NET WORK LISER

Discussion Paper Series

IZA DP No. 19028

October 2026

Do Criminally Accused Politicians Widen the Gender Gap in Science Choice? Evidence from India

Soham Sahoo
Indian Institute of Management Bangalore
and IZA@LISER

Sudipa Sarkar
National Law School of India University
and University of Warwick

The IZA Discussion Paper Series (ISSN: 2365-9793) ("Series") is the primary platform for disseminating research produced within the framework of the IZA@LISER Network, an unincorporated international network of labour economists coordinated by the Luxembourg Institute of Socio-Economic Research (LISER). The Series is operated by LISER, a Luxembourg public establishment (établissement public) registered with the Luxembourg Business Registers under number J57, with its registered office at 11, Porte des Sciences, 4366 Esch-sur-Alzette, Grand Duchy of Luxembourg.

Any opinions expressed in this Series are solely those of the author(s). LISER accepts no responsibility or liability for the content of the contributions published herein. LISER adheres to the European Code of Conduct for Research Integrity. Contributions published in this Series present preliminary work intended to foster academic debate. They may be revised, are not definitive, and should be cited accordingly. Copyright remains with the author(s) unless otherwise indicated.

Do Criminally Accused Politicians Widen the Gender Gap in Science Choice? Evidence from India*

Abstract

Gender gaps in educational participation can mask inequality in what women and men study, with consequences for later education and employment. We study whether criminal political representation contributes to such sorting at India's higher-secondary level, where science is the principal route to tertiary STEM education. Using nationally representative surveys, candidate affidavits, election results, and school-availability data, we instrument criminal political representation using close contests between candidates who differ in accusation status. We find that criminal political representation widens female disadvantage in science choice but has no discernible effect on the enrolment gender gap. A one-standard-deviation increase in the share of criminally accused politicians widens the science gap by 5–7 percentage points, with girls shifting toward arts/humanities. The adverse effect is larger where fewer schools offer science. Criminal political representation also increases perceived harassment of unmarried girls, while prior harassment is associated with a wider female disadvantage in science choice. Political selection can thus alter the composition of female human capital even when educational participation is unchanged.

JEL classification

I24, J16, D72

Keywords

criminally accused politicians, close elections, instrumental variables, safety perceptions, gender gap, STEM education, India

Corresponding author

Soham Sahoo
soham.sahoo@iimb.ac.in

* *Acknowledgments*: An earlier version of this paper was presented at the "Gender equality in STEM worldwide – innovative approaches to make STEM inclusive" workshop at Heidelberg University in October 2025 and at the Development Economics Workshop (DEW) at Durham University in June 2026, where we received constructive comments. We thank Nishith Prakash for helpful feedback and Nanditha Ajith for excellent research assistance. The authors acknowledge AI assistance in language editing, coding, and formatting using ChatGPT, Codex, and Claude. The authors reviewed all AI-assisted outputs and take full responsibility for the final content.

1 Introduction

Educational expansion changes economic opportunity through both the amount and the composition of human capital. Students who complete the same number of years of schooling can acquire very different skills, enter different occupations, and receive markedly different earnings depending on the fields they study. Evidence from a range of settings shows that choices made across academic fields during secondary school and college shape subsequent educational pathways, employment, and earnings (Card and Payne, 2021; Dahl et al., 2023; Sloane et al., 2021). These choices are also gendered: women remain underrepresented in science and other technical fields that provide access to many high-return occupations (Kahn and Ginther, 2017). As gender gaps in school enrolment narrow, the allocation of students across fields becomes an increasingly consequential source of inequality. Barriers that redirect girls away from specialised fields can therefore misallocate female talent well before women enter the labour market.

Safe mobility is an important input into access to specialised human capital. Science and technical education require particular teachers, laboratories, and curricular offerings and are often available in fewer institutions than general education. Pursuing these fields may therefore require students to travel farther or attend a school outside their immediate locality. The risks associated with public spaces make this access cost gender-specific. Women adjust their routes, destinations, and use of public space in response to harassment and perceived insecurity. Borker (2021), for example, shows that female students trade college quality for safer travel, while Beaman et al. (2026) show that fear of harassment is an important consideration in women's decisions about whether and where to work. Amaral et al. (2025) further demonstrate experimentally that visible policing reduces severe street harassment and women's withdrawal from public locations. When families perceive the surrounding environment to be unsafe, they may continue to educate their daughters while directing them towards fields that can be pursued closer to home. Aggregate enrolment can consequently improve even as girls remain excluded from specialised and economically valuable forms of education.

India provides a particularly revealing setting for studying this form of educational (mis)allocation. Students complete ten years of general education before selecting a stream at the higher-secondary level, typically among science, commerce, humanities, and technical or vocational education. This choice, usually made at the age of 15 or 16, determines eligibility for many tertiary programmes. In particular, the science stream provides the principal route into university-level science, technology, engineering, and mathematics (STEM) education. Girls remain substantially less likely than boys to study science, even after accounting for prior academic performance and household background (Sahoo and Klasen, 2021; Kumar and Sahoo, 2024). The survey data used in this paper display the same descriptive pattern (Figure 1). Stream choice is economically consequential. Evi-

dence from India links science-related education to higher earnings and more favourable employment outcomes (Roychowdhury, 2021; Jain et al., 2022). Recent evidence from India further shows that women who study science or commerce are more likely to participate in the labour market, obtain salaried employment, and earn higher wages than comparable women who study humanities (Sahoo and Klasen, 2026). Higher-secondary stream choice is thus a critical gateway through which early educational constraints can translate into persistent gender inequality in occupations and earnings.

Local safety is partly shaped by the political environment. We study this political source of educational constraints through the election of criminally accused politicians in India. Candidates facing criminal accusations are legally permitted to contest elections and hold public office unless disqualified following a conviction, and a substantial share of elected state legislators face criminal charges, including serious accusations involving murder, kidnapping, rape, and assault (Vaishnav, 2017). State legislators can influence policing, bureaucratic incentives, and the implementation of public services within the areas they represent. Previous research shows that criminally accused politicians reduce economic activity and household welfare and impede socioeconomic development (Chemin, 2012; Cheng and Urpelainen, 2019; Prakash et al., 2019). Recent evidence further shows that politicians accused of serious crimes worsen local crime environments, including crimes against women, and reduce female labour-force participation (Prakash et al., 2026). These findings raise an earlier question with potentially more persistent consequences: does criminal political representation affect the type of human capital that girls acquire before entering the labour market?

This paper asks whether the election of criminally accused politicians shifts girls away from the science stream at the higher-secondary level. The proposed safety-and-access mechanism has two empirical implications. First, the effect should be particularly pronounced where science is offered by relatively few schools, because access to the stream is then more likely to require mobility beyond the immediate locality, making safety concerns more salient in stream choice. Second, criminal political representation should worsen households' perceptions of the safety environment faced by girls, and such perceptions should be associated with a wider gender gap in science choice. The analysis therefore connects political selection, local safety, and access to specialised schooling in shaping girls' educational choices.

We combine three rounds of nationally representative education data from the National Sample Survey with information on candidates' criminal accusations and state legislative election outcomes. The main analysis links individual educational outcomes to the district-level proportion of elected Members of the Legislative Assembly who face criminal accusations. The election of accused politicians is endogenous because the places that elect them may differ systematically in institutional quality, social norms, economic conditions, and underlying crime. We address this concern using an instrumental vari-

ables strategy based on close electoral contests between candidates who differ in criminal-accusation status. Specifically, we instrument the district-level share of accused legislators with the share of constituencies in which the relevant accused candidate narrowly defeats the comparison candidate. Close contests generate plausibly exogenous variation in which candidate type wins. We conduct the analysis separately for any criminal accusation and for serious accusations. We then combine the education and election data with administrative measures of the local availability of schools offering science. We also use the India Human Development Survey (IHDS) to examine whether the distinction between science-stream choice and educational continuation is corroborated in a separate dataset and to study perceived safety as a potential mechanism. In the safety analysis, we first estimate the effect of criminal political representation on households' perceived harassment of unmarried girls and then examine whether perceptions of harassment measured earlier are associated with subsequent gender differences in science-stream choice.

We find that criminal political representation reallocates girls within higher-secondary education. It has little discernible effect on the female–male gap in overall higher-secondary enrolment but substantially widens the gender gap in science-stream participation, with similar effects under the any- and serious-accusation definitions. A one-standard-deviation increase in criminal political representation widens the female–male gap in science choice by roughly 5–7 percentage points. These effects are sizeable relative to the raw gender gap of roughly ten percentage points observed in recent education surveys. Importantly, the decline in girls' relative science participation is almost entirely offset by an increase in arts/humanities, with no discernible change in commerce. The contrast between the enrolment and stream-choice results therefore locates the adjustment within education: girls remain enrolled but are redirected away from the science pathway. The same qualitative distinction also appears in the IHDS, where criminal political representation widens the female disadvantage in science-stream choice under both definitions, with a more precise estimate for serious accusations, while having no discernible effect on the gender gap in post-secondary enrolment. This compositional response is consequential because choosing another stream closes access to many tertiary STEM programmes and can redirect subsequent occupational trajectories, yet it would be missed by measures that track only enrolment or years of schooling. The main pattern is robust across several tests, including alternative specifications, sample definitions, and a placebo test using subsequent elections.

Consistent with the role of mobility and access, the adverse gender-differentiated effect of criminal political representation is larger in districts where science is offered by relatively few schools and attenuates as local availability expands. This access gradient is most consistently evident for politicians facing serious criminal accusations, although we also find evidence of it for the broader measure of criminality when availability is measured among secondary schools. The pattern suggests that political insecurity becomes more

consequential for girls' educational choices when access to science depends on a more limited set of institutions. Political conditions and the local availability of specialised schooling therefore jointly shape girls' entry into the science pathway.

The mechanism evidence provides a more direct link between criminal political representation and the safety constraints highlighted above. Using close-election variation, we find that a higher share of criminally accused legislators increases households' perceptions that unmarried girls face harassment in their village or neighbourhood. We then exploit the panel structure of the IHDS to ask whether local safety conditions measured earlier are related to subsequent educational choices. Areas with greater prior perceived harassment exhibit a larger subsequent female disadvantage in science-stream participation, together with a relative shift toward arts/humanities. This relationship does not appear to simply reflect a more general link between local insecurity and the gender gap in science choice: the association with harassment remains essentially unchanged after allowing other local crime perceptions to have different relationships with science choice for girls and boys. Because the safety measure is observed before the educational outcome, the timing reduces concerns that households' reports are responding to the stream choices we study. The pattern is also consistent with earlier evidence showing that politicians accused of serious crimes increase reported crimes against women (Prakash et al., 2026). Taken together, the results connect political selection to the local safety environment and, in turn, to the gendered allocation of students across educational fields.

These findings advance the literature on safety and women's economic choices. Existing research shows that women alter college choice, transportation, mobility, and participation in public spaces in response to harassment and perceived insecurity (Borker, 2021; Aguilar et al., 2021; Amaral et al., 2025; Kondylis et al., 2025; Beaman et al., 2026). Related evidence from South Asia shows that local crime and violence affect children's learning, the timing of marriage, women's participation in employment, and household preferences over daughters and sons (Chakraborty et al., 2018; Siddique, 2022; Sarkar, 2024; Sharma and Sunder, 2025; Fakir et al., 2026; Mukherjee et al., 2026). We add two elements to this literature. First, we show that safety constraints operate at an earlier and highly consequential educational margin: the field that girls study even when they remain enrolled. Second, we identify criminal political representation as an upstream political determinant of the perceived safety environment and show that its consequences for educational choice are amplified when specialised schooling is less locally available. The results therefore connect the literature on women's responses to insecurity with the political and institutional forces that can generate those constraints in the first place.

The paper also extends research on political selection and development. Studies of criminal politicians have documented effects on economic growth, household welfare, infrastructure, crime, and labour supply (Chemin, 2012; Cheng and Urpelainen, 2019; Prakash et al., 2019, 2026). A related political-economy literature shows that politicians'

characteristics and partisan connections affect educational resources and the establishment of educational institutions (Lahoti and Sahoo, 2020; Kukreja, 2024; Kalliyil and Sahoo, 2026). We add to these literatures by showing that political selection also affects how students are allocated across fields of study, with gender-differentiated consequences for entry into science. This identifies educational specialisation as a distinct margin through which political institutions can shape human-capital formation.

Finally, the analysis contributes to research on gender segregation across fields of study. Existing work highlights expected returns and preferences as determinants of field-of-study choice (Wiswall and Zafar, 2015; Dasgupta and Sharma, 2022), while parental influence and social identity provide additional explanations for gendered sorting across fields (Carlana and Corno, 2024; Kumar and Sahoo, 2024). Evidence from secondary education also points to the role of academic preparation and access to specialised programmes (Sahoo and Klasen, 2021; Ngo and Dustan, 2024). Recent evidence from India further shows that affirmative action expanding women's access to elite STEM colleges increases girls' likelihood of choosing the science stream at the higher-secondary level (Gupta, 2026). We add a political-economy dimension to this literature by showing that local political conditions can shape gendered sorting across fields of study, and that the effect of political insecurity depends on the local supply of specialised schooling. This highlights an interaction between institutional environment and educational access that has received relatively little attention in explanations of gender segregation across fields.

The findings also point to a broader complementarity between educational access and public safety. Expanding the local availability of science education may reduce the mobility required to pursue the stream, while improvements in women's safety may broaden the set of institutions and fields that girls are able to consider. Because these constraints can affect what girls study without necessarily reducing their educational participation, policies focused only on overall enrolment may leave an important dimension of educational inequality unchanged.

The rest of the paper proceeds as follows. Section 2 describes the institutional setting, including higher-secondary stream choice and the role of state legislators and criminal accusations. Section 3 describes the data and construction of the key variables. Section 4 presents the empirical strategy. Section 5 assesses the validity of the close-election identification strategy. Section 6 reports the main findings, heterogeneity by access to science education, and evidence on mechanisms. Section 7 presents robustness checks. Section 8 concludes.

2 Institutional Background

This section describes two institutional features that are central to the analysis: the organisation of higher-secondary education and stream choice in India, and the role of

state legislators and criminal accusations in the political system.

2.1 Higher-secondary education and stream choice in India

School education in India broadly follows a 10+2 structure. Students complete ten years of largely general education through the secondary level (Class 10), followed, for those continuing in school, by two years of higher-secondary education (Classes 11–12).[1] Higher secondary is the first stage at which students specialise in a field of study. The principal streams are science, commerce, and arts/humanities, with a smaller proportion of students pursuing technical or vocational alternatives.

Stream choice at this stage substantially constrains subsequent educational options. Admission to many tertiary programmes depends on the subjects studied at the higher-secondary level. In particular, studying science at higher secondary provides the principal route to subsequent education in science, technology, engineering, mathematics, medicine, and related fields. Students specialising in commerce or arts/humanities generally do not satisfy the higher-secondary subject prerequisites for these programmes. Existing studies of science education in India therefore treat higher-secondary science choice as an important early educational specialisation with consequences for subsequent education and labour-market trajectories (Roychowdhury, 2021; Jain et al., 2022).

Access to different streams also depends on what schools offer. Schools that provide education through Class 10 vary in whether they extend to the higher-secondary level, and higher-secondary schools vary in the streams they offer. Science is particularly unevenly available across higher-secondary schools: administrative school-census data show substantial variation across districts in the share of higher-secondary schools offering science (Kumar and Sahoo, 2024). A student wishing to study science may therefore need to attend a different school after completing Class 10, even when higher-secondary education in another stream is locally available. Consistent with this, students studying science are more likely than those studying arts/humanities to attend an educational institution at least 5 km from their residence (Appendix Table A.2). This institutional distinction between access to higher-secondary schooling and access to science is central to the setting studied in this paper.

[1] In Indian usage, school years are commonly referred to as "Classes" rather than "Grades"; Class 10, for example, corresponds to Grade 10. The broad field of study chosen at the higher-secondary level is referred to as a "stream" in India and is closest to a high-school major or academic track. Similarly, "commerce" refers broadly to the business- and accounting-oriented field of study. We retain the Indian terms throughout the paper.

2.2 State legislators, criminal accusations, and the local political environment

India is a federal parliamentary democracy in which legislative authority is divided between the national Parliament and state legislatures. Each state has an elected Legislative Assembly composed of Members of the Legislative Assembly (MLAs). MLAs represent single-member assembly constituencies and are elected under a first-past-the-post system, with the candidate receiving the largest number of votes winning the seat. Legislative Assemblies normally have five-year terms unless dissolved earlier. Administrative districts generally encompass multiple assembly constituencies and are therefore associated with several elected MLAs.

State governments have substantial responsibility for local administration, including law enforcement. Under the Constitution of India, both "public order" and "police" fall within the State List.[2] MLAs do not formally command the police or the civil service, but elected politicians interact closely with the administrative machinery within their constituencies. Political influence over bureaucratic assignments, postings, and transfers is well documented in India (Iyer and Mani, 2012), providing a channel through which elected state politicians can influence the local bureaucratic and policing environment.

An important feature of Indian elections is the public disclosure of candidates' criminal cases. Following a 2003 Supreme Court ruling, candidates contesting elections are required to file sworn affidavits disclosing, among other characteristics, pending criminal cases and prior convictions.[3] A pending criminal case does not by itself disqualify a candidate from contesting an election. Disqualification under the Representation of the People Act, 1951 instead applies following conviction for specified offences and under other statutory conditions. Consequently, candidates facing pending criminal cases can contest elections and, if elected, serve as MLAs. Throughout the paper, "criminally accused" therefore refers to politicians facing disclosed criminal accusations and does not imply that they have been convicted.

Criminal accusations against politicians vary substantially in their nature. Some arise from activities such as political protests and public-order violations, while others involve offences such as murder, kidnapping, rape, assault, and criminal intimidation. We therefore distinguish between any criminal accusation and serious criminal accusations, following the classification in Prakash et al. (2026).[4]

[2]Constitution of India, Seventh Schedule, List II (State List), Entries 1 and 2.

[3]These disclosures are made through affidavits submitted with the nomination papers to the Election Commission of India.

[4]We classify an accusation as serious if it involves an offence under Sections 299–377 or 503–510 of the Indian Penal Code (IPC), corresponding to Chapters XVI and XXII. The IPC has since been replaced by the Bharatiya Nyaya Sanhita (BNS), 2023; we retain the IPC classification because the criminal cases

The ability of MLAs to influence the police and bureaucracy provides a potential channel through which criminal political representation may affect local safety. Politicians facing criminal charges may have stronger incentives to keep policing weak or manipulable, while ties to local criminal networks can weaken deterrence by shielding connected offenders from enforcement (Vaishnav, 2017). Consistent with this channel, Kim and Lee (2022) find that senior police officers serve shorter tenures in districts where politicians facing serious criminal charges narrowly win office. Such weakening of enforcement may be particularly consequential for harassment of women and girls. Harassment is often socially tolerated and its incidence can depend strongly on visible policing and officers' willingness to sanction perpetrators; experimental evidence from India shows that police presence can reduce severe street harassment through deterrence (Amaral et al., 2025). This makes harassment of women and girls an especially relevant dimension of the local safety environment for the educational choices studied here. Families may respond to such risks by restricting adolescent girls' mobility rather than withdrawing them from education altogether. Because science is less widely available across higher-secondary schools, such mobility restrictions can affect girls' stream choice even when overall educational participation is unchanged.

3 Data Description

We combine nationally representative household surveys with electoral and administrative data to study whether criminal political representation affects gender differences in science-stream choice at the higher-secondary level. The main analysis links repeated cross-sectional education data from the National Sample Survey (NSS) to district-level measures of political criminality constructed from candidate affidavits and state legislative election results. We further combine these data with administrative measures of the local availability of schools offering science to examine how access to specialised schooling conditions the effects of political criminality. Finally, we use the India Human Development Survey (IHDS) both to examine whether the main education results are corroborated in a separate dataset and to study the proposed safety mechanism using household perceptions of the harassment of unmarried girls.

3.1 Education outcomes: National Sample Survey data

Our main education outcomes come from three rounds of the National Sample Survey (NSS) education surveys conducted in 2007–08, 2014, and 2017–18. These are large,

in our data are recorded under the IPC.

nationally representative household surveys that collect information on individuals' demographic characteristics, current enrolment, level of education, and, for those enrolled at the higher-secondary level, their subject or stream of study. The surveys also record the distance from the student's residence to the educational institution attended, which we use descriptively to compare travel distances across higher-secondary streams.

As discussed in Section 2.1, higher-secondary science is the principal route to subsequent tertiary STEM education. We therefore use enrolment in the science stream as our main measure of entry into the STEM educational pathway. The main analysis is restricted to individuals aged 15–18, corresponding closely to the ages at which students enter and complete higher-secondary education. We examine two outcomes. First, higher-secondary enrolment is an indicator for whether an individual in the full age-15–18 sample is currently enrolled at the higher-secondary level. This estimation sample contains 77,859 individuals. Second, among individuals currently enrolled at the higher-secondary level, we define an indicator equal to one for enrolment in the science stream and zero for enrolment in the other academic streams. The science-stream estimation sample contains 23,657 students. Table 1 reports summary statistics separately for these two estimation samples.

3.2 Political criminality and election data

Information on criminal accusations against political candidates comes from sworn affidavits submitted to the Election Commission of India (ECI) and digitised by the Association for Democratic Reforms (ADR).[5] The affidavits report whether a candidate faces criminal accusations and provide information on the nature of the charges. We construct separate measures for any criminal accusation and serious criminal accusations using the classification described in Section 2.2.

We combine the affidavit data with official results from state legislative assembly elections. Constituency-level election outcomes are mapped to districts and aggregated to construct our main measure of political exposure: the share of elected Members of the Legislative Assembly (MLAs) in a district who face criminal accusations.[6] We construct this measure separately for any criminal accusation and for serious criminal accusations. Because state assembly elections occur asynchronously and MLAs normally serve five-year terms, we identify the legislators in office in each year and construct annual district-level measures of political representation. In the main specification, we use the measure for the year preceding the corresponding NSS survey round.

[5]We obtain the affidavit data from ADR's MyNeta portal. Election results are compiled from ECI statistical reports and the Trivedi Centre for Political Data's Lok Dhaba database.

[6]For the election years relevant to our analysis, the political data contain 7,496 constituency-election observations, corresponding to 1,065 district-election-year observations across 576 districts.

For the close-election design, we identify constituencies in which candidates with different criminality status contest against one another. Under the any-accusation definition, these are contests between an accused and a non-accused candidate; under the serious-accusation definition, they are contests between a seriously accused candidate and a candidate without a serious accusation. For each relevant contest, we record the vote-share margin between these candidates and whether the accused candidate wins. These constituency-level measures are used to construct the district-level close-election instruments described in Section 4.2. For the future-election placebo in Section 7, we extend the political series through 2023.

3.3 School availability and access to science education

To examine whether the effect of criminal political representation depends on local access to science education, we use administrative school census data to construct district-level measures of the availability of schools offering science at the higher-secondary level.[7] We consider two measures: the percentage of secondary schools that extend to the higher-secondary level and offer science there, and the percentage of higher-secondary schools that offer the science stream. These measures capture variation across districts in how widely science education is locally available.

The school-availability measures are aggregated at the district-year level and matched to the NSS data by district and survey round. In the main heterogeneity analysis, school availability is measured two years prior to the NSS round, while political exposure is measured one year prior to the survey.

3.4 India Human Development Survey: education and perceived safety

We use the two rounds of the India Human Development Survey (IHDS), conducted in 2004–05 and 2011–12, for two complementary purposes.[8] The survey has a panel structure, following households across rounds. We first use the second round to examine whether the main education results are corroborated in a separate dataset.[9] We then use information on households' perceptions of local crime and safety in both rounds to examine perceived safety as a potential mechanism.

[7] We use data from the District Information System for Education (DISE). DISE is an annual census of registered schools and contains information on schools' location, classes taught, streams offered, and other characteristics. We aggregate the school-level data to the district-year level and use measures of school availability lagged by two years relative to the NSS survey round.

[8] Further details about this survey can be found at: https://ihds.umd.edu/.

[9] Because the first IHDS round was conducted in 2004–05, comparable affidavit-based information on the criminal backgrounds of state legislators is not available for the preceding electoral cycle across states. Analyses using criminal political representation as the explanatory variable therefore rely on the second IHDS round.

For the education analysis, we restrict the second-round sample to individuals aged 15–18. Science-stream choice is constructed from the course-of-study information collected for current or recent students studying above Class 10. We classify these courses into four mutually exclusive categories: science, arts/humanities, commerce, and other. The science-stream estimation sample contains 4,440 individuals. We also construct an indicator for post-secondary enrolment, equal to one if an individual is currently enrolled and has completed at least ten years of schooling. This estimation sample contains 12,186 individuals. Table A.1 reports summary statistics for the IHDS estimation samples.

For the first safety exercise, we use households' second-round reports on whether unmarried girls are harassed in their village or urban neighbourhood. The corresponding analysis is conducted at the household level using household-head observations and contains 37,638 households. The survey also records perceptions of theft, breaking-in, and threat or attack, which we use as comparison outcomes to distinguish the harassment result from a broader shift in households' perceptions of local crime.

For the second safety exercise, we use perceptions reported in the first round to construct a measure of the prior local safety environment. Household responses on harassment of unmarried girls are aggregated to the village or urban-neighbourhood level using a leave-one-out procedure, so that a household's own report does not enter the locality measure assigned to it. We construct analogous leave-one-out locality measures for the three other crime perceptions. The panel structure allows these earlier locality-level perceptions to be linked to individuals observed in the second round, when science-stream choices are measured. The preferred district-fixed-effects sample contains 4,113 individuals. Because the perception measure is constructed from the first round, it predates the educational outcomes used in this analysis.

3.5 Merging and unit of analysis

For the main analysis, individual-level observations from the NSS education surveys are merged with district-level measures of political criminality and school availability using district identifiers and the corresponding survey period.[10] The resulting dataset remains at the individual level, while the main political and school-availability variables vary at the district-year level.

For the IHDS analyses, individual education records from the second round are linked to district-level political variables for the education analysis. Household-head observations from the second round are linked to the same district-level political variables for the perceived-harassment analysis. For the lagged-safety analysis, locality-level perceptions

[10]State and district identifiers are harmonised across survey rounds and data sources using a common concordance; where necessary, NSS districts (and states) are combined to match the political-data geography (e.g., Telangana is mapped to Andhra Pradesh to maintain comparability across rounds).

constructed from the first round are linked to individuals observed in the second round. The unit of observation is therefore the individual for the education analyses and the household for the household perception analysis.

4 Empirical Strategy

This section describes the empirical strategy. We first define the baseline relationship between criminal political representation and the gender gap in science-stream choice and discuss the endogeneity of political criminality. We then present the instrumental-variables strategy based on close electoral contests. Finally, we describe the specifications used to examine heterogeneity by local access to science education and to conduct the complementary IHDS analyses of educational outcomes and perceived safety.

4.1 Baseline specification

Let $Science_{idt}$ denote an indicator equal to one if individual i in district d and survey round t is enrolled in the science stream at the higher-secondary level, and zero if enrolled in another higher-secondary stream. Let $Female_i$ be an indicator equal to one for female students. Let $CrimShare_{d,t-1}$ denote the share of elected Members of the Legislative Assembly (MLAs) in district d who face criminal accusations, measured one year prior to the NSS survey round. We construct $CrimShare_{d,t-1}$ separately for any criminal accusation and for serious criminal accusations.

We consider the following baseline specification:

$$\begin{aligned} Science_{idt} = &\alpha + \beta_1 CrimShare_{d,t-1} + \beta_2 Female_i \\ &+ \beta_3 \left(Female_i \times CrimShare_{d,t-1}\right) + X'_{idt}\gamma + \delta_d + \lambda_t + \varepsilon_{idt}, \end{aligned} \tag{1}$$

where X_{idt} includes indicators for caste and religion, age fixed effects, years of schooling of the household head, and household size. District fixed effects δ_d absorb time-invariant differences across districts, while λ_t denotes NSS survey-round fixed effects.

The coefficient β_1 captures the relationship between criminal political representation and science-stream choice for male students, while $\beta_1 + \beta_3$ gives the corresponding relationship for female students. Our primary coefficient of interest is β_3, which captures the differential effect on girls relative to boys and hence the effect of criminal political representation on the gender gap in science-stream choice. We estimate the same specification for higher-secondary enrolment using the full sample of individuals aged 15–18, with the dependent variable replaced by an indicator for current higher-secondary enrolment.

Equation (1) defines the parameter of interest but does not by itself address the endogenous selection of criminally accused politicians. We discuss this identification problem and our close-election instrumental-variables (IV) strategy in the next subsection.

4.2 Identification strategy

The share of criminally accused politicians in a district is unlikely to be exogenous. Districts that elect such politicians may differ systematically in voter preferences, candidate characteristics, institutional quality, social norms, or underlying crime, all of which may also affect educational choices. We therefore exploit quasi-random variation generated by close electoral contests between candidates who differ in criminality status. When such an election is decided by a sufficiently small margin, whether the accused or comparison candidate wins is plausibly as good as random (Lee, 2008). Since the outcome of each close contest is quasi-random, aggregating these outcomes across constituencies provides plausibly exogenous variation in the district-level share of criminally accused MLAs. Related work in the literature has similarly used close contests between candidates of different types to generate quasi-random variation in district-level political representation (Clots-Figueras, 2012; Bhalotra et al., 2014; Lahoti and Sahoo, 2020; Prakash et al., 2026). In the baseline specification, we classify a relevant contest as close when the vote-share margin between the two candidates is less than 3 percentage points. This threshold follows related applications of the same aggregated close-election IV strategy (Nellis and Siddiqui, 2018; Lahoti and Sahoo, 2020; Bhalotra et al., 2021; Prakash et al., 2026).[11] We also examine the sensitivity of the results to alternative thresholds of 2, 4, and 5 percentage points.

Let $CrimCloseShare_{d,t-1}$ denote the share of constituencies in district d in which a criminally accused candidate narrowly defeats the relevant comparison candidate. Under the any-accusation definition, the relevant contests are between an accused and a non-accused candidate; under the serious-accusation definition, they are between a seriously accused candidate and a candidate without a serious accusation. We use $CrimCloseShare_{d,t-1}$ as an instrument for $CrimShare_{d,t-1}$. Correspondingly, $Female_i \times CrimCloseShare_{d,t-1}$ instruments $Female_i \times CrimShare_{d,t-1}$.

Although the outcome of a close election is plausibly quasi-random, the occurrence of such a contest need not be. We therefore control for the incidence of relevant close contests, as well as for the occurrence and vote margins of contests between candidates who differ in criminality status. Let $CloseShare_{d,t-1}$ denote the share of constituencies in the district with a close contest between candidates who differ in criminality status. Within each district-election year, contests between candidates who differ in criminality status are indexed in ascending order of the absolute vote-share margin; $I_{jd,t-1}$ indicates the presence of the jth such contest, and $M_{jd,t-1}$ denotes its signed vote-share margin.

[11] At this threshold, 360 of 1,065 district-election-year observations (33.8 percent) contain at least one relevant close contest under the any-accusation definition; the corresponding figure under the serious-accusation definition is 278 (26.1 percent). These district-election-year observations contain 517 and 367 close constituency-level contests, respectively.

The specification includes $I_{jd,t-1}$ and its interaction with a polynomial in the vote margin, $F(M_{jd,t-1})$. The main specification uses a linear function of the margin; quadratic and cubic functions are considered as robustness checks.

The second-stage equation is

$$\begin{aligned} Science_{idt} = \alpha &+ \beta_1 CrimShare_{d,t-1} + \beta_2 Female_i + \beta_3 \left(Female_i \times CrimShare_{d,t-1}\right) \\ &+ \gamma CloseShare_{d,t-1} + \sum_j \mu_j I_{jd,t-1} + \sum_j \pi_j \left[I_{jd,t-1} \times F(M_{jd,t-1})\right] \\ &+ X'_{idt}\gamma_1 + \delta_d + \lambda_t + \varepsilon_{idt}. \end{aligned} \tag{2}$$

Because both $CrimShare_{d,t-1}$ and its interaction with $Female_i$ are treated as endogenous, the 2SLS specification contains two first-stage equations:

$$\begin{aligned} CrimShare_{d,t-1} = \omega_1 &+ \lambda_{11} CrimCloseShare_{d,t-1} \\ &+ \lambda_{12}\big(Female_i \times CrimCloseShare_{d,t-1}\big) \\ &+ \kappa_1 Female_i + \tau_1 CloseShare_{d,t-1} + \sum_j \nu_{1j} I_{jd,t-1} \\ &+ \sum_j \sigma_{1j}\big[I_{jd,t-1} \times F(M_{jd,t-1})\big] + X'_{idt}\gamma_2 + \delta_d + \lambda_t + \eta_{idt}. \end{aligned} \tag{3}$$

and

$$\begin{aligned} Female_i \times CrimShare_{d,t-1} = \omega_2 &+ \lambda_{21} CrimCloseShare_{d,t-1} \\ &+ \lambda_{22}\big(Female_i \times CrimCloseShare_{d,t-1}\big) \\ &+ \kappa_2 Female_i + \tau_2 CloseShare_{d,t-1} + \sum_j \nu_{2j} I_{jd,t-1} \\ &+ \sum_j \sigma_{2j}\big[I_{jd,t-1} \times F(M_{jd,t-1})\big] \\ &+ X'_{idt}\gamma_3 + \delta_d + \lambda_t + \xi_{idt}. \end{aligned} \tag{4}$$

The identifying assumption is that, conditional on the incidence of relevant close contests, the vote-margin controls, the individual and household covariates, and the fixed effects, the outcomes of close elections affect educational choices only through their effect on criminal political representation. We assess the quasi-randomness of close-election outcomes using density, continuity, and balance tests in Section 5.

All main specifications include the controls defined in Equation (1), district and NSS survey-round fixed effects, and use NSS sampling weights. Standard errors are clustered at the district level, corresponding to the level at which political exposure varies. The same 2SLS specification is used for higher-secondary enrolment, with the dependent variable and estimation sample changed as described in Section 3.

4.3 Heterogeneity by access to science education

To examine whether the effect of criminal political representation depends on local access to science education, we estimate reduced-form specifications that interact close victories by accused candidates with the district-level availability of schools offering science.[12] As described in Section 3, we consider two measures of access: the percentage of secondary schools that extend to the higher-secondary level and offer science there, and the percentage of higher-secondary schools that offer the science stream. School availability is measured two years prior to the NSS round, while political exposure is measured one year prior to the survey.

Let $Access_{d,t-2}$ denote either measure of science availability. We estimate the following specification, which includes all constituent main effects and lower-order interactions:

$$\begin{aligned} Science_{idt} = \alpha &+ \phi_1 CrimCloseShare_{d,t-1} + \phi_2 Female_i + \phi_3 Access_{d,t-2} \\ &+ \phi_4 \left(Female_i \times CrimCloseShare_{d,t-1}\right) \\ &+ \phi_5 \left(CrimCloseShare_{d,t-1} \times Access_{d,t-2}\right) + \phi_6 \left(Female_i \times Access_{d,t-2}\right) \\ &+ \phi_7 \left(Female_i \times CrimCloseShare_{d,t-1} \times Access_{d,t-2}\right) \\ &+ X'_{idt}\gamma + \mathbf{C}'_{d,t-1}\eta + \delta_d + \lambda_t + \varepsilon_{idt}, \end{aligned} \tag{5}$$

where $\mathbf{C}_{d,t-1}$ contains the close-contest incidence and vote-margin controls used in the main 2SLS specification. The coefficient ϕ_7 captures how the effect of close victories by accused candidates on the female–male gap in science-stream choice varies with local access to science education. At a given level of school availability, A, the estimated effect of close victories by accused candidates on the female–male gap in science-stream choice is $\phi_4 + \phi_7 A$. We therefore use the estimated interaction terms to trace this gender-differentiated effect across the observed distribution of school availability. The regressions include the same individual and household controls, district and NSS survey-round fixed effects, sampling weights, and district-clustered standard errors as the main specification.

4.4 IHDS corroboration and mechanism analysis

We use the IHDS for one corroboration exercise and two complementary analyses of perceived safety. First, we examine whether the distinction between science-stream choice and educational continuation found in the NSS is also present in an independent dataset.

[12] A 2SLS specification with heterogeneous effects would require treating the interactions of criminal political representation with gender and school availability, including the triple interaction, as endogenous and instrumenting them with the corresponding interactions of the close-election instrument. We instead report the reduced form, which directly examines whether the gender-differentiated effect of accused-candidate victories in close elections varies with local science availability and avoids relying on potentially weak higher-order first stages.

Second, we examine whether criminal political representation affects households' perceptions of the harassment of unmarried girls. Third, we examine whether perceptions of harassment measured before the educational outcome are associated with subsequent gender differences in science-stream choice.

For the education analysis, we use individuals aged 15–18 from the second round of the IHDS and estimate a district-fixed-effects IV specification corresponding to the main gender-differentiated analysis. Because the district-level share of accused MLAs is absorbed by the district fixed effects, the endogenous regressor is $Female_i \times CrimShare_d$, which we instrument with $Female_i \times CrimCloseShare_d$. The regressions include the female indicator, age and indicators for caste and religion, and district fixed effects, with standard errors clustered at the district level. We estimate this specification separately for science-stream choice and post-secondary enrolment, using the samples described in Section 3.

We next examine whether criminal political representation affects households' perceptions of the safety environment faced by girls. Let $Harassment^{II}_{hdr}$ denote household h's perception of the harassment of unmarried girls in district d and region r, measured in the second round of the IHDS. We estimate

$$Harassment^{II}_{hdr} = \alpha + \rho_1 CrimShare_d + \mathbf{C}'_d\rho_2 + \mathbf{H}'_{hd}\rho_3 + \delta_r + \nu_{hdr}, \tag{6}$$

where $CrimShare_d$ is the district-level share of criminally accused MLAs, $\mathbf{C}_d$ contains the close-election controls described in Section 4.2, $\mathbf{H}_{hd}$ contains baseline household covariates, and δ_r denotes region fixed effects. We instrument $CrimShare_d$ with the corresponding district-level close-win accused share, $CrimCloseShare_d$, and estimate the specification separately for any and serious criminal accusations. The analysis uses household-head observations and standard errors clustered at the district level. We also estimate the same specification for household perceptions of theft, breaking-in, and threat or attack to examine whether the effect on perceived harassment is accompanied by a broader shift in households' perceptions of local crime.

The final exercise relates science-stream choice observed in the second round to local perceptions of harassment measured in the first round.[13] As described in Section 3, first-round household reports are aggregated at the village or urban-neighbourhood level using a leave-one-out procedure and linked to individuals observed in the second round. Let $Harassment^{I}_{id}$ denote the resulting prior locality-level measure assigned to individual i

[13]We do not instrument prior perceived harassment with criminal political representation. Comparable affidavit-based political data are not available for the electoral cycle preceding IHDS-I; moreover, such an approach would require the strong exclusion restriction that criminal political representation affects subsequent science choice only through perceived harassment.

in district d. We estimate

$$Science_{id} = \alpha + \theta_1 Harassment^I_{id} + \theta_2 Female_i + \theta_3 \left(Female_i \times Harassment^I_{id}\right) + X'_{id}\gamma + \delta_d + \varepsilon_{id}. \tag{7}$$

where X_{id} includes age, indicators for caste and religion, baseline household characteristics, other first-round locality-level crime perceptions, and baseline measures of norms and social capital. Our preferred specification includes district fixed effects δ_d, and standard errors are clustered at the primary sampling unit level. We also examine specifications without geographic fixed effects and with primary-sampling-unit fixed effects. The coefficient θ_3 captures whether the female–male gap in science-stream choice varies with prior perceived harassment. As an additional comparison, we allow an index of the other locality-level crime perceptions to have a gender-differentiated association with science choice.

The panel structure and temporal ordering in Equation (7) reduce concerns that reported perceptions of local safety respond to the subsequent educational choices being studied. We interpret the resulting association as evidence consistent with perceived safety as a mechanism. Taken together, the two safety exercises provide complementary evidence on the proposed channel.

5 Validity of the Identification Strategy

This section assesses the validity of the close-election identification strategy. We examine the strength of the instruments and whether close-election outcomes appear quasi-random with respect to observable constituency, candidate, and district characteristics.

5.1 Instrument relevance

Table 2 reports the first-stage estimates for the two endogenous variables in the main specification. Panel A uses the district-level share of accused MLAs as the dependent variable, while Panel B uses its interaction with the female indicator. In each case, the excluded instruments are the close-win accused share and its interaction with the female indicator.

The first-stage relationships are strong. In Panel A, the close-win accused share is a large and statistically significant predictor of the overall share of accused MLAs under both definitions of criminality and for both estimation samples. In Panel B, the interaction between the female indicator and the close-win accused share strongly predicts the corresponding endogenous interaction. The Sanderson–Windmeijer conditional first-stage F statistics (Sanderson and Windmeijer, 2016) range from 44.01 to 136.89 for the accused-share equation and from 75.97 to 122.98 for the interaction equation. The

Kleibergen–Paap F statistics (Kleibergen and Paap, 2006) for the system range from 17.92 to 55.72.[14] These diagnostics indicate that weak identification is not a concern in the baseline specifications.

Figure A.1 provides a graphical illustration of the underlying first-stage relationship using constituency-election observations, under the any- and serious-accusation definitions. The two panels plot the district-level share of accused MLAs against the vote margin between candidates who differ in criminality status. Consistent with the regression estimates, both panels show an upward discontinuity in the district-level share of accused legislators when the accused candidate narrowly wins rather than narrowly loses the election.

5.2 Quasi-randomness of close-election outcomes

The identifying strategy requires that, within sufficiently close contests between candidates who differ in criminality status, whether the accused candidate narrowly wins or loses is not systematically related to characteristics that independently affect educational outcomes. Following standard validity checks for close-election designs (Eggers et al., 2015), we assess this assumption using several complementary tests.

We first examine whether there is evidence of manipulation or sorting around the zero vote-margin threshold. Following McCrary (2008), Figure A.2 plots the density of the vote margin under the any- and serious-accusation definitions. The distributions are smooth around the threshold and provide no evidence of a discontinuous change in density when the accused candidate moves from narrowly losing to narrowly winning.[15]

We next examine whether selected predetermined characteristics change discontinuously at the threshold. Figures A.3 and A.4 present continuity checks under the any- and serious-accusation definitions, respectively. The panels cover electorate size, the number of voters, constituency reservation status, the proportion of female candidates, average candidate education, average log candidate wealth, and the proportions of candidates facing serious or any criminal accusations. The graphical evidence does not reveal systematic discontinuities in these characteristics around the zero vote-margin threshold.

Table A.3 complements the graphical evidence by comparing observable characteristics across close elections won and lost by accused candidates, separately for the serious- and any-accusation definitions. Panel A considers characteristics of the constituency and candidate pool that are predetermined with respect to the close-election outcome.

[14] Because the specification contains two endogenous regressors, we report the Sanderson–Windmeijer conditional F statistic for each first-stage equation as well as the Kleibergen–Paap statistic for the system.

[15] The McCrary density test yields p-values of 0.216 for the any-accusation definition and 0.415 for the serious-accusation definition.

These are generally similar across the two sides of the threshold, although a few isolated differences emerge. Panel B instead considers characteristics of the elected candidate. Because these characteristics are realized jointly with the election outcome, differences across the threshold do not by themselves indicate a failure of the close-election design; rather, they raise the possibility that criminality is bundled with other attributes of the elected politician. Accused winners are less likely to be female and differ in wealth and ruling-party affiliation.[16] As shown below, the predetermined observable characteristics are also jointly unable to predict whether the accused candidate narrowly wins a close contest. We nevertheless account directly for the observed imbalances in the robustness analysis by augmenting the main specification with the characteristics that differ significantly across close-election outcomes.

We also examine whether predetermined observable characteristics jointly predict whether the accused candidate wins a close election. Table A.4 regresses an indicator for an accused candidate winning a close contest on a broad set of constituency, candidate-pool, and district characteristics. Although a small number of individual coefficients are statistically significant, the covariates are jointly insignificant under both definitions of criminality: the joint F-test has a p-value of 0.210 for serious accusations and 0.578 for any accusation. Thus, the predetermined observable characteristics considered in the table do not jointly predict whether the accused candidate narrowly wins.

Finally, Table A.5 compares district-level characteristics across district-election years with relatively more versus fewer accused winners in close contests. These characteristics are also broadly similar across the two groups, with only isolated statistically significant differences. Taken together, the density, continuity, balance, and predictability tests provide no systematic evidence that the outcomes of close elections are selected on observable characteristics. These results support the use of close-election outcomes as a source of quasi-random variation in criminal political representation.

6 Findings

We begin by documenting the gender differences in educational specialisation at the higher-secondary stage. Figure 1 shows the raw distribution of higher-secondary students across streams by gender in the surveys used in our analysis. Across all three NSS rounds and in IHDS-II, females are less likely than males to study science and more likely to study

[16] Recent work on close-election designs based on politician characteristics emphasizes that such designs identify the effect of electing one type of candidate rather than the isolated effect of the defining characteristic holding all other candidate attributes fixed (Marshall, 2024; Bertoli and Hazlett, 2025). We therefore interpret our estimates as the effect of criminal political representation, rather than the isolated effect of criminal accusations abstracted from other characteristics associated with accused politicians.

arts/humanities. In the NSS, the female disadvantage in science ranges from roughly 7 to 10 percentage points across rounds, equivalent to about one-quarter of the average science participation rate. These descriptive patterns motivate our analysis of whether criminal political representation contributes to gendered sorting across fields of study and whether local safety concerns help explain this relationship.

6.1 Main results

Our main analysis shows that criminal political representation has a gender-differentiated effect on science-stream choice, while having little discernible effect on the gender gap in overall higher-secondary enrolment. In Table 3, the coefficient on the interaction between the female indicator and the share of accused MLAs is negative and statistically significant under both definitions of criminality. The estimated interaction is -0.261 for any criminal accusation and -0.272 for serious criminal accusations, indicating very similar effects across the two definitions. Thus, a 10-percentage-point increase in the share of accused MLAs widens the female–male gap in science choice by approximately 2.6–2.7 percentage points. Evaluated at a one-standard-deviation increase in the respective measures of political criminality, the implied widening of the gender gap is approximately 5–7 percentage points.[17] These magnitudes are substantial relative to existing gender differences in science-stream choice. To get a sense of the scale, the average female disadvantage in science choice across the three NSS rounds is about 9 percentage points (Figure 1), so the estimated 5–7 percentage-point widening is roughly 60–75 percent as large as this observed gap. By contrast, the corresponding interaction coefficients for higher-secondary enrolment in Table 3 are small and statistically insignificant. The evidence therefore points to an effect on what girls study relative to boys, rather than on whether they remain enrolled at the higher-secondary level.

The decline in girls' relative science participation is mirrored by a shift toward arts or humanities rather than commerce. As shown in Table 4, the negative gender-differentiated effect on science is accompanied by a positive and statistically significant effect of similar magnitude on arts/humanities under both definitions of criminality, while the effect on commerce is small and statistically insignificant. The near-offsetting effects on science and arts/humanities indicate that the response is primarily a reallocation across higher-secondary streams, concentrated along the science–arts/humanities margin.

The IHDS provides corroborating evidence for this distinction between educational specialisation and continuation in a separate dataset. Table 5 shows that the estimated

[17] These magnitudes combine the interaction coefficients in Table 3 with the standard deviations of the corresponding political-exposure measures reported in Table 1. Specifically, the implied effects are approximately $0.261 \times 0.254 = 0.066$ for any accusation and $0.272 \times 0.201 = 0.055$ for serious accusations.

female interaction for science-stream choice is negative under both definitions of criminality and statistically significant for serious accusations. By contrast, the corresponding interactions for post-secondary enrolment are small and statistically insignificant. Appendix Table A.6 extends the science result across the four mutually exclusive IHDS course-of-study categories. The negative interaction for science is accompanied by positive interactions for the other stream categories, although these estimates are imprecise. Taken together, the IHDS evidence reproduces the central qualitative pattern in a separate dataset: criminal political representation is associated with a widening gender gap in science choice without a corresponding widening of the gender gap in educational continuation.

Downstream economic implications. We next provide an illustrative calculation of the potential economic magnitude of the estimated reallocation away from science.[18] The total effects for females imply that a one-standard-deviation increase in criminal political representation reduces girls' probability of studying science by approximately 7.9 percentage points under the any-accusation definition and 15.1 percentage points under the serious-accusation definition. Combining these estimates with labour-market benchmarks from Sahoo and Klasen (2026) yields illustrative declines of approximately 0.75–1.43 percentage points in later female labour-force participation[19] and, conditional on employment, 0.83–1.58 percentage points in salaried employment. The participation effects correspond to about 2.8–5.3 percent of the mean female labour-force participation rate. The implied reduction in annual earnings is approximately 1.1–2.1 percent. These estimates also imply that the widening of the gender gap in science choice caused by criminal political representation could translate into women earning roughly 0.8–1.0 percent less relative to men through this channel.[20] Appendix Table A.7 reports the calculations. These calculations provide a sense of the economic magnitude of the estimated effects, rather than estimates of subsequent causal effects.

[18]The corroborating IHDS estimates also indicate a widening of the female disadvantage in science choice, with a more precise estimate under the serious-accusation definition. We nevertheless base the illustrative calculations on the NSS estimates, which constitute the main analysis of the paper.

[19]Prakash et al. (2026) independently estimate a coefficient of -0.144 for the effect of seriously accused political representation on female labour-force participation. Using the 0.201 standard deviation of the serious-accusation measure in our sample, their coefficient implies a decline of approximately 2.9 percentage points in female labour-force participation. Our 1.43 percentage-point stream-reallocation benchmark is therefore about one-half of this independently estimated effect. The comparison suggests that changes in educational specialisation could constitute a quantitatively important pathway linking criminal political representation to subsequent female economic participation.

[20]This calculation combines the estimated 5.5–6.6 percentage-point widening of the gender gap in science choice from a one-standard-deviation increase in criminal political representation with the 0.151 differential earnings benchmark reported by Sahoo and Klasen (2026).

6.2 Heterogeneity by access to science education

The heterogeneity exercise examines whether the gender-differentiated effect of close victories by accused candidates depends on local access to science education. Table 6 shows that the female–male gap in science choice associated with close victories by accused candidates becomes less negative as science availability increases. This pattern is present under both definitions of criminality when access is measured by the percentage of secondary schools that extend to the higher-secondary level and offer science there, and is stronger for serious criminal accusations. When access is instead measured by the percentage of higher-secondary schools offering science, the same attenuation is evident for serious accusations, while the corresponding pattern for any accusation is weaker. Figure 2 further shows that the statistically significant adverse effects are concentrated at the lower end of the science-availability distribution, while the estimates attenuate and become indistinguishable from zero as availability rises. Overall, the access gradient is therefore most consistently evident for serious criminal accusations.

This pattern is consistent with the proposed mechanism that safety constraints become more consequential when pursuing science requires access to a relatively limited set of schools. Related evidence from India shows that geographic access can have gender-differentiated educational consequences: Muralidharan and Prakash (2017) find that reducing the travel cost of attending secondary school increases girls' enrolment, with larger effects among those living farther from school. While their analysis concerns the enrolment margin, our results point to a related constraint operating at a later stage of educational choice – affecting what girls study rather than whether they remain enrolled.

6.3 Mechanisms

We next utilize the IHDS data to examine perceived safety as a potential channel through which criminal political representation affects girls' educational choices. Table 7 reports the first safety exercise. Instrumenting the district-level share of accused MLAs with the corresponding close-win accused share, we find that criminal political representation increases households' perceived harassment of unmarried girls. The estimated effect is positive under both definitions of criminality and statistically significant at the 5 percent level for any accusation and at the 10 percent level for serious accusations. Appendix Table A.8 applies the same specification to other perceptions of local crime. We find no corresponding effect on perceived theft, breaking-in, or threat or attack. The contrast suggests that the harassment result is not simply capturing a broad-based deterioration across households' perceptions of local crime, but is more closely related to the safety environment faced by unmarried girls.

The second exercise examines whether perceptions of harassment measured earlier in IHDS-I are related to subsequent gender differences in science-stream choice. Ta-

ble 8 shows that the female–male gap in science participation becomes more negative as prior perceived harassment rises. This pattern is present across all three specifications, including the ones with district and primary-sampling-unit (PSU) fixed effects. The association is also robust to allowing broader perceptions of local crime to have a gender-differentiated relationship with science choice. Specifically, including an index of perceptions of theft, breaking-in, and threat or attack and allowing its association with science choice to differ by gender leaves the estimated relationship between prior harassment and the female–male science gap similar in magnitude and statistically significant (Appendix Table A.9). This suggests that it is households' concerns about girls' safety rather than their perceptions of the broader local crime environment that are driving the gender gap in science choice. Additionally, Appendix Table A.10 shows where this difference appears in the composition of study: the larger female disadvantage in science is accompanied by a relative shift toward arts/humanities, while there is little corresponding change in the gender gaps in commerce or other streams.

This evidence also complements earlier administrative evidence on the consequences of criminal political representation. Using crime data from the National Crime Records Bureau (NCRB), Prakash et al. (2026) show that politicians accused of serious crimes increase reported crimes against women. The two sources capture different aspects of the local safety environment: the NCRB records reported gender-related crimes, while the IHDS measures households' perceptions of harassment of unmarried girls. The fact that both point toward a deterioration in the environment faced by women and girls strengthens the safety interpretation of our education results.

Finally, the timing of the second IHDS exercise provides an additional advantage. The harassment measure is constructed from the first round of the survey and therefore predates the science-stream choices observed in the second round. This reduces concerns that reported safety conditions respond to the educational choices being studied. Taken together, the two safety exercises provide complementary evidence on the proposed channel.

7 Robustness Analysis

We assess the robustness of the main results along several dimensions, including sensitivity to alternative controls, functional-form choices, close-election thresholds, estimation specifications, a future-election placebo, and alternative sample definitions.

Controls for observable imbalances. The main estimates are robust to controlling directly for the observable characteristics that are imbalanced across close-election outcomes. As discussed in Section 5, balance is not exact for every constituency and candidate characteristic. We therefore augment the baseline specification with the set of significantly imbalanced characteristics specific to each criminality definition. After

adding these controls, criminal political representation continues to widen the female–male gap in science-stream choice by essentially the same amount as in the baseline specification under both definitions of criminality (Table A.11). There remains no discernible effect on the gender gap in higher-secondary enrolment. These results indicate that the main findings are not driven by the observable imbalances documented in the close-election sample.

Alternative vote-margin controls and close-election thresholds. The main results are robust both to alternative functional forms for the vote-margin controls and to alternative definitions of a close election. Allowing the vote margin to enter quadratically or cubically leaves the estimated widening of the gender gap in science-stream choice very similar to that obtained with the baseline linear specification (Table A.12). The same pattern holds when the definition of a close election is varied from the baseline 3-percentage-point threshold to thresholds of 2, 4, and 5 percentage points (Table A.13): criminal political representation continues to widen the female–male gap in science choice, with no corresponding effect on the gender gap in higher-secondary enrolment.[21]

Alternative estimation specifications. The reduced-form estimates show that narrow victories by accused candidates widen the female–male gap in science-stream choice but have little effect on the corresponding enrolment gap (Table A.14). OLS estimates display the same qualitative pattern: greater criminal political representation is associated with a larger female disadvantage in science choice, but not in higher-secondary enrolment (Table A.15). The IV results are also stable when we absorb state-specific shocks across survey rounds using state-by-NSS-round fixed effects (Table A.16) and when the regressions are estimated without NSS sampling weights (Table A.17). Across these specifications, the central result remains that criminal political representation changes girls' relative participation in science without producing a comparable change in their relative participation in higher-secondary education.

Future-election placebo. We next conduct a temporal falsification exercise using political outcomes from elections that occur after the educational outcomes are observed. For each NSS round, we replace the baseline political exposure with the first state assembly election held after completion of the survey and reconstruct both the accused-share treatment and the close-election instruments using that future election. Since political representation determined in a subsequent election cannot affect educational choices observed earlier, we should not expect to find an effect in this placebo exercise. Table A.18 shows no statistically significant female interaction for science-stream choice or higher-secondary enrolment under either definition of criminality. The total effects for

[21]The 2-percentage-point serious-accusation science specification should be interpreted cautiously because the Kleibergen–Paap statistic falls to 5.79; the estimates at the wider 4- and 5-percentage-point thresholds do not face the same weak-identification concern.

females are likewise statistically insignificant in all four specifications. Thus, the placebo estimates provide additional support for the interpretation of the main results.

Alternative sample definitions. The main finding is also robust to alternative sample definitions. Expanding the sample from ages 15–18 to ages 15–20 continues to show a larger female disadvantage in science-stream choice under both definitions of criminality, while the gender gap in higher-secondary enrolment remains largely unaffected (Table A.19). We also address the possibility that the main science result arises from conditioning on students who have already reached higher-secondary education. To do so, we define an unconditional science outcome in the full age-15–18 sample, equal to one for individuals who are both enrolled at the higher-secondary level and studying science and zero otherwise. Criminal political representation also widens the female–male gap in this unconditional measure under both definitions of criminality (Table A.20). The main finding is therefore not driven solely by conditioning the stream-choice analysis on higher-secondary students.

8 Conclusion

This paper shows that criminal political representation affects the composition of girls' education at a critical stage of educational specialisation. Exploiting quasi-random variation from close electoral contests, we find that a higher share of criminally accused legislators substantially widens the female–male gap in science-stream choice, while having little discernible effect on the gender gap in overall higher-secondary enrolment. A one-standard-deviation increase in criminal political representation widens the science gender gap by roughly 5–7 percentage points. The decline in girls' relative science participation is accompanied by a corresponding increase in arts/humanities rather than commerce. The corroborating analysis yields a similar pattern, with a statistically significant widening of the science gender gap for serious accusations. Political criminality therefore changes what girls study even when it does not reduce their overall participation in education.

The results also point to an important interaction between political institutions, safety, and access to specialised schooling. The adverse gender-differentiated effect on science choice is larger where science is available in relatively few schools, with the access gradient most consistently evident for politicians facing serious criminal accusations. The evidence on perceived safety is also consistent with this channel. Criminal political representation increases households' perceived harassment of unmarried girls. Higher prior perceptions of harassment are in turn associated with a larger subsequent female disadvantage in science-stream choice and a relative shift toward arts/humanities. These findings are consistent with earlier research using administrative crime data showing that politicians accused of serious crimes increase reported crimes against women (Prakash et al., 2026). The

analyses based on administrative crime records and household survey data on perceived harassment capture different dimensions of the local safety environment, but both point toward a deterioration in the environment faced by women and girls under criminal political representation. Thus, improvements in overall educational enrolment can coexist with substantial constraints on the type of human capital girls acquire. Reducing gender inequality in high-return educational pathways may therefore require attention not only to the availability of specialised schooling, but also to the political and safety environment in which educational choices are made.

References

Aguilar, A., Gutiérrez, E., and Soto Villagrán, P. (2021). Benefits and unintended consequences of gender segregation in public transportation: Evidence from Mexico City's subway system. *Economic Development and Cultural Change*, 69(4):1379–1410.

Amaral, S., Borker, G., Fiala, N., Kumar, A., Prakash, N., and Sviatschi, M. M. (2025). Sexual harassment in public spaces and police patrols: Experimental evidence from urban India. *The Quarterly Journal of Economics*, 140(4):3191–3231.

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

Bertoli, A. and Hazlett, C. (2025). Seeing like a district: Understanding what close-election designs for leader characteristics can and cannot tell us. *Political Analysis*, 33(4):393–411.

Bhalotra, S., Clots-Figueras, I., Cassan, G., and Iyer, L. (2014). Religion, politician identity and development outcomes: Evidence from India. *Journal of Economic Behavior & Organization*, 104:4–17.

Bhalotra, S., Clots-Figueras, I., and Iyer, L. (2021). Religion and abortion: The role of politician identity. *Journal of Development Economics*, 153:102746.

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

Card, D. and Payne, A. A. (2021). High school choices and the gender gap in STEM. *Economic Inquiry*, 59(1):9–28.

Carlana, M. and Corno, L. (2024). Thinking about parents: Gender and field of study. *AEA Papers and Proceedings*, 114:254–258.

Chakraborty, T., Mukherjee, A., Rachapalli, S. R., and Saha, S. (2018). Stigma of sexual violence and women's decision to work. *World Development*, 103:226–238.

Chemin, M. (2012). Welfare effects of criminal politicians: A discontinuity-based approach. *The Journal of Law and Economics*, 55(3):667–690.

Cheng, C.-Y. and Urpelainen, J. (2019). Criminal politicians and socioeconomic development: Evidence from rural India. *Studies in Comparative International Development*, 54(4):501–527.

Clots-Figueras, I. (2012). Are female leaders good for education? Evidence from India. *American Economic Journal: Applied Economics*, 4(1):212–244.

Dahl, G. B., Rooth, D.-O., and Stenberg, A. (2023). High school majors and future earnings. *American Economic Journal: Applied Economics*, 15(1):351–382.

Dasgupta, A. and Sharma, A. (2022). Preferences or expectations: Understanding the gender gap in major choice. *Oxford Economic Papers*, 74(4):1167–1194.

Eggers, A. C., Fowler, A., Hainmueller, J., Hall, A. B., and Snyder, J. M. (2015). On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races. *American Journal of Political Science*, 59(1):259–274.

Fakir, A. M. S., Bharati, T., and Khan, A. S. (2026). The impact of media coverage of sexual violence on women's work decisions in rural Bangladesh. *Journal of Population Economics*, 39:19.

Gupta, R. (2026). Fixing the leaky pipeline: Affirmative action in local elite colleges and subject choice. *Southern Economic Journal*, 93(1):457–473.

Iyer, L. and Mani, A. (2012). Traveling agents: Political change and bureaucratic turnover in India. *Review of Economics and Statistics*, 94(3):723–739.

Jain, T., Mukhopadhyay, A., Prakash, N., and Rakesh, R. (2022). Science education and labor market outcomes in a developing economy. *Economic Inquiry*, 60(2):741–763.

Kahn, S. and Ginther, D. K. (2017). Women and STEM. NBER Working Paper 23525, National Bureau of Economic Research.

Kalliyil, M. and Sahoo, S. (2026). The political economy of privatization of education: Role of local politicians in India. *Journal of Economic Behavior & Organization*, 241:107325.

Kim, J. E. and Lee, A. (2022). Crime, politics and policing: Evidence from India. Working paper, University of Rochester.

Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1):97–126.

Kondylis, F., Legovini, A., Vyborny, K., Zwager, A., and Andrade, L. (2025). Demand for "safe space": Avoiding harassment and complying with norms. *Journal of Development Economics*, 174:103392.

Kukreja, R. (2024). The political economy of education: Politician criminality and higher education institutions. *European Journal of Political Economy*, 85:102555.

Kumar, A. and Sahoo, S. (2024). Caste, gender, and intersectionality in stream choice: Evidence from higher secondary education in India. *Education Economics*, 32(1):20–46.

Lahoti, R. and Sahoo, S. (2020). Are educated leaders good for education? Evidence from India. *Journal of Economic Behavior & Organization*, 176:42–62.

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. house elections. *Journal of Econometrics*, 142(2):675–697.

Marshall, J. (2024). Can close election regression discontinuity designs identify effects of winning politician characteristics? *American Journal of Political Science*, 68(2):494–510.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714.

Mukherjee, T., Bhattacharya, S., and Mukhopadhyay, I. (2026). Unsafe daughters: Sexual violence in public spheres and intrahousehold preference for sons. *Journal of Population Economics*, 39(1):1–41.

Muralidharan, K. and Prakash, N. (2017). Cycling to school: Increasing secondary school enrollment for girls in India. *American Economic Journal: Applied Economics*, 9(3):321–350.

Nellis, G. and Siddiqui, N. (2018). Secular party rule and religious violence in Pakistan. *American Political Science Review*, 112(1):49–67.

Ngo, D. and Dustan, A. (2024). Preferences, access, and the STEM gender gap in centralized high school assignment. *American Economic Journal: Applied Economics*, 16(4):257–287.

Prakash, N., Rockmore, M., and Uppal, Y. (2019). Do criminally accused politicians affect economic outcomes? Evidence from India. *Journal of Development Economics*, 141:102370.

Prakash, N., Sahoo, S., Saraswat, D., and Sindhi, R. (2026). When criminality begets crime: The role of elected politicians in India. *The Journal of Law, Economics, and Organization*, 42(1):61–89.

Roychowdhury, P. (2021). (Em)Powered by science? Estimating the relative labor-market returns to majoring in science in high school in India. *Economics of Education Review*, 82:102118.

Sahoo, S. and Klasen, S. (2021). Gender segregation in education: Evidence from higher secondary stream choice in India. *Demography*, 58(3):987–1010.

Sahoo, S. and Klasen, S. (2026). Education choices and the gender gap in labour market outcomes: The role of STEM in India. *The Journal of Development Studies*, pages 1–23.

Sanderson, E. and Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, 190(2):212–221.

Sarkar, S. (2024). Local crime and early marriage: Evidence from India. *The Journal of Development Studies*, 60(5):763–787.

Sharma, S. and Sunder, N. (2025). Crime and human capital in India. *World Development*, 196:107161.

Siddique, Z. (2022). Media-reported violence and female labor supply. *Economic Development and Cultural Change*, 70(4):1337–1365.

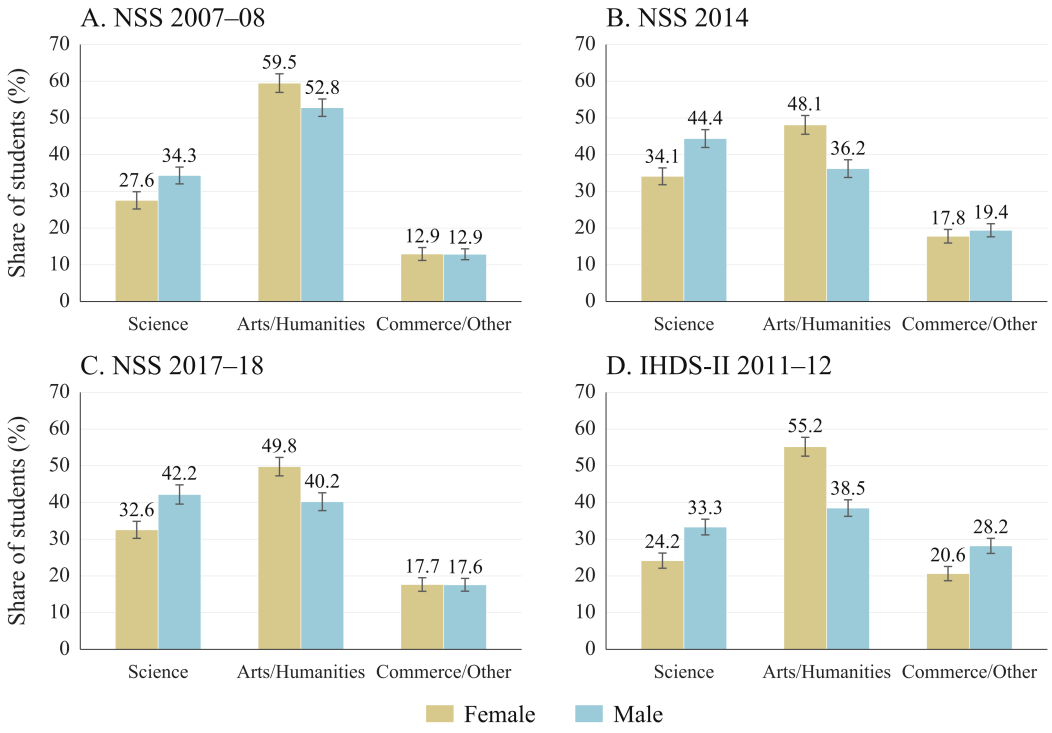
Sloane, C. M., Hurst, E. G., and Black, D. A. (2021). College majors, occupations, and the gender wage gap. *Journal of Economic Perspectives*, 35(4):223–248.

Vaishnav, M. (2017). *When Crime Pays: Money and Muscle in Indian Politics*. Yale University Press, New Haven, CT.

Wiswall, M. and Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, 82(2):791–824.

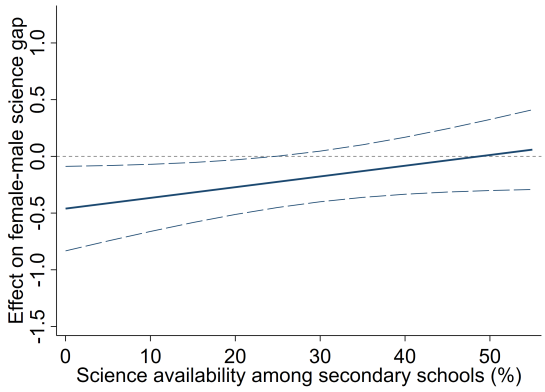
Main figures and tables

Figure 1: Higher-secondary stream choice by gender across surveys

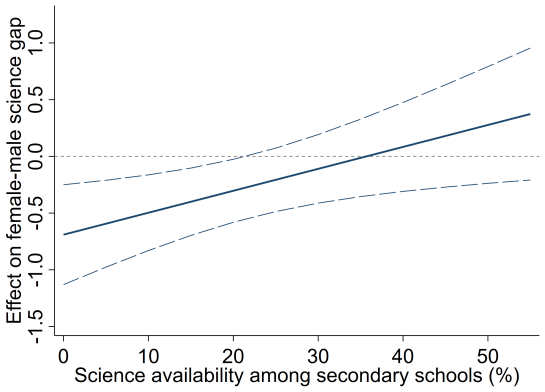


Notes: The figure reports higher-secondary stream shares for individuals aged 15–18 in the respective stream-choice samples, separately by gender. The National Sample Survey (NSS) panels use nationally representative education surveys conducted in 2007–08, 2014, and 2017–18; IHDS-II refers to the second round of the nationally representative India Human Development Survey, conducted in 2011–12. Whiskers indicate 95% confidence intervals. Commerce/Other includes all valid streams other than science and arts/humanities.

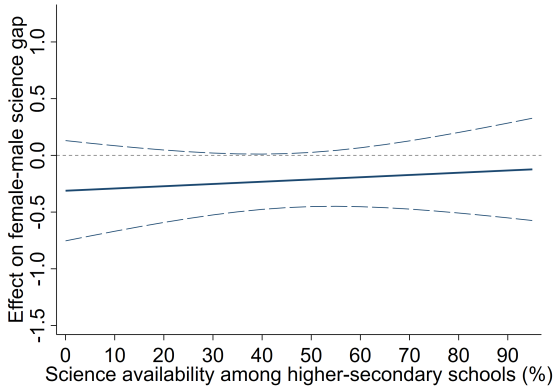
Figure 2: School science availability and the gender-differentiated effect on science choice



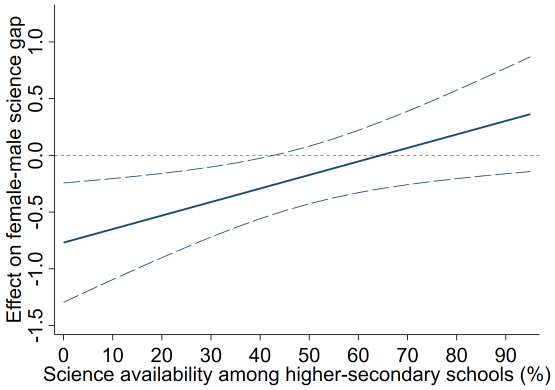
(a) Secondary schools, any accusation



(b) Secondary schools, serious accusation



(c) Higher-secondary schools, any accusation



(d) Higher-secondary schools, serious accusation

Notes: Each panel plots the estimated gender-differentiated reduced-form effect of close victories by accused candidates on science-stream choice at different levels of district science availability. Panels (a)–(b) use the percentage of secondary schools that extend to the higher-secondary level and offer science there; panels (c)–(d) use the percentage of higher-secondary schools that offer the science stream. Solid lines show marginal effects and dashed lines show 95% confidence intervals; the horizontal dashed line denotes zero. Estimates come from the specification in Equation (5), which includes the close-contest incidence and vote-margin controls used in the main specification, all constituent main effects and lower-order interactions, as well as the same individual and household covariates, district and NSS survey-round fixed effects, sampling weights, and district-clustered standard errors as the main analysis. School availability is measured two years before the NSS round and the close-win accused share one year before the round.

Table 1: Summary statistics: NSS estimation samples (ages 15–18)

	Observations	Mean	SD	Min	Max
Panel A: Science-stream estimation sample					
A1. Outcome and individual characteristics					
Science stream	23,657	0.405	0.491	0.000	1.000
Female	23,657	0.461	0.498	0.000	1.000
Other Backward Class	23,657	0.415	0.493	0.000	1.000
Scheduled Caste	23,657	0.150	0.357	0.000	1.000
Scheduled Tribe	23,657	0.112	0.316	0.000	1.000
Age	23,657	16.641	0.930	15.000	18.000
Years of schooling, household head	23,657	7.819	5.051	0.000	17.000
Household size	23,657	5.340	2.208	1.000	68.000
A2. Political variables					
Share of accused MLAs (any accusation)	23,657	0.333	0.254	0.000	1.000
Close-win accused share (any accusation)	23,657	0.042	0.086	0.000	1.000
Share of relevant close contests (any accusation)	23,657	0.071	0.110	0.000	1.000
Share of accused MLAs (serious accusation)	23,657	0.179	0.201	0.000	1.000
Close-win accused share (serious accusation)	23,657	0.026	0.067	0.000	1.000
Share of relevant close contests (serious accusation)	23,657	0.045	0.087	0.000	1.000
A3. NSS rounds					
NSS round 2014 indicator	23,657	0.338	0.473	0.000	1.000
NSS round 2017–18 indicator	23,657	0.572	0.495	0.000	1.000
Panel B: Higher-secondary estimation sample					
B1. Outcome and individual characteristics					
Higher-secondary enrolment	77,859	0.313	0.464	0.000	1.000
Female	77,859	0.438	0.496	0.000	1.000
Other Backward Class	77,859	0.409	0.492	0.000	1.000
Scheduled Caste	77,859	0.180	0.384	0.000	1.000
Scheduled Tribe	77,859	0.140	0.347	0.000	1.000
Age	77,859	16.485	1.128	15.000	18.000
Years of schooling, household head	77,859	6.007	5.014	0.000	17.000
Household size	77,859	5.591	2.328	1.000	68.000
B2. Political variables					
Share of accused MLAs (any accusation)	77,859	0.333	0.257	0.000	1.000
Close-win accused share (any accusation)	77,859	0.042	0.085	0.000	1.000
Share of relevant close contests (any accusation)	77,859	0.071	0.110	0.000	1.000
Share of accused MLAs (serious accusation)	77,859	0.188	0.208	0.000	1.000
Close-win accused share (serious accusation)	77,859	0.028	0.070	0.000	1.000
Share of relevant close contests (serious accusation)	77,859	0.049	0.092	0.000	1.000
B3. NSS rounds					
NSS round 2014 indicator	77,859	0.329	0.470	0.000	1.000
NSS round 2017–18 indicator	77,859	0.520	0.500	0.000	1.000

Notes: Statistics are reported separately for the two main NSS estimation samples, both restricted to ages 15–18. Panel A describes students currently enrolled at the higher-secondary level and used for the science-stream outcome; the science indicator equals one for the science stream and zero for the other academic streams. Panel B describes the full age-15–18 sample used for the higher-secondary enrolment outcome. Political variables are measured at the district level. The share of accused MLAs is measured one year before the NSS round. The close-win accused share is the share of constituencies in which the relevant accused candidate narrowly wins, and the close-contest share is the share of constituencies with a close contest between candidates who differ in accusation status. Close elections use the 3-percentage-point threshold. NSS 2007–08 is the omitted survey-round indicator.

Table 2: First-stage estimates

	Science stream		HS enrolment	
	(1) Any	(2) Serious	(3) Any	(4) Serious
Panel A: Dependent variable = share of accused MLAs				
Close-win accused-share instrument	1.016***	0.778***	1.152***	0.900***
	(0.142)	(0.133)	(0.138)	(0.128)
Female × close-win instrument	0.019	0.039*	0.010	0.012
	(0.020)	(0.021)	(0.010)	(0.009)
Sanderson–Windmeijer conditional F	77.60	44.01	136.89	113.01
Panel B: Dependent variable = Female × share of accused MLAs				
Close-win accused-share instrument	-0.005	-0.242***	0.050	-0.139*
	(0.083)	(0.088)	(0.071)	(0.074)
Female × close-win instrument	0.987***	1.341***	0.991***	1.279***
	(0.113)	(0.128)	(0.096)	(0.117)
Sanderson–Windmeijer conditional F	75.97	109.42	108.60	122.98
Kleibergen–Paap F	39.40	17.92	55.72	25.02
Observations	23,657	23,657	77,859	77,859
Individual/household controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
NSS round fixed effects	Yes	Yes	Yes	Yes

Notes: The table reports the two first-stage equations corresponding to the 2SLS estimates in Table 3. Panel A uses the district-level share of accused MLAs as the dependent variable; Panel B uses its interaction with the female indicator. The excluded instruments are the close-win accused share and its interaction with the female indicator. Close elections use the 3-percentage-point threshold. Columns (1)–(2) use the science-stream estimation sample and columns (3)–(4) use the higher-secondary enrolment sample. All specifications include the close-contest incidence and vote-margin controls, individual and household covariates, district and NSS survey-round fixed effects, and NSS sampling weights used in the main specification. Standard errors are clustered at the district level. The Sanderson–Windmeijer statistics are conditional first-stage F statistics for the corresponding endogenous regressor; the Kleibergen–Paap statistic is the system weak-identification statistic. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Criminally accused politicians, science-stream choice, and higher-secondary enrolment

	Science stream		HS enrolment	
	(1) Any	(2) Serious	(3) Any	(4) Serious
Share of accused MLAs	-0.052	-0.479*	-0.012	0.080
	(0.191)	(0.275)	(0.090)	(0.129)
Female × share of accused MLAs	-0.261**	-0.272**	-0.042	-0.011
	(0.130)	(0.120)	(0.055)	(0.054)
Female	0.012	-0.027	0.016	0.003
	(0.049)	(0.027)	(0.021)	(0.013)
Effect for females	-0.312*	-0.751***	-0.054	0.069
	(0.189)	(0.278)	(0.090)	(0.121)
Mean dependent variable	0.405	0.405	0.313	0.313
Kleibergen–Paap F statistic	39.40	17.92	55.72	25.02
Observations	23,657	23,657	77,859	77,859
Individual/household controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
NSS round fixed effects	Yes	Yes	Yes	Yes

Notes: The table reports 2SLS estimates for individuals aged 15–18. The science-stream outcome is estimated among students currently enrolled at the higher-secondary level; the higher-secondary enrolment outcome uses the full age-15–18 sample. The district-level share of accused MLAs and its interaction with the female indicator are instrumented by the close-win accused share and its interaction with the female indicator, using a 3-percentage-point close-election threshold. Political representation is measured one year before the NSS round. All specifications include the close-contest incidence and vote-margin controls, indicators for caste and religion, age fixed effects, years of schooling of the household head, household size, district fixed effects, and NSS survey-round fixed effects. NSS sampling weights are used and standard errors are clustered at the district level. "Effect for females" is the sum of the coefficient on the share of accused MLAs and the Female × share interaction; its standard error and significance level are obtained from the corresponding linear-combination test. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Reallocation across higher-secondary streams

	Science	Arts/Humanities	Commerce
Panel A: Any accusation			
Female × share of accused MLAs	-0.261**	0.305**	-0.044
	(0.130)	(0.127)	(0.076)
Panel B: Serious accusation			
Female × share of accused MLAs	-0.272**	0.258**	0.014
	(0.120)	(0.119)	(0.063)
Mean dependent variable	0.405	0.423	0.172
Observations	23,657	23,657	23,657
Individual/household controls	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes
NSS round fixed effects	Yes	Yes	Yes

Notes: The table reports the coefficient on Female × share of accused MLAs from separate 2SLS regressions for the mutually exclusive science, arts/humanities, and commerce streams among higher-secondary students aged 15–18. The instruments, close-election controls, individual and household covariates, fixed effects, sampling weights, and clustering are identical to the main specification in Table 3. Standard errors clustered at the district level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Criminally accused politicians, science-stream choice, and post-secondary enrolment: IHDS

	Science stream		Post-secondary enrolment	
	(1) Any	(2) Serious	(3) Any	(4) Serious
Panel A: Second-stage estimates				
Female × share of accused MLAs	-0.354	-0.377**	0.063	0.037
	(0.219)	(0.176)	(0.098)	(0.073)
Female	0.012	-0.017	-0.002	0.010
	(0.069)	(0.040)	(0.032)	(0.020)
Panel B: First-stage estimates				
Female × close-win instrument	0.855***	1.064***	0.962***	1.088***
	(0.162)	(0.141)	(0.161)	(0.129)
Mean dependent variable	0.289	0.289	0.382	0.382
Kleibergen–Paap F statistic	27.98	56.66	35.67	71.33
Observations	4,440	4,440	12,186	12,186
Individual controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes

Notes: The table reports 2SLS estimates from IHDS-II for individuals aged 15–18. Columns (1)–(2) use science-stream choice from the course-of-study information collected for current or recent students studying above Class 10; columns (3)–(4) use an indicator equal to one if an individual is currently enrolled and has completed at least ten years of schooling. Because the specifications include district fixed effects, the district-level share of accused MLAs is absorbed; the endogenous regressor is Female × share of accused MLAs, instrumented by Female × close-win accused share. Panel B reports the coefficient on this excluded instrument from the corresponding first-stage regression. Close elections use the 3-percentage-point threshold. All specifications include age and indicators for caste and religion. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Heterogeneity in science-stream choice by local access to science education

Panel A: Percentage of secondary schools with higher-secondary science

	Science-stream choice	
	(1) Any	(2) Serious
Female × close-win accused share	-0.461**	-0.690***
	(0.190)	(0.225)
Female × close-win accused share × science availability	0.009*	0.019**
	(0.005)	(0.008)
Observations	23,657	23,657

Panel B: Percentage of higher-secondary schools offering science

	Science-stream choice	
	(1) Any	(2) Serious
Female × close-win accused share	-0.312	-0.769***
	(0.225)	(0.268)
Female × close-win accused share × science availability	0.002	0.012**
	(0.004)	(0.005)
Observations	23,657	23,657
Individual/household controls	Yes	Yes
District fixed effects	Yes	Yes
NSS round fixed effects	Yes	Yes

Notes: The table reports results from two separate sets of reduced-form regressions for science-stream choice among higher-secondary students aged 15–18. Panel A measures local science availability as the percentage of secondary schools that extend to the higher-secondary level and offer science there. Panel B measures local science availability as the percentage of higher-secondary schools in the district that offer the science stream. Within each panel, columns (1) and (2) use the any-accusation and serious-accusation definitions, respectively. The reported coefficients are for Female × close-win accused share and its interaction with the relevant science-availability measure. Close elections use the 3-percentage-point threshold. Each regression also includes the close-contest incidence and vote-margin controls used in the main specification, all constituent main effects and lower-order interactions, the same individual and household covariates as the main specification, district and NSS survey-round fixed effects, and sampling weights. Standard errors are clustered at the district level. School availability is measured two years before the NSS round and the close-win accused share one year before the round. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Criminally accused politicians and perceived harassment of unmarried girls: IHDS

	Perceived harassment	
	(1) Any	(2) Serious
Panel A: Second-stage estimates		
Share of accused MLAs	0.345**	0.342*
	(0.159)	(0.207)
Panel B: First-stage estimates		
Close-win accused-share instrument	0.966***	0.791***
	(0.162)	(0.150)
Mean dependent variable	0.202	0.202
Kleibergen–Paap F statistic	35.37	27.71
Observations	37,638	37,638
Baseline household controls	Yes	Yes
Region fixed effects	Yes	Yes

Notes: The table reports 2SLS estimates using household-head observations from IHDS-II. The dependent variable is the household-level perception of harassment of unmarried girls in the respondent's village or urban neighbourhood. The district-level share of accused MLAs is instrumented by the close-win accused share, using a 3-percentage-point close-election threshold. Under the any-accusation definition, the relevant close contests are between accused and non-accused candidates; under the serious-accusation definition, they are between seriously accused candidates and candidates without a serious accusation. All specifications include close-election controls, baseline household covariates, and region fixed effects. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Prior perceived harassment and the gender gap in science-stream choice: IHDS

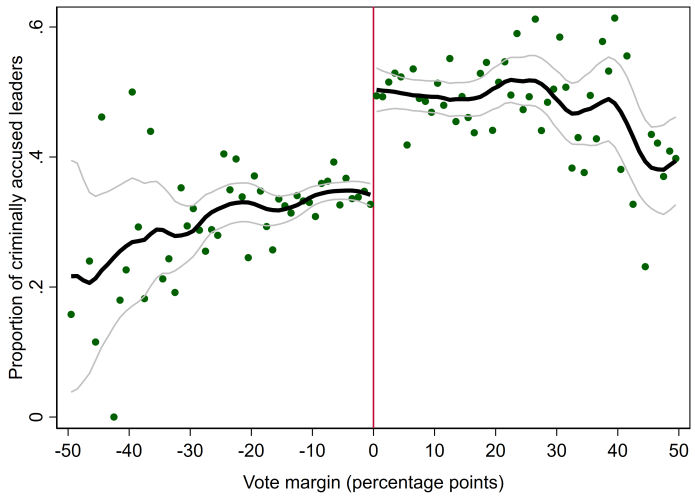
	Science-stream choice		
	(1) Baseline	(2) District FE	(3) PSU FE
Prior perceived harassment	0.074	0.110	-0.079
	(0.070)	(0.084)	(0.564)
Female	-0.078***	-0.069***	-0.059***
	(0.017)	(0.017)	(0.020)
Female × prior perceived harassment	-0.152*	-0.217**	-0.179*
	(0.082)	(0.084)	(0.108)
Mean dependent variable	0.290	0.290	0.293
Observations	4,212	4,113	3,610
Controls	Yes	Yes	Yes
District fixed effects	No	Yes	No
PSU fixed effects	No	No	Yes

Notes: The table reports OLS regression estimates where the dependent variable is an indicator for science-stream choice among individuals aged 15–18 in IHDS-II. Prior perceived harassment is the leave-one-out village- or urban-neighbourhood-level average of household perceptions of harassment of unmarried girls measured in IHDS-I. All specifications include age, caste and religion indicators, baseline household characteristics, the other IHDS-I leave-one-out locality crime-perception measures, and baseline norms and social-capital controls. Column (1) is the baseline specification with the full control set; column (2), the preferred specification, additionally includes district fixed effects; and column (3) instead includes primary-sampling-unit (PSU) fixed effects. Standard errors are clustered at the primary sampling unit level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

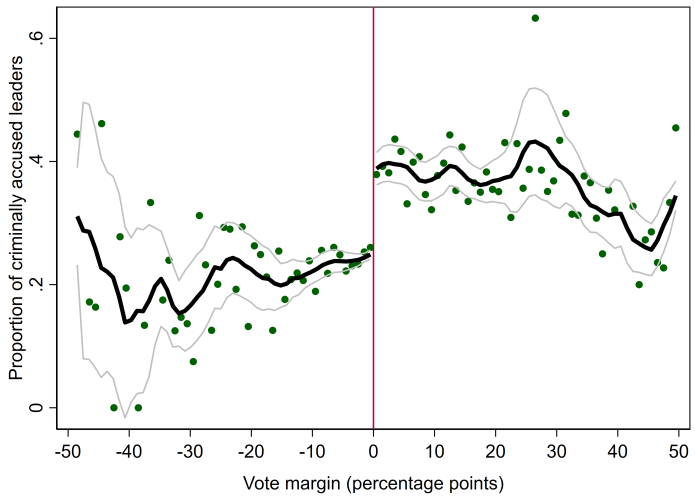
Online Supplementary Appendices

A Supplementary Figures and Tables

Figure A.1: Close-election first-stage relationship



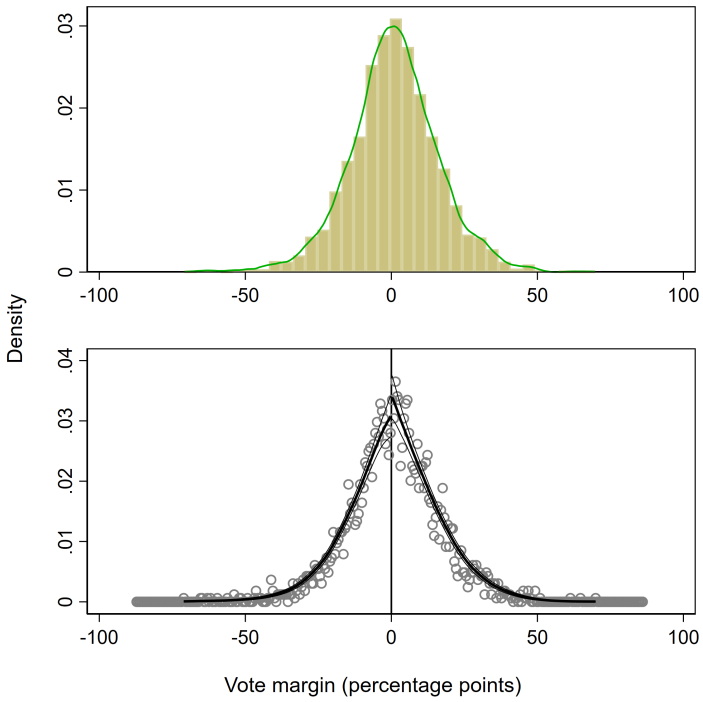
(a) Any accusation



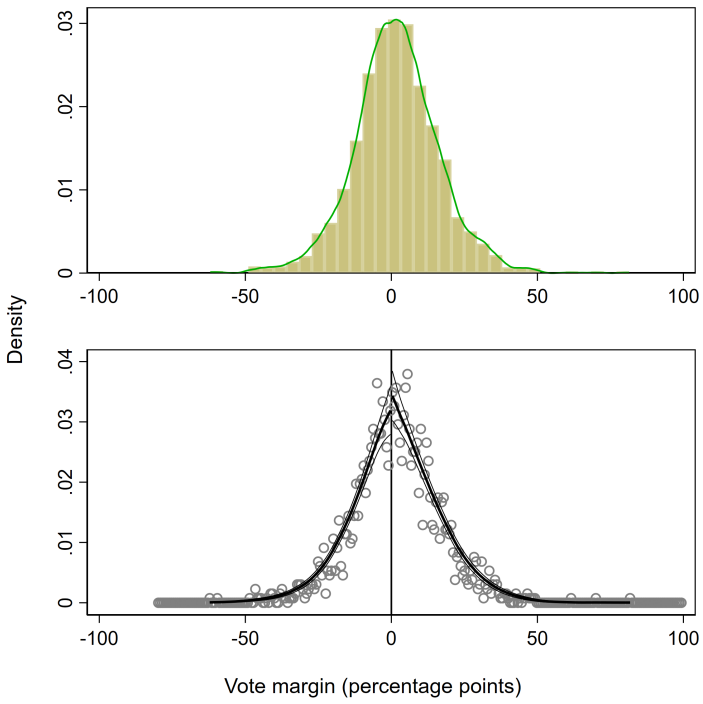
(b) Serious accusation

Notes: The panels plot the district-level share of accused MLAs against the constituency-level vote margin between the accused candidate and the relevant comparison candidate. The margin is positive when the accused candidate wins and negative when the comparison candidate wins. The data are aggregated into one-percentage-point vote-margin bins. The curves are local polynomial regressions with 95% confidence intervals, fitted separately on either side of the zero vote-margin threshold.

Figure A.2: Density and continuity of the close-election vote margin



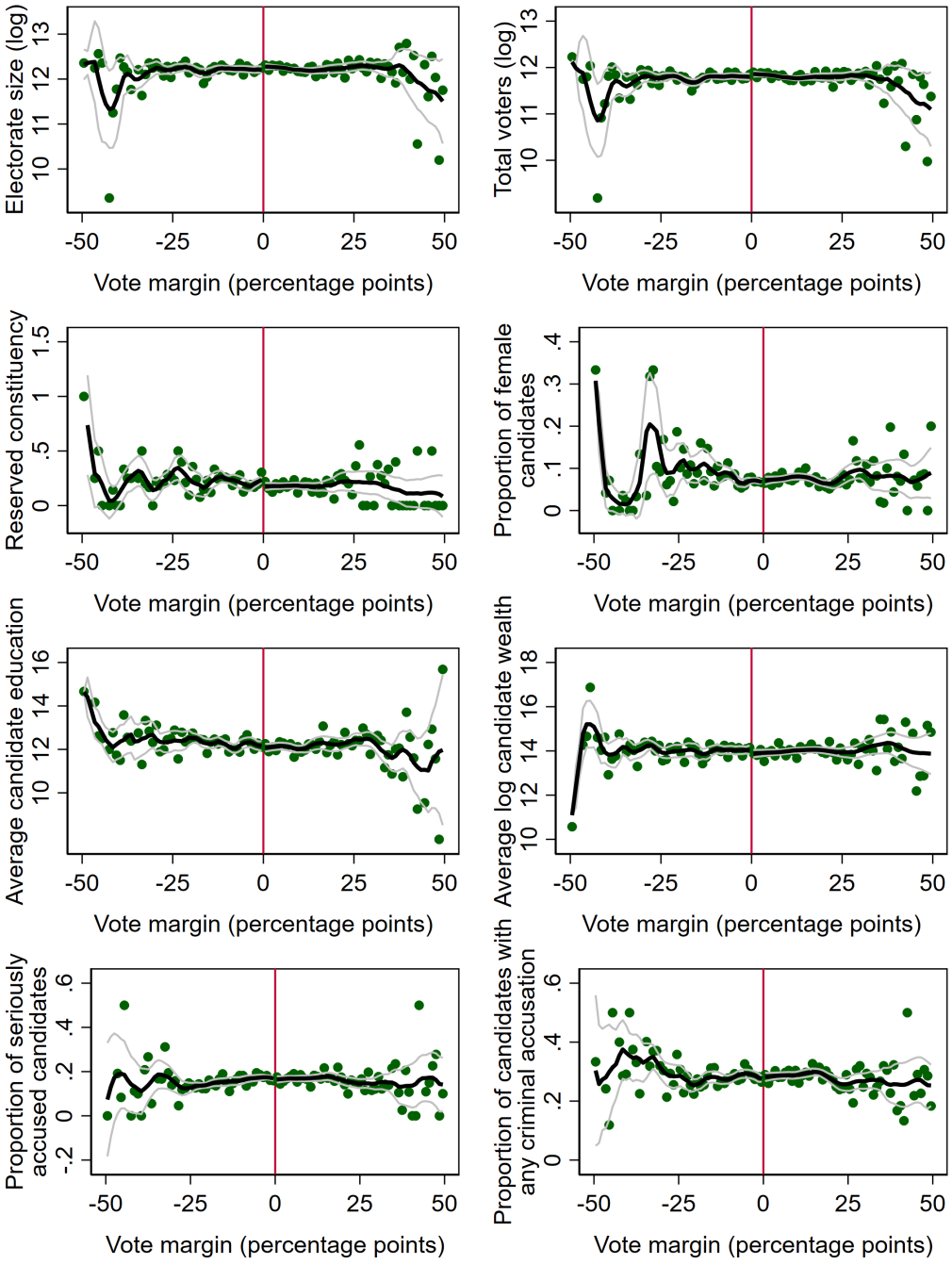
(a) Any accusation



(b) Serious accusation

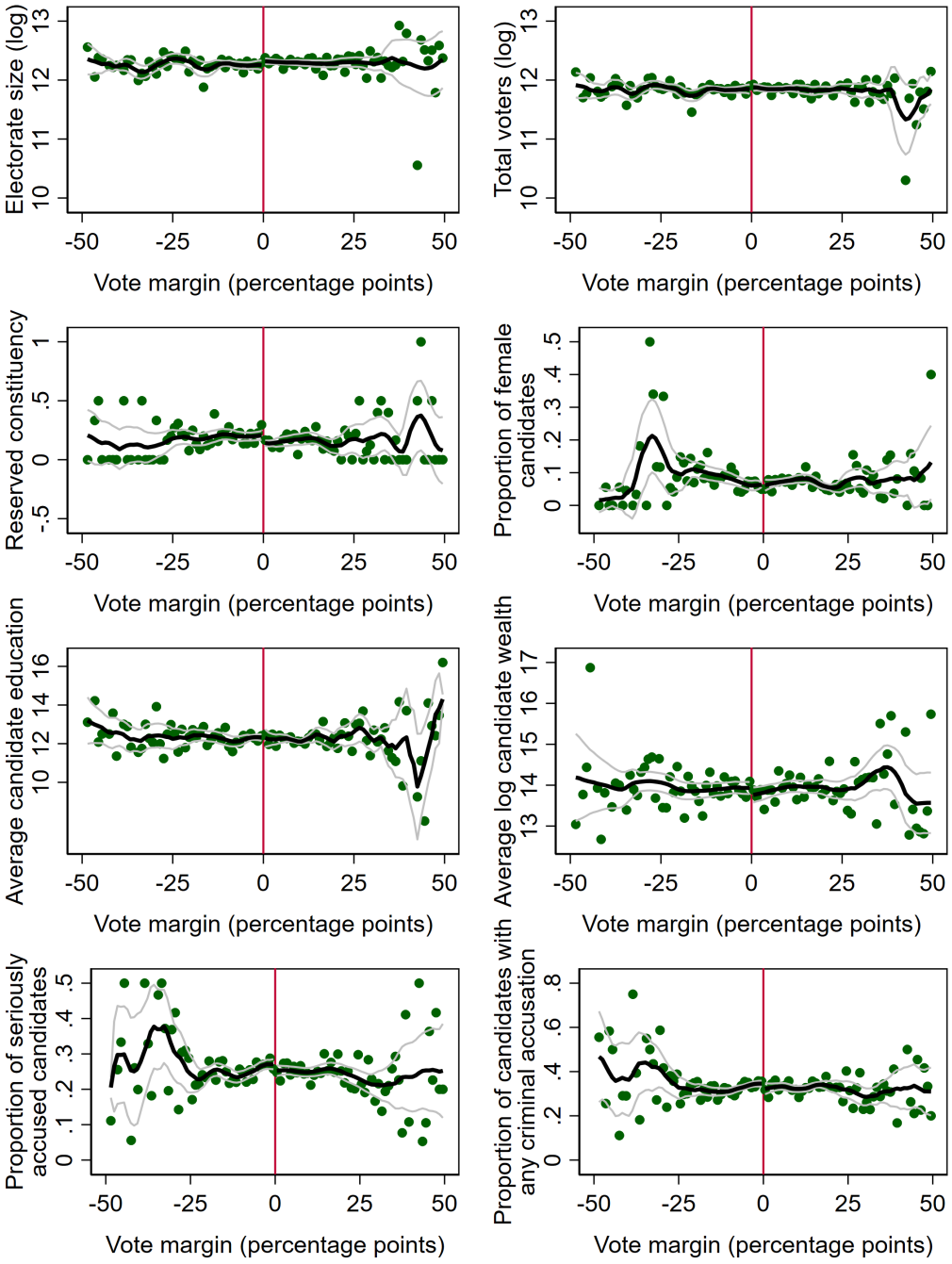
Notes: The horizontal axis is the vote-share margin between the accused candidate and the relevant comparison candidate, centered at zero. Positive values indicate an accused-candidate victory and negative values a comparison-candidate victory. In each panel, the upper plot shows the density of the vote margin and the lower plot shows the McCrary density test around the zero threshold.

Figure A.3: Continuity checks for predetermined characteristics: any accusation



Notes: The horizontal axis is the vote-share margin between the accused candidate and the candidate without any accusation; positive values indicate that the accused candidate wins. The vertical axes show the predetermined constituency and candidate-pool characteristics named in each panel. The curves are local polynomial regressions with 95% confidence intervals, fitted separately on either side of the zero vote-margin threshold.

Figure A.4: Continuity checks for predetermined characteristics: serious accusation



Notes: The horizontal axis is the vote-share margin between the seriously accused candidate and the candidate without a serious accusation; positive values indicate that the seriously accused candidate wins. The vertical axes show the predetermined constituency and candidate-pool characteristics named in each panel. The curves are local polynomial regressions with 95% confidence intervals, fitted separately on either side of the zero vote-margin threshold.

Table A.1: Summary statistics: IHDS estimation samples

Variable	Observations	Mean	SD	Min	Max
Panel A: Science-stream estimation sample (ages 15–18)					
Science	4,440	0.289	0.453	0.000	1.000
Arts/Humanities	4,440	0.466	0.499	0.000	1.000
Commerce	4,440	0.162	0.369	0.000	1.000
Other	4,440	0.084	0.277	0.000	1.000
Female	4,440	0.485	0.500	0.000	1.000
Age	4,440	16.872	0.974	15.000	18.000
Share of accused MLAs (any accusation)	4,440	0.314	0.239	0.000	1.000
Close-win accused share (any accusation)	4,440	0.042	0.087	0.000	1.000
Share of accused MLAs (serious accusation)	4,440	0.217	0.216	0.000	1.000
Close-win accused share (serious accusation)	4,440	0.031	0.080	0.000	1.000
Panel B: Post-secondary-enrolment estimation sample (ages 15–18)					
Post-secondary enrolment	12,186	0.382	0.486	0.000	1.000
Female	12,186	0.477	0.499	0.000	1.000
Age	12,186	16.455	1.106	15.000	18.000
Share of accused MLAs (any accusation)	12,186	0.314	0.249	0.000	1.000
Close-win accused share (any accusation)	12,186	0.046	0.094	0.000	1.000
Share of accused MLAs (serious accusation)	12,186	0.222	0.225	0.000	1.000
Close-win accused share (serious accusation)	12,186	0.035	0.091	0.000	1.000
Panel C: Household-perception analysis sample					
Perceived theft	37,638	0.039	0.194	0.000	1.000
Perceived breaking-in	37,637	0.010	0.102	0.000	1.000
Perceived threat/attack	37,636	0.019	0.137	0.000	1.000
Perceived harassment of unmarried girls	37,638	0.202	0.401	0.000	1.000
Share of accused MLAs (any accusation)	37,638	0.306	0.244	0.000	1.000
Close-win accused share (any accusation)	37,638	0.043	0.090	0.000	1.000
Share of accused MLAs (serious accusation)	37,638	0.215	0.217	0.000	1.000
Close-win accused share (serious accusation)	37,638	0.033	0.085	0.000	1.000
Panel D: Prior-harassment science-choice estimation sample (ages 15–18)					
Science	4,113	0.290	0.454	0.000	1.000
Arts/Humanities	4,113	0.462	0.499	0.000	1.000
Commerce	4,113	0.161	0.368	0.000	1.000
Other	4,113	0.086	0.281	0.000	1.000
Female	4,113	0.482	0.500	0.000	1.000
Prior perceived harassment	4,113	0.116	0.196	0.000	1.000

Notes: Panel A uses the science-stream estimation sample from the IHDS education analysis. Panel B uses the corresponding post-secondary-enrolment sample. Panel C uses the household-head estimation sample for the perceived-harassment analysis; the other perception outcomes have slightly fewer observations because of item nonresponse. Panel D uses the preferred district-fixed-effects sample for the prior-harassment science-choice analysis. Political variables are summarized within the corresponding any- or serious-accusation estimation sample. The four stream indicators are mutually exclusive and exhaustive among individuals with course-of-study information.

Table A.2: Share of higher-secondary students attending an educational institution at least 5 km from residence, by stream and gender

	Science	Arts/Humanities	Difference (pp)
Girls	32.9	26.3	6.6
Boys	36.3	32.9	3.5

Notes: The table reports percentages of higher-secondary students aged 15–18 whose current educational institution is at least 5 km from their residence, pooling the 2007–08, 2014, and 2017–18 National Sample Survey (NSS) education rounds. The sample is restricted to students enrolled in science or arts/humanities with non-missing distance information. The final column reports the science–arts/humanities difference in percentage points. NSS sampling weights are used.

Table A.3: Balance of constituency and candidate characteristics in close elections

	Seriously Accused (SA)			Any Accused (AA)		
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	**Non SA**	**SA**	**Diff**	**Non AA**	**AA**	**Diff**
Panel A: Constituency Characteristics						
Total voters in constituency (log)	11.84	11.87	0.03	11.79	11.85	0.07*
	(0.39)	(0.35)	(0.04)	(0.44)	(0.36)	(0.04)
Total electors in constituency (log)	12.26	12.33	0.07	12.20	12.27	0.07*
	(0.45)	(0.42)	(0.05)	(0.49)	(0.43)	(0.04)
Reserved constituency	0.22	0.15	-0.07*	0.22	0.17	-0.05
	(0.41)	(0.36)	(0.04)	(0.41)	(0.37)	(0.03)
Reserved constituency for SC	0.13	0.09	-0.04	0.15	0.10	-0.05*
	(0.34)	(0.29)	(0.03)	(0.35)	(0.29)	(0.03)
Reserved constituency for ST	0.08	0.06	-0.03	0.07	0.07	-0.00
	(0.27)	(0.23)	(0.03)	(0.26)	(0.26)	(0.02)
Proportion of female candidates	0.06	0.06	-0.00	0.07	0.07	0.00
	(0.10)	(0.09)	(0.01)	(0.10)	(0.11)	(0.01)
Average age of candidates	46.09	45.34	-0.75	46.11	45.75	-0.36
	(5.68)	(4.83)	(0.55)	(5.49)	(4.99)	(0.46)
Average years of education of candidates	12.28	12.29	0.01	12.13	12.08	-0.05
	(1.95)	(1.82)	(0.20)	(2.08)	(1.78)	(0.17)
Average log of wealth of candidates	13.86	13.82	-0.04	14.00	13.91	-0.09
	(1.34)	(1.44)	(0.15)	(1.44)	(1.37)	(0.12)
Proportion of seriously accused candidates	0.27	0.25	-0.02	0.17	0.16	-0.01
	(0.16)	(0.15)	(0.02)	(0.18)	(0.16)	(0.02)
Proportion of criminally accused candidates	0.35	0.31	-0.03*	0.27	0.28	0.00
	(0.18)	(0.18)	(0.02)	(0.15)	(0.14)	(0.01)
Congress party contested	0.70	0.69	-0.01	0.70	0.73	0.03
	(0.46)	(0.46)	(0.05)	(0.46)	(0.44)	(0.04)
BJP contested	0.67	0.67	0.01	0.70	0.70	-0.00
	(0.47)	(0.47)	(0.05)	(0.46)	(0.46)	(0.04)
Panel B: Candidate Characteristics						
Female winner	0.09	0.01	-0.08***	0.11	0.03	-0.08***
	(0.28)	(0.07)	(0.02)	(0.31)	(0.17)	(0.02)
Winner's age	49.84	48.52	-1.32	50.43	50.21	-0.22
	(9.61)	(10.09)	(1.03)	(9.87)	(10.06)	(0.88)
Winner's years of education	13.55	13.13	-0.42	13.26	13.06	-0.20
	(3.85)	(3.19)	(0.37)	(3.88)	(3.20)	(0.31)
Winner's net wealth (log)	15.21	15.63	0.42*	15.61	15.95	0.34*
	(2.64)	(2.16)	(0.25)	(2.31)	(1.82)	(0.18)
Winner from ruling party	0.59	0.40	-0.19***	0.57	0.45	-0.12***
	(0.49)	(0.49)	(0.05)	(0.50)	(0.50)	(0.04)
Winner is incumbent	0.29	0.23	-0.05	0.30	0.27	-0.03
	(0.45)	(0.42)	(0.05)	(0.46)	(0.45)	(0.04)
Votes received by winner (log)	10.88	10.87	-0.00	10.82	10.88	0.05
	(0.43)	(0.38)	(0.04)	(0.49)	(0.40)	(0.04)
Votes received by runner-up (log)	10.84	10.83	-0.00	10.78	10.84	0.05
	(0.44)	(0.38)	(0.04)	(0.49)	(0.40)	(0.04)
Observations	171	196		234	283	

Notes: The table compares observable characteristics across relevant close elections using the 3-percentage-point vote-margin threshold. For serious accusations, "SA" denotes contests narrowly won by the seriously accused candidate and "Non SA" contests narrowly won by the comparison candidate without a serious accusation. For any accusation, "AA" denotes contests narrowly won by the accused candidate and "Non AA" contests narrowly won by the non-accused comparison candidate. The level columns report means with standard deviations in parentheses; "Diff" reports the difference between accused-winner and comparison-winner contests, with standard errors in parentheses. There are 367 relevant close contests under the serious-accusation definition and 517 under the any-accusation definition. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Predictors of an accused candidate winning a close election

Variable	(1) Serious crimes	(2) Any crimes
Total voters in constituency (log)	-0.030	0.125
	(0.162)	(0.127)
Total electors in constituency (log)	0.091	-0.039
	(0.146)	(0.114)
Reserved constituency for SC	-0.177*	-0.143*
	(0.091)	(0.076)
Reserved constituency for ST	-0.279*	-0.071
	(0.157)	(0.132)
Proportion of female candidates	-0.136	-0.021
	(0.327)	(0.228)
Average age of candidates	-0.009	-0.003
	(0.007)	(0.005)
Average years of education of candidates	-0.004	-0.000
	(0.016)	(0.013)
Average log of wealth of candidates	0.005	-0.023
	(0.024)	(0.019)
Proportion of seriously accused candidates	-0.269	-0.269
	(0.362)	(0.191)
Proportion of criminally accused candidates	-0.169	0.322
	(0.272)	(0.249)
Congress party contested	-0.085	0.056
	(0.067)	(0.057)
BJP contested	-0.057	-0.035
	(0.068)	(0.054)
Regional party contested	-0.162	-0.052
	(0.102)	(0.088)
Left party contested	-0.066	-0.018
	(0.059)	(0.051)
Independent candidate contested	0.034	0.039
	(0.078)	(0.067)
Proportion of urban population (2004–05)	0.067	0.134
	(0.209)	(0.157)
Proportion of female population (2004–05)	2.952**	1.864
	(1.341)	(1.160)
Proportion of ST population (2004–05)	0.414	0.205
	(0.378)	(0.308)
Proportion of SC population (2004–05)	0.129	0.181
	(0.371)	(0.300)
Proportion of OBC population (2004–05)	-0.238	0.011
	(0.156)	(0.127)
Male primary education completion rate (2004–05)	0.187	0.547
	(0.653)	(0.560)
Female primary education completion rate (2004–05)	-0.538	-0.736*
	(0.521)	(0.433)
Constant	-0.679	-1.000
	(1.175)	(0.960)
Observations	350	491
R-squared	0.066	0.037
F statistic	1.242	0.913
p-value	0.210	0.578

Notes: The dependent variable is an indicator for whether the relevant accused candidate wins a close election. Column (1) uses close contests between seriously accused candidates and candidates without a serious accusation; column (2) uses close contests between accused and non-accused candidates. Close elections use the 3-percentage-point vote-margin threshold. The regressors are the observable constituency, candidate, and district characteristics listed in the table. Robust standard errors are reported in parentheses. The F statistic and p-value at the bottom test the joint significance of the included characteristics. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: District-level comparison of close elections by type of accused candidate

	Seriously Accused			**Any Accused**		
Variable	(1) Less	(2) More	(3) Diff	(4) Less	(5) More	(6) Diff
Proportion of urban population (2004–05)	0.21	0.20	-0.01	0.22	0.22	0.00
	(0.19)	(0.15)	(0.02)	(0.18)	(0.16)	(0.02)
Proportion of female population (2004–05)	0.49	0.49	0.00	0.49	0.49	0.00
	(0.03)	(0.02)	(0.00)	(0.03)	(0.02)	(0.00)
Proportion of ST population (2004–05)	0.08	0.08	-0.00	0.08	0.07	-0.00
	(0.15)	(0.13)	(0.02)	(0.14)	(0.13)	(0.02)
Proportion of SC population (2004–05)	0.19	0.21	0.01	0.19	0.20	0.00
	(0.09)	(0.09)	(0.01)	(0.10)	(0.09)	(0.01)
Proportion of OBC population (2004–05)	0.46	0.43	-0.03	0.45	0.45	0.00
	(0.22)	(0.21)	(0.03)	(0.23)	(0.22)	(0.03)
Male primary education completion rate (2004–05)	0.67	0.65	-0.02	0.68	0.67	-0.01
	(0.13)	(0.12)	(0.02)	(0.13)	(0.12)	(0.01)
Female primary education completion rate (2004–05)	0.51	0.48	-0.03*	0.52	0.50	-0.02
	(0.17)	(0.16)	(0.02)	(0.16)	(0.15)	(0.02)
Proportion of constituencies reserved for SC	0.16	0.16	0.00	0.16	0.15	-0.01
	(0.12)	(0.10)	(0.01)	(0.12)	(0.10)	(0.01)
Proportion of constituencies reserved for ST	0.10	0.08	-0.01	0.09	0.08	-0.01
	(0.25)	(0.21)	(0.03)	(0.24)	(0.19)	(0.02)
Proportion of seriously accused winners in non-close elections	0.20	0.21	0.01	0.29	0.31	0.03
	(0.18)	(0.18)	(0.02)	(0.22)	(0.21)	(0.02)
Proportion of non-seriously/non accused winners in non-close elections	0.62	0.61	-0.01	0.53	0.48	-0.06**
	(0.20)	(0.20)	(0.03)	(0.23)	(0.23)	(0.03)
Total constituencies in the district	9.49	9.81	0.32	9.51	8.58	-0.93
	(9.87)	(7.97)	(1.12)	(9.47)	(5.11)	(0.83)
Observations	119	137		140	181	

Notes: The unit of observation is a district-election year containing at least one relevant close contest, defined using the 3-percentage-point vote-margin threshold. Columns (1)–(3) use the serious-accusation definition and columns (4)–(6) use the any-accusation definition. "Less" and "More" classify district-election years according to whether accused winners are fewer or more numerous than comparison-candidate winners across the relevant close contests. The level columns report means with standard deviations in parentheses; "Diff" reports the difference between the "More" and "Less" groups, with standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Criminally accused politicians and gender differences across higher-secondary streams: IHDS

	Science	Arts/Humanities	Commerce	Other
Panel A: Any accusation				
Female × share of accused MLAs	-0.354	0.190	0.090	0.074
	(0.219)	(0.232)	(0.117)	(0.074)
Panel B: Serious accusation				
Female × share of accused MLAs	-0.377**	0.217	0.068	0.091
	(0.176)	(0.201)	(0.090)	(0.058)
Mean dependent variable	0.289	0.466	0.162	0.084
Observations	4,440	4,440	4,440	4,440
Individual controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes

Notes: The table reports the coefficient on Female × share of accused MLAs from separate 2SLS regressions for four mutually exclusive and exhaustive IHDS-II course-of-study categories among individuals aged 15–18. The specification and common estimation sample are the same as for the science-stream regressions in Table 5. Because district fixed effects absorb the district-level share of accused MLAs, Female × share of accused MLAs is instrumented by Female × close-win accused share. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Illustrative downstream labour-market implications

	Benchmark	Any accusation	Serious accusation
Panel A: Estimated change in girls' science choice			
Female total effect on science choice		−0.312	−0.751
SD of criminal political representation		0.254	0.201
Implied decline in girls' science participation		**7.9 pp**	**15.1 pp**
Panel B: Illustrative downstream implications			
Effect on female labour-force participation	9.5 pp	**0.75 pp**	**1.43 pp**
As % of female LFP mean (27%)		**2.8%**	**5.3%**
Effect on female salaried employment, conditional on employment	10.5 pp	**0.83 pp**	**1.58 pp**
Effect on female annual earnings	0.136 log points	**1.1%**	**2.1%**

Notes: Panel A combines the female total effects from the baseline science-choice regressions with a one-standard-deviation change in the corresponding measure of criminal political representation. The benchmark effects in Panel B are from Sahoo and Klasen (2026). The implied downstream magnitudes multiply the estimated decline in girls' science participation by the corresponding benchmark effect. Salaried employment is conditional on employment. These calculations are illustrative magnitude benchmarks and should not be interpreted as estimates of downstream causal effects.

Table A.8: Criminally accused politicians and household perceptions of local crime: IHDS

	Theft	Breaking-in	Threat/attack	Harassment
Panel A: Any accusation				
Share of accused MLAs	-0.005	0.006	0.008	0.348**
	(0.042)	(0.013)	(0.034)	(0.158)
Panel B: Serious accusation				
Share of accused MLAs	0.011	0.013	0.038	0.346*
	(0.047)	(0.018)	(0.040)	(0.206)
Observations	37,635	37,635	37,635	37,635
Baseline household controls	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes

Notes: The table reports 2SLS estimates using household-head observations from IHDS-II and follows the specification in Table 7, changing only the dependent variable across columns. The outcomes are household-level perceptions of theft, breaking-in, threat or attack, and harassment of unmarried girls in the respondent's village or urban neighbourhood. In each regression, the district-level share of accused MLAs is instrumented by the corresponding close-win accused share using a 3-percentage-point close-election threshold. All four outcomes are estimated on the common sample with non-missing observations for each perception measure; the harassment coefficient therefore differs slightly from Table 7. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Prior perceived harassment, other local crime perceptions, and the gender gap in science-stream choice: IHDS

	Science-stream choice	
	(1)	(2)
Prior perceived harassment	0.111	0.103
	(0.083)	(0.086)
Female	-0.069***	-0.067***
	(0.017)	(0.018)
Female × prior perceived harassment	-0.216**	-0.202**
	(0.084)	(0.094)
Prior other-crime perception index	-0.332	-0.240
	(0.234)	(0.304)
Female × prior other-crime perception index		-0.184
		(0.357)
Controls	Yes	Yes
District fixed effects	Yes	Yes
Observations	4,113	4,113

Notes: The dependent variable is an indicator for science-stream choice among individuals aged 15–18 in IHDS-II. Prior perceived harassment is the leave-one-out village- or urban-neighbourhood-level average of household perceptions of harassment of unmarried girls measured in IHDS-I. The prior other-crime perception index is the simple average of the corresponding leave-one-out locality-level perceptions of theft, breaking-in, and threat or attack. Column (1) includes both prior perceived harassment and the other-crime perception index. Column (2) additionally allows the association between the other-crime perception index and science choice to differ by gender. All specifications include age, caste and religion indicators, baseline household characteristics, and baseline norms and social-capital controls. District fixed effects are included, and standard errors are clustered at the primary sampling unit level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: Prior perceived harassment and gender differences across higher-secondary streams: IHDS

	Science	Arts/Humanities	Commerce	Other
Female × prior perceived harassment	-0.217**	0.231***	-0.031	0.017
	(0.084)	(0.081)	(0.052)	(0.037)
Mean dependent variable	0.290	0.462	0.161	0.086
Observations	4,113	4,113	4,113	4,113
Controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes

Notes: The table reports the coefficient on Female × prior perceived harassment from separate regressions for four mutually exclusive and exhaustive IHDS-II course-of-study categories among individuals aged 15–18. Prior perceived harassment is the leave-one-out village- or urban-neighbourhood-level average of household perceptions of harassment of unmarried girls measured in IHDS-I. The regressions use the common district-fixed-effects sample and the same controls as the preferred specification in Table 8; the constituent main effects for Female and prior perceived harassment are included but not reported. Standard errors are clustered at the primary sampling unit level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.11: Robustness controlling for definition-specific imbalanced characteristics

	Science stream		HS enrolment	
	(1) Any	(2) Serious	(3) Any	(4) Serious
Share of accused MLAs	-0.100	-0.537**	-0.010	0.085
	(0.193)	(0.272)	(0.088)	(0.137)
Female × share of accused MLAs	-0.259**	-0.271**	-0.042	-0.011
	(0.130)	(0.121)	(0.055)	(0.054)
Female	0.011	-0.027	0.016	0.003
	(0.049)	(0.027)	(0.021)	(0.013)
Effect for females	-0.359*	-0.807***	-0.052	0.074
	(0.191)	(0.275)	(0.087)	(0.128)
Mean dependent variable	0.405	0.405	0.313	0.313
Kleibergen–Paap F statistic	33.54	19.87	55.27	22.81
Observations	23,657	23,657	77,859	77,859
Individual/household controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
NSS round fixed effects	Yes	Yes	Yes	Yes

Notes: The table re-estimates the 2SLS specification in Table 3 after adding the observable constituency and candidate characteristics that are significantly imbalanced for the corresponding accusation definition in Table A.3. The additional controls therefore differ between the any- and serious-accusation specifications. All other treatment, instrument, timing, covariate, fixed-effect, weighting, and clustering choices are unchanged from the main specification. "Effect for females" is the sum of the coefficient on the share of accused MLAs and the Female × share interaction. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.12: Alternative vote-margin polynomial controls

	Science stream		HS enrolment	
	(1) Any	(2) Serious	(3) Any	(4) Serious
Panel A: Linear				
Share of accused MLAs	-0.052	-0.479*	-0.012	0.080
	(0.191)	(0.275)	(0.090)	(0.129)
Female × share of accused MLAs	-0.261**	-0.272**	-0.042	-0.011
	(0.130)	(0.120)	(0.055)	(0.054)
Female	0.012	-0.027	0.016	0.003
	(0.049)	(0.027)	(0.021)	(0.013)
Effect for females	-0.312*	-0.751***	-0.054	0.069
	(0.189)	(0.278)	(0.090)	(0.121)
Mean dependent variable	0.405	0.405	0.313	0.313
Kleibergen–Paap F statistic	39.40	17.92	55.72	25.02
Observations	23,657	23,657	77,859	77,859
Panel B: Quadratic				
Share of accused MLAs	-0.102	-0.433	0.000	0.071
	(0.204)	(0.278)	(0.091)	(0.127)
Female × share of accused MLAs	-0.256**	-0.276**	-0.042	-0.010
	(0.129)	(0.120)	(0.055)	(0.054)
Female	0.010	-0.026	0.015	0.002
	(0.049)	(0.027)	(0.021)	(0.013)
Effect for females	-0.358*	-0.709**	-0.042	0.060
	(0.196)	(0.283)	(0.091)	(0.118)
Mean dependent variable	0.405	0.405	0.313	0.313
Kleibergen–Paap F statistic	27.71	16.57	56.54	24.77
Observations	23,657	23,657	77,859	77,859
Panel C: Cubic				
Share of accused MLAs	-0.137	-0.370	0.005	0.102
	(0.242)	(0.260)	(0.095)	(0.127)
Female × share of accused MLAs	-0.248*	-0.277**	-0.043	-0.011
	(0.129)	(0.120)	(0.055)	(0.054)
Female	0.007	-0.026	0.016	0.002
	(0.049)	(0.027)	(0.021)	(0.013)
Effect for females	-0.385*	-0.647**	-0.038	0.091
	(0.230)	(0.265)	(0.095)	(0.118)
Mean dependent variable	0.405	0.405	0.313	0.313
Kleibergen–Paap F statistic	17.40	20.10	49.20	27.04
Observations	23,657	23,657	77,859	77,859
Individual/household controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
NSS round fixed effects	Yes	Yes	Yes	Yes

Notes: The table re-estimates the 2SLS specification in Table 3 using alternative polynomial controls for the close-election vote margin. Panels A, B, and C use linear, quadratic, and cubic vote-margin controls, respectively; the close-election threshold remains 3 percentage points. All other treatment, instrument, timing, covariate, fixed-effect, weighting, and clustering choices follow the main specification. "Effect for females" is the sum of the coefficient on the share of accused MLAs and the Female × share interaction. The Kleibergen–Paap F statistic is reported for each specification. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.13: Alternative close-election thresholds

	Science stream		HS enrolment	
	(1) Any	(2) Serious	(3) Any	(4) Serious
Panel A: 2-percentage-point threshold				
Share of accused MLAs	-0.168	-0.941*	-0.202	0.045
	(0.276)	(0.513)	(0.126)	(0.184)
Female × share of accused MLAs	-0.356**	-0.299**	-0.006	0.013
	(0.181)	(0.151)	(0.076)	(0.068)
Female	0.047	-0.022	0.002	-0.003
	(0.068)	(0.033)	(0.029)	(0.016)
Effect for females	-0.524**	-1.239**	-0.208	0.058
	(0.248)	(0.495)	(0.135)	(0.161)
Mean dependent variable	0.405	0.405	0.313	0.313
Kleibergen–Paap F statistic	21.53	5.79	30.53	9.55
Observations	23,657	23,657	77,859	77,859
Panel B: 4-percentage-point threshold				
Share of accused MLAs	-0.110	-0.534**	0.068	0.162
	(0.205)	(0.266)	(0.102)	(0.118)
Female × share of accused MLAs	-0.269**	-0.298***	-0.031	-0.017
	(0.115)	(0.109)	(0.049)	(0.048)
Female	0.015	-0.022	0.011	0.004
	(0.044)	(0.025)	(0.019)	(0.012)
Effect for females	-0.379*	-0.832***	0.037	0.145
	(0.198)	(0.268)	(0.103)	(0.112)
Mean dependent variable	0.405	0.405	0.313	0.313
Kleibergen–Paap F statistic	20.33	19.17	28.95	30.23
Observations	23,657	23,657	77,859	77,859
Panel C: 5-percentage-point threshold				
Share of accused MLAs	0.008	-0.265	0.061	0.115
	(0.190)	(0.203)	(0.106)	(0.094)
Female × share of accused MLAs	-0.228**	-0.232**	-0.048	-0.059
	(0.108)	(0.108)	(0.046)	(0.046)
Female	-0.000	-0.036	0.018	0.013
	(0.042)	(0.025)	(0.018)	(0.012)
Effect for females	-0.220	-0.496**	0.013	0.057
	(0.188)	(0.207)	(0.107)	(0.089)
Mean dependent variable	0.405	0.405	0.313	0.313
Kleibergen–Paap F statistic	23.38	40.43	24.53	61.55
Observations	23,657	23,657	77,859	77,859
Individual/household controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
NSS round fixed effects	Yes	Yes	Yes	Yes

Notes: The table re-estimates the 2SLS specification in Table 3 using alternative definitions of a close election. Panels A, B, and C use vote-margin thresholds of 2, 4, and 5 percentage points, respectively, with the corresponding close-win instruments and close-election controls. All other treatment, timing, covariate, fixed-effect, weighting, and clustering choices follow the main specification. "Effect for females" is the sum of the coefficient on the share of accused MLAs and the Female × share interaction. The Kleibergen–Paap F statistic is reported for each specification. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.14: Reduced-form estimates

	Science stream		HS enrolment	
	(1) Any	(2) Serious	(3) Any	(4) Serious
Close-win accused share	-0.051	-0.307	-0.016	0.073
	(0.189)	(0.209)	(0.104)	(0.116)
Female × close-win accused share	-0.258**	-0.383**	-0.042	-0.013
	(0.131)	(0.153)	(0.054)	(0.068)
Female	-0.074***	-0.074***	0.002	0.000
	(0.012)	(0.012)	(0.006)	(0.005)
Mean dependent variable	0.405	0.405	0.313	0.313
Observations	23,657	23,657	77,859	77,859
Individual/household controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
NSS round fixed effects	Yes	Yes	Yes	Yes

Notes: The table reports reduced-form regressions in which the outcomes are regressed directly on the close-win accused share and its interaction with the female indicator. The outcome definitions, estimation samples, close-contest incidence and vote-margin controls, individual and household covariates, district and NSS survey-round fixed effects, and sampling weights follow the main specification. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.15: OLS estimates

	Science stream		HS enrolment	
	(1) Any	(2) Serious	(3) Any	(4) Serious
Share of accused MLAs	-0.015	-0.024	-0.040*	-0.038
	(0.057)	(0.062)	(0.024)	(0.029)
Female × share of accused MLAs	-0.141***	-0.156***	0.020	0.019
	(0.040)	(0.047)	(0.020)	(0.023)
Female	-0.034**	-0.052***	-0.008	-0.005
	(0.017)	(0.014)	(0.008)	(0.007)
Mean dependent variable	0.405	0.405	0.313	0.313
Observations	23,657	23,657	77,859	77,859
Individual/household controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
NSS round fixed effects	Yes	Yes	Yes	Yes

Notes: The table reports OLS estimates of the baseline relationship between the district-level share of accused MLAs and the two educational outcomes, without instrumenting the accused share or its interaction with the female indicator. The outcome definitions and estimation samples match the main analysis. Specifications include the individual and household covariates, district fixed effects, and NSS survey-round fixed effects used in the baseline model. NSS sampling weights are used and standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.16: Robustness with state × NSS-round fixed effects

	Science stream		HS enrolment	
	(1) Any	(2) Serious	(3) Any	(4) Serious
Share of accused MLAs	-0.071	-0.433	-0.034	0.141
	(0.226)	(0.291)	(0.100)	(0.135)
Female × share of accused MLAs	-0.267**	-0.278**	-0.043	-0.012
	(0.132)	(0.125)	(0.055)	(0.054)
Female	0.014	-0.027	0.016	0.003
	(0.050)	(0.028)	(0.021)	(0.013)
Effect for females	-0.338	-0.711**	-0.077	0.129
	(0.225)	(0.296)	(0.099)	(0.126)
Mean dependent variable	0.405	0.405	0.313	0.313
Kleibergen–Paap F statistic	20.48	15.38	49.24	23.52
Observations	23,657	23,657	77,859	77,859
Individual/household controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
State × NSS-round fixed effects	Yes	Yes	Yes	Yes

Notes: The table re-estimates the 2SLS specification in Table 3 while absorbing state × NSS-round fixed effects in addition to district fixed effects. All other treatment, instrument, timing, covariate, weighting, and clustering choices follow the main specification. "Effect for females" is the sum of the coefficient on the share of accused MLAs and the Female × share interaction. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.17: Unweighted main specification

	Science stream		HS enrolment	
	(1) Any	(2) Serious	(3) Any	(4) Serious
Share of accused MLAs	-0.026	-0.370	-0.030	0.072
	(0.145)	(0.227)	(0.064)	(0.088)
Female × share of accused MLAs	-0.189**	-0.257***	0.034	-0.041
	(0.078)	(0.071)	(0.033)	(0.035)
Female	0.002	-0.015	0.024**	0.043***
	(0.027)	(0.014)	(0.012)	(0.008)
Effect for females	-0.215	-0.626***	0.003	0.031
	(0.143)	(0.230)	(0.062)	(0.080)
Mean dependent variable	0.405	0.405	0.313	0.313
Kleibergen–Paap F statistic	33.16	19.81	75.70	26.38
Observations	23,657	23,657	77,859	77,859
Individual/household controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
NSS round fixed effects	Yes	Yes	Yes	Yes

Notes: The table re-estimates the 2SLS specification in Table 3 without NSS sampling weights. All treatment, instrument, timing, covariate, fixed-effect, and clustering choices are otherwise unchanged. "Effect for females" is the sum of the coefficient on the share of accused MLAs and the Female × share interaction. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.18: Future-election placebo

	Science stream		HS enrolment	
	(1) Any	(2) Serious	(3) Any	(4) Serious
Future share of accused MLAs	-0.089	-0.115	0.069	0.006
	(0.216)	(0.223)	(0.115)	(0.099)
Female × future share of accused MLAs	0.120	-0.031	0.039	0.034
	(0.172)	(0.124)	(0.102)	(0.086)
Female	-0.139*	-0.078**	-0.017	-0.009
	(0.079)	(0.037)	(0.044)	(0.023)
Effect for females	0.031	-0.146	0.107	0.040
	(0.197)	(0.207)	(0.088)	(0.078)
Mean dependent variable	0.407	0.407	0.312	0.312
Kleibergen–Paap F statistic	25.25	44.42	30.19	58.69
Observations	23,186	23,186	76,606	76,606
Individual/household controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
NSS round fixed effects	Yes	Yes	Yes	Yes

Notes: The table re-estimates the 2SLS specification in Table 3, replacing the baseline political exposure with the first state assembly election held after completion of the relevant NSS round. The future share of accused MLAs and its interaction with the female indicator are instrumented by the corresponding future close-win accused share and its interaction with the female indicator, using a 3-percentage-point close-election threshold. All other covariate, fixed-effect, weighting, and clustering choices follow the main specification. "Effect for females" is the sum of the coefficient on the future share of accused MLAs and the Female × future share interaction. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.19: Age-range robustness: expanding the main sample to ages 15–20

	Science stream		HS enrolment	
	(1) Any	(2) Serious	(3) Any	(4) Serious
Share of accused MLAs	-0.032	-0.378	0.028	0.081
	(0.178)	(0.262)	(0.071)	(0.099)
Female × share of accused MLAs	-0.222*	-0.240**	-0.049	0.006
	(0.124)	(0.113)	(0.054)	(0.048)
Female	-0.004	-0.035	0.010	-0.009
	(0.047)	(0.026)	(0.020)	(0.012)
Effect for females	-0.254	-0.618**	-0.020	0.087
	(0.176)	(0.265)	(0.074)	(0.094)
Mean dependent variable	0.394	0.394	0.248	0.248
Kleibergen–Paap F statistic	48.33	17.61	57.69	24.27
Observations	25,458	25,458	106,023	106,023
Individual/household controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
NSS round fixed effects	Yes	Yes	Yes	Yes

Notes: The table re-estimates the 2SLS specification in Table 3 after expanding the main age range from 15–18 to 15–20. The science-stream outcome is estimated among higher-secondary students in the expanded age range and the higher-secondary enrolment outcome uses the full expanded sample. Political exposure remains measured one year before the NSS round and close elections use the 3-percentage-point threshold; the remaining treatment, instrument, covariate, fixed-effect, weighting, and clustering choices follow the main specification. "Effect for females" is the sum of the coefficient on the share of accused MLAs and the Female × share interaction. Standard errors are clustered at the district level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.20: Unconditional participation in higher-secondary science

	(1) Any accusation	(2) Serious accusation
Share of accused MLAs	0.010	-0.097
	(0.060)	(0.081)
Female × share of accused MLAs	-0.091**	-0.078**
	(0.039)	(0.038)
Female	0.011	-0.006
	(0.015)	(0.009)
Effect for females	-0.081	-0.175**
	(0.060)	(0.082)
Mean dependent variable	0.123	0.123
Kleibergen–Paap F	55.72	25.02
SW F: accused-share equation	136.89	113.01
SW F: interaction equation	108.60	122.98
Observations	77,859	77,859
Individual/household controls	Yes	Yes
District fixed effects	Yes	Yes
NSS round fixed effects	Yes	Yes

Notes: The dependent variable equals one for individuals aged 15–18 who are currently enrolled at the higher-secondary level and study science, and zero otherwise in the full age-15–18 sample. The table otherwise follows the 2SLS specification in Table 3, including the one-year-lagged political exposure, 3-percentage-point close-election threshold, close-election controls, individual and household covariates, district and NSS survey-round fixed effects, and sampling weights. Standard errors are clustered at the district level. "Effect for females" is the sum of the coefficient on the share of accused MLAs and the Female × share interaction. The table reports the system Kleibergen–Paap F statistic and the Sanderson–Windmeijer conditional F statistic for each endogenous regressor. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.