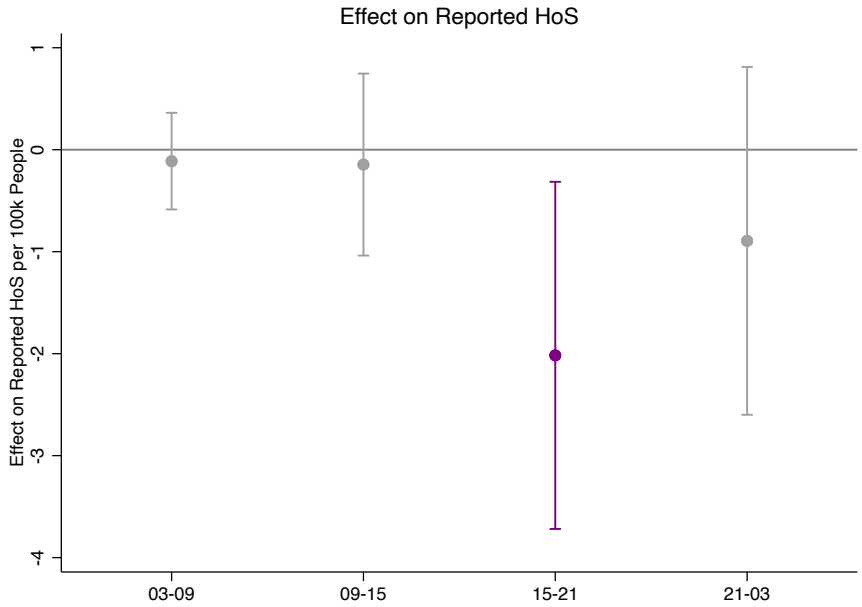


Figure 4.1: Effect of a station opening on reported harassment by time of occurrence

Notes: Each point reports the estimated effect of a station opening on reported harassment crimes against women per 100,000 residents within the indicated six-hour interval. Estimates compare the 0–1 km ring with the 1–2 km ring using the Callaway and Sant'Anna (2021) estimator. Bars denote 95% confidence intervals, with standard errors clustered at the Metro-station level. Incidents are assigned to intervals using their recorded time of occurrence.

[10] Appendix Section B.6 documents coverage of the time-of-occurrence field and Appendix Table A.12 reports a supplementary comparison of the narrower 4 hour intervals (00:00–04:00, 04:00–08:00, 08:00–12:00, 12:00–16:00, 16:00–20:00, 20:00–24:00) windows.

4.2 Where: The Effect Is Largest in Initially Inactive Areas

The second test asks whether the effect depends on how much activity surrounded a station before it opened. I divide stations at the median of nighttime luminosity—a proxy for baseline local activity—in their immediate (0–1 km) catchments in 2016 and re-estimate the ring specification separately for stations in lower- and higher-activity areas. Because the first station in this analysis opens in May 2017, luminosity in 2016 is measured before treatment for every station.

The reduction is substantially larger around stations in initially dim areas. The estimated effect in the 0–1 km ring relative to the 1–2 km ring is -5.45 reports per 100,000 residents in lower-activity areas, statistically significant at the 1% level, compared with -1.45 in higher-activity areas, which is not statistically distinguishable from zero (Figure 4.2). The low-minus-high difference is -4.00, with a standard error of 2.28, and is statistically significant at the 8% level. The same pattern appears when I compare the bottom and top quartiles of baseline luminosity (Table A.13). A station's 2016 baseline nightlight intensity is not significantly correlated with its pre-opening harassment rate in the 0–1 km ring (Pearson correlation 0.169, $p = 0.210$; Spearman's $\rho = 0.056$, $p = 0.682$; $N = 57$), indicating that the low/high-activity split is not simply re-partitioning stations by their baseline level of reported harassment.

This heterogeneity is consistent with diminishing marginal returns to informal surveillance. Where streets were already active, an additional station may add relatively little to the number and continuity of potential observers. Where baseline activity was low, the same opening can produce a much larger change in the conditions under which harassment occurs. Conversely, a uniform formal-security explanation provides no immediate reason for the effect to depend so strongly on activity before the station opened. The result does not rule out formal security, but it indicates that a channel whose effect depends on the pre-existing street environment is important.

This heterogeneity also matters for the spatial allocation of transit investment. Conventional evaluations emphasize ridership, travel-time savings, and improvements in network accessibility. Because the safety gains identified here are largest where baseline activity was low, they need not be greatest in the locations that perform best according to these conventional criteria. Accounting for this place-based safety benefit could therefore change where new lines and stations appear most socially valuable. The estimates do not by themselves identify the optimal placement of transit infrastructure, which would require valuing the safety gains and weighing them against construction costs and other network benefits. They do show that omitting this benefit may distort both the estimated welfare return to transit expansion and its preferred spatial allocation.

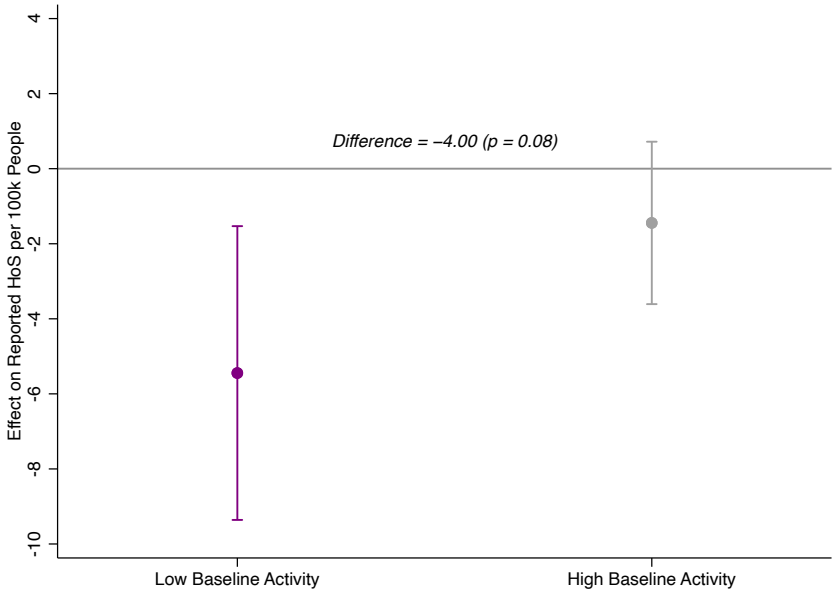


Figure 4.2: Effect heterogeneity by baseline local activity

Notes: Each point reports the Callaway and Sant'Anna (2021) estimate comparing reported harassment crimes against women per 100,000 residents in the 0–1 km and 1–2 km rings, estimated separately for stations below and above the median of 2016 nighttime luminosity in the 0–1 km catchment. Bars denote 95% confidence intervals, with standard errors clustered at the Metro-station level. The low-minus-high difference is −4.00 (standard error 2.28). Appendix Table A.13 reports alternative comparisons.

4.3 How: Changes in the Surrounding Public Environment

The third test examines whether the environment surrounding a station changes along the dimensions predicted by informal surveillance. I use Safetipin audits conducted between 2013 and 2024 and compare audit outcomes within 0–1 km of a station with those in the 1–2 km ring before and after it opens. Each specification includes station-by-ring fixed effects and audit-date fixed effects, and standard errors are clustered at the Metro-station level. I standardize each outcome by subtracting its mean and dividing by its standard deviation among audits conducted in the treated 0–1 km ring before the station opened.

Figure 4.3 shows that station openings are accompanied by substantial changes in the surrounding public environment. Visibility—which records windows, entrances, vendors, and other potential "eyes" facing the street—increases by 0.45 standard deviations. Lighting rises by 0.42 standard deviations, and auditors' perceived ability to return to the location alone at night (*Feeling*) improves by 0.49 standard deviations. Openness, visible police or private security, and the gender diversity of people using the space also improve significantly. By contrast, walkpath quality—a comparatively fixed feature not central to the proposed mechanism—does not change. Because these measures are collected independently of police crime reports, they establish that the observed decline in reported harassment coincides with a broader transformation of the surrounding public space.[11]

4.4 Assessing Formal Security and Changes in Reporting

The evidence permits a more precise assessment of the two alternative explanations. Safetipin audits show that visible police or private security increases after a station opens, suggesting that formal security may contribute to the decline. The analysis therefore cannot, and need not, exclude a complementary role for formal enforcement. A formal-security-only explanation is less consistent with the full pattern, however. In particular, a uniform increase in security at new stations does not explain why the reduction in harassment is substantially larger in neighborhoods with low baseline activity, or why several dimensions of informal watchability improve at the same time. The evidence instead suggests that formal and informal surveillance may operate together, with informal surveillance contributing more where baseline activity and ambient surveillance were lower.

[11] I omit Safetipin's *People* parameter from the main presentation because it is a coarse ordinal measure of the number of individuals visible at an audit point rather than a measure of pedestrian density or activity within the station catchment. I also do not interpret Safetipin's composite safety score as separate mechanism evidence because it incorporates both this parameter and access to public transportation, which may respond mechanically to a Metro opening. Appendix Table A.14 reports estimates for all individual parameters and the provider-defined composite score.

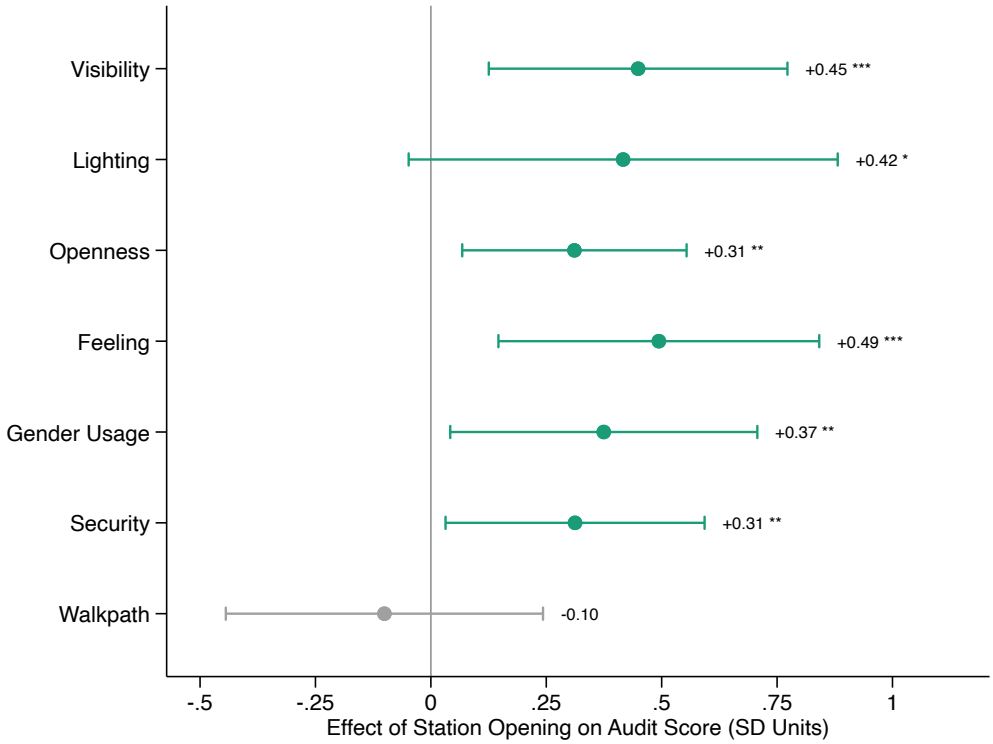


Figure 4.3: Effect of a station opening on Safetipin audit outcomes

Notes: Each point reports the coefficient from a separate difference-in-differences regression comparing Safetipin audit outcomes within 0–1 km of a new station with outcomes in the 1–2 km ring before and after the station opens. Specifications include station-by-ring and audit-date fixed effects and use audits within 0–2 km of a station ($N = 3{,}655$). Outcomes are standardized using the mean and standard deviation among pre-opening audits in the treated 0–1 km ring. Bars denote 95% confidence intervals, with standard errors clustered at the Metro-station level. Appendix Table A.14 reports estimates for all outcomes. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Changes in reporting require a different qualification. Because the outcome is constructed from FIRs, the estimates cannot mechanically separate changes in the incidence of harassment from changes in victims' willingness to report or police registration practices. Three results nevertheless weigh against a pure reporting explanation. First, the decline is concentrated in the evening and does not occur during the morning travel peak, which is difficult to reconcile with a simple account based on hurried commuters reporting less. Second, the ring-level outcome-specificity results in Figure 3.5 show no corresponding change in domestic violence, other crimes against women constructed to exclude harassment, or non-gender-specific street crimes. Third, the Safetipin audits document independent improvements in observable street conditions and perceived safety. None of these findings alone eliminates every possible reporting response, but together they are difficult to reconcile with unchanged safety and a decline arising solely from reporting or registration.

4.5 Interpreting the Mechanism Evidence

The mechanism evidence does not assign the entire decline in reported harassment to a single channel. It instead reveals a coherent pattern: harassment falls during the evening period in which Metro activity reaches residential neighborhoods; the reduction is largest where baseline activity was low; and independently collected audits show improvements in lighting, visibility, openness, gender usage, security, and perceived safety. The increase in visible security suggests a complementary role for formal enforcement, but the dependence of the effect on prior activity and the broader changes in street watchability point to informal surveillance as an important part of the response. Together with the outcome-specificity results, these patterns are also difficult to explain as a pure reporting artifact. Metro expansion therefore appears to generate a place-based safety return by changing the conditions under which women use the public spaces surrounding stations.

4.6 A Suggestive Downstream Response: Hostel Locations

I finally examine whether the accommodation market exhibits a corresponding gender-specific pattern. Using the MagicBricks hostel listings described in Section 2.3, for each hostel, I measure its distance to its nearest station among those that opened between 2016 and 2019 and compare this distance across men-only hostels and women-only and mixed-gender hostels. I classify a hostel as pre-opening if its reported operating-since year precedes the opening year of its nearest station and as during-or-post-opening if it began operating in or after that year.

Figure 4.4 shows that women-only and mixed-gender hostels beginning operations during

or after a station's opening are located approximately 0.44 km closer to the station than those beginning operations beforehand. The corresponding difference among men-only hostels is 0.01 km. This pattern is consistent with safer and better-connected neighborhoods becoming relatively more attractive locations for accommodation serving women.

The comparison is descriptive rather than causal. The data are a cross-section of hostels listed on MagicBricks in May 2026 and do not include establishments that closed or were not listed at the time of collection. Moreover, the operating-since field identifies only the year in which a hostel began operating, and the same station opening improves transport access as well as potentially improving safety. The evidence therefore neither isolates safety from connectivity nor establishes a change in the gender composition of residents. It provides suggestive evidence of a possible downstream response in the local accommodation market.

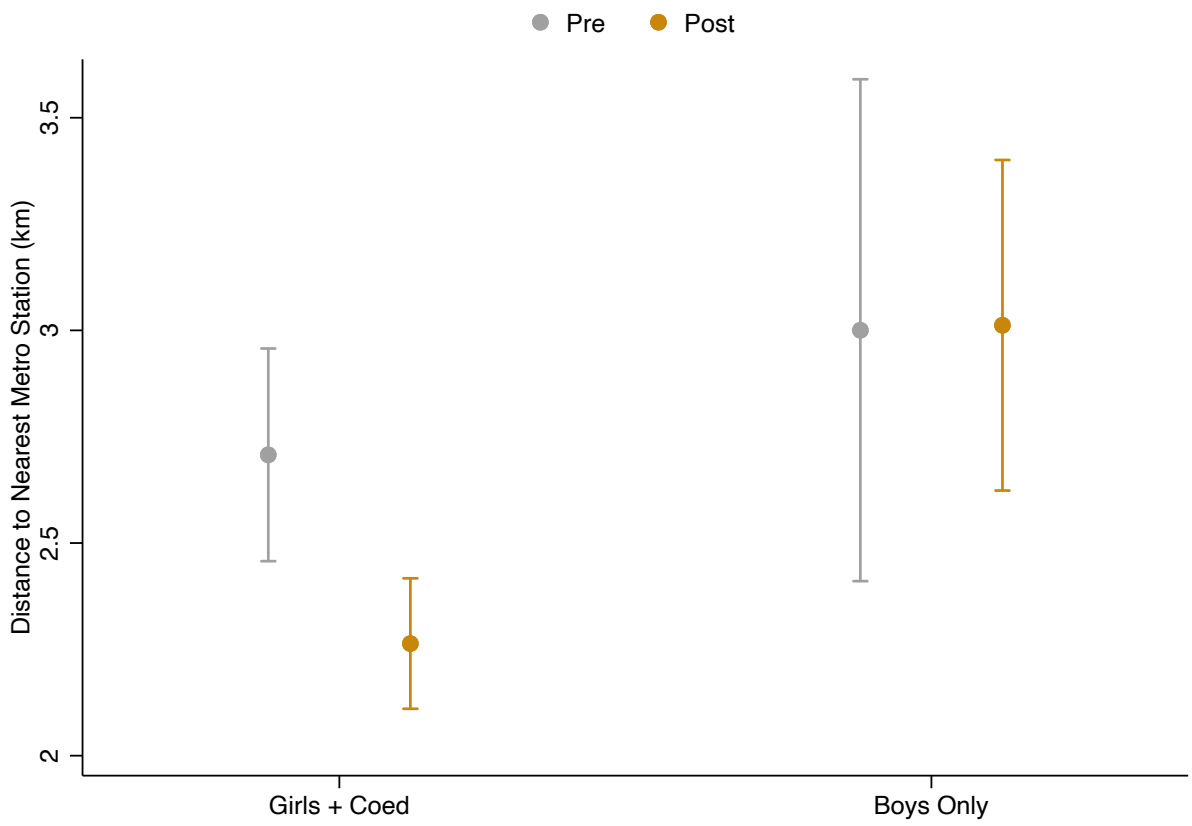


Figure 4.4: Hostel proximity to new Metro stations, by operating period and gender served

Notes: The figure reports mean distance to the nearest Metro station opening between 2016 and 2019 for MagicBricks hostel listings classified as men-only ("Boys only") or women-only and mixed-gender ("Girls + Coed"). "Pre-opening" denotes hostels whose reported operating-since year precedes the station's opening year; "during or post-opening" denotes hostels reporting that they began operating in or after that year. Bars denote 95% confidence intervals. The comparison is descriptive and does not identify the causal effect of a station opening on hostel entry or location.

5 Conclusion

Mass transit is typically evaluated by how it connects locations and moves people across a city. This paper shows that it can also transform the places it connects. I study the staggered arrival of Delhi Metro stations between 2016 and 2019 and estimate how a jurisdiction's first station affects reported harassment of women in the surrounding public space. The analysis is enabled by a new high-frequency crime dataset constructed from nearly 300,000 scanned, bilingual police reports. Converting these records into structured data, harmonizing changes in police-jurisdiction boundaries, and geocoding incident locations make it possible to examine both changes across jurisdictions and their precise spatial distribution around new stations.

A jurisdiction's first Metro station reduces reported harassment by approximately 32% relative to the pre-opening mean among treated jurisdictions. The decline is concentrated within one kilometer of a station, with no evidence of an offsetting increase in adjacent areas or within the Metro system. Nor is it accompanied by comparable changes in other crimes against women or non-gender-specific street crime. The effect is therefore neither a general decline in recorded crime nor a spatial redistribution of harassment away from station entrances.

The mechanism evidence points to changes in the surrounding public environment as an important part of the explanation. Harassment falls primarily during the evening, when Metro passenger flows disperse into residential neighborhoods, and the reduction is largest around stations in areas with low baseline activity. Independently collected safety audits show corresponding improvements in lighting, visibility, openness, the presence of women and children, and women's perceived ability to use these spaces alone at night. Visible formal security also increases, suggesting that formal and informal surveillance may operate together. Because the main outcome is reported crime, changes in reporting or registration cannot be separated completely from changes in incidence. Yet the timing of the decline, its concentration in initially inactive areas, its specificity to harassment, and the independently measured transformation of the street environment are difficult to reconcile with a pure reporting explanation. Suggestive evidence from hostel listings also points to a possible downstream response in the accommodation market serving women.

These findings have implications for how cities evaluate and allocate transit investment. Conventional assessments emphasize ridership, travel-time savings, accessibility, and agglomeration benefits. The safety return identified here is instead place-based and appears largest in neighborhoods with relatively little prior activity. Accounting for it may therefore change not only the estimated social return to transit expansion but also where new lines

and stations appear most valuable. The results do not by themselves determine the optimal placement of transit infrastructure: doing so would require valuing these safety gains and weighing them against construction costs and other network benefits. They nevertheless show that infrastructure designed without an explicit gender objective can generate important gender-specific welfare gains by changing women's ability to use public space.

References

Akbar, P., Couture, V., Duranton, G., and Storeygard, A. (2023). Mobility and congestion in urban india. *American Economic Review*, 113(4):1083–1111.

Aloul, S., Naffa, R., and Mansour, M. (2018). Gender in public transportation: A perspective of women users of public transportation. Technical report, SADAQA and Friedrich-Ebert-Stiftung.

Amaral, S., Bhalotra, S., and Prakash, N. (2022). Gender, crime and punishment: Evidence from women police stations in india. Revise and Resubmit, *Economic Development and Cultural Change*.

Anderson, M. L. (2014). Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *American Economic Review*, 104(9):2763–2796.

Asher, S., Lunt, T., Matsuura, R., and Novosad, P. (2021). Development research at high geographic resolution: An analysis of night-lights, firms, and poverty in india using the shrug open data platform. *The World Bank Economic Review*, 35(4):845–871.

Asher, S. and Novosad, P. (2020). Rural roads and local economic development. *American Economic Review*, 110(3):797–823.

Beaman, L. A., Jayachandran, S., and Sharma, A. (2026). The city through her eyes: Gender-based harassment in public spaces in urban india. Working Paper 35738, National Bureau of Economic Research.

Borker, G. (2021). Safety first: Perceived risk of street harassment and educational choices of women. Working Paper 9731, World Bank Policy Research Working Paper.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies*, 91(6):3253–3285.

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Card, D. and Dahl, G. B. (2011). Family violence and football: The effect of unexpected emotional cues on violent behavior. *The Quarterly Journal of Economics*, 126(1):103–143.

Colella, F., Lalive, R., Sakalli, S. O., and Thoenig, M. (2019). Inference with arbitrary clustering. IZA Discussion Paper 12584, IZA Institute of Labor Economics.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1):1–45.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

DeAngelo, G., Gittings, K., Grossman, D. S., and Khalil, U. (2019). Urban transport and crime: Evidence from unanticipated mass transit strikes. *Economic Inquiry*, 57(3):1718–1737.

Delhi Metro Rail Corporation (2020). Annual report 2019–20. Delhi Metro Rail Corporation Ltd., New Delhi. Directors' Report, Status of the Delhi MRTS Project up to Phase III, p. 6.

Gardner, J. (2022). Two-stage differences in differences. Working paper. arXiv:2207.05943.

Gibbons, S. and Machin, S. (2005). Valuing rail access using transport innovations. *Journal of Urban Economics*, 57(1):148–169.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Heblich, S., Redding, S. J., and Sturm, D. M. (2020). The making of the modern metropolis: Evidence from london. *The Quarterly Journal of Economics*, 135(4):2059–2133.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review*, 102(2):994–1028.

Hossain, M. A., Mahajan, K., and Sekhri, S. (2022). Access to toilets and violence against women. *Journal of Environmental Economics and Management*, 114:102695.

Iyer, L., Mani, A., Mishra, P., and Topalova, P. (2012). The power of political voice: Women's political representation and crime in india. *American Economic Journal: Applied Economics*, 4(4):165–193.

Jacobs, J. (1961). *The Death and Life of Great American Cities*. Random House, New York.

Jassal, N. (2020). Gender, law enforcement, and access to justice: Evidence from all-women police stations in india. *American Political Science Review*, 114(4):1035–1054.

Khanna, G., Medina, C., Nyshadham, A., Ramos-Menchelli, D., Tamayo, J., and Tiew, A. (2022). Spatial mobility, economic opportunity, and crime. Revise and Resubmit, *American Economic Review*.

Kyba, C. C. M. et al. (2017). Artificially lit surface of earth at night increasing in radiance and extent. *Science Advances*, 3(11):e1701528.

Luca, D. L., Owens, E., and Sharma, G. (2015). Can alcohol prohibition reduce violence against women? *American Economic Review*, 105(5):625–629.

Miller, A. R. and Segal, C. (2019). Do female officers improve law enforcement quality? effects on crime reporting and domestic violence. *The Review of Economic Studies*, 86(5):2220–2247.

Murty, M. N., Dhavala, K. K., Ghosh, M., and Singh, R. (2006). Social cost-benefit analysis of delhi metro. Working paper, Institute of Economic Growth.

National Crime Records Bureau (2023). Crime in india 2023. Ministry of Home Affairs, Government of India, New Delhi. Table 3A.1, Crimes Against Women (State/UT-wise) - 2021-2023.

Office of the Registrar General and Census Commissioner, India (2011). Primary census abstract, census of india 2011. Ministry of Home Affairs, Government of India, New Delhi. District-level data for the National Capital Territory of Delhi (files DDW_PCA0701–0709), https://censusindia.gov.in.

Phillips, D. C. and Sandler, D. (2015). Does public transit spread crime? evidence from temporary rail station closures. *Regional Science and Urban Economics*, 52:13–26.

Ratho, A. and Jain, S. (2021). Women on the move: The impact of safety concerns on women's mobility. Technical report, Observer Research Foundation, New Delhi.

Román, M. O. et al. (2018). NASA's black marble nighttime lights product suite. *Remote Sensing of Environment*, 210:113–143.

Sukhtankar, S., Kruks-Wisner, G., and Mangla, A. (2022). Policing in patriarchy: An experimental evaluation of reforms to improve police responsiveness to women in india. *Science*, 377(6602):191–198.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Tsivanidis, N. (2026). Evaluating the impact of urban transit infrastructure: Evidence from bogotá's transmilenio. *American Economic Review*, 116(2):418–463.

UN Women and ICRW (2013). Safe cities free of violence against women and girls: Baseline findings, delhi. UN Women and International Center for Research on Women.

Viswanath, K. and Basu, A. (2015). Safetipin: An innovative mobile app to collect data on women's safety in indian cities. *Gender & Development*, 23(1):45–60.

Winston, C. and Maheshri, V. (2007). On the social desirability of urban rail transit systems. *Journal of Urban Economics*, 62(2):362–382.

Xavier, B. D., Varghese, V., Chikaraishi, M., and Fujiwara, A. (2025). Measuring the impact of built environment factors on station-level contributions to link-level crowding using a novel crowding contribution index. *Scientific Reports*, 15:43535.

Zárate, R. D. (2022). Spatial misallocation, informality, and transit improvements: Evidence from mexico city. Policy research working paper, World Bank.

A Additional Tables

Table A.1: Summary statistics of the construction phases of Delhi Metro

Phase	Line	Route	Stations	Length (km)	Date
Phase-I	Red Line	Shahdara – Tis Hazari	6	8.35	Dec 2002
		Tis Hazari – Inderlok	4	4.87	Oct 2003
		Inderlok – Rithala	8	8.84	Apr 2004
	Yellow Line	Vishwavidyalaya – Kashmere Gate	4	4.06	Dec 2004
		Kashmere Gate – Central Secretariat	6	6.62	Jul 2005
	Blue Line	Barakhamba – Dwarka	22	22.74	Dec 2005
		Dwarka – Dwarka Sector 9	6	6.47	Apr 2006
		Barakhamba – Indraprastha	3	2.80	Nov 2006
Phase-II	Red Line	Shahdara – Dilshad Garden	3	3.09	Jun 2008
	Yellow Line	Vishwavidyalaya – Jahangirpuri	5	6.36	Feb 2009
		HUDA City Centre – Qutab Minar	10	15.82	Jun 2010
		Central Secretariat – Qutab Minar	9	11.76	Sep 2010
	Blue Line	Indraprastha – Yamuna Bank	1	2.17	May 2009
		Yamuna Bank – Noida City Centre	10	12.85	Nov 2009
		Dwarka Sector 9 – Dwarka Sector 21	2	2.76	Oct 2010
		Anand Vihar ISBT – Vaishali	2	2.57	Jan 2010
	Green Line	Inderlok – Kirti Nagar – Mundka	16	18.46	
	Orange Line	New Delhi – IGI Airport (T-3) – Dwarka Sector-21	6	22.70	Feb 2011
	Violet Line	Central Secretariat – Sarita Vihar	13	15.34	Oct 2010
		Sarita Vihar – Badarpur	3	4.82	Jan 2011
Phase-III	Red Line	Dilshad Garden – Shaheed Sthal (New Bus Adda)	8	9.63	Mar 2019
	Yellow Line	Jahangirpuri – Samaypur Badli	3	4.38	Nov 2015
	Blue Line	Noida City Centre – Noida Electronic City	6	6.80	Mar 2019
	Violet Line	Central Secretariat – Janpath – Mandi House	2		Jun 2014
		ITO station			Jun 2015
		Sarai – Escorts Mujesar (Faridabad)	9	13.56	Sep 2015
		Delhi Gate – Kashmere Gate	3		May 2017
		Escorts Mujesar – Raja Nahar Singh (Ballabhgarh)	2	3.35	Nov 2018
	Pink Line	Majlis Park – Durgabhai Deshmukh South Campus	12	21.57	Mar 2018
		Durgabhai Deshmukh South Campus – Lajpat Nagar	6	8.53	Aug 2018
		Trilokpuri – Sanjay Lake – Shiv Vihar	15	17.86	Oct 2018
		Lajpat Nagar – Mayur Vihar Pocket-1	5	9.63	Dec 2018
	Magenta Line	Botanical Garden – Kalkaji Mandir	9	12.64	Dec 2017
		Kalkaji Mandir – Janakpuri West	16	24.82	May 2018
	Grey Line	Dwarka – Najafgarh	4	4.98	Oct 2019

Notes: Routes highlighted in purple contain the metro stations that opened between 2016 and 2019 and are the first stations within their respective police jurisdictions; these are the stations used in the analysis. *Source:* https://www.delhimetrorail.com/pages/en/present-network.

Table A.2: Summary Statistics

	Mean	Std. Dev.	Min	Max
Panel A: Jurisdiction–month panel (jurisdiction design)				
Reported harassment	3.89	9.45	0.00	205.26
Population (2011 Census)	111,995	109,586	3,898	644,122
Panel B. Ring design, 0–1 km vs. 1–2 km (57 stations)				
Reported harassment	4.57	18.06	0.00	484.26
Other crimes against women	2.65	8.04	0.00	151.07
of which: domestic violence	1.79	6.43	0.00	151.07
Non-gender street crime	35.54	83.08	0.00	1420.01
of which: theft	32.29	92.88	0.00	1631.50
Ring population	39,498	51,961	207	454,527
Panel C. Safetipin safety audits (3,655 audit points, 56 stations)				
Safety score (composite, 0–10)	7.12	2.34	0.20	9.80
Lighting	1.62	0.81	0.00	3.00
Openness	1.58	0.76	0.00	3.00
Visibility	1.53	0.90	0.00	3.00
People	1.61	0.82	0.00	3.00
Security	1.10	0.98	0.00	3.00
Walkpath	1.70	1.04	0.00	3.00
Public transport	1.62	0.87	0.00	3.00
Gender usage	1.35	0.91	0.00	3.00
Feeling	1.75	0.81	0.00	3.00
Panel D. Hostels (1,282 hostels, 52 nearest-station clusters)				
Distance to nearest metro (km)	2.56	2.37	0.03	17.34
Women's / coed (indicator)	0.74	0.44	0.00	1.00

Notes: Crime variables are monthly crime reports per 100,000 residents. Panel A covers the 79 police jurisdictions used in the jurisdiction-level design (25 treated, 54 never-treated) over January 2016–December 2019 ($N = 3{,}792$ jurisdiction-months). Panel B describes the station-ring sample used in the spatial design; each observation is a station–ring–month cell, and statistics pool the treated (0–1 km) and the control ring (1–2 km). Ring population is the residential count within the band. Panel C reports Safetipin safety-audit scores. The composite safety score ranges 0–10; each of the nine component dimensions is scored 0–3. Variables enter the regressions standardized (mean 0, SD 1) within the pre-treatment sample; raw scores are shown here for interpretability. $N = 3{,}655$ audit points across 56 stations; only 56 of the 57 ring-design stations have at least one Safetipin audit point within 0–2,km, and the remaining station is excluded from this panel. Panel D covers 1,282 hostels, each assigned to its nearest metro station (52 stations); only 52 of the 57 stations are the nearest station to at least one hostel in the cleaned MagicBricks sample, and the other stations have no hostel assigned to them as nearest.

Table A.3: Balance: Treated vs. Control Police Jurisdictions (Census 2011)

Census 2011 variable	Treated mean	Control mean	Difference	p-value
Total population	110,843	112,528	−1,684	0.95
Households	22,425	22,752	−327	0.95
Male population	59,202	60,361	−1,159	0.94
Female population	51,642	52,167	−525	0.97
Male scheduled castes	7,624	11,203	−3,579	0.18
Female scheduled castes	6,791	9,999	−3,208	0.18
Male literate	47,513	47,977	−463	0.97
Female literate	36,715	36,255	460	0.96
Male workers	30,853	31,751	−898	0.90
Female workers	5,348	5,109	238	0.83

Notes: Means of Census 2011 characteristics for treated jurisdictions (first metro station opens 2016–2019) and control jurisdictions. The p-value is for a test of equality of means. No difference is statistically significant, consistent with station siting being unrelated to jurisdiction demographics.

Table A.4: Robustness: Alternative Comparison Groups

	Not-yet-treated (baseline)	Never-treated	Treated after 2019
	(1)	(2)	(3)
Post 1st metro station	-2.533^{***}	-2.537^{***}	-2.511^{**}
	(0.976)	(0.976)	(1.022)
Jurisdiction FE	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes
Observations	3,792	3,792	1,872

Notes: Callaway and Sant'Anna (2021) estimates; the outcome is reported harassment crimes against women per 100,000. Column (1) is the baseline, which uses not-yet-treated (and never-treated) jurisdictions as controls; column (2) restricts controls to never-treated jurisdictions; column (3) adds jurisdictions receiving their first station after 2019 as untreated controls. Standard errors, clustered at the police-jurisdiction level, are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.5: Robustness: Alternative Heterogeneity-Robust Estimators

	Sun & Abraham (1)	Borusyak et al. (2)	de Ch. & D'Haultf. (3)	Gardner (2-stage) (4)
Post 1st metro station	-2.537^{***}	-2.465^{**}	-1.963^{**}	-2.465^{*}
	(0.964)	(1.028)	(0.961)	(1.274)
Jurisdiction FE	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes
Observations	3,792	3,792	3,792	3,792

Notes: Each column re-estimates the average post-opening effect on reported harassment crimes against women per 100,000 using a different heterogeneity-robust estimator (Sun and Abraham, 2021; Borusyak et al., 2024; de Chaisemartin and D'Haultfœuille, 2020; Gardner, 2022). Standard errors, clustered at the police-jurisdiction level, are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.6: Robustness: Functional Form and Count Models

	Poisson (PPML) (1)	IHS outcome (2)
Post 1st metro station	-0.302^{**}	-0.716^{***}
	(0.134)	(0.207)
Implied % change	-26%	—
Jurisdiction FE	Yes	Yes
Month-year FE	Yes	Yes
Observations	3,792	3,792

Notes: Column (1) is a Poisson (PPML) regression of the harassment *count* on the treatment indicator with population as exposure, so the coefficient is a log rate-ratio; the implied proportional effect is $100 \times (e^{\beta} - 1) = -26\%$. Column (2) applies the inverse hyperbolic sine (IHS) transformation to the outcome, which is defined at zero and so retains the zero-harassment months; the reported effect is the post-opening average, estimated against a flat, insignificant pre-period (-0.069, s.e. 0.228). Standard errors, clustered at the police-jurisdiction level, are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.7: Robustness: Sample Composition

	ATT	Std. err.	Observations
Baseline	-2.533^{***}	(0.976)	3,792
Leave-one-cohort-out (range)	[−2.91, −1.67]	—	—
Leave-one-jurisdiction-out (range)	[−2.78, −1.79]	—	—
Exclude high-crime jurisdictions	-2.506^{***}	(0.699)	3,402
Redefine treatment (2nd station, >1 km from pre-2016 station)	-2.003^{***}	(0.706)	4,416

Notes: The outcome is reported harassment crimes against women per 100,000; each row re-estimates the Callaway and Sant'Anna (2021) average post-opening effect on a modified sample. The leave-one-out rows report the range of the point estimate across all iterations—dropping each of the seven opening cohorts, or each treated jurisdiction, in turn—and every iteration is negative and significant at the 5% level. "Exclude high-crime jurisdictions" drops jurisdictions above the 90th percentile of the 2016 (pre-treatment) harassment rate per 100,000, computed across all jurisdictions. "Redefine treatment" adds back the 13 jurisdictions that already contained a station before 2016, timing treatment by their first station to open during 2016–2019 and retaining only those whose new station lies more than one kilometer from the nearest pre-2016 station, so that it activates a previously unserved catchment. Standard errors, clustered at the police-jurisdiction level, are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.8: Inference: Spatial and Serial Correlation

	TWFE	Conley spatial-HAC SE		
	(clustered)	3 km	5 km	10 km
	(1)	(2)	(3)	(4)
Post 1st metro station	-2.910^{*}	-2.910^{**}	-2.910^{**}	-2.910^{**}
	(1.727)	(1.212)	(1.240)	(1.244)
Jurisdiction FE	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes
Observations	3,792	3,792	3,792	3,792

Notes: All four columns report the same two-way fixed-effects estimate of the effect of a jurisdiction's first metro station on reported harassment crimes against women per 100,000; they differ only in how the standard error is computed. Column (1) clusters at the police-jurisdiction level. Columns (2)–(4) report Conley (1999) spatial-HAC standard errors, computed using Colella et al. (2019), allowing spatial correlation among jurisdictions whose centroids lie within the indicated distance and serial correlation up to twelve months. The static two-way fixed-effects specification is reported because the Conley estimator requires it; the efficient Callaway and Sant'Anna (2021) estimator used throughout the paper yields a point estimate of −2.533 with a standard error of 0.976 (Table 3.1). Standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.9: Change in Reported Harassment by Distance Band Around New Stations

Distance band to nearest new station	Post − Pre change (per 100,000)
0–1 km	-2.434^{**}
	(0.975)
1–2 km	0.529
	(0.782)
2–3 km	0.552
	(0.545)
3–4 km	0.381
	(0.485)
4–5 km	0.387
	(0.700)
Station fixed effects	Yes
Month-year fixed effects	Yes
Observations	11,712
Stations	57

Notes: Each entry is the post-opening change in reported harassment crimes against women per 100,000 residents in the given distance band, from the saturated specification in Equation (5) (`dist_bin` × `post` with station and month-year fixed effects), recovered as the average marginal effect of `Post` within each band. Standard errors, clustered at the Metro-station level, are in parentheses. Areas within 1 km of a station open before 2016 are excluded. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.10: Effect on alternative crime outcomes per 100,000 residents

	Reported harassment	Other CAW	Domestic violence	Non-gendered street crimes	Theft
Post 1st metro station	-3.227^{**}	−0.455	−0.459	10.078	4.690
	(1.260)	(0.939)	(0.884)	(6.332)	(5.088)
Station × band fixed effects	Yes	Yes	Yes	Yes	Yes
Month-year fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	5,472	5,472	5,472	5,472	5,472

Notes: Each column reports the Callaway and Sant'Anna (2021) ring ATT (0–1 km treated vs. the 1–2 km control ring) for a different outcome, expressed as reported crimes per 100,000 residents. "Reported harassment" is the paper's main outcome. "CAW" denotes crimes against women; "Other CAW" excludes harassment from the broader crimes-against-women category. "Non-gendered crimes" aggregates non-gender-specific street crime (for example, theft and robbery); "Theft" reports this single component separately. Standard errors, clustered at the Metro-station level, are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.11: Effect on reported harassment by time of day: six-hour bands

	3am–9am	9am–3pm	3pm–9pm	9pm–3am
Post 1st metro station	−0.112	−0.146	−2.017**	−0.894
	(0.242)	(0.455)	(0.868)	(0.870)
Station × band fixed effects	Yes	Yes	Yes	Yes
Month-year fixed effects	Yes	Yes	Yes	Yes
Observations	5,472	5,472	5,472	5,472

Notes: Each column reports the Callaway and Sant'Anna (2021) ring ATT (0–1 km treated vs. the 1–2 km control ring) for reported harassment crimes against women per 100,000 residents, based on each crime's time of occurrence within the stated six-hour window. Standard errors, clustered at the Metro-station level, are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.12: Effect on reported harassment by time of day: four-hour bands

	12–4am	4–8am	8am–12pm	12–4pm	4–8pm	8pm–12am
Post 1st metro station	0.137	−0.070	0.030	−0.274	−1.625**	−1.390
	(0.141)	(0.197)	(0.158)	(0.407)	(0.718)	(1.011)
Station × band fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Month-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,472	5,472	5,472	5,472	5,472	5,472

Notes: Each column reports the Callaway and Sant'Anna (2021) ring ATT (0–1 km treated vs. the 1–2 km control ring) for reported harassment crimes against women per 100,000 residents, based on each crimes's time of occurrence within the stated four-hour window. This finer-grained cut is a check that the pattern in Table A.11 is not an artifact of where the six-hour cutoffs fall. Standard errors, clustered at the Metro-station level, are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.13: Effect heterogeneity by baseline economic activity

	Low activity	High activity	Difference
Panel A: Below vs. above median			
Post-opening ATT	-5.445^{***}	-1.445	-4.000^{*}
	(1.997)	(1.104)	(2.282)
Observations	2,784	2,688	—
Stations	29	28	—
Panel B: Bottom vs. top quartile			
Post-opening ATT	-2.223^{***}	-1.022	-1.201
	(0.773)	(0.775)	(1.095)
Observations	1,440	1,440	—
Stations	15	15	—
Station × band fixed effects	Yes	Yes	Yes
Month-year fixed effects	Yes	Yes	Yes

Notes: Each cell reports the Callaway and Sant'Anna (2021) ring ATT (0–1 km treated vs. the 1–2 km control ring). Panel A estimates the specification separately for stations below and above the median of baseline nightlight intensity; Panel B restricts to the bottom and top quartiles of the same distribution, excluding the middle 50% of stations. Because Panel A and Panel B are estimated on different subsamples of the 57 ring-design stations, each column reports its own observation and station count within its panel; the two station counts in a panel sum to that panel's total, and Panel B's total is smaller than Panel A's because the middle 50% of stations is dropped. The "Difference" column is the low-minus-high contrast, with standard error $\sqrt{\mathrm{se}_{\mathrm{low}}^2 + \mathrm{se}_{\mathrm{high}}^2}$, since the two groups are disjoint sets of stations; it is a computed contrast rather than a separate regression, so no observation count is reported for it. The quartile split shows the same direction of heterogeneity as the median split, though the difference is smaller in magnitude and not statistically significant. Standard errors, clustered at the Metro-station level, are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.14: Effect of a Station Opening on Safetipin Audit Scores

	Visibility	Lighting	Openness	Feeling	Gender usage
Treated × post	0.449***	0.417*	0.311**	0.494***	0.375**
	(0.165)	(0.237)	(0.124)	(0.177)	(0.170)
	Security	Public transport	People†	Walkpath†	Safety score‡
Treated × post	0.312**	0.443**	0.115	−0.101	0.283*
	(0.143)	(0.210)	(0.185)	(0.175)	(0.155)
Station × ring FE			Yes		
Audit-date FE			Yes		
Observations			3,655		

Notes: Each column is a separate difference-in-differences regression of the indicated Safetipin parameter on the treated-ring × post interaction, with station-by-ring and audit-date fixed effects, on audits within 0–2 km of a station. Parameters are standardized on the pre-opening treated distribution, so estimates are in standard deviations and comparable across columns. Standard errors, clustered at the Metro-station level, are in parentheses. † dimensions the mechanism does not predict should respond; ‡ composite index of all parameters. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B Construction of the Crime Data

B.1 From Scanned Reports to Structured Records

This appendix describes the construction of the First Information Report (FIR) dataset, which transforms the raw, unstructured reports published by the Delhi Police into a structured, machine-readable panel. I first scrape all available FIRs from the Delhi Police website using the command-line tool *wget*.[12] Across all station-years, the number of reports retrieved equals essentially 100% of the highest FIR serial number (Table B.1); retrieval marginally exceeds this maximum in some years because sensitive reports omitted from the searchable index are nonetheless captured, so the scrape recovers the full published record and a small number of reports beyond it. A sample FIR is shown in Figure B.1.

Table B.1: Completeness of FIR extraction

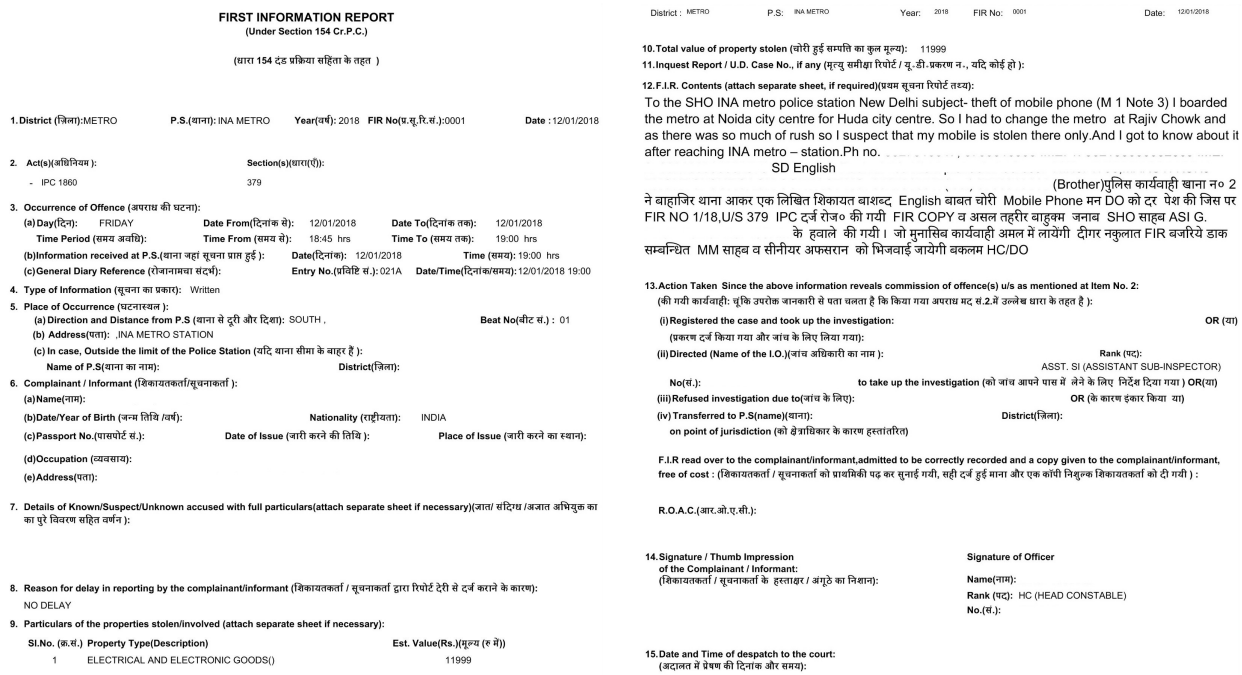
Year	Reports retrieved	Registered	Coverage
2016	102,218	102,183	100.0%
2017	90,009	89,986	100.0%
2018	82,568	82,532	100.0%
2019	83,007	82,954	100.1%
Total	357,802	357,655	100.0%

Notes: Within each police station and year, First Information Reports are numbered sequentially from one, so the highest FIR serial number ("Registered") gives the number filed that year. "Reports retrieved" is the number successfully scraped. Coverage is the ratio of the two, aggregated across all police stations and consolidating stations across the 2018 reorganization of Delhi's police districts. Coverage marginally above 100% in some years arises because certain sensitive FIRs are withheld from the portal's searchable index— lowering the highest *visible* serial number—yet are still retrieved, so the count retrieved can slightly exceed the highest indexed number.

The reports are published as scanned PDFs—images of the filed forms rather than machine-readable text—and are bilingual, with fields recorded in both English and Hindi (Figure B.2). Crucially, every FIR follows the same standardized fifteen-section form, and I exploit this fixed layout rather than attempting free-form text recognition. Each page is rendered to a high-resolution grayscale image (500 DPI) and passed to the Tesseract OCR engine (via *pytesseract*) using a combined English–Hindi language model. To locate the fields, I first read the numbered section headers of the form from the recognized page layout (using

[12] http://59.180.234.21:8080/citizen/firSearch.htm

Figure B.1: A Sample First Information Report



FIRST INFORMATION REPORT
(Under Section 154 Cr.P.C.)

(धारा 154 दंड प्रक्रिया संहिता के तहत)

1. District (जिला):METRO P.S.(थाना): INA METRO Year(वर्ष): 2018 FIR No(प्र.सू.रि.सं.):0001 Date : 12/01/2018

2. Act(s)(अधिनियम): Section(s)(धारा(एँ)):
- IPC 1860 379

3. Occurrence of Offence (अपराध की घटना):
(a) Day(दिन): FRIDAY Date From(दिनांक से): 12/01/2018 Date To(दिनांक तक): 12/01/2018
Time Period (समय अवधि): Time From (समय से): 18:45 hrs Time To (समय तक): 19:00 hrs
(b) Information received at P.S.(थाना जहां सूचना प्राप्त हुई): Date(दिनांक): 12/01/2018 Time (समय): 19:00 hrs
(c) General Diary Reference (रोजनामचा संदर्भ): Entry No.(प्रविष्टि सं.): 021A Date/Time(दिनांक/समय): 12/01/2018 19:00

4. Type of Information (सूचना का प्रकार): Written

5. Place of Occurrence (घटनास्थल):
(a) Direction and Distance from P.S (थाना से दूरी और दिशा): SOUTH , Beat No(बीट सं.) : 01
(b) Address(पता): ,INA METRO STATION
(c) In case, Outside the limit of the Police Station (यदि थाना सीमा के बाहर हैं):
Name of P.S(थाना का नाम): District(जिला):

6. Complainant / Informant (शिकायतकर्ता/सूचनाकर्ता):
(a) Name(नाम):
(b) Date/Year of Birth (जन्म तिथि /वर्ष): Nationality (राष्ट्रीयता): INDIA
(c) Passport No.(पासपोर्ट सं.): Date of Issue (जारी करने की तिथि): Place of Issue (जारी करने का स्थान):
(d) Occupation (व्यवसाय):
(e) Address(पता):

7. Details of Known/Suspect/Unknown accused with full particulars(attach separate sheet if necessary)(ज्ञात/ संदिग्ध /अज्ञात अभियुक्त का का पूरे विवरण सहित वर्णन):

8. Reason for delay in reporting by the complainant/informant (शिकायतकर्ता / सूचनाकर्ता द्वारा रिपोर्ट देरी से दर्ज कराने के कारण):
NO DELAY

9. Particulars of the properties stolen/involved (attach separate sheet if necessary):

Sl.No. (क्र.सं.)	Property Type(Description)	Est. Value(Rs.)(मूल्य (रु में))
1	ELECTRICAL AND ELECTRONIC GOODS()	11999

District : METRO P.S: INA METRO Year: 2018 FIR No: 0001 Date: 12/01/2018

10. Total value of property stolen (चोरी हुई सम्पति का कुल मूल्य): 11999

11. Inquest Report / U.D. Case No., if any (मृत्यु समीक्षा रिपोर्ट / यू.डी.प्रकरण न., यदि कोई हो):

12. F.I.R. Contents (attach separate sheet, if required)(प्रथम सूचना रिपोर्ट तथ्य):
To the SHO INA metro police station New Delhi subject- theft of mobile phone (M 1 Note 3) I boarded the metro at Noida city centre for Huda city centre. So I had to change the metro at Rajiv Chowk and as there was so much of rush so I suspect that my mobile is stolen there only.And I got to know about it after reaching INA metro – station.Ph no.
SD English
(Brother)पुलिस कार्यवाही थाना न० 2
ने बाहाजिर थाना आकर एक लिखित शिकायत बाशब्द English बाबत चोरी Mobile Phone मन DO को दर पेश की जिस पर FIR NO 1/18,U/S 379 IPC दर्ज रोज० की गयी FIR COPY व असल तहरीर बाहुक्म जनाब SHO साहब ASI G.
के हवाले की गयी। जो मुनासिब कार्यवाही अमल में लायेंगी दीगर नकुलात FIR बजरिये डाक
सम्बन्धित MM साहब व सीनीयर अफसरान को भिजवाई जायेगी बकलम HC/DO

13. Action Taken Since the above information reveals commission of offence(s) u/s as mentioned at Item No. 2:
(की गयी कार्यवाही: चूंकि उपरोक्त जानकारी से पता चलता है कि किया गया अपराध मद सं.2.में उल्लेख धारा के तहत है):
(i) Registered the case and took up the investigation: OR (या)
(प्रकरण दर्ज किया गया और जांच के लिए लिया गया):
(ii) Directed (Name of the I.O.)(जांच अधिकारी का नाम): Rank (पद):
ASST. SI (ASSISTANT SUB-INSPECTOR)
No(सं.): to take up the investigation (को जांच आपने पास में लेने के लिए निर्देश दिया गया) OR(या)
(iii) Refused investigation due to(जांच के लिए): OR (के कारण इंकार किया या)
(iv) Transferred to P.S(name)(थाना): District(जिला):
on point of jurisdiction (को क्षेत्राधिकार के कारण हस्तांतरित)

F.I.R read over to the complainant/informant,admitted to be correctly recorded and a copy given to the complainant/informant, free of cost : (शिकायतकर्ता / सूचनाकर्ता को प्राथमिकी पढ़ कर सुनाई गयी, सही दर्ज हुई माना और एक कॉपी निशुल्क शिकायतकर्ता को दी गयी) :

R.O.A.C.(आर.ओ.ए.सी.):

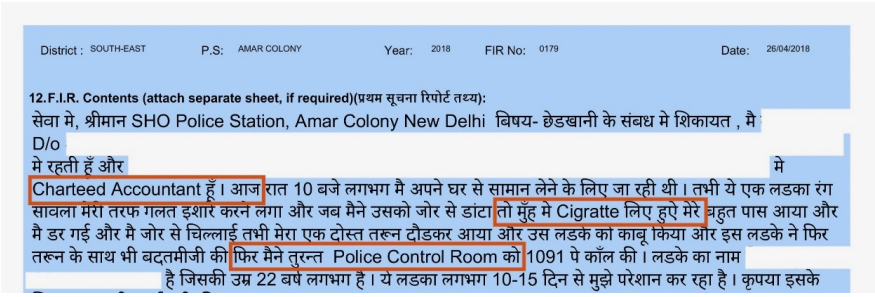
14. Signature / Thumb Impression of the Complainant / Informant:
(शिकायतकर्ता / सूचनाकर्ता के हस्ताक्षर / अंगूठे का निशान):
Signature of Officer
Name(नाम):
Rank (पद): HC (HEAD CONSTABLE)
No.(सं.):

15. Date and Time of despatch to the court:
(अदालत में प्रेषण की दिनांक और समय):

Notes: A representative two-page First Information Report as published on the Delhi Police website, in its original scanned, bilingual form.

BeautifulSoup over Tesseract's hOCR output); these numbered anchors pin down the position of each section on the page even when scans are shifted or span multiple pages. I then define bounding boxes relative to those anchors (Figure B.3) and extract the text within each box, parsing it into tabular form (Figure B.4).

Figure B.2: Scanned, Bilingual Nature of the Reports



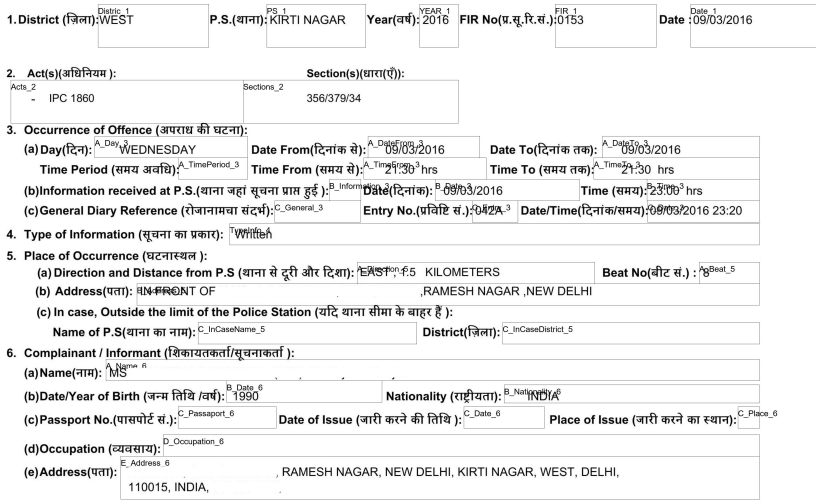
District : SOUTH-EAST P.S: AMAR COLONY Year: 2018 FIR No: 0179 Date: 26/04/2018

12. F.I.R. Contents (attach separate sheet, if required)(प्रथम सूचना रिपोर्ट तथ्य):
सेवा मे, श्रीमान SHO Police Station, Amar Colony New Delhi बिषय- छेडखानी के संबध मे शिकायत , मै
D/o
मे रहती हूँ और मे
Charteed Accountant हूँ । आज रात 10 बजे लगभग मै अपने घर से सामान लेने के लिए जा रही थी । तभी ये एक लडका रंग सावला मेरी तरफ गलत इशारे करने लगा और जब मैने उसको जोर से डांटा तो मुँह मे Cigratte लिए हुऐ मेरे बहुत पास आया और मै डर गई और मै जोर से चिल्लाई तभी मेरा एक दोस्त तरून दौडकर आया और उस लडके को काबू किया और इस लडके ने फिर तरून के साथ भी बदतमीजी की फिर मैने तुरन्त Police Control Room को 1091 पे कॉल की । लडके का नाम
है जिसकी उम्र 22 बर्ष लगभग है । ये लडका लगभग 10-15 दिन से मुझे परेशान कर रहा है । कृपया इसके

Notes: The reports are scanned images rather than machine-readable text, with fields recorded in both English and Hindi.

Two features of the extraction limit measurement error at the source. First, the numeric fields on which the analysis depends—the FIR number, the incident and registration dates, and the statutory (IPC) section numbers—are recognized under a digit-only character constraint, so that recognition cannot introduce stray letters into a date or a section code; free-text

Figure B.3: Bounding Boxes around the Extracted Fields



Notes: Bounding boxes are drawn around each field of the scanned form to isolate the relevant text prior to optical character recognition. The boxes are anchored to the form's numbered section headers, so that they track each field even when a scan is shifted or runs onto a second page.

Figure B.4: Data Parsed into Tabular Form

	A	B	C	D	E	F	G	H
1			1			R	F	G
2		District	P.S.	Year	FIR No	Date	Act(s)	Section(s)
3								
4	763	METRO	161 AIRPORT METRO	2019	11	24/09/201	IPC 1860 N	379/511/3
5	891	METRO	Metro Police Station Azadpur	2019	6	18/10/201	- IPC 1860	379
6	863	METRO	KASHMIRI GATE METRO	2019	78	31/10/201	- IPC 1860	379/511
7	907	METRO	Metro Police Station Ghitorni	2019	11	23/08/201	- IPC 1860	379/511
8	967	METRO	Metro Police Station Pragati Ram? al a tam	2019	7	13/04/201	- IPC 1860	379
9	103	METRO	RAJA GARDEN METRO	2019	2	15/01/201	- IPC 1860	323/341
10	866	METRO	KASHMIRI GATE METRO	2019	81	11/11/201	- IPC 1860	379/511
11	836	METRO	KASHMIRI GATE METRO	2019	51	15/07/201	- IPC 1860	379/411
12	978	METRO	Metro Police Station Pragati Ram? al a tam	2019	18	19/11/201	IPC 1860 T	509/354(A
13	777	METRO	JANAK PURI METRO	2019	3	15/03/201	- IPC 1860	354/354(D
14	107	METRO	SHASHTRI PARK METRO	2019	21	24/09/201	- IPC 1860	365
15	840	METRO	KASHMIRI GATE METRO	2019	55	30/07/201	- IPC 1860	379/511
16	794	METRO	KASHMIRI GATE METRO	2019	9	12/02/201	- IPC 1860	379
17	106	METRO	SHASHTRI PARK METRO	2019	9	10/03/201	- IPC 1860	384
18	111	METRO	YAMUNA DEPOT METRO	2019	27	06/12/201	- IPC 1860	379/411
19	818	METRO	KASHMIRI GATE METRO	2019	33	08/05/201	- IPC 1860	379
20	103	METRO	RAJA GARDEN METRO	2019	4	17/01/201	- IPC 1860	379
21	808	METRO	KASHMIRI GATE METRO	2019	23	31/03/201	- IPC 1860	379/411

Notes: The text extracted from each bounding box, parsed into structured, machine-readable records.

fields are similarly cleaned of spurious characters. Second, the fields that define the unit of analysis and the timing of treatment are printed, structured entries in fixed positions on the form—in particular, a report's jurisdiction is taken from the printed police-station field rather than from a free-text address—and are therefore recognized far more reliably than handwritten narrative text. Fields that run across multiple pages are concatenated, and reports whose layout cannot be parsed are flagged and set aside rather than silently mis-recorded, as detailed in Section B.2.

The parsed records are then cleaned, geocoded to the location of each reported crime (Section B.5), and classified into crime categories using the Indian Penal Code sections recorded on each report (Table B.3). This yields the crime-level panel of roughly 300,000 FIRs used throughout the paper.

B.2 Raw Corpus and Sample Construction

Table B.2 accounts for the flow from the raw corpus to the estimation samples. Of the 357,802 FIR PDFs downloaded, 356,140 (99.54%) were successfully processed by the OCR pipeline; the small remainder failed layout parsing—typically the lowest-quality scans—and were logged and set aside rather than silently mis-recorded. Restricting to reports that carry a valid incident date and an identifiable Indian Penal Code (IPC 1860) offense leaves 296,869 FIRs.[13] This set of roughly 300,000 substantive crime reports is the basis for both research designs.

The two designs then impose different requirements. The jurisdiction-level design assigns each report to a territorial police jurisdiction using the printed police-station field, and retains the 286,677 reports (96.57%) filed in territorial units, excluding those recorded by police units that do not correspond to a geographic jurisdiction (such as railway police, the crime branch, and other specialized units). The ring-based spatial design instead requires the incident location to be geocoded from the recorded address; 261,493 of the IPC 1860 reports (88.08%) geocoded successfully—87.84% across all OCR-processed reports.

B.3 Police Jurisdictions: Boundary Harmonization and Sample Construction

The unit of analysis in the jurisdiction-level design is the police jurisdiction. Two features of Delhi's policing geography require attention before the reports can be organized into a

[13] In reports listing multiple acts, the OCR occasionally returned garbled act names (for example, `"IPC 1860 TL DIONATIEAATINN MAAC"`); I manually corrected these to recover the IPC classification and retain the report. After correction, only 2 of the 296,869 reports have a missing section.

Table B.2: FIR corpus and sample construction

	FIRs	Share
FIR PDFs downloaded	357,802	—
Successfully processed by OCR	356,140	99.54%
Valid date and identifiable IPC 1860 offense	296,869	83.36%
Jurisdiction-level sample		
Filed in a territorial police jurisdiction	286,677	96.57%
Ring (spatial) sample		
Successfully geocoded	261,493	88.08%

Notes: Shares are computed relative to the natural base at each step: 99.54% of downloaded PDFs; 83.36% of OCR-processed reports; and, for the two design samples, 96.57% and 88.08% of the 296,869 reports with a valid date and an IPC 1860 offense. Across all OCR-processed reports, 87.84% geocoded successfully. Reports failing OCR layout parsing were logged and excluded. Multi-act reports with garbled act names were manually corrected to recover the IPC classification.

consistent panel.

First, police-station boundaries were redrawn several times between 2016 and 2019, typically as new stations were carved out of existing ones. To hold the unit of observation fixed across the sample period, I combine any stations involved in such a boundary change into a single jurisdiction defined consistently throughout, yielding 147 harmonized territorial units. This step is a pure aggregation: it merges the affected stations' reports into one unit and retains every report, so it removes no crimes from the sample.

Second, not every police unit corresponds to a residential territory. I exclude non-territorial units—crime, special, economic-offences, and vigilance cells—which have no fixed jurisdiction or resident population and so cannot anchor a per-capita crime rate. I also exclude transit police stations (railway and metro), which police the transport network rather than a residential catchment; crimes occurring on the metro are handled by dedicated metro police stations and are examined separately in the displacement analysis of Section 3.2. Crimes filed at these excluded units are dropped from the analysis panel and are not reassigned to a territorial jurisdiction.

Because the combination step retains every report, the excluded units are the only source of any shortfall in crime coverage. The 147 territorial jurisdictions that remain account for 97% of all reported crimes and 99% of reported harassment crimes against women in the corpus. The construction thus removes only a small share of the crime universe while ensuring that every unit in the panel has a stable boundary and a well-defined resident population.

B.4 Crime Categories

I classify each FIR into crime categories from the Indian Penal Code (IPC 1860) sections recorded on the report; Table B.3 lists the full mapping. The outcome of primary interest, reported harassment crimes against women, comprises offenses against women in public space—obscene acts and words in public, assault or criminal force with intent to outrage a woman's modesty, sexual harassment, stalking, and voyeurism (IPC 294, 354, 354A–354D, and 509). The comparison categories used in the paper—all crimes against women, domestic violence, and non-gender-specific street crimes such as theft and robbery—are defined analogously from their respective sections. Defining categories from legal sections rather than free-text descriptions keeps the classification transparent and replicable.

Table B.3: Crime Categories and Corresponding Indian Penal Code (1860) Sections

Category	IPC 1860 sections
Reported harassment crimes against women	294, 354, 354A, 354B, 354C, 354D, 509
All crimes against women	294, 304B, 305, 306, 312, 313, 314, 354, 354A, 354B, 354C, 354D, 366, 366A, 366B, 372, 373, 375, 376, 376A, 376AB, 376B, 376C, 376D, 376DA, 376DB, 376E, 376F, 493, 496, 497, 498, 498A, 509, 376/511
Domestic violence	304B, 305, 306, 312, 313, 314, 376B, 493, 496, 498A
Non-gender-specific street crimes	283, 319, 320, 321, 322, 323, 324, 325, 326A, 326B, 334, 335, 336, 337, 338, 349, 350, 351, 352, 355, 356, 378, 379, 425, 426, 427, 431, 432, 434, 510
Theft	378, 379, 380, 381, 382

Notes: Crime categories are constructed from the IPC 1860 sections recorded on each First Information Report.

B.5 Geocoding

Addresses are geocoded from the recorded incident location using the ArcGIS geocoding service, with the bilingual (Hindi–English) address text first converted to UTF-8; the small share of addresses the service could not resolve were geocoded by hand. Each police station's reports are geocoded within that station's own jurisdiction shapefile, so every point falls inside the correct police jurisdiction and the jurisdiction-level estimates do not depend on geocoding precision. Within a jurisdiction, several reports can share an identical coordinate when they refer to the same street or landmark—a named road, a market, or a major metro interchange—described in slightly different words, and where an address is incomplete the

geocoder may return a street centroid. Because the ring design compares one-kilometre distance bands, this sub-kilometre imprecision does not affect a report's band assignment; and because any such error is spread diffusely across streets and localities throughout the city rather than concentrated around any single station, it adds noise to the spatial design rather than biasing the estimated effect at any one station.

Of the roughly 300,000 First Information Reports, 88% are geocoded to a point location; for reported harassment crimes against women the rate is higher, at 93%. Table B.4 shows that coverage is stable across the sample period—there is no deterioration around the years in which stations open—and similar across jurisdictions of different treatment status. This matters for interpretation in two ways. First, the jurisdiction-level design does not depend on geocoding at all: each FIR is assigned to the police jurisdiction of the station at which it is filed, so the jurisdiction-level harassment counts include every report, geocoded or not. Geocoding coverage therefore cannot bias the main results. Second, geocoding enters only the ring-based design, which requires the point location of each crime; there, coverage remains high—93% for reported harassment—and is lower only for a small number of peripheral stations, so incomplete geocoding is unlikely to distort the ring estimates.

Table B.4: Geocoding Coverage

	All crimes	Harassment
Panel A: Coverage by year		
2016	0.870	0.930
2017	0.882	0.943
2018	0.882	0.936
2019	0.879	0.913
Panel B: Coverage by treatment status		
Control (never/not-yet-treated)	0.904	0.953
Treated	0.892	0.931
Already-treated (Excluded from analysis)	0.903	0.923

Notes: Share of First Information Reports geocoded to a point location. Panel A reports coverage by calendar year for all crimes and for reported harassment crimes against women; Panel B reports coverage by jurisdiction treatment status. Coverage is high and stable over time and similar across treatment groups.

I construct station catchments as Voronoi polygons around all 57 stations used in the ring-based design, so that every location is assigned to its nearest station and no crime or resident is attributed to more than one station where 5-kilometer buffers would otherwise overlap. I then remove any area within one kilometer of a Metro station that opened before 2016, so that exposure to the pre-existing network is not attributed to a new station opening. Within each station's remaining catchment, population is allocated to the 0–1, 1–2, 2–3, 3–4,

and 4–5 kilometer distance bands using the same area-weighted procedure applied to police jurisdictions in Section 2.3: where a Census town or village straddles more than one band, its population is apportioned to each band in proportion to the share of its area falling within that band.

B.6 Time of Occurrence

Each First Information Report records the time at which the offense is reported to have occurred, which I use to assign incidents to time-of-day windows in the mechanism analysis (Section 4.1). The time is missing for 12.93% of FIRs. Crucially for interpretation, this coverage is balanced: as Table B.5 shows, it is high and stable across the sample period, with no deterioration around the years in which stations open, and similar across distance bands. Because the recording of the time field—including any default or imprecise entries—does not differ across treatment groups or over time, it differences out of the within-window comparisons and cannot bias the time-of-day estimates; it enters only as noise.

Reported times cluster heavily on round values—the hour and the half-hour account for the large majority of entries—consistent with approximate recall rather than precise recording. The analysis therefore groups incidents into four six-hour windows (03:00–09:00, 09:00–15:00, 15:00–21:00, and 21:00–03:00), a resolution coarse enough that this rounding does not affect a report's assignment to a window. I report the estimated effect for each of these six-hour windows in Appendix Table A.11. I additionally check narrower four-hour windows; results are reported in Appendix Table A.12.

Table B.5: Coverage of the Reported Time-of-Occurrence Field

Panel A: By calendar year		*Panel B: By distance band*	
Year	Harassment	Band	Harassment
2016	0.852	≤1 km	0.866
2017	0.873	1–2 km	0.876
2018	0.871		
2019	0.881		

Notes: Sample excludes all First Information Reports within 1 km of metro stations that opened before 2016. Share of First Information Reports with the time of occurrence of crime reported. Panel A reports harassment coverage by calendar year; Panel B reports harassment coverage by distance band. Coverage is high and stable over time and similar across distance bands.

C Additional Figures

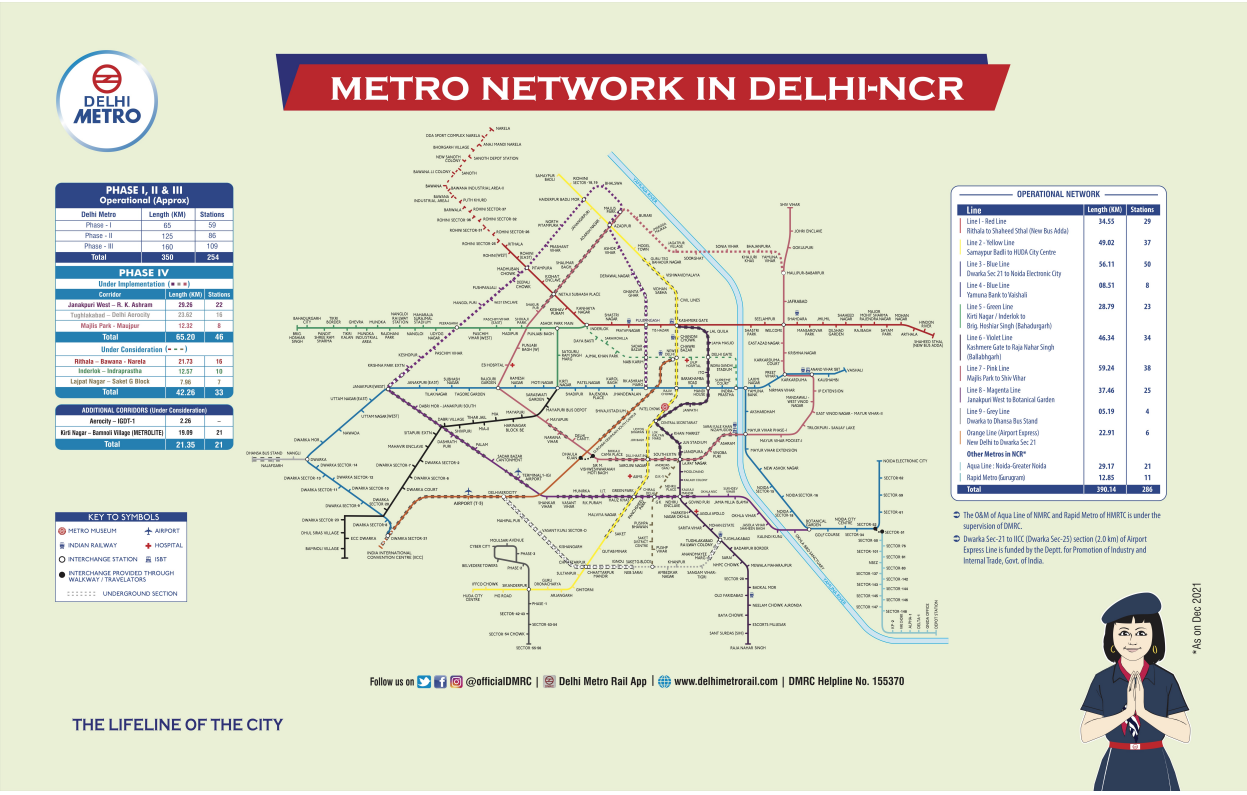


Figure C.1: Delhi Metro Network in 2021

Source: https://www.delhimetrorail.com/pages/en/network_map.

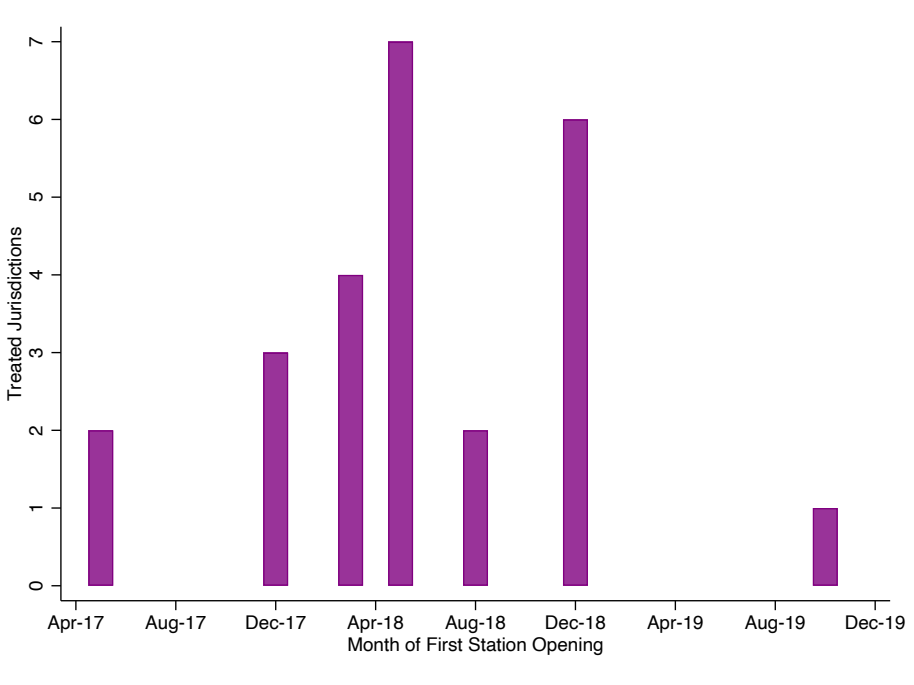


Figure C.2: Timing of first metro station openings across police jurisdictions

Notes: Number of police jurisdictions receiving their first metro station in each opening month, among the 25 treated jurisdictions. Openings are staggered across seven cohorts between 2017 and 2019.

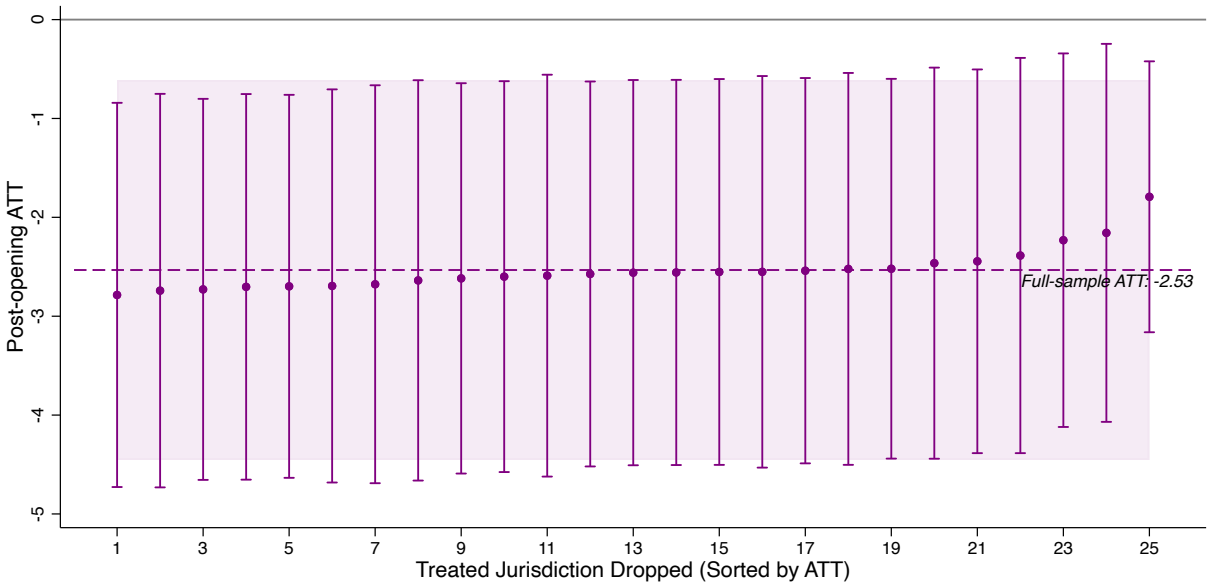


Figure C.3: Leave-one-out over treated jurisdictions

Notes: Each point is the average post-opening ATT re-estimated after dropping one of the 25 treated jurisdictions, sorted by magnitude, with 95% pointwise confidence intervals, standard errors clustered at the police-jurisdiction level. The dashed line and shaded band show the full-sample estimate and its 95% confidence interval. Every leave-one-out estimate is negative, significant, and within the full-sample interval.