Measuring Gender and Religious Bias in the Indian Judiciary

Working Paper

Author(s):
Ash, Elliott; Asher, Sam; Bhowmick, Aditi; Chen, Daniel L.; Devi, Tanaya; Goessmann, Christoph; Novosad, Paul; Siddiqi, Bilal

Publication date:
2021-01

Permanent link:
https://doi.org/10.3929/ethz-b-000472802

Originally published in:
Center for Law & Economics Working Paper Series 2021(03)

ETH zürich

Center for Law & Economics Working Paper Series

Number 03/2021

Measuring Gender and Religious Bias in the Indian Judiciary

Elliott Ash, Sam Asher, Aditi Bhowmick,
Daniel Chen, Tanaya Devi, Christoph Goessmann,
Paul Novosad, Bilal Siddiqi

January 2021

Measuring Gender and Religious Bias in the Indian Judiciary

Elliott Ash, Sam Asher, Aditi Bhowmick,
Daniel Chen, Tanaya Devi, Christoph Goessmann,
Paul Novosad, Bilal Siddiqi*

January 11, 2021

Abstract

We study judicial in-group bias in Indian criminal courts, collecting data on over 80 million legal case records from 2010–2018. We exploit quasi-random assignment of judges and changes in judge cohorts to examine whether defendant outcomes are affected by being assigned to a judge with a similar religious or gender identity. We estimate tight zero effects of in-group bias. The upper end of our 95% confidence interval rejects effect sizes that are one-fifth of those in most of the prior literature.

JEL codes: J15, J16, K4, O12

1 Introduction

Structural inequalities across groups defined by gender, religion, and ethnicity exist in almost all societies. Governments often try to remedy these inequalities through

*Author Details: Elliott Ash, ETH Zurich: ashe@ethz.ch; Sam Asher, John Hopkins: sasher2@jhu.edu; Aditi Bhowmick, Princeton: aditib@princeton.edu; Daniel L. Chen, Toulouse and World Bank: daniel.chen@iast.fr; Tanaya Devi, Harvard: tdevi@g.harvard.edu; Christoph Goessmann, ETH Zurich: christoph.goessmann@gess.ethz.ch; Paul Novosad, Dartmouth: pnovosad@dartmouth.deu; Bilal Siddiqi, UC Berkeley: bilal.siddiqi@berkeley.edu. We thank Alison Campion, Rebecca Cai, Nikhitha Cheeti, Kritarth Jha, Romina Jafarian, Ornelie Manzambi, Chetana Sabnis, and Jonathan Tan for helpful research assistance. A special thanks to Sandeep Bhupatiraju for contributions in preparation of the data. We thank the World Bank Program on Data and Evidence for Justice Reform and the UC Berkeley Center for Effective Global Action for financial support.

policies, such as anti-discrimination statutes or affirmative action, which must then be enforced by the legal system. A challenging problem is that the legal system itself may have unequal representation. It remains an open question whether legal systems in developing countries are effective at pushing back against structural inequality or whether they serve to entrench it (e.g. Aldashev et al., 2012).

This paper examines bias in India's lower courts, asking whether judges deliver more favorable treatment to defendants who match their identities. Judicial bias along gender, religious, or ethnic lines appears to be nearly universal in richer countries, having been identified in a wide range of settings around the world.[1] However, it has not been widely studied in the courts of lower-income countries. In-group bias of this form has been identified in other contexts in India, such as among loan officers (Fisman et al., 2020) and school-teachers (Hanna and Linden, 2012). The judicial setting is of particular interest, given the premise that individuals who are discriminated against in informal settings should receive equal treatment under the law (Sandefur and Siddiqi, 2015).

We focus on the dimensions of gender and religion in India's lower courts, where unequal representation is a recognized issue. Women represent 48% of the Indian population but only 28% of district court judges. Similarly, India's 200 million Muslims represent 14% of the population but only 7% of lower court judges. There is growing evidence that India's Muslims and women do not enjoy equal access to economic or other opportunities (Ito, 2009; Bertrand et al., 2010; Hnatkovska et al., 2012; Hanna and Linden, 2012; Jayachandran, 2015; Borker, 2017; Asher et al., 2020). We examine whether unequal representation in the courts has a direct effect on the judicial outcomes of Muslims and women, in the form of judges delivering better outcomes to criminal defendants who match their gender or religion.

Our analysis draws upon a new dataset of 80 million court records covering 2010–2018 from India eCourts, an online platform documenting the complete set of cases heard in India's district courts.These cases cover the universe of India's 7,000+ district and subordinate trial courts, staffed by over 80,000 judges. We are releasing anonymized data on these cases, opening the door to many new analyses of the judicial process in the world's largest democracy and largest common-law legal system.

We enrich the dataset by classifying judges and defendants to gender and religious (Muslim and non-Muslim) identity groups based on their names. An automated pro-

[1] See, for example, Shayo and Zussman (2011), Didwania (2018), Arnold et al. (2017), Abrams et al. (2012), Alesina and La Ferrara (2014), and others below.

cess uses a deep neural network applied to the sequence of characters in names. The distinctive nature of female and Muslim names allows us to classify individuals with over 97% out-of-sample accuracy on both dimensions.[2]

The main research question is whether judges tend to treat defendants differently when they share the same gender or religion. We focus on the subset of cases filed under India's criminal codes ($N = 5.5$ million), where acquittal and conviction rates, as well as judicial delay, are readily interpretable as positive or negative outcomes. We implement two different identification strategies to generate causal estimates of how judge identity affects a defendant's outcome.

First, we exploit the arbitrary rules by which cases are assigned to judges, generating as-good-as-random variation in judge identity. Our preferred specification includes court, charge, and month-year fixed effects. Effectively, we compare the outcomes of two defendants with the same identity classification, charged under the same criminal section, in the same court and in the same month, but who are assigned to judges with different identities.

Second, we exploit judicial turnover events that change the gender and religion balance of judges serving in a district court, exogenously changing the probability that a defendant matches identity with the judge overseeing their case. We use a regression discontinuity specification which measures the difference in judicial outcomes for defendants whose cases are heard immediately before and immediately after a transition that makes the bench more or less similar along identity dimensions.

In both of these specifications, we find a robust null estimate of in-group bias among Indian judges. Judges of different genders do not treat defendants differently according to their gender, nor do judges display favoritism on the basis of religion. This is true both in terms of outcomes (i.e. acquittals and convictions) and in terms of process (i.e. speed of decision). In a subset of specifications, we find a very small in-group gender bias, which is marginally positive and not robust. However, the size of this effect, even in the marginally significant specifications, is an order of magnitude smaller than nearly all prior estimates of in-group bias based on similar identification strategies in the literature.[3] The upper end of our 95% confidence interval rejects a 0.7 percentage point effect size in the worst case; studies using the same identification strategies in

[2]We do not examine bias on the dimensions of income or caste because we do not yet have an algorithm that can classify these dimensions with high accuracy.

[3]The exception is Lim et al. (2016), who find zero effects of in-group gender bias and marginal effects of in-group racial bias among judges in Texas state district courts.

other contexts have routinely found bias effects ranging from 5 to 20 percentage points.

Our estimates do not rule out bias in the Indian legal system entirely; we observe only a subset of the legal process and we measure only in-group bias by gender and religion. For example, it is possible that both Muslim and non-Muslim judges discriminate against Muslims (as found for Black defendants in Arnold et al. (2017)). It is also possible that arrests and/or charges disproportionately target Muslims, or that judges exhibit bias based on defendant caste or income. However, the bias that we study has been widely reported in other studies with large effect sizes, and the public discussion of discrimination against Muslims and women in India in many ways parallels discussion of marginalized groups in other countries.

Relative to the prior literature, we make several contributions. First, we demonstrate an absence of bias in an important context with substantial religious and gender cleavages. Second, the sample of our study is an order of magnitude larger than earlier studies, allowing us to measure bias much more precisely than prior work. Third, to our knowledge this is the first large-scale study of judicial bias in a low- or middle-income country and it makes available a dataset which may have substantial utility to future scholars.

These results add to a literature on biased decision-making in the legal system. Most prior work is on the U.S. legal system, where disparities have been documented at many levels.[4] The closest paper to ours is Shayo and Zussman (2011), who analyze the effect of assigning a Jewish versus an Arab judge in Israeli small claims court. They find robust evidence of in-group bias, where Jewish judges favor Jewish defendants (and Arab judges favor Arab defendants) by an average 17–19 percentage-point margin, an effect ten times larger than the upper bound of our confidence interval on either religion or gender bias. Several more studies use one of our two identification strategies

[4]These include racial disparities in the execution of stop-and-frisk programs (Goel et al., 2016), motor vehicle searches by police troopers (Anwar and Fang, 2006), bail decisions (Arnold et al., 2017), charge decisions (Rehavi and Starr, 2014), and judge sentence decisions (Mustard, 2001; Abrams et al., 2012; Alesina and La Ferrara, 2014; Kastellec, 2013). African-American judges have been found to vote differently from Caucasian-American judges on issues where minorities are disproportionately affected, such as affirmative action, racial harassment, unions, and search and seizure cases (Scherer, 2004; Chew and Kelley, 2008; Kastellec, 2011). In a similar manner, a number of papers have documented the effect of judges' gender in sexual harassment cases (Boyd et al., 2010; Peresie, 2005). A smaller set of papers use information on both the identity of the defendant and the decision-maker. Anwar et al. (2012) look at random variation in the jury pool and find that having a black juror in the pool decreases conviction rates for black defendants. A similar result from Israel is documented by Grossman et al. (2016), who find that the effect of including even one Arab judge on the decision-making panel substantially influences trial outcomes of Arab defendants. Didwania (2018) find in-group bias in that prosecutors charge same-gender defendants with less severe offenses.

to generate point estimates that are directly comparable to ours, and of these only Lim et al. (2016) find a null in-group effect of judge ethnicity or gender.[5]

In the Indian context, there is a growing body of evidence on the legal system, mostly focusing on judicial efficacy and economic performance (Chemin, 2009; Rao, 2019), or on corruption in the Indian Supreme Court (Aney et al., 2017). A recent working paper finds that judges are more prone to deny bail if they were exposed to communal riots in their early childhood (Bharti and Roy, 2020). However, we are aware of no prior large-scale empirical research on unequal legal treatment on either the gender or religion dimension in India, a topic of substantial policy relevance.

Beyond the issue of in-group bias, we add to the growing literature on courts in developing countries. Well-functioning courts are widely considered a central component of effective, inclusive institutions, with judicial equity and rule of law seen as key indicators of a country's institutional quality (Rodrik, 2000; Le, 2004; Rodrik, 2005; Pande and Udry, 2006; Visaria, 2009; Lichand and Soares, 2014; Ponticelli and Alencar, 2016; World Bank, 2017). A handful of important cross-country studies have recovered some broad stylized facts on the causes and consequences of different broad features of legal systems (Djankov et al., 2003; La Porta et al., 2004, 2008). But largely due to a lack of data, there has been a relative paucity of within-country court- or case-level research on the delivery of justice in lower-income settings.

The rest of the paper is organized as follows. After outlining the institutional context (Section 2) and data sources (Section 3), we articulate our empirical approach (Section 4). Section 5 reports the results. Section 6 compares the results to the previous literature and concludes.

2 Background

2.1 Institutional Context

India's population is characterized by cross-cutting divisions between gender and religion. Women's rights and their status in society are under intense political debate.

[5] Gazal-Ayal and Sulitzeanu-Kenan (2010) find positive in-group bias in bail decisions when Arab and Jewish defendants are randomly assigned to a judge of the same ethnicity. Knepper (2018) and Sloane (2019) leverage random assignment of cases in the U.S. to judges and prosecutors respectively, finding significant in-group bias in trial outcomes. Depew et al. (2017) exploit random assignment of judges to juvenile crimes in Louisiana and find *negative* in-group bias in sentence lengths and likelihood of being placed in custody. It is notable that of all these studies, Lim et al. (2016) has one of the largest sample sizes (N=250,000).

Muslims in India (14% of the population) have historically had intermediate socioeconomic outcomes worse than upper caste groups but better than lower caste groups. However, they have been protected by few of the policies and reservations targeted to Scheduled Castes and Tribes. In recent decades, many successful political parties have been accused of implicitly or explicitly discriminating against Muslims.

Women constitute 48% of the population, and remain vulnerable to precarious social practices such as female infanticide, child marriage, and dowry deaths despite existing legislation outlawing all of the above. Prior to the pandemic, India accounted for one-third of all child marriages globally (Cousins, 2020). As of 2020, India also accounts for nearly one-third of the 142.6 million missing females in the world (Erken et al., 2020). The unambiguously marginalized status of Indian women and Muslims motivates the exploration of the role of gender and religion in the context of India's criminal justice system in this study.

India's judicial system is organized in a jurisdictional hierarchy that is similar to other common-law systems. There is a Supreme Court, 25 state High Courts, and 672 district courts below them. Beneath the district courts, there are about 7000 subordinate courts. The district courts and subordinate courts collectively constitute India's lower judiciary. These courts represent the preliminary point of entry of almost all criminal cases in India.[6]

These courts are staffed by over 81,000 judges. Due to common law institutions where court rulings serve as binding precedent in future cases, judges in India are important policymakers. Indian judges are arguably even more powerful than their U.S. counterparts because they do not share decision authority with juries, which were banned in 1959. Therefore fair and efficient decision-making by judges is an important issue for governance.

There is an active debate in India around reforming the court system. Problems under discussion include a reputation for corruption (Dev, 2019) as well as a substantial backlog of cases (Trusts, 2019). In 2015, Prime Minister Modi attempted to implement a series of reforms giving his administration more control over judge selection through the creation of a National Judicial Appointments Commission. However, the effort to move away from the collegium system of judicial appointment was reversed by the Supreme Court, citing breach of judicial independence.

[6]We define criminal cases as all cases filed either under the Indian Penal Code Act or the Code of Criminal Procedure Act.

2.2 Case Assignment To Judges

The procedure of case assignment to judges is important for this study, because our main empirical strategy hinges on exogenous assignment of judges to cases. To better understand the case assignment process, we consulted with several criminal lawyers who practice in India's district courts, senior research fellows at the Vidhi Center for Legal Policy, as well as a number of working court clerks in courts around the country.

Criminal cases are assigned to judges as follows. First, a crime is reported at a particular local police station, where a First Information Report (FIR) is filed. Each police station lies within the territorial jurisdiction of a specific district courthouse, which will receive the case. The case will then be assigned to a judge sitting in that courthouse. If there is just one judge working there, that judge will get the case.

When there are multiple judges, a rules-based process fully determines the judge assignment. Each judge sits in a specific courtroom in a court for several months at a time. A courtroom is assigned for every police station and every charge. For example, at a given police station, every murder charge will go to the same courtroom; a larceny charge might go to a different courtroom, as might a murder charge reported at a different police station. Judges typically spend two to three years in a given court, during which they rotate through several of the courtrooms.[7]

The police station-charge lists thus leave little discretion for charges to be seen by specific judges. Since the timing of the first course appearance is unknown when charges are filed (given judicial delays), even if a defendant or prosecutor had discretion over which police station filed the charges, the rotation of judges between courtrooms would make it difficult to target a specific judge. Finally, the judiciary explicitly condemns the practice of "judge shopping" or "forum shopping", where litigants select particular judges seeking a favorable outcome. One of the earliest cases in which the Indian Supreme Court condemned the practice of shopping is the case of *M/s Chetak Construction Ltd. v. Om Prakash & Ors.*, 1998(4) SCC 577, where the Court ruled against a litigant trying select a favorable judge, writing that judge shopping "must be crushed with a heavy hand." This decision has been cited heavily in subsequent judgments.

Finally, it should be noted that in the most recent years (since 2013), some courts have adopted a random assignment lottery mechanism implemented through the eCourts platform, making judge selection very unlikely. The eCourts assignment mechanism

[7] Severe cases (with severity defined by the section or act under which the charge was filed) require judges with higher levels of seniority; thus a case in a given district in some cases may be seen only by a subset of judges in that district.

was intended to be used throughout the country but in practice it has not been widely adopted to date. In Section 4, we present formal tests of the exogenous assignment of judges to cases in our dataset.

3 Data

3.1 Case Records

We obtained 81.2 million case records from the Indian eCourts platform – a semi-public system put in place by the Indian government as a "national data warehouse for case data including the orders/judgments for courts across the country."[8] The publicly available information includes the filing, registration, hearing, and decision dates for each case, as well as petitioner and respondent names, the position of the presiding judge, the acts and sections under which the case was filed, and the final decision or disposition.[9]

The database covers India's lower judiciary – all courts including and under the jurisdiction of District and Sessions courts. In this paper, we focus on cases filed either under the Indian Penal Code or the Code of Criminal Procedure for two reasons. First, there is only a single litigant, rather than two, providing a clear definition of identity match between judge and defendant. Second, it is relatively straightforward to identify good and bad outcomes for criminal defendants, and much more difficult to do so for litigants in civil cases. This constraint filters out 69% of the dataset, leaving us with 25.2 million criminal case records.

3.2 Judge Information

We also obtained data on judges pertaining to all courts in the Indian lower judiciary from the eCourts platform. The data for each judge includes the judge's name, their position or designation, and the start and end date of the judge's appointment to each court.[10]

We joined the case-level data with the judge-level data based on the judge's designation and the initial case filing date. In this process, another 17% of the initial

[8] https://ecourts.gov.in/ecourts_home/static/about-us.php, accessed Oct 14 2020

[9] We illustrate such a record in Appendix Figure A1.

[10] See Appendix Figure A2 for a sample page from which we extract the judge data. The data does not include the room in the court to which a judge is assigned.

observations are dropped. The remaining dataset where cases are linked to a unique judge consists of 11.0 million cases. Further, we drop all cases where both judge gender and religion could not be deduced. We also drop cases where both defendant gender and religion cannot be inferred from the information available. The analysis dataset for the randomized case assignment experiment approach consists of 8.0 million cases.

For our alternative event study empirical approach, we joined cases to courts based on the court location and decision date associated with each case. The resulting analysis dataset for this approach comprises of 17 million cases – 68% of the initial universe of criminal cases.

3.3 Assigning Religion and Gender Identity

The eCourts platform does not provide demographic metadata on judges and defendants. However, gender and religious identity can be determined quite accurately in India based on individuals' names. We train a machine classifier on a large database of labeled names and then use it to assign these characteristics in the legal data.

We have access to two databases of names with associated demographic labels. First, to classify gender, we use a dataset of 13.7 million names with labeled gender from the Delhi voter rolls. Second, to classify religion, we use a database of 1.4 million names with a religion label for individuals who sat for the National Railway Exam.

Summary tabulations on these datasets are provided in Appendix Table A1. For gender, we observe two categories: female or male. For religion, we observe five categories: Hindu, Muslim, Christian, Buddhist, and Other. Our classifier takes a two-label specification: Muslim or non-Muslim. We do not distinguish between the non-Muslim religion categories because of their small number and because their names are not as distinctive as Muslim names. Each name record is therefore assigned two binary labels: male/female, and Muslim/Non-Muslim.

Before applying the classifier, we pre-process the name strings by transliterating characters from Hindi to Latin, and normalizing capitalization, punctuation and spacing. We then apply a neural net classifier to predict the identity label based on the name string, similar to the approach in Chaturvedi and Chaturvedi (2020). We use a bidirectional Long Short-Term Memory (LSTM) model applied directly to the sequence of name-string characters. LSTMs are a gated recurrent neural network architecture that takes as input sequential data and retains memory of previous inputs as it handles new items in a sequence. LSTMs are particularly useful in understanding text sequences because the meaning of an individual letter or word is often dependent on the context of other letters and words that both precede and follow it. "Bidirectional" means that the classifier reads the sequence both backwards and forwards when trying to assign a label.[11]

We use hold-out test sets within the labeled databases to assess the out-of-sample performance of the LSTM classifiers for gender and religion. The classifiers perform well along the standard metrics, including our preferred metrics which adjust for imbalance in the class shares. We report balanced accuracy, the average accuracy (recall) for each of the two identity categories, and F1, the harmonic mean of precision and recall.[12] For gender, the balanced accuracy is .975 with F1 = .976. For religion, the balanced accuracy is .98 and the F1 = .99.

We then apply the trained classifier to the eCourts case records. The judge names tend to be complete (first and last name) and often include salutations indicating gender. Our algorithm can classify the names of 96% of the 81,232 judges (22,413 unique names) appearing in the case dataset according to gender (female/male), and 98% according to religion (Muslim/non-Muslim). The information on litigant names is of lower quality, often missing either the first name or last name. We are able to classify 80% of litigants by religion, and 74% by gender. Cases with unclassified labels are dropped from analyses requiring those labels.

To verify the accuracy of the LSTM classification within the new domain of the court records, we conduct a manual verification of random subsets of names classified by gender and religion, stratified across all states. We can confirm an accuracy rate of 97% for both the gender and religion classification based on manual verification. As an additional validation step, we compare the LSTM-classified Muslim defendant share

[11] The neural net architecture is as follows. The model takes as input a sequence of characters and outputs a probability distribution across name classes. The characters are input to an embedding layer, which was initialized randomly rather than using pre-trained weights. The embedded vectors are input to a bidirectional LSTM layer, then to a single dense hidden layer, and finally to the output layer, which uses sigmoid activation to output a probability across the binary classes. To avoid overfitting, we used dropout between layers and used early stopping during training, which ceases network training when validation loss stops improving. To account for the imbalance in the sample, we used class weights during the training.

[12] Balanced accuracy and F1 are preferred as metrics to standard accuracy when the labels to be predicted are not balanced. While gender is roughly balanced in the voter rolls data, religion is heavily imbalanced with Muslims only comprising one-tenth of the sample. Therefore a model could achieve 90% accuracy in predicting religion by guessing non-Muslim. Balanced accuracy addresses this issue by rewarding good accuracy for both classes: we calculate the accuracy for each class and then average, rather than taking the accuracy measure across the whole sample. F1 addresses this issue by rewarding higher precision, which penalizes false positives, and higher recall, which penalizes false negatives.

by state to the state-wise Muslim population shares reported in the 2011 Population Census, and can show they are highly correlated (Appendix Figure A3).

3.4 Case Outcome Specification

We define the defendant's outcome (represented by Y below) as a case-level indicator variable that takes the value 1 if the outcome is desirable for the defendant. Our primary specification uses an indicator for defendant acquittal. A secondary specification uses an indicator for any outcome other than conviction. Unfortunately, there are many cases where eCourts does not provide a clear indication of whether the outcome is desirable. For instance, a case outcome may be described in the metadata as "disposed," , with no additional judgment information uploaded for the case. For a case like this, we define the outcome as neither acquitted nor convicted; that is, the positive outcome variable takes the value of 0 when Y=*acquitted*, and the value of 1 when Y=*not convicted* (Table A2). About 60% of case dispositions can be clearly designated as good or bad, while 40% are ambiguous; we show that our results are robust to alternate definitions of positive outcomes.

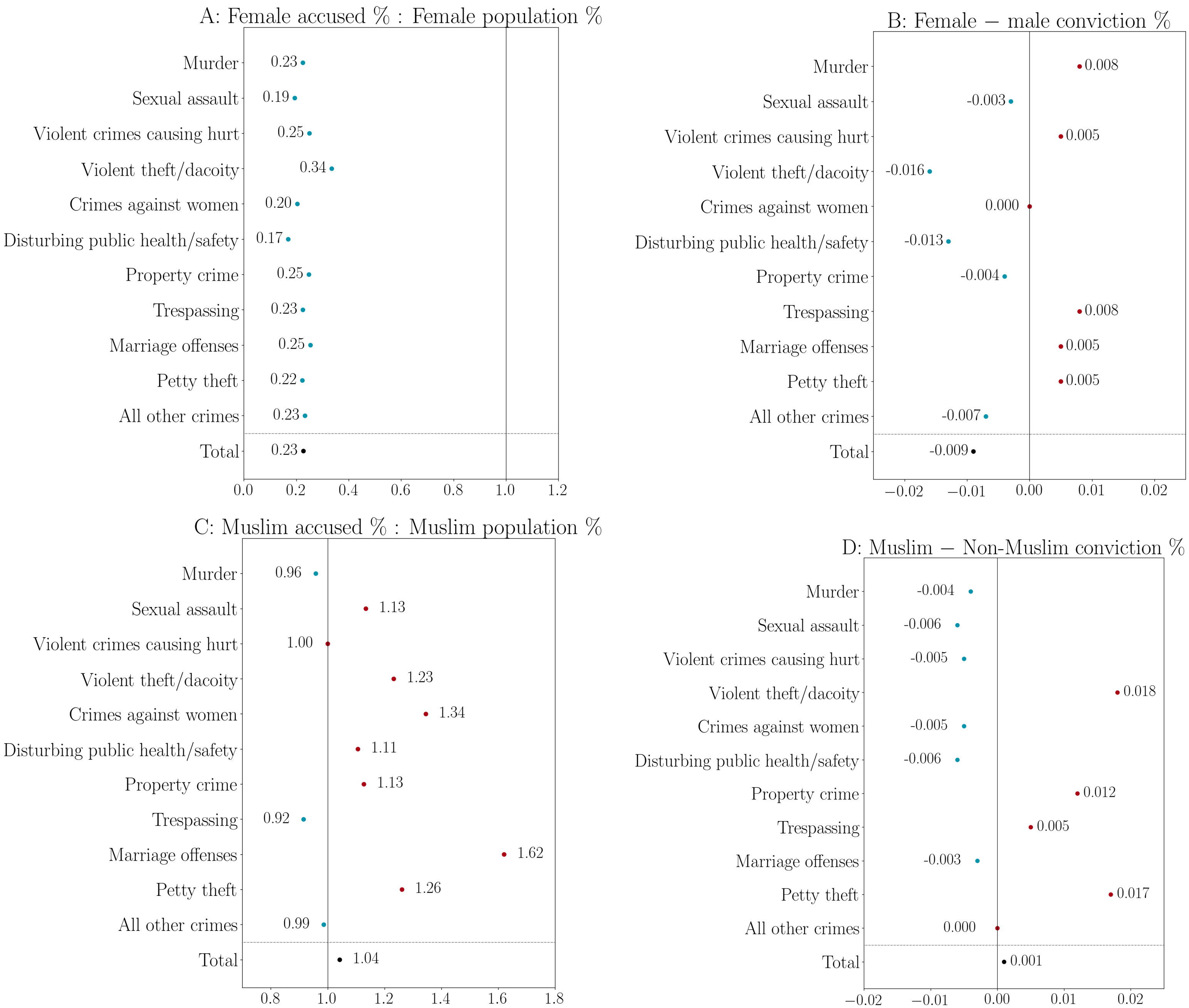
In about 40% of cases, the judge presiding over the initial case filing does not reach a decision; a decision is reached by a future judge or else the case remains undecided. Our analysis is focused entirely on the first judge to see the case; because decision deferral may be endogenous, we cannot treat the assignment of the second judge as random. Judicial delay is also a major policy issue in India; getting a decision at all is therefore an outcome of interest in and of itself. We define a variable *decision* as an indicator for whether the first judge to preside over the case reaches a decision on the case. We discuss our treatment of cases that pass to other judges in Section 4. In our main analysis, we exclude cases that were not decided at all, but our findings are robust to alternate choices.

Table 1: Coding of outcome variables

Outcome in e–Courts Data	**1(Any Decision)**	**1(Acquitted)**	**1(Not Convicted)**
No decision by any judge	0	–	–
Acquitted	1	1	1
Neither acquitted nor convicted	1	0	1
Convicted	1	0	0

Notes: The outcome variables were coded based on the trial outcome recorded in the disposition variable associated with each case record. Under any of the three trial outcome definitions, a value of 1 always represents a positive outcome.

Figure 1: Summary statistics by crime category and defendant identity



Notes: Panels A & C show the ratio of share of accused Muslim or female over the population share of Muslims or females respectively, for each crime category. Panels B & D show the difference in mean conviction rates between defendant groups within crime categories

3.5 Summary Statistics on Case Outcomes

Figure 1 presents descriptive statistics of charges and convictions by gender and religious identity of defendants, respectively.[13] These summary measures are descriptive in nature, and are not directly informative of bias in the judicial system because we do not know the share of Muslim and female defendants who commit crimes, or who are in fact guilty upon being charged with crimes.

Figure 1 Panel A shows that the share of females charged under all crime categories is substantially lower than their population share. Men are three to five times more likely to be charged with crimes under any classification. Panel B shows that the conviction rate varies by crime, but overall it is about 1 percentage point lower for women (the "Total" category, at the bottom). Crimes are ordered by maximal punishment, from most to least severe.

Panel C shows that Muslims are disproportionately represented in the universe of criminal charges for most offenses. In particular, they are 34% more likely to be charged with crimes against women, 23% more likely to be charged with robbery, and 62% more likely to be charged with marriage offenses. Muslims are less likely to face charges for murder. In Panel D, we see no aggregate differences for Muslims in conviction rates, although these vary across crime types. Conditional on being charged, Muslim defendants are substantially more likely to be convicted than non-Muslims with robbery, property crime, and theft, but less likely to be convicted of obscenity, murder, or crimes against women.

Table 2 shows descriptive statistics of judges in the analysis sample. About 28% of judges are female, and 7.5% of judges are Muslim. On average, Muslim and female judges have similar conviction and decision rates to non-Muslim and male judges.

4 Empirical Strategy

Our objective is to estimate whether defendants experience different outcomes depending on the identity of the judge presiding over their case. To estimate a causal effect of judge identity, we need to effectively control for any factors other than defendant identity that could affect both judge identity and the case outcome. For instance, if female judges see less severe cases on average, and less severe cases have different conviction rates, we do not want to attribute that difference to a female judge effect. Similarly,

[13]The corresponding point estimates are reported in Appendix Tables A3 and A4.

Table 2: Outcome probability, by judge identity

		Judge gender		Judge religion	
	(1)	(2)	(3)	(4)	(5)
	Total	Female	Male	Muslim	Non-Muslim
Female judge	0.285	1.000	0.000	0.277	0.283
	(0.002)	(0.000)	(0.000)	(0.009)	(0.003)
Muslim judge	0.073	0.073	0.075	1.000	0.000
	(0.001)	(0.003)	(0.002)	(0.000)	(0.000)
Tenure length	487.940	488.574	494.013	473.531	490.079
	(2.386)	(4.548)	(2.849)	(9.077)	(2.482)
Decisions					
Decision (given first filing)	0.591	0.589	0.587	0.605	0.590
	(0.002)	(0.003)	(0.002)	(0.007)	(0.002)
Acquitted	0.236	0.231	0.242	0.237	0.236
	(0.002)	(0.003)	(0.002)	(0.006)	(0.002)
Convicted	0.055	0.067	0.049	0.062	0.053
	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)
N	34,911	9,567	23,981	2,526	31,902

Notes: Coefficients represent means for each variable in the sample, collapsed to the judge level. Standard errors have been reported in parentheses.

Muslim defendants and judges may be more predominant in parts of the country with different base conviction rates.

We use two empirical strategies to isolate the causal effect of judge identity. First, we rely on the exogenous assignment of judges to cases, which produces as-good-as-random assignment of defendants to judges, conditional on charge and district. Second, we use a regression discontinuity design to exploit changes in the staffing of judges sitting in a given court, which creates exogenous changes in the likelihood of judge-defendant identity matches. These different identification strategies also have largely different samples, because random assignment is most relevant in large courts, while staffing changes are most likely to substantially affect the identity composition of small courts.

We formalize each approach in the following subsections. For ease of exposition, we describe the empirical strategy investigating gender bias; the specification and considerations for estimating religious identity bias are identical.

4.1 Random Assignment of Judges to Cases

As with much of the prior empirical literature, judge assignment in district courts is not strictly random, but follows a set of rules that gives defendants and prosecutors virtually no control over which judge oversees the case. As described in Section 2, a case is assigned to a room in a court (and thus the judge in that room) when charges are filed, based on the police station and charge type, giving prosecutors and defendants little control over the judge's identity. From a defendant's perspective, the judge assignment is as good as random; for simplicity and consistency with the prior literature, we describe the approach as random assignment below, and we follow a standard empirical strategy used by other papers using similar types of judge assignment to estimate judicial bias (Shayo and Zussman, 2011).

Random assignment of judges to cases is empirically important because of the concern that judges could treat defendants differently not because of their identity, but because of other case characteristics that are correlated with judge identity. For example, if Muslim judges could systematically choose to sit in cases with Muslim defendants who had committed less serious crimes, we might see in-group differences, but they would be due to differences in the underlying cases of Muslim defendants matched to Muslim judges, rather than due to bias. Alternately, Muslim defendants and judges may be more likely to appear in some parts of the country than others; of those regions are characterized by different crime distributions, we might again mistakenly attribute those differences to in-group bias.

Our ideal experiment would take two defendants identical in all ways, charged with identical crimes in the same police station on the same date, and then assign them to judges with different identities. In practice, the Indian court system runs this experiment whenever a defendant is charged in a jurisdiction with multiple judges of different identities on the bench.

We use a canonical regression approach to test for the effect of judge identity on case outcomes, as used by Shayo and Zussman's (2011) analysis of judicial in-group bias in Israel. We model outcome $Y_{i,s,c,t}$ (e.g. 1=acquitted) for case i with charge s, filed in court c at time t as:

$$\begin{aligned} Y_{i,s,c,t} = \alpha + \beta_1 \text{judge\_male}_{i,s,c,t} + \beta_2 \text{def\_male}_{i,s,c,t} + \\ \beta_3 \text{judge\_male}_{i,s,c,t} * \text{def\_male}_{i,s,c,t} + \phi_{c,t} + \zeta_s + \delta\chi_{i,s,c,t} + \epsilon_{i,s,c,t} \end{aligned} \tag{1}$$

$$\begin{aligned} Y_{i,s,c,t} = \alpha + \beta_1 \text{judge\_nonMuslim}_{i,s,c,t} + \beta_2 \text{def\_nonMuslim}_{i,s,c,t} + \\ \beta_3 \text{judge\_nonMuslim}_{i,s,c,t} * \text{def\_nonMuslim}_{i,s,c,t} + \phi_{c,t} + \zeta_s + \delta\chi_{i,s,c,t} + \epsilon_{i,s,c,t} \end{aligned} \tag{2}$$

where judge_male and judge_nonMuslim are binary variables that indicate whether a judge is male or non-Muslim, respectively. Similarly, def_male and def_nonMuslim indicate the defendant's identity. $\phi_{c,t}$ is a court-month or court-year fixed effect, and ζ_s is an act and section fixed effect. $\chi_{i,s,c,t}$ includes controls for defendant religion, judge religion, and an interaction term of judge gender and defendant religion in the gender analysis. In the religion analysis, $\chi_{i,s,c,t}$ represents controls for defendant gender, judge gender, and an interaction term of judge religion and defendant gender.

The charge section fixed effect ensures that we are comparing defendants charged with similar crimes. The court-time fixed effect ensures that we are comparing defendants who are being charged in the same court at the same time. Our primary specification uses a court-month fixed effect; a secondary specification uses a court-year fixed effect. The court-year fixed effect allows a much larger sample, at some potential bias. Judges on the bench may not hear new cases in some months because they are tied up with previous cases or away from work; it is unlikely that prosecutors or defendants can time their filings to match these absences, nor do we find evidence of disproportionate identity matching in balance tests of either specification below. Court-time periods with no variation in judge identity are retained to increase precision of fixed effects and controls but they do not affect the coefficients of interest. We drop court-time periods where only one judge appears, though they may appear in the regression discontinuity setup.

There are three causal effects of interest. β_1 describes the causal effect on a female defendant of having a male judge assigned to her case, rather than a female judge. $\beta_1 + \beta_3$ describes the causal effect on a *male* defendant of having a male judge assigned to his case. The difference between these effects (or β_3) is the own-gender bias — it tells us whether individuals receive better outcomes when a judge matching their gender identity is randomly assigned to their case. Appendix Table A5 presents a visual summary of the meanings of these coefficients in a difference-in-differences setup. Since all three causal effects are of interest, we report all three coefficients in the regression tables. The coefficient meanings are analogous in Equation 2. Standard errors are clustered at the judge level, since judge assignment is the level of randomization.

4.2 Balance Tests

To test the validity of the random assignment of cases to judges, we run the following empirical balance test in the analysis sample:

$$\text{judge\_female}_{i,s,c,t} = \alpha + \beta_1 \text{def\_Muslim}_{i,s,c,t} + \beta_2 \text{def\_female}_{i,s,c,t} + \gamma \phi_{c,t} + \zeta_s + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t} \tag{3}$$

$$\text{judge\_Muslim}_{i,s,c,t} = \eta + \gamma_1 \text{def\_Muslim}_{i,s,c,t} + \gamma_2 \text{def\_female}_{i,s,c,t} + \gamma \phi_{c,t} + \zeta_s + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t}, \tag{4}$$

Variables are defined as above. Our causal identification strategy relies on β_1 and γ_1 to be equal to zero.

The result is shown in Table 3. Male and female defendants are effectively equally likely to be assigned to female judges, and similarly, Muslim and non-Muslim defendants are equally likely to be assigned to Muslim judges. We do find that Muslim defendants are 0.12 percentage points more likely to have their cases heard by female judges. This difference is economically small but it is statistically significant in part due to the very large sample. Of the eight prior studies we found that exploit random judge assignment, none of them are statistically powered to rule out an effect of this size in their balance tests, and all report point estimates larger in magnitude than 0.12 percentage points. Nevertheless, to ensure that this small difference in assignment to female judges does not influence our result on Muslims, we control for judge and defendant gender in the religion regressions (and for judge and defendant religion in the gender regressions).

Table 3: Balance test for assignment of judge identity

	(1) Female judge	(2) Female judge	(3) Muslim judge	(4) Muslim judge
Panel A: Sample with clear acquitted/convicted outcomes				
Female defendant	-0.0007**	-0.0004	0.0001	-0.00034
	(0.0003)	(0.0005)	(0.0002)	(0.0003)
Muslim defendant	0.0005	0.0012*	0.0001	-0.0001
	(0.0005)	(0.0006)	(0.0003)	(0.0004)
Observations	3,105,245	3,149,781	3,165,276	3,210,450
Fixed Effect	Court-month	Court-year	Court-month	Court-year
Panel B: Sample including observations with no decision				
Female defendant	-0.0006*	-0.0003	0.0000	-0.0001
	(0.0003)	(0.0004)	(0.0002)	(0.0002)
Muslim defendant	0.0006	0.0012**	0.0003	0.0003
	(0.0004)	(0.0005)	(0.0002)	(0.0003)
Observations	5,287,744	5,320,717	5,371,715	5,405,141
Fixed Effect	Court-month	Court-year	Court-month	Court-year

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports results from a formal test of random assignment of judges to cases in the study sample. For specification details, see Equations 3 and 4. Panel A reports results from the test on the sample of case observations that have a case decision. Panel B reports analogous results for the entire sample of cases, including those with no decisions. Columns 1–2 report the likelihood of being assigned to a female judge relative to a male judge using court-month, and court-year fixed effects. Columns 3–4 report the likelihood of being assigned to a Muslim judge relative to a non-Muslim judge using court-month, and court-year fixed effects. Heteroskedasticity robust standard errors are reported below point estimates.

Overall, the balance test indicates that defendants of a given identity do not face different odds of encountering judges with the same identity. This null supports the essential assumption underlying causal identification of judge identity on case outcomes.

4.3 Regression discontinuity approach using court transitions

The sample for the randomized assignment design requires courts with many judges to which a defendant can plausibly by assigned. In this section, we describe a complementary identification strategy that focuses on courts with a smaller number of judges on the bench. We exploit our high-frequency outcome data, along with discrete changes in the set of judges working in a given court, to provide additional evidence on the the same questions of in-group bias.

We define a court transition as any instance when a judge begins or ends their tenure in a court. For each court transition, we calculate the change in the shares of female and Muslim judges before and after the transition. We then examine whether the outcomes of defendants with specific social identities change following the transition.

We analyze three types of court transitions, defined respective to an identity group (female or Muslim), which are listed in Table 4. We define a *pro-defendant* court transition as a transition that results in a court whose judge composition is at least 50 percentage points more similar to the defendant's identity than it was before the transition. For example, in a court with two judges, if one male judge is replaced by a female judge, we describe this as a favorable transition for female defendants. An *against-defendant* transition is the reverse; replacing a male judge by a female judge in the court above would be an against-defendant transition for male defendants.. A *composition-neutral* transition is a judge entry or exit that has zero effect on the balance of the court for the identity in question, such as when a male judge is replaced by another male judge. Judge transitions that result in a non-zero but less than 50 percentage point change in the identity makeup of the court are dropped from the sample. For example, in a court with ten judges, moving from four female judges to five female judges would not be included as an analyzed transition. This approach maximizes statistical power by focusing on transition which have a large effect on the likelihood of a defendant-judge identity match.[14]

[14] Results are similar if we use different thresholds for positive or negative transitions. Using a lower threshold results in a larger sample but a smaller first stage effect of the transition on the likelihood of a defendant-judge identity match. A 50% threshold maximizes power to detect an in-group bias effect.

Table 4: Types of Judge Transitions

Event	Description
Pro-defendant transition	Share of judges belonging to defendant's identity increases by $\geq$ 50 percentage points in the court
Against defendant transition	Share of judges belonging to defendant's identity decreases by $\geq$ 50 percentage points in the court
Composition neutral transition	Share of judges belonging to defendant's identity remains unchanged by a transition
Dropped from sample	Share of judges belonging to defendant's identity changes by 1–49 percentage points

We use a regression discontinuity specification to examine whether defendants experience different kinds of outcomes after each type of judge transition. We use time in days as the running variable and the court transition as the event date. Our local linear regression includes cases decided within a given number of days of the transition date (the bandwidth) and controls for the running variable on either side of the threshold. The treatment is having the case heard in the post-transition period for each event.

We set the baseline bandwidth at 25 days, but the estimates are not sensitive to varying the bandwidth. The sample is limited to courts and dates where the justices on the bench have been in position for at least the same number of weeks as the specification bandwidth before and after the transition. This ensures that each case appears only once in the sample — either before a judge transition or after.

The outcome $Y_{i,s,c,t}$ is a binary variable indicating a positive outcome for the defendant in case i, court c, time t, charged under section s. The estimating equation is given by:

$$Y_{i,s,c,t} = \alpha + \gamma_1 \text{pro\_post}_{c,t} + \gamma_2 \text{against\_post}_{c,t} + \gamma_3 \text{neutral\_post}_{c,t} + \chi_{i,s,c,t} + \epsilon_{i,s,c,t}, \tag{5}$$

where pro_post is a post-event indicator for a pro-defendant transition, against_post for an against-defendant transition, and neutral_post for a composition-neutral transition. $\chi_{i,s,c,t}$ includes all of the linear trends in the forcing variable (date relative to transition), court-time fixed effects, and charge section fixed effects. Standard errors are clustered by transition events. This regression effectively stacks three standard regression discontinuity estimations, estimating a treatment effect for each type of transition. We could also estimate each of the three regression discontinuity specifications separately, but pooling allows us to estimate one set of fixed effects and clusters.

If there is in-group judicial bias, we expect γ_1 (post-event effect for pro-defendant transitions) to be positive and γ_2 (against-defendant transitions) to be negative. We don't expect any effect on average for γ_3.

Identification of these causal parameters comes from the standard assumptions of regression discontinuity designs. Appendix Figure A5 shows that the distribution of cases is flat around positive, neutral, and negative transition events for both men and women, supporting the assumption of no manipulation of case timing around the transition date. This test is analogous to the McCrary test (McCrary, 2008). As further support for absence of manipulation, Appendix Figure A6 shows that there is no variation in the average charge severity of cases seen just before and just after these transitions.

5 Results

5.1 Effect of assignment to judge types

The first two rows of Table 5 Panel A present the impact, for female and male defendants respectively, of being randomly assigned to a male judge; these are β_1 and $\beta_1 + \beta_3$ in Equation 1. The third row shows the difference between these two coefficients (β_3), which is the own-gender bias. The outcome variable is an indicator for defendant acquittal. Columns 1–3 show results using court-month fixed effects, while Columns 4–6 use court-year fixed effects. Within each set of three columns, the second column adds additional demographic controls, while the third column adds judge fixed effects.

Male judges consistently deliver more acquittals than female judges. The point estimate on this effect is nearly identical for male and defendants across all specifications. We interpret this as a null effect. [15]

Panel B shows the effect of filing judge gender on a binary variable indicating whether a case has been decided in our sample period at all. We find no evidence that assignment to a female judge results in any difference in time to resolution for either male or female defendants. In short, we find that while male judges are somewhat more lenient on average in terms of lower acquittal rates, we do not find substantial gender bias in any dimension.

[15] Appendix Table A7 shows estimates when we exclude closed cases for which we are unable to determine the outcome. While we find marginally significant bias effects (in the expected direction), the point estimate on the bias term is never higher than 0.7pp.

Table 5: Impact of assignment to a male judge on defendant outcomes

	Outcome variable: Acquittal rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	0.018***	0.020***	—	0.007*	0.009**	—
	(0.007)	(0.008)		(0.004)	(0.004)	
Male judge on male defendant	0.019***	0.020***	—	0.008***	0.009***	—
	(0.007)	(0.008)		(0.003)	(0.004)	
Difference = Own gender bias	0.001	0.000	-0.001	0.000	0.000	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Reference group mean	0.210	0.210	0.210	0.211	0.211	0.211
Observations	3,180,438	3,105,245	3,103,800	3,224,595	3,149,781	3,1475,62
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

	Outcome variable: Any decision at all					
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	0.004	0.006	—	0.008*	0.011**	—
	(0.009)	(0.009)		(0.005)	(0.005)	
Male judge on male defendant	0.001	0.003	—	0.006	0.008	—
	(0.008)	(0.009)		(0.004)	(0.005)	
Difference = Own gender bias	-0.002	-0.002	-0.002	-0.003	-0.003	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Reference group mean	0.601	0.6	0.6	0.601	0.6	0.6
Observations	5,403,930	5,287,744	5,286,462	5,436,465	5,320,717	5,318,974
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

Notes: Standard errors in parentheses. $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.
Reference group: Female judges, female defendants.
Specification: $Y_{i,c,t} = \alpha + \beta_1 judge\_male_{i,c,t} + \beta_2 def\_male_{i,c,t} + \beta_3 judge\_male_{i,c,t} * def\_male_{i,c,t} + \phi_{c,t} + \delta\chi_{i,c,t} + \epsilon$

Table 6: Impact of assignment to a non-Muslim judge on defendant outcomes

	Outcome variable: Acquittal rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Muslim judge on Muslim defendant	-0.002	-0.001	—	-0.001	-0.002	—
	(0.012)	(0.013)		(0.006)	(0.007)	
Non-Muslim judge on non-Muslim defendant	0.000	0.001	—	0.001	0.001	—
	(0.011)	(0.012)		(0.005)	(0.006)	
Difference = Own religion bias	0.002	0.002	0.002	0.002	0.002	0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Reference group mean	0.202	0.205	0.205	0.202	0.205	0.205
Observations	3,478,261	3,165,276	3,163,786	3,521,995	3,210,450	3,208,135
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

	Outcome variable: Any decision at all					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Muslim judge on Muslim defendant	-0.024**	-0.026**	—	-0.012*	-0.016**	—
	(0.012)	(0.012)		(0.007)	(0.008)	
Non-Muslim judge on non-Muslim defendant	-0.024**	-0.027**	—	-0.013**	-0.017**	—
	(0.012)	(0.012)		(0.006)	(0.007)	
Difference = Own religion bias	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Reference group mean	0.601	0.600	0.600	0.602	0.600	0.600
Observations	5,884,349	5,371,715	5,370,370	5,916,119	5,405,141	5,403,305
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

Notes: Standard errors in parentheses. $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.
Reference group: Muslim judges, Muslim defendants.
Specification: $Y_{i,c,t} = \alpha + \beta_1 judge\_nonmuslim_{i,c,t} + \beta_2 def\_nonmuslim_{i,c,t} + \beta_3 judge\_nonmuslim_{i,c,t} * def\_nonmuslim_{i,c,t} + \phi_{c,t} + \delta \chi_{i,c,t} + \epsilon$

Table 6 presents analogous results for Muslim and non-Muslim defendants randomly assigned to Muslim and non-Muslim judges; all panels and columns have the same interpretation as the prior table. The effect of judge religion on the acquittal rate is again a precise zero. The point estimates on any form of bias are never higher than 0.6pp. The estimates rule out an own-religion bias of of 1.0–1.5pp with 95% confidence.

Panel B shows that Muslim judges are 1.2 to 2.7 percentage points more likely to each a decision on a case. This effect holds equally for Muslim and non-Muslim defendants; the own-religion bias estimate is a precise null. These results are robust to alternate specifications.[16]

Our estimates thus far show that judges do not provide substantively better outcomes for own-gender and own-religion defendants, on average. An alternate hypothesis is that judges specifically discriminate against cross-identity defendants when the victim matches their own identity.[17] This is an interesting and important possibility, especially since there is only correlational evidence of victim-oriented group bias in the previous literature.

We cannot systematically test for differential bias based on victim identity in judicial decisions because we do not have systematic data on victim identity. An important exception is in crimes against women — here, the gender identity of the victim is revealed by the criminal charge itself. These crimes include causing miscarriage, death caused by act with intent to cause miscarriage, assault, kidnapping, or abducting a woman, rape, and marriage offenses.[18]

The nature of cases requires that the defendant is male, so the only coefficient that we can estimate is β_1: the effect of a male defendant being assigned to a female judge. We show in Appendix Table A10 that there is little evidence of bias even here. The point estimate on the acquittal rate is larger than earlier estimates (male judges are 3.4 percentage points more likely to acquit), but it is marginally statistically significant and not robust to alternate specifications. In short, we find no evidence of in-group

[16] Appendix Table A8 reports analogous regressions with conviction as the outcome. Appendix Table A9 shows estimates that exclude ambiguous case outcomes. While we find marginally significant bias effects (in the expected direction) in a handful of specifications, the majority are statistically insignificant and the point estimate on the bias term is never higher than 1pp.

[17] Baldus et al. (1997) analyze death penalty statistics and find that a death sentence is more likely for black defendants and non-black victims. The more recent Baumgartner et al. (2015) find a similar relationship – that capital punishment is more likely for Black-on-White homicides than for Black-on-Black homicides. In a vignette study, ForsterLee et al. (2006) find reverse discrimination: Australian mock jurors gave a harsher sentence with racial-minority victims.

[18] Respectively, these crime categories correspond to sections 312–313, 314, 354, 366, 375–376, and 498 of the Indian Penal Code.

bias based on identity of the crime victim.

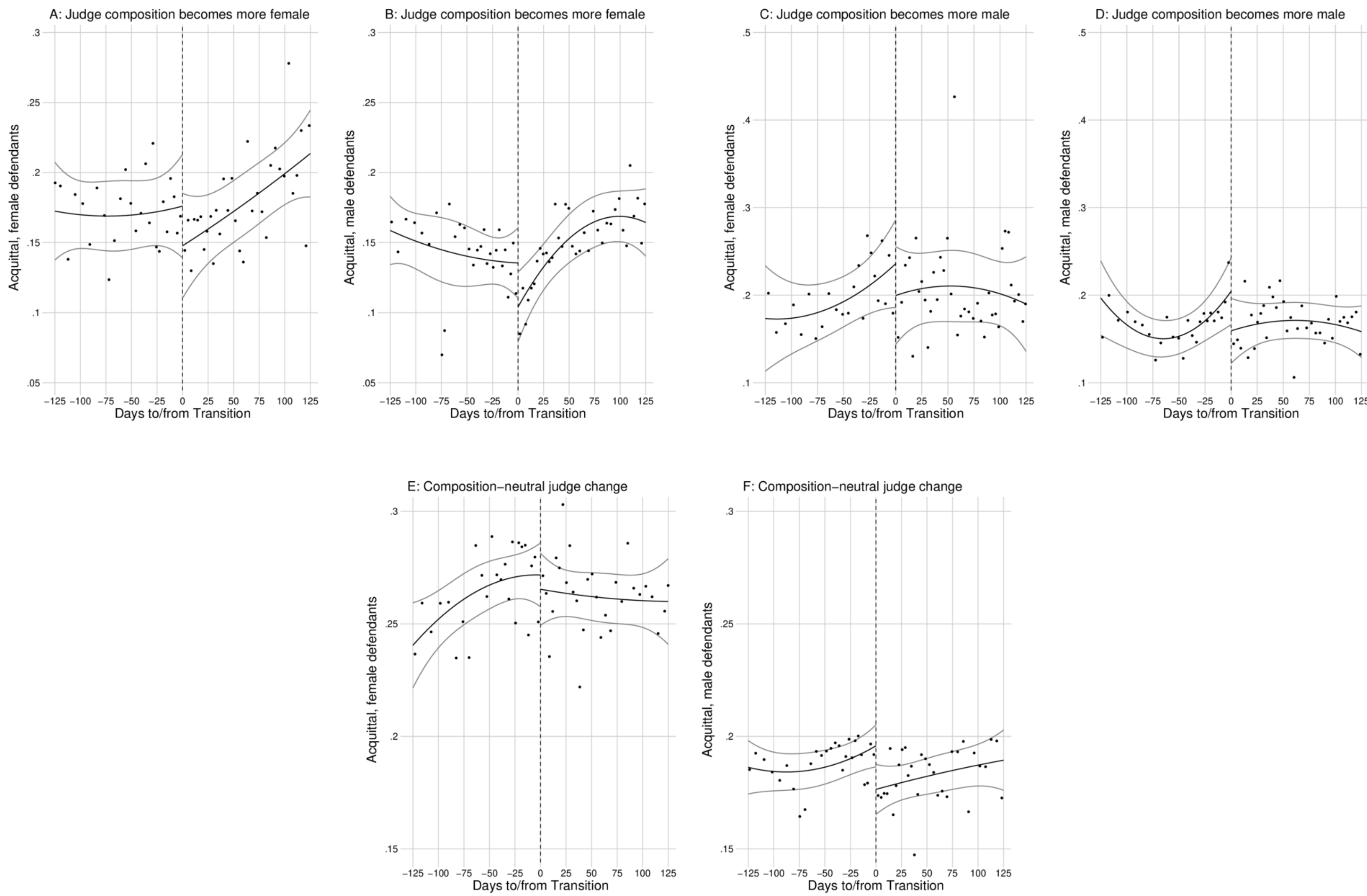
5.2 Effect of changes in court composition

In this section, we estimate judicial bias by exploiting exogenous changes in courtroom staffing. We use a regression discontinuity specification (Equation 5) to test whether defendant outcomes change immediately after the composition of judges in the court becomes more or less similar to the defendant in terms of identity.

We first show our results graphically. Figure 2 shows average acquittal rates before and after a judge joins or leaves the staff in a courtroom. The horizontal axis shows the number of days before and after the transition (with negative and positive numbers respectively), while the vertical axis shows the acquittal rate for defendants of a given gender. Panels A and B show the effects of transitions that increase the female judge share (by at least 50%) on female and male defendants respectively. Panels C and D show composition-neutral transitions as a reference or control group.

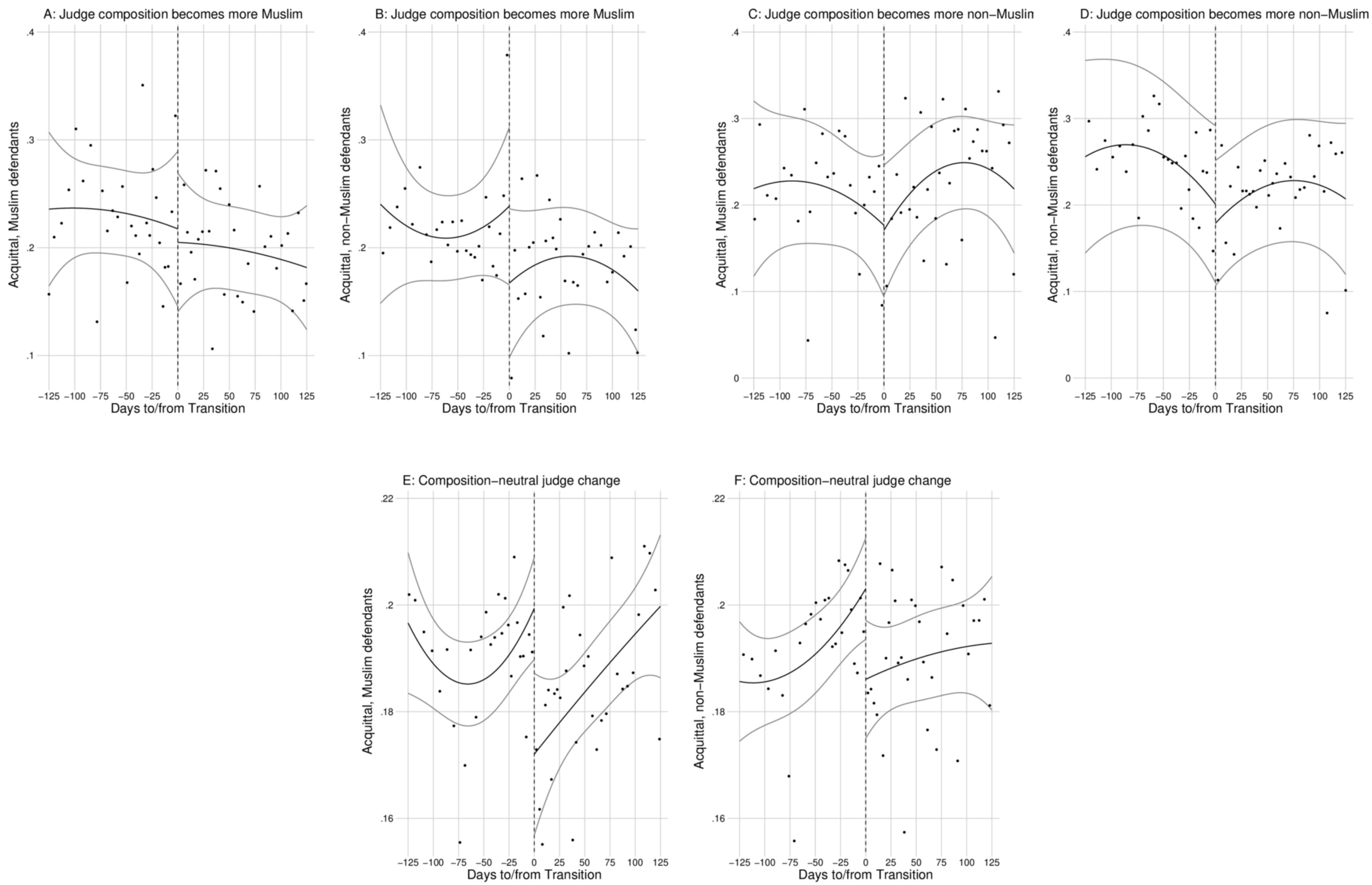
In almost all cases, we see a small decline in the acquittal rate in the weeks immediately after the transition; importantly, this also holds for null transitions. This tells us that courts become briefly more strict in the weeks after a staffing change, even if the gender composition does not change. If judges show bias toward members of their own gender, then we should see a differential break in the acquittal rate when the court becomes more or less matched to the defendant in identity terms. In other words, we should see a relative increase in the acquittal rate when the court becomes more similar to the defendant (Panel A) and the opposite when it becomes less similar (Panel B). In fact, we find that none of the switches that change gender composition have effects that are significantly different from judge changes that are composition-neutral.

Figure 2: Event Discontinuity effect: Transitions that change the likelihood of a same gender judge assignment



Notes: The figure shows acquittal rates of cases decided before and after a transition in the court. Panels A & B show the effect of a transition that increases the likelihood of getting assigned to a female judge. Panels C & D show the effect of a transition that increases the likelihood of getting assigned to a male judge. Panels E & F represent the effect of a transition that leaves the gender composition of the court unchanged.

Figure 3: Event Discontinuity effect: Transitions that change the likelihood of a same religion judge assignment



Notes: The figure shows acquittal rates of cases decided before and after a transition in the court. Panels A & B show the effect of a transition that increases the likelihood of getting assigned to a Muslim judge. Panels C & D show the effect of a transition that increases the likelihood of getting assigned to a non-Muslim judge. Panels E & F represent the effect of a transition that leaves the religion composition of the court unchanged.

Table 7: Impact of judge transitions that affect court composition on acquittal rates

	Gender composition changes		Religion composition changes	
	(1)	(2)	(3)	(4)
	Acquitted	Acquitted	Acquitted	Acquitted
Pro-defendant transition	-0.007	0.003	0.019	0.012
	(0.013)	(0.011)	(0.018)	(0.017)
Against-defendant transition	-0.016	-0.009	-0.003	-0.002
	(0.011)	(0.010)	(0.019)	(0.017)
Constant composition judge transition	-0.008	-0.012**	-0.014**	-0.014**
	(0.006)	(0.005)	(0.006)	(0.006)
Observations	360768	407385	494428	545566
No. of transitions	1840	2878	2316	3476
Mean Acquittal rate	0.192	0.192	0.192	0.192
Fixed effect	Court-month	Court-year	Court-month	Court-year

Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of judge transitions that affect court composition on acquittal rates of defendants. Columns 1–2 estimate the impact of a court transition that increases or decreases the share of judges belonging to the defendant's gender, using different fixed effects. Columns 3–4 estimate the analogous impact on acquittal rates for court transitions that change the religion share of judges in the court (see Section 4 for specification details). Heteroskedasticity robust standard errors are reported below point estimates.

Figure 3 shows a similar graph examining whether changing the Muslim / Non-Muslim composition of the judges' bench differentially affects outcomes for Muslims and non-Muslims. There is no evidence of a differential change in acquittal rates based on the religion of the defendant or direction of the religious composition change in the court.

To formally test these hypotheses, we use a single estimation that calculates all these regression discontinuites simultaneously. We pool results from the six graphs into point estimates for three transition types: pro-defendant, against-defendant, and constant-composition. Table 7 reports the results. Constant composition transitions are associated with about a 1 percentage point decline in the acquittal rate; this effect is not driven by judge identity, because the average judge identity on the bench has not changed.

If there is same-identity bias, then the coefficient on pro-defendant transition should be higher than the coefficient on the constant-composition transition, and the against-defendant coefficient should be lower. However, we find no statistically significantly different effect on either of these transitions. The standard errors are slightly larger than in the randomized assignment regressions, because the regression discontinuity approach has less power. We can rule out a 2.5 percentage point bias effect on the gender dimension and a 4 percentage point bias effect on the religion dimension. We find similar results for alternate specifications and outcome variables. [19]

The null results are consistent with the findings on randomized judge assignment; we reject the hypothesis that judges issue more favorable rulings for defendants with the same gender or religious identity. This specification is not only a robustness test; it also shows that the result holds up in smaller courtrooms, whereas the random assignment tests are mostly identified off of larger courtrooms.

6 Conclusion

In providing fair justice, courts in developing countries face a number of special challenges, including cultural mismatch from transplanted legal codes, informal justice-system substitutes, citizen skepticism toward formal courts, insufficient (human) capital investments in the court system, the inability of many individuals to pay for high-quality representation, implicit or explicit bias among members of the judiciary, and corruption (Djankov et al., 2003; La Porta et al., 2008). Yet with a few exceptions (Ponticelli and Alencar, 2016, for example), these characteristics of developing-country courts have been documented only anecdotally.

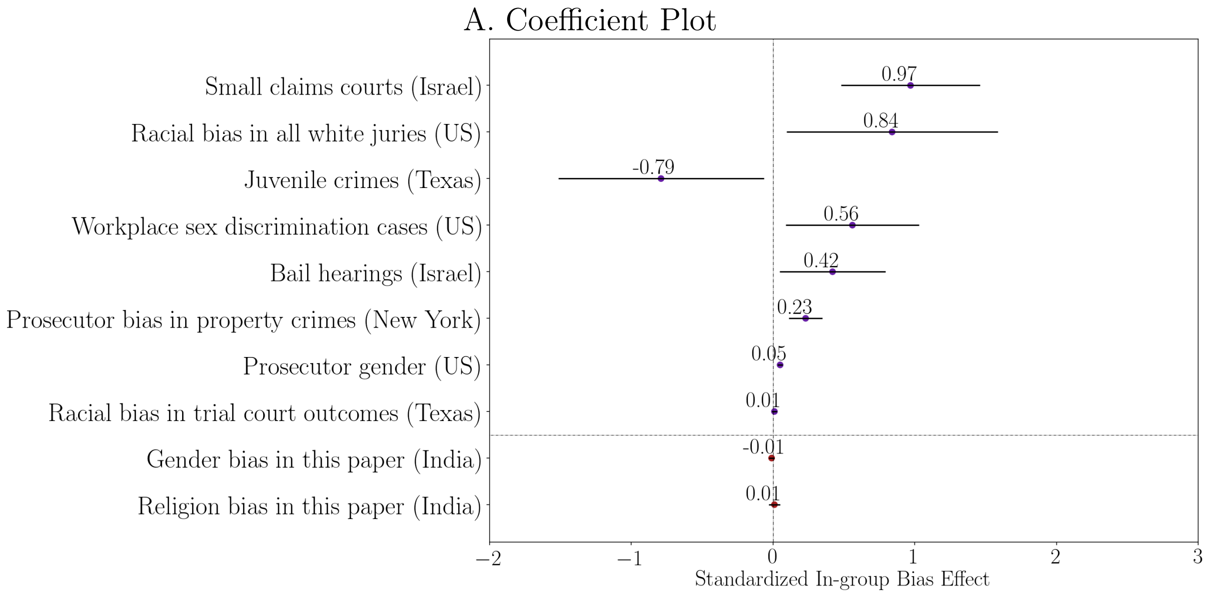
In this paper, we make available a large-scale dataset for the analysis of court proceedings in India and present evidence of negligible judicial in-group bias in criminal cases. The null estimate of in-group bias presented here contrasts with findings in the previous literature, which has tended to find large effects. Figure 4A compares our point estimates with estimates of bias from the studies most similar to ours that we were able to find.[20] Effect sizes are standardized by dividing each in-group bias effect by the sample standard deviation of the outcome variable. The high end of our confidence interval is an order of magnitude smaller than nearly all prior studies.

The most straightforward interpretation of these findings is that, unlike judges analyzed in the other papers, India's district court judges do not exhibit in-group bias along the pertinent identity margins. Our research is consistent with judges taking their role seriously and working hard to provide justice on fair terms to all litigants, or with judicial institutions that constrain discretion and protect defendants from biased

[19] Appendix Figure A7 shows no effect on case delay. Appendix Table A11 there is no bias effect for other outcome definitions, such as the conviction rate.

[20] We included every study we could find that focused on measuring in-group bias among judges on a race/ethnicity, gender, or religious dimension, using either random assignment to judges or rotations in judge cohorts as an identification strategy.

Figure 4: Comparison with judicial bias estimates in other contexts



Notes: This figure reports point estimates of in-group bias from other studies in the relevant literature. From top to bottom, the coefficients of in-group bias (Panel A) correspond to Shayo and Zussman (2011), Anwar et al. (2012), Depew et al. (2017), Knepper (2018), Gazal-Ayal and Sulitzeanu-Kenan (2010), Sloane (2019), Didwania (2018), Lim et al. (2016), and the present study respectively. Panel B plots reported effect magnitudes (Y axis) against effect standard errors. All effect sizes are standardized (dividing outcome variables by their standard deviation) to allow comparison across studies.

decision-making. It is also possible that the social distance between (normally) upper class judges and (normally) lower class criminal litigants may mitigate a sense of shared identity between judges and litigants. Yet it is also consistent with corruption that is blind to religious and gender identity. Rich as they are, our data do not allow us to differentiate between these very different mechanisms; exploring these possibilities is an important area for future work.

Another interpretation of why our India results stand out from the literature is that there could be publication bias in studies of judicial in-group bias. Focusing on the subset of bias studies that use one of our two identification strategies, Figure 4B plots the effect size of each study against the standard error of the main estimated effect.[21] In the absence of publication bias or a design-based mechanical correlation (such as adaptive sampling), the standard error should not be correlated with the effect size (Gerber et al., 2001; Levine et al., 2009; Slavin and Smith, 2009; Kühberger et al., 2014). In fact, a regression of effect size on standard error is highly significant ($\hat{\beta} = 2.25, p = 0.001$), which may suggest that studies finding a lack of bias are less likely to be published.

It is worth emphasizing again that we have not ruled out bias in the Indian criminal justice system as a whole. We have focused on two kinds of bias which have been widely documented in other countries and we have focused on the singular contributions of judges to criminal-justice outcomes. The legal system could still be biased against Muslims and women overall, through geographic distribution of policing, discrimination in investigations, police/prosecutor decisions to file cases, the severity of charges applied, the severity of penalties imposed, the appeals process, and others. It is also possible that bias takes a more subtle form, such as discrimination conditional on the interaction between defendant, victim, and type of crime. More research, and in particular more data, are needed to study the entire justice process in India and other developing countries.

[21] When papers report multiple specifications for the main effect, we used the effect size described most prominently in the text or described by the authors as the "main specification." When papers had multiple outcomes, we used the outcome most similar to the acquittal or conviction rate, as in this study. If these were unavailable, we used the outcome most prominently described in the paper's abstract and introduction.

References

Abrams, D. S., Bertrand, M., and Mullainathan, S. (2012). Do judges vary in their treatment of race? *The Journal of Legal Studies*, 41(2):347 – 383.

Aldashev, G., Chaara, I., Platteau, J.-P., and Wahhaj, Z. (2012). Using the Law to Change the Custom. *Journal of Development Economics*, 97(2):182–200.

Alesina, A. and La Ferrara, E. (2014). A test of racial bias in capital sentencing. *The American Economic Review*, 104(11):3397–3433.

Aney, M. S., Dam, S., and Ko, G. (2017). Jobs for justice (s): Corruption in the supreme court of india. *Available at SSRN 3087464*.

Anwar, S., Bayer, P., and Hjalmarsson, R. (2012). The impact of jury race in criminal trials. *The Quarterly Journal of Economics*, 127(2):1–39.

Anwar, S. and Fang, H. (2006). An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review*, 96(1):127–151.

Arnold, D., Dobbie, W., and Yang, C. S. (2017). Racial bias in bail decisions. Technical report, National Bureau of Economic Research.

Asher, S., Novosad, P., and Rafkin, C. (2020). Intergenerational Mobility in India: New Methods and Estimates Across Time, Space, and Communities. Working Paper.

Baldus, D. C., Woodworth, G., Zuckerman, D., and Weiner, N. A. (1997). Racial discrimination and the death penalty in the post-furman era: An empirical and legal overview with recent findings from philadelphia. *Cornell L. Rev.*, 83:1638.

Baumgartner, F. R., Grigg, A. J., and Mastro, A. (2015). #blacklivesdon'tmatter: race-of-victim effects in us executions, 1976–2013. *Politics, Groups, and Identities*, 3(2):209–221.

Bertrand, M., Hanna, R., and Mullainathan, S. (2010). Affirmative action in education: Evidence from engineering college admissions in India. *Journal of Public Economics*, 94(1-2):16–29.

Bharti, N. and Roy, S. (2020). The early origins of judicial bias in bail decisions: Evidence from early childhoodexposure to hindu-muslim riots in india.

Borker, G. (2017). Safety first: Perceived risk of street harassment and educational choices of women. *Job Market Paper, Department of Economics, Brown University*, pages 12–45.

Boyd, C., Epstein, L., and Martin, A. D. (2010). Untangling the causal effects of sex on judging. *American Journal of Political Science*, 54(2):389–411.

Chaturvedi, R. and Chaturvedi, S. (2020). It's all in the name: A character based approach to infer religion. *arXiv preprint arXiv:2010.14479*.

Chemin, M. (2009). Do judiciaries matter for development? Evidence from India. *Journal of Comparative Economics*, 37(2):230–250.

Chew, P. K. and Kelley, R. E. (2008). Myth of the color-blind judge: An empirical analysis of racial harassment cases. *Wash. UL Rev.*, 86:1117.

Cousins, S. (2020). 2· 5 million more child marriages due to covid-19 pandemic. *The Lancet*, 396(10257):1059.

Depew, B., Eren, O., and Mocan, N. (2017). Judges, juveniles, and in-group bias. *The Journal of Law and Economics*, 60(2):209–239.

Dev, A. (2019). Corruption has india's supreme court veering on the edge.

Didwania, S. H. (2018). Gender-based favoritism among criminal prosecutors. *Columbia Law & Economics Workshop*.

Djankov, S., La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2003). Courts. *Quarterly Journal of Economics*, 118(2):453–517.

Erken, A., Chalasani, S., Diop, N., Liang, M., Weny, K., Baker, D., Baric, S., Guilmoto, C., Luchsinger, G., Mogelgaard, K., and et al. (2020). *DEFYING THE PRACTICES THAT HARM WOMEN AND GIRLS AND UNDERMINE EQUALITY State of World Population 2020 This report was developed under the auspices of the UNFPA Division of Communications and Strategic Partnerships. EDITOR-IN-CHIEF*.

Fisman, R., Sarkar, A., Skrastins, J., and Vig, V. (2020). Experience of communal conflicts and intergroup lending. *Journal of Political Economy*, 128(9):3346–3375.

ForsterLee, R., ForsterLee, L., Horowitz, I. A., and King, E. (2006). The effects of defendant race, victim race, and juror gender on evidence processing in a murder trial. *Behavioral Sciences & the Law*, 24(2):179–198.

Gazal-Ayal, O. and Sulitzeanu-Kenan, R. (2010). Let my people go: Ethnic in-group bias in judicial decisions—evidence from a randomized natural experiment. *Journal of Empirical Legal Studies*, 7(3):403–428.

Gerber, A. S., Green, D. P., and Nickerson, D. (2001). Testing for publication bias in political science. *Political Analysis*, pages 385–392.

Goel, S., Rao, J. M., Shroff, R., et al. (2016). Precinct or prejudice? understanding racial disparities in new york city's stop-and-frisk policy. *The Annals of Applied Statistics*, 10(1):365–394.

Grossman, G., Gazal-Ayal, O., Pimentel, S. D., and Weinstein, J. M. (2016). Descriptive representation and judicial outcomes in multiethnic societies. *American Journal of Political Science*, 60(1):44–69.

Hanna, R. N. and Linden, L. L. (2012). Discrimination in grading. *American Economic Journal: Economic Policy*, 4(4):146–68.

Hnatkovska, V., Lahiri, A., and Paul, S. (2012). Castes and labor mobility. *American Economic Journal: Applied Economics*, 4(2):274–307.

Ito, T. (2009). Caste discrimination and transaction costs in the labor market: Evidence from rural North India. *Journal of Development Economics*, 88(2):292–300.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *economics*, 7(1):63–88.

Kastellec, J. P. (2011). Panel composition and voting on the us courts of appeals over time. *Political Research Quarterly*, 64(2):377–391.

Kastellec, J. P. (2013). Racial diversity and judicial influence on appellate courts. *American Journal of Political Science*, 57(1):167–183.

Knepper, M. (2018). When the shadow is the substance: Judge gender and the outcomes of workplace sex discrimination cases. *Journal of Labor Economics*, 36(3):623–664.

Kühberger, A., Fritz, A., and Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PloS one*, 9(9):e105825.

La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2008). The economic consequences of legal origins. *Journal of Economics Literature*, 46(2):285–332.

La Porta, R. L., de Silanes, F. L., Pop-Eleches, C., and Shleifer, A. (2004). Judicial checks and balances. *Journal of Political Economy*, 112(2):pp.445–470.

Le, Q. V. (2004). Political and economic determinants of private investment. *Journal of International Development*, 16(4):589–604.

Levine, T. R., Asada, K. J., and Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlated in meta-analyses: Evidence and implications of a publication bias against nonsignificant findings. *Communication Monographs*, 76(3):286–302.

Lichand, G. and Soares, R. R. (2014). Access to justice and entrepreneurship: Evidence from brazil?s special civil tribunals. *The Journal of Law and Economics*, 57(2):459–499.

Lim, C. S., Silveira, B. S., and Snyder, J. M. (2016). Do judges' characteristics matter? ethnicity, gender, and partisanship in texas state trial courts. *American Law and Economics Review*, 18(2):302–357.

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

Mustard, D. B. (2001). Racial, ethnic, and gender disparities in sentencing: Evidence from the us federal courts. *The Journal of Law and Economics*, 44(1):285–314.

Pande, R. and Udry, C. (2006). Institutions and Development: A View from Below.

Peresie, J. L. (2005). Female judges matter: Gender and collegial decisionmaking in the federal appellate courts. *The Yale Law Journal*, 114(7):1759–1790.

Ponticelli, J. and Alencar, L. S. (2016). Court enforcement, bank loans, and firm investment: evidence from a bankruptcy reform in brazil. *The Quarterly Journal of Economics*, 131(3):1365–1413.

Rao, M. (2019). Judges, lenders, and the bottom line: Court-ing firm growth in india.

Rehavi, M. M. and Starr, S. B. (2014). Racial disparity in federal criminal sentences. *Journal of Political Economy*, 122(6):1320–1354.

Rodrik, D. (2000). Institutions for high-quality growth: what they are and how to acquire them. *Studies in comparative international development*, 35(3):3–31.

Rodrik, D. (2005). Growth strategies. *Handbook of economic growth*, 1:967–1014.

Sandefur, J. and Siddiqi, B. (2015). Delivering justice to the poor: Theory and experimental evidence from liberia. Working Paper.

Scherer, N. (2004). Blacks on the bench. *Political Science Quarterly*, 119(4):655–675.

Shayo, M. and Zussman, A. (2011). Judicial ingroup bias in the shadow of terrorism. *The Quarterly Journal of Economics*, 126(3):1447–1484.

Slavin, R. and Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational evaluation and policy analysis*, 31(4):500–506.

Sloane, C. (2019). Racial bias by prosecutors: Evidence from random assignment. Technical report, Working paper, Texas A&M. Available at: https://github.com/carlywillsloan ....

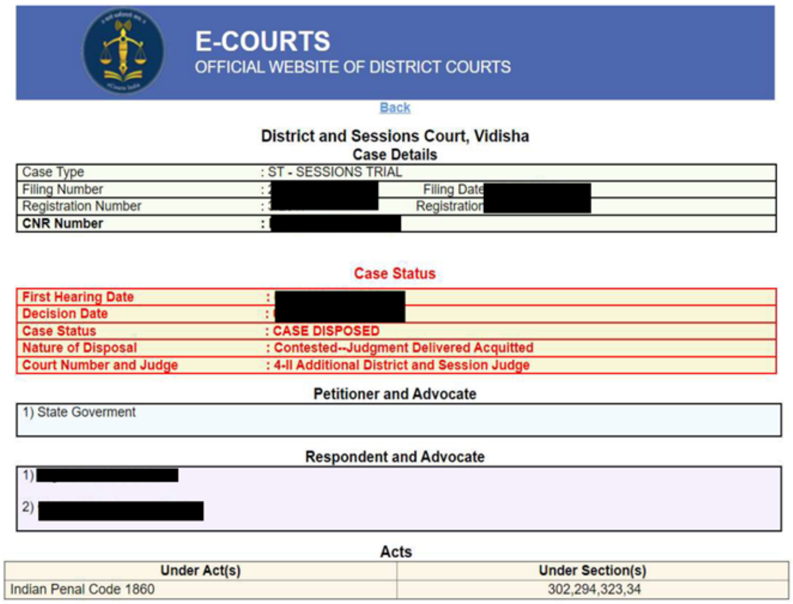
Trusts, T. (2019). India justice report: Ranking states on police, judiciary, prisons & legal aid.

Visaria, S. (2009). Legal reform and loan repayment: The microeconomic impact of debt recovery tribunals in india. *American Economic Journal: Applied Economics*, 1(3):59–81.

World Bank (2017). *World Development Report: Governance and the Law*. World Bank Group.

A Appendix

Figure A1: India eCourts Case Record Sample



E-COURTS
OFFICIAL WEBSITE OF DISTRICT COURTS

Back

District and Sessions Court, Vidisha
Case Details

Case Type	: ST - SESSIONS TRIAL		
Filing Number	: [redacted]	Filing Date	[redacted]
Registration Number	: [redacted]	Registration	[redacted]
CNR Number	: [redacted]		

Case Status

First Hearing Date	: [redacted]
Decision Date	: [redacted]
Case Status	: **CASE DISPOSED**
Nature of Disposal	: **Contested--Judgment Delivered Acquitted**
Court Number and Judge	: **4-II Additional District and Session Judge**

Petitioner and Advocate

1) State Goverment

Respondent and Advocate

1) [redacted]

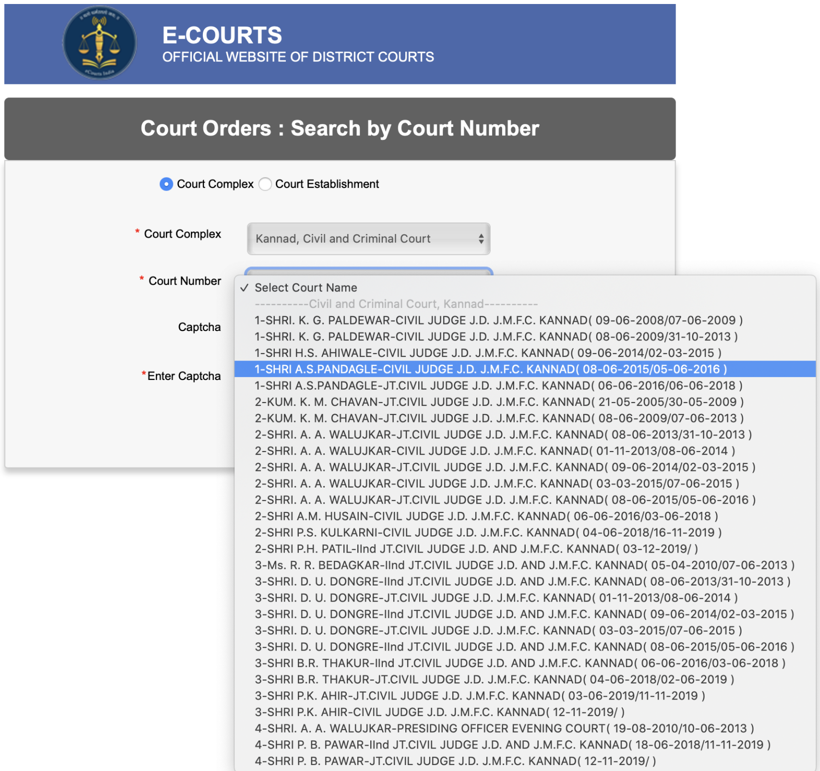
2) [redacted]

Acts

Under Act(s)	Under Section(s)
Indian Penal Code 1860	302,294,323,34

Notes: The figure displays an anonymized version of a sample court record from https://ecourts.gov.in/ for the District and Sessions Court of Vidisha. The record is comprising 'Case Details' variables, as well as information on the 'Case Status' – for instance the disposition information. The 'Petitioner and Advocate' and 'Respondent and Advocate' sections contain the litigant names that we use for assigning gender and religion. The 'Acts' section contains the data that allows us to discriminate between civil and criminal cases. We use the 'Under Section(s)' column to infer the corresponding crime categories.

Figure A2: India eCourts Sample Judge Information inside the Search Engine



Notes: Sample view of the eCourts court order search engine. We scraped the judge information implicitly given in the 'Court Number' drop-down list of the search mask on – in this case – `https://services.ecourts.gov.in/ecourtindia_v4_bilingual/cases/s_order.php?state=D&state_cd=1&dist_cd=19` to obtain judge names and tenures.

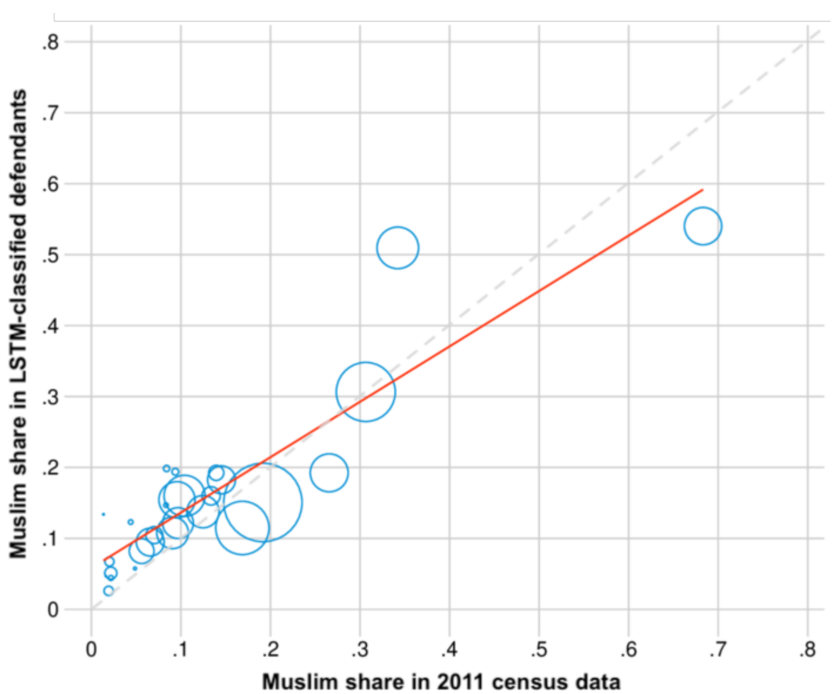
Table A1: Summary of Name Classifier Training Datasets

Panel A: Delhi voter rolls names		
Gender	Instances	Percentage
Female	6,138,337	44.8%
Male	7,556,138	55.2%
Total	13,694,475	100.0%

Panel B: National Railway exam names		
Religion	Instances	Percentage
Buddhist	1,910	0.1%
Christian	11,194	0.8%
Hindu	1,174,076	84.8%
Muslim	163,861	11.8%
NA	33,882	2.4%
Total	1,384,923	100.0%

Notes: Panels A & B of this table show the distribution of identities in the underlying training datasets of the gender and religion LSTM name classification models respectively.

Figure A3: Religion Classifier Outputs and Local Muslim Population Shares



Notes: This figure illustrates that the states in which LSTM classified defendant Muslim share is high also happen to be the states with a higher underlying Muslim population share. This figure is a qualitative illustration of the accuracy of the neural network classification.

Table A2: Outcome variables mapped to dispositions

	Mapping of Dispositions to Outcomes		
Disposition name	Acquitted	Convicted	Decision
258 crpc [acquitted]	X		X
Acquitted	X		X
Allowed	X		X
Committed			X
Compromise			X
Convicted		X	X
Decided			X
Dismissed			X
Disposed			X
Fine			X
Judgement			X
Other			X
Plead guilty		X	X
Prison		X	X
Referred to lok adalat			X
Reject			X
Remanded			X
Transferred			X
Withdrawn			X
Missing			

Notes: This table illustrates the classification of the disposition types of the sample case records into three outcome variables. If a case has a disposition at all, the indicator variable *Decision* equals 1, and 0 otherwise. Conditional on having a disposition, if the disposition is clearly acquitted, the outcome variable *Acquitted* takes the value 1, and 0 otherwise. The outcome variable for conviction has been coded analogously.

Table A3: Summary of charges, by gender of defendant

	(1) Female share	(2) Female share/ population share/	(3) Female conviction rate	(4) Male conviction rate	(5) Difference (3) - (4)	(6) Number of cases
Murder	0.108	0.225	0.037	0.029	0.008	1,298,000
Sexual assault	0.093	0.194	0.033	0.036	-0.003	285,791
Violent crimes causing hurt	0.120	0.250	0.095	0.090	0.005	1,955,000
Violent theft/dacoity	0.161	0.335	0.009	0.025	-0.016	416,426
Crimes against women	0.098	0.204	0.020	0.020	0.000	783,089
Disturbed pub. health/tranquility	0.081	0.169	0.144	0.157	-0.013	2,311,000
Property Crime	0.119	0.248	0.041	0.045	-0.004	3,276,000
Trespass	0.108	0.225	0.048	0.040	0.008	581,108
Marriage offenses	0.122	0.254	0.014	0.009	0.005	331,911
Petty theft	0.107	0.223	0.084	0.079	0.005	1,021,000
Other crimes	0.112	0.233	0.065	0.072	-0.007	8,884,000
Total	0.109	0.227	0.069	0.078	-0.009	18,560,000

Notes: Column 1 of this table reports the share of female defendants for each crime category. Column 2 reports the ratio of the female share for each crime and the female population share in India. Column 3 reports the conviction rate for females accused of each crime category. Column 4 reports the analogous conviction rates for males. Column 5 reports the difference in female and male conviction rates for each crime category. Column 6 reports the total number of case records in each crime category.

Table A4: Summary of charges, by religion of defendant

	(1) Muslim share	(2) Muslim share/ population share/	(3) Muslim conviction rate	(4) Non-Muslim conviction rate	(5) Difference (3) - (4)	(6) Number of cases
Murder	0.136	0.958	0.027	0.031	-0.004	1,390,000
Sexual assault	0.161	1.134	0.031	0.037	-0.006	306,688
Violent crimes causing hurt	0.142	1.000	0.088	0.093	-0.005	2,110,000
Violent theft/dacoity	0.175	1.232	0.037	0.019	0.018	445,925
Crimes against women	0.191	1.345	0.016	0.021	-0.005	837,751
Disturbed pub. health/tranquility	0.157	1.106	0.149	0.155	-0.006	2,535,000
Property Crime	0.160	1.127	0.055	0.043	0.012	3,473,000
Trespass	0.130	0.915	0.046	0.041	0.005	622,220
Marriage offenses	0.230	1.620	0.008	0.011	-0.003	352,219
Petty theft	0.179	1.261	0.097	0.080	0.017	1,080,000
Other crimes	0.140	0.986	0.072	0.072	0.000	9,455,000
Total	0.148	1.042	0.078	0.077	0.001	19,860,000

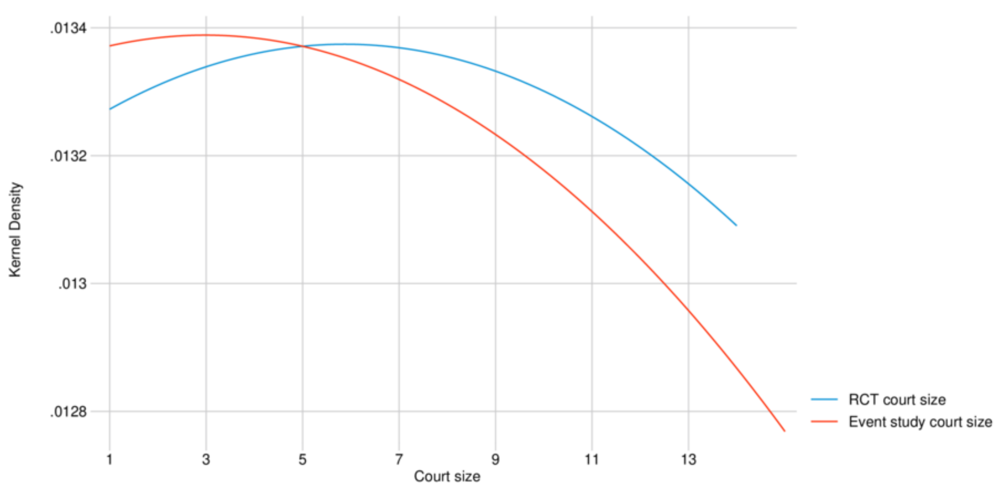
Notes: Column 1 of this table reports the share of Muslim defendants for each crime category. Column 2 reports the ratio of the Muslim share for each crime and the Muslim population share in India. Column 3 reports the conviction rate for Muslims accused of each crime category. Column 4 reports the analogous conviction rates for non-Muslims. Column 5 reports the difference in Muslim and non-Muslim conviction rates for each crime category. Column 6 reports the total number of case records in each crime category.

Table A5: Summary of Coefficients for Case-Assignment Analysis

	Female Defendant	**Male Defendant**	**Defendant Difference Female - Male**
Female Judge	$\beta_1 + \beta_2 + \beta_3$	β_1	$\beta_2 + \beta_3$
Male Judge	β_2	–	β_2
Judge Difference Female - Male	$\beta_1 + \beta_3$	β_1	**Diff-in-Diffs:** β_3

Notes: This table delineates the meaning of each coefficient that appears in Equation 3. The coefficients have analogous meanings in the religion bias analysis specification — Equation 4.

Figure A4: Distribution of court size in the analysis samples



Note: The median courts in the RCT and event study samples have 5 judges and 2 judges respectively

Table A6: Impact of assignment to a male judge on non-conviction

	Outcome variable: Not convicted					
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	-0.008	-0.006	—	-0.006	-0.005	—
	(0.008)	(0.008)		(0.004)	(0.004)	
Male judge on male defendant	-0.006	-0.005	—	-0.004	-0.002	—
	(0.008)	(0.009)		(0.004)	(0.004)	
Difference = Own gender bias	0.001	0.001	0.001	0.002	0.002	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Reference group mean	0.922	0.922	0.922	0.923	0.923	0.923
Observations	3,180,438	3,105,245	3,103,800	3,224,595	3,149,781	3,147,562
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reference group: female judges, female defendants

Specification: $Y_{i,c,t} = \alpha + \beta_1 judge\_male_{i,c,t} + \beta_2 def\_male_{i,c,t} + \beta_3 judge\_male^*_{i,c,t} def\_male_{i,c,t} + \phi_{c,t} + \delta\chi_{i,c,t} + \epsilon$

Table A7: Impact of assignment to a male judge on acquittal rates, dropping ambiguous outcomes

	Outcome variable: Acquittal rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	0.019**	0.022**	—	0.007	0.010*	—
	(0.009)	(0.01)		(0.005)	(0.006)	
Male judge on male defendant	0.026***	0.028***	—	0.013***	0.016***	—
	(0.009)	(0.01)		(0.004)	(0.005)	
Difference = Own gender bias	0.006*	0.006*	0.005	0.007**	0.006*	0.006*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Reference group mean	0.358	0.358	0.358	0.358	0.358	0.358
Observations	1,720,095	1,673,560	1,672,153	1,772,961	1,726,710	1,724,269
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reference group: Female judges, female defendants

Specification: $Y_{i,c,t} = \alpha + \beta_1 judge\_male_{i,c,t} + \beta_2 def\_male_{i,c,t} + \beta_3 judge\_male^{*}_{i,c,t} def\_male_{i,c,t} + \phi_{c,t} + \delta \chi_{i,c,t} + \epsilon$

Table A8: Impact of assignment to a non-Muslim judge on non-conviction

	Outcome variable: Not convicted					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Muslim judge on Muslim defendant	-0.001	-0.008	—	-0.004	-0.010	—
	(0.013)	(0.013)		(0.007)	(0.006)	
Non-Muslim judge on non-Muslim defendant	-0.001	-0.007	—	-0.004	-0.008	—
	(0.013)	(0.012)		(0.006)	(0.006)	
Difference = Own religion bias	0.000	0.001	0.003	0.001	0.002	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Reference group mean	0.905	0.907	0.907	0.906	0.908	0.908
Observations	3,478,261	3,165,276	3,163,786	3,521,995	3,210,450	3,208,135
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reference group: Muslim judges, Muslim defendants

Specification: $Y_{i,c,t} = \alpha + \beta_1 judge\_nonmuslim_{i,c,t} + \beta_2 def\_nonmuslim_{i,c,t} + \beta_3 judge\_nonmuslim^{*}_{i,c,t} def\_nonmuslim_{i,c,t} + \phi_{c,t} + \delta \chi_{i,c,t} + \epsilon$

Table A9: Impact of assignment to a non-Muslim judge on acquittal rates, dropping ambiguous outcomes

	Outcome variable: Acquittal rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Muslim judge on Muslim defendant	-.008	-.007	—	-.004	-.003	—
	(0.016)	(0.018)		(0.008)	(0.01)	
Non-Muslim judge on non-Muslim defendant	-.004	-.003	—	.001	.003	—
	(0.015)	(0.017)		(0.007)	(0.009)	
Difference = Own religion bias	.004	.004	.002	.005	.006	.004
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Reference group mean	0.339	0.343	0.343	0.34	0.344	0.344
Observations	1,867,992	1,697,981	1,696,540	1,921,274	1,752,109	1,749,574
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Reference group: Non-Muslim judges, non-Muslim defendants

Specification: $Y_{i,c,t} = \alpha + \beta_1 judge\_nonmuslim_{i,c,t} + \beta_2 def\_nonmuslim_{i,c,t} + \beta_3 judge\_nonmuslim^{*}_{i,c,t} def\_nonmuslim_{i,c,t} + \phi_{c,t} + \delta \chi_{i,c,t} + \epsilon$

Table A10: Effect of female judge on acquittal rates for crimes against women

Panel A: Court-month fixed effect			
	(1)	(2)	(3)
	Any decision	Acquitted	Not convicted
Female judge on Male defendant (β_1)	-0.031*	-0.038	0.002
	(0.017)	(0.029)	(0.004)
Observations	2,26,104	1,09,940	1,09,940

Panel B: Court-year fixed effect			
	(1)	(2)	(3)
	Any decision	Acquitted	Not convicted
Female judge on Male defendant (β_1)	-0.007	-0.009	0.005**
	(0.007)	(0.010)	(0.002)
Observations	2,69,702	1,46,210	1,46,210

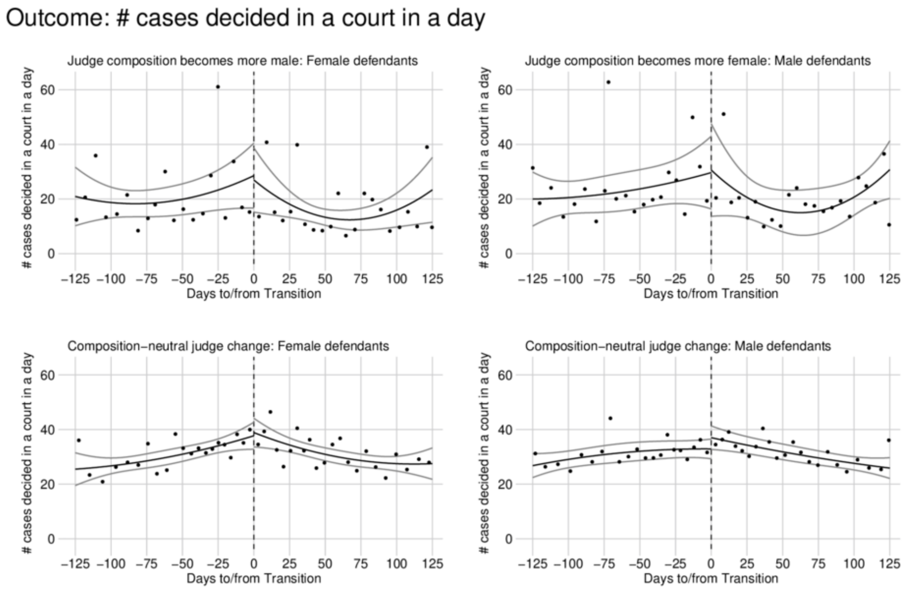
Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

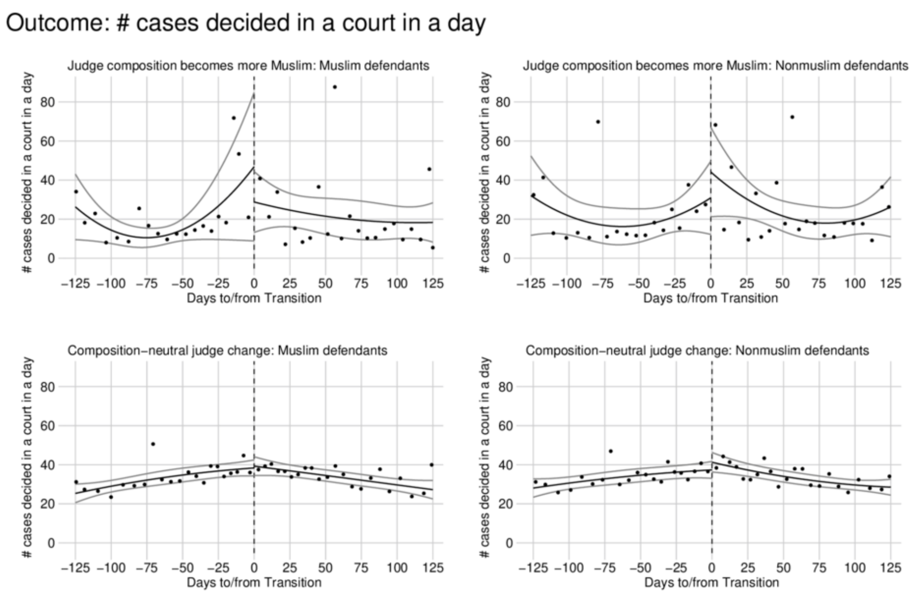
The share of of female defendants accused of crimes against women is negligible, therefore, only male defendants are included in this sub-sample analysis.

Figure A5: Distribution of Cases Is Even at Transition Events

Panel A. Gender



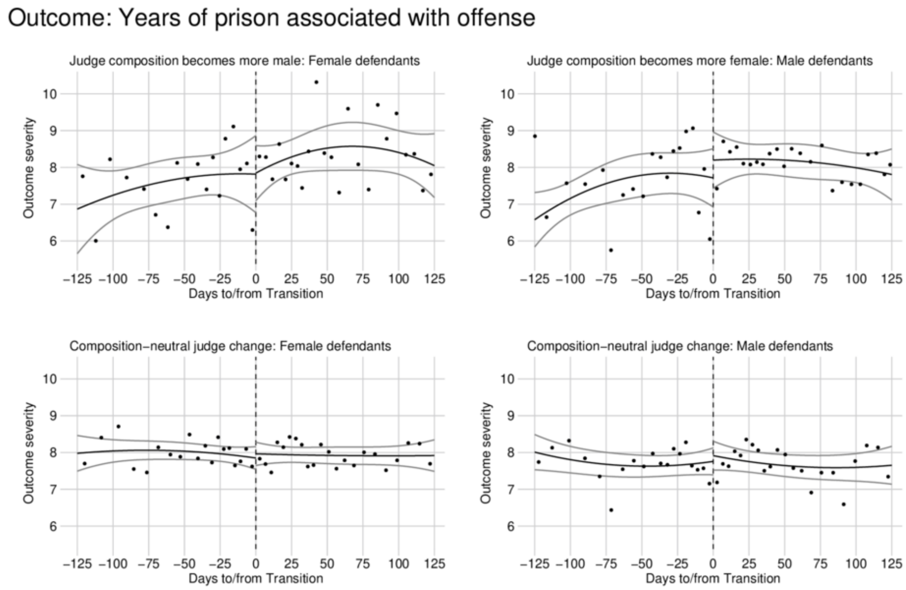
Panel B. Religion



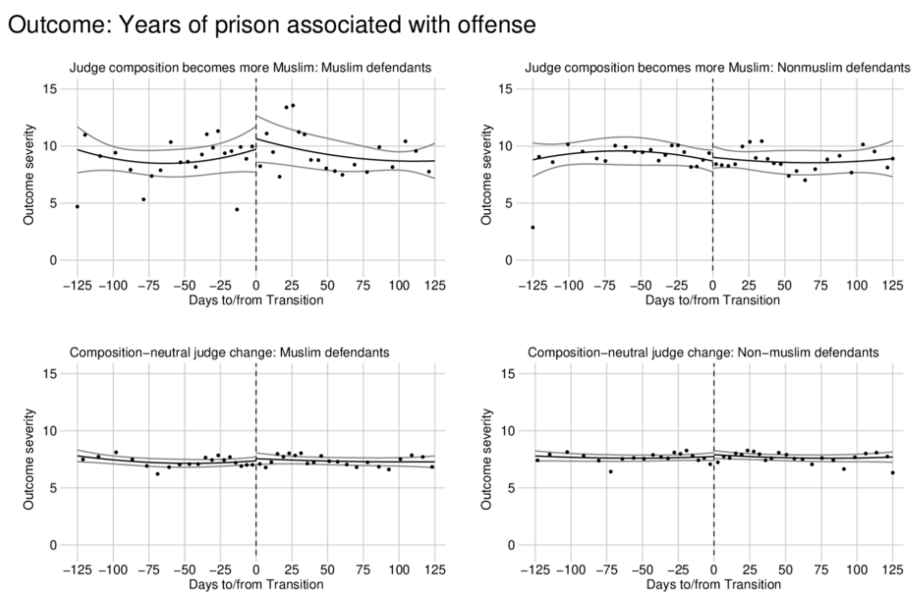
Notes: Panels A and B illustrate that the number of cases decided in a court does not vary when the composition of judges in a court becomes more female or more Muslim respectively.

Figure A6: Judge Transition Events do not Affect Charge Severity

Panel A. Gender



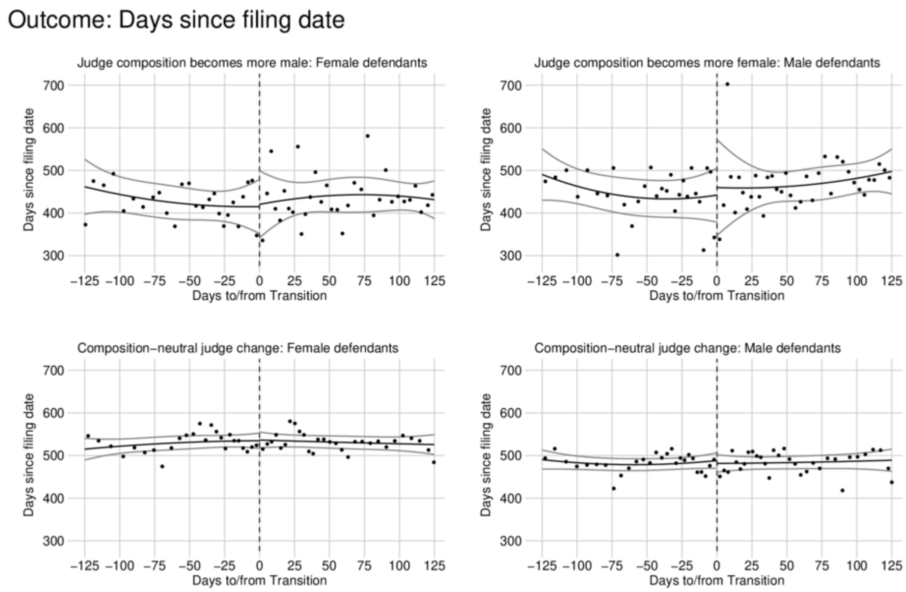
Panel B. Religion



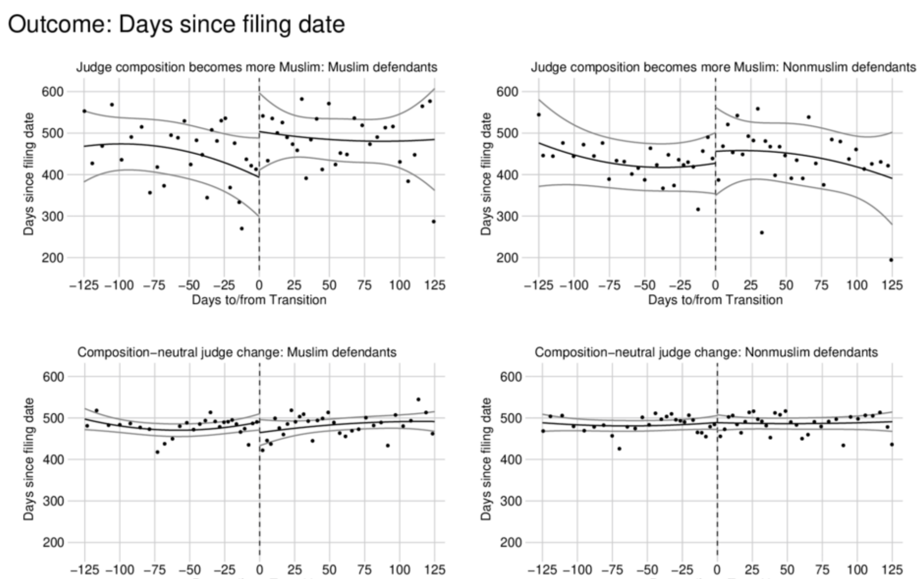
Notes: Panels A and B illustrate that the severity of cases decided, as measured by years of punishment associated with the charge of each case, does not vary when the composition of judges in a court becomes more female or more Muslim respectively.

Figure A7: Effect of Judge Composition Change on Case Delay

Panel A. Gender



Panel B. Religion



Notes: Panels A and B illustrate that the time it takes for a case to reach completion does not vary when the composition of judges in a court becomes more female or more Muslim respectively.

Table A11: Impact of judge transitions that affect court composition on non conviction rates

	Gender composition changes		Religion composition changes	
	(1)	(2)	(3)	(4)
	not convicted	not convicted	not convicted	not convicted
Pro-defendant transition	-0.017**	-0.009	-0.009	-0.007
	(0.008)	(0.007)	(0.011)	(0.008)
Against-defendant transition	-0.017**	-0.008	0.001	0.002
	(0.007)	(0.006)	(0.010)	(0.009)
Constant composition judge transition	0.005	0.003	0.002	0.000
	(0.004)	(0.004)	(0.003)	(0.003)
Observations	360768	407385	494428	545566
No. of transitions	1840	2878	2316	3476
Mean non-convicted	0.920	0.920	0.919	0.919
Fixed effect	Court-month	Court-year	Court-month	Court-year

Notes: This table presents regression discontinuity estimates from the main estimating equation of the effect of judge transitions that affect court composition on non-conviction rates of defendants. Columns 1–2 estimate the impact of a court transition that increases or decreases the share of judges belonging to the defendant's gender, using different fixed effects. Columns 3–4 estimate the analogous impact on acquittal rates for court transitions that change the religion share of judges in the court (see Section 4 for specification details). Heteroskedasticity robust standard errors are reported below point estimates.