Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Baseline leniency versus severity responsiveness in prosecutorial charging decisions: Evidence from China's drunk-driving cases

Yuhao Wu

yuhaowu@pku.edu.cn

Peking University

Article

Keywords:

Posted Date: March 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8764192/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Baseline leniency versus severity responsiveness in prosecutorial charging decisions: Evidence from China's drunk-driving cases

Abstract

Prosecutorial charging decisions are a pivotal gateway into the criminal process, yet charging-stage consistency is difficult to assess because comparable, large-scale data are rarely observable. Drawing on a nationwide corpus of prosecutorial case documents from China's high-volume drunk-driving enforcement, this study asks whether local jurisdictions differ mainly in baseline willingness to prosecute or in how strongly they translate legally salient case seriousness into charging outcomes. We estimate multilevel logistic models that allow both baseline prosecution propensity and responsiveness to blood alcohol concentration (BAC) to vary across basic-level jurisdictions, while accounting for defendant characteristics, prior record, case circumstances, procedural factors, and year effects. The results indicate pronounced geographic heterogeneity on both dimensions. Jurisdictions differ not only in their overall inclination to file public prosecutions for otherwise comparable cases, but also in the steepness of the BAC–prosecution relationship, implying heterogeneous practical thresholds. Baseline propensity and severity responsiveness tend to move together, yet they do not collapse into a single "toughness" ranking: which jurisdictions appear stricter depends on case severity. Disparities are most consequential near the margin, where small differences in local screening thresholds produce large differences in who enters the criminal process. These findings highlight charging as a key site where spatial inequality can be generated and suggest that standardization and oversight should focus on severity-conditioned decision rules, especially for borderline cases.

Introduction

Prosecutorial charging decisions—whether the state initiates a formal public prosecution and on what basis—shape the entire downstream trajectory of criminal justice. In most systems, prosecutors control the gateway into adjudication, structure the factual and legal frame that courts ultimately see, and set the bargaining space that follows. Because this authority is both expansive and difficult to supervise in routine practice, charging discretion has long been treated as a recurrent stress test for rule-of-law commitments to consistency and equal treatment (Vorenberg, 1981; Davis, 2007; Sklansky, 2018).

Despite its centrality, prosecutorial decision-making remains empirically less visible than later-stage outcomes such as sentencing. A persistent theme in the literature is that the "black

box" of prosecution is hard to study not only because the decision is institutionally complex, but also because systematic charging records—especially those linkable to rich case facts—are often unavailable to researchers (Miller and Wright, 2008; Wright et al., 2022). This blind spot is not merely descriptive. If heterogeneity is generated upstream at the point of charging, reforms focused on later stages may leave the main sources of unequal treatment intact. Moreover, scholarship on "hydraulic" discretion argues that constraining judges through sentencing rules can shift practical leverage toward prosecutors by raising the stakes of charge selection and charge dismissal, further increasing the importance of understanding charging-stage variation (Stuntz, 2001; Starr and Rehavi, 2013).

A second, under-specified issue concerns the structure of geographic variation in charging. Jurisdictional differences are often treated as a single dimension—some places are simply more punitive than others. Yet research on court communities and organizational decision-making suggests a more nuanced possibility: local legal institutions may differ not only in their baseline tendency to proceed, but also in how strongly they condition decisions on legally salient indicators of seriousness (Johnson, 2006; Johnson et al., 2008). Translating this insight to charging implies a conceptually useful decomposition: baseline leniency, a jurisdiction's propensity to prosecute even for otherwise "average" cases, versus severity responsiveness, the steepness with which prosecution probability increases as case severity rises. These two mechanisms have different normative implications for "like cases treated alike," and different governance implications for standardization. If baseline leniency dominates, "strict" and "lenient" places should look strict or lenient across the severity distribution. If severity responsiveness differs, however, "where prosecution is stricter" becomes conditional on case severity, and the largest divergence should appear among borderline cases where small threshold shifts change who enters the criminal process.

China's drunk-driving enforcement provides an unusually strong setting for testing this decomposition because a continuous severity indicator—blood alcohol concentration (BAC)—is embedded in legal and policy frameworks, and because the offence is handled at mass scale. The dangerous-driving offence was introduced through the Eighth Amendment to China's Criminal Law in 2011, bringing drunk driving into the criminal law framework and contributing to a sharp increase in high-volume case processing (NPC Standing Committee, 2011). In the ensuing decade, dangerous driving became one of the most frequently handled offences and, in prosecutorial statistics, has been described as the most prosecuted crime in recent years—an institutional context in which "equal treatment" pressures and capacity pressures are simultaneously acute (Supreme People's Procuratorate, 2021; Supreme People's Procuratorate, 2023).

The legal–policy trajectory also foregrounds the kind of threshold governance that motivates this study. National guidance in 2013 set a baseline BAC criterion for defining

drunk driving and articulated seriousness-related factors to guide differentiated treatment (Supreme People's Court et al., 2013). Over time, practice and supplementary local policies created concerns about inconsistent standards across places, which national authorities later explicitly identified as a problem requiring clarification and unification (Supreme People's Procuratorate, 2024). In 2023, a new multi-agency opinion further specified how BAC and surrounding circumstances should be operationalized, including pathways for treating some low-severity cases as "obviously minor" rather than proceeding with full criminal prosecution (Supreme People's Court et al., 2023). Meanwhile, China's Criminal Procedure Law provides a general legal basis for discretionary non-prosecution in cases with minor circumstances, reinforcing the institutional relevance of prosecutorial screening as a tool of "minor-offence governance" (National People's Congress, 2018).

Empirically, this study leverages large-scale procuratorial documents—rather than court judgments—to observe charging outcomes directly at the prosecution stage. We compile a nationwide corpus of 1,836,317 case documents for drunk-driving dangerous-driving cases, including decisions to initiate public prosecution and decisions not to prosecute. The jurisdictional unit is operationalized at the basic level (county/district), corresponding to the local procuratorial authority recorded in the documents and aligning closely with the most granular unit of local criminal-justice governance in practice. For computational tractability and balanced representation across local units, the main multilevel models are estimated on a stratified analytical sample of 470,693 cases drawn from 2,694 basic-level jurisdictions, while controlling for defendant demographics, prior record, case circumstances (e.g., accident and injury indicators, hit-and-run, unlicensed driving, speeding/overloading/overcapacity), procedural factors (e.g., defense counsel and plea-related indicators), and year fixed effects.

Our empirical strategy uses multilevel logistic models in which both the intercept and the BAC slope vary by jurisdiction and are allowed to be correlated. This design directly operationalizes the paper's central question: Do jurisdictions differ mainly in baseline leniency, or in how they translate BAC-defined severity into charging decisions? The decomposition also yields interpretable quantities suited for a general social-science audience, including how widely predicted prosecution probabilities spread across jurisdictions at different points of the BAC distribution and whether "strict" jurisdictions remain strict across severity levels—or instead change rank as BAC increases.

Literature review and theoretical framework

Prosecutorial charging as discretionary governance

A long-standing view in legal and socio-legal scholarship is that prosecution is not a mechanical relay between police and courts, but a governance institution that actively structures who is translated into a defendant and under what legal narrative (Vorenberg, 1981;

Davis, 2007; Langer and Sklansky, 2017). Because declinations and charge selections are typically difficult to review through ordinary judicial processes, charging discretion has repeatedly been framed as a stress test for rule-of-law ideals of equal treatment and consistency in the everyday operation of criminal justice (Sklansky, 2016; Sklansky, 2018).

This concern is not purely normative; it is also distributional. A substantial line of work in U.S. criminal law and empirical legal studies argues that when judges are constrained through sentencing rules, the system's distributive work can shift upstream—toward prosecutors' decisions about what to file, what to dismiss, and what to bargain over (Stuntz, 2001; Starr and Rehavi, 2013). The implication for empirical research is straightforward: understanding disparity requires observing the gatekeeping stage rather than treating sentencing as the sole or primary generator of unequal outcomes.

Yet charging remains comparatively under-measured. The central obstacle is infrastructural as much as conceptual: systematic charging records are often fragmented, locally stored, or not linkable to rich case facts, making large-scale measurement difficult even in jurisdictions with otherwise extensive court data (Miller and Wright, 2008; Wright et al., 2022). Recent work on prosecutorial transparency and "open prosecution" underscores that the charging stage has historically been data-poor relative to sentencing, and that this opacity has limited both accountability and cumulative empirical knowledge (Garrett, 2023). This data constraint has shaped the literature: much more is known about geographic disparity in sentencing than about whether "like cases are treated alike" at the earlier charging stage.

Empirical findings on charging: seriousness, uncertainty, and organizational variation

Where charging data exist, empirical research consistently shows that prosecutorial decisions are shaped by legally salient factors such as offense seriousness, prior record, and indicators related to evidentiary strength and case provability. A classic theoretical articulation emphasizes uncertainty: prosecutors are risk managers operating under bounded information and institutional incentives, so they often calibrate decisions to both seriousness and perceived likelihood of conviction (Albonetti, 1987).

Empirical studies across diverse settings support this general pattern while also documenting heterogeneity that is not reducible to case facts alone. Quantitative research finds that prosecutors' filing, dismissal, and charge-reduction practices vary systematically across local jurisdictions and organizational contexts, even after conditioning on legally relevant predictors (Franklin, 2010; Shermer and Johnson, 2010). In addition, multi-stage perspectives emphasize that disparity can accumulate through successive discretionary sites—police, prosecutors, courts—so that charging decisions can be a major contributor to later-stage inequality (Kutateladze et al., 2014).

More recent research strengthens two points that are particularly relevant for the present

study. First, charging practices can exhibit large dispersion even when decision-makers are presented with substantively similar cases, suggesting that "prosecutorial heterogeneity" can be substantial at the individual or office level (Wright et al., 2022). Second, organizational accounts increasingly treat charging as an institutional process—shaped by office policies, supervision, workloads, and local professional norms—rather than solely as individualized legal judgment (Frederick and Stemen, 2012; Bowman, 2024). Together, these findings motivate an approach that treats local variation not as residual noise but as a substantive object of inquiry.

Geographic disparity and local legal institutions

A complementary body of theory explains inter-jurisdictional variation through the social organization of criminal case processing. The "court community" and "courtroom workgroup" traditions argue that repeated interaction among prosecutors, defense counsel, judges, and other courtroom actors produces stable local routines—shared understandings of what counts as a "normal" case, how to dispose of cases efficiently, and how to manage caseload pressure (Eisenstein et al., 1988; Nardulli et al., 1988; Flemming et al., 1992). These traditions were developed largely in relation to plea bargaining and sentencing, but their core insight generalizes: even under formally uniform law, local institutions can generate patterned divergence in practice.

Public administration theory provides a mechanism for why divergence may concentrate around decision margins. The street-level bureaucracy framework argues that frontline officials facing limited resources and high volumes develop coping strategies—categorization, routinization, and informal thresholds—that effectively become policy in action (Lipsky, 1980). Applied to criminal case processing, this perspective predicts that local variation should be most consequential for borderline cases, where small differences in internal thresholds decide whether a case is escalated into formal prosecution or disposed of informally. Related scholarship on high-volume "mass" case processing in misdemeanors similarly emphasizes that systems often operate through practical heuristics that shift burdens and consequences upstream of trial (Feeley, 1979; Kohler-Hausmann, 2018; Natapoff, 2018).

These theoretical traditions converge on two claims that matter for measurement. First, geographic disparity is plausibly generated by stable local institutional differences rather than random error. Second, local divergence may be severity-conditional, with the strongest divergence appearing near the margin where discretion is most action-consequential.

From one-dimensional "toughness" to a two-parameter charging function

A limitation in both policy discussion and some empirical work is the tendency to treat geographic disparity as one-dimensional: some jurisdictions are "tougher" while others are "more lenient." Methodologically, however, a major development in sentencing and court-community research has been to treat contextual disparity as multi-dimensional by allowing

not only baseline levels but also the operation of key predictors to vary across places (Ulmer and Johnson, 2004; Johnson, 2006; Ulmer, 2012). This move matters because it distinguishes two qualitatively different forms of contextual variation: **（1）Baseline leniency (level differences)**: jurisdictions differ in their general propensity to prosecute otherwise comparable cases—what a one-dimensional account would label "toughness" or "leniency." **（2）Severity responsiveness (slope differences):** jurisdictions differ in how strongly they translate a legally salient seriousness signal into charging decisions—i.e., the steepness of the severity-to-charging gradient and the implied severity thresholds at which prosecution becomes highly likely.

This distinction is not merely statistical. It carries different normative and governance implications. Baseline leniency variation is consistent with accounts emphasizing office culture, supervisory practices, resources, and local policy priorities. Severity responsiveness variation points to heterogeneous thresholds and conditional strictness: a jurisdiction may look "lenient" at one part of the seriousness distribution but "strict" at another. In practical terms, if slope heterogeneity is substantial, then "where prosecution is stricter" becomes severity-dependent, and equal-treatment concerns become especially salient for borderline cases.

The present study's contribution is to operationalize this two-parameter account at the charging stage—where empirical visibility is typically limited—using a setting in which seriousness is measured by a continuous, legally meaningful metric (BAC).

China: prosecution-centred case processing, minor-offence governance, and discretionary non-prosecution

China is theoretically informative for charging research because prosecutorial discretion is both institutionally central and closely tied to contemporary governance of high-volume, relatively low-level crimes. A growing English-language literature describes shifts toward prosecution-centredness in criminal case processing, particularly under reforms associated with negotiated/streamlined case handling and plea-related leniency arrangements, in which procuratorates play a larger role in structuring outcomes and courts may operate with constrained space for independent re-examination (He, 2023; Li, 2022; Wu and Lin, 2024).

At the same time, scholarship on China's procuratorial system emphasizes organizational and accountability dynamics—supervision, evaluation, and managerial incentives—that can shape how local procuratorates manage risk, workload, and consistency, providing plausible mechanisms for geographic divergence even under formally national rules (Mou, 2017; Mou, 2022).

Within this broader environment, minor-offence governance makes prosecutorial screening especially salient. Chinese criminal procedure formally authorizes non-prosecution

in minor cases and thereby creates a legal channel through which prosecutors can dispose of borderline cases without full adjudication (NPC Standing Committee, 2018). In a high-volume domain like drunk-driving dangerous driving—where BAC provides a salient seriousness signal and central authorities have repeatedly issued guidance aimed at standardization—this institutional design creates a sharp empirical question: do local procuratorial jurisdictions converge in practice, or do they diverge in baseline propensity and/or in severity-to-charging thresholds (NPC Standing Committee, 2011; Supreme People's Procuratorate et al., 2013; Supreme People's Procuratorate et al., 2023)?

Empirical hypotheses

Synthesizing the international literature on prosecutorial discretion, local legal institutions, and thresholding under high-volume constraints, and linking it to China's prosecution-centred minor-offence governance, this study advances the following empirical hypotheses:

H1 (Two-dimensional geographic heterogeneity). Local procuratorial jurisdictions exhibit substantial heterogeneity in both (a) baseline propensity to prosecute and (b) severity responsiveness to BAC, rather than differing only along a single "toughness" dimension (Ulmer and Johnson, 2004; Johnson, 2006; Wright et al., 2022).

H2 (Margin concentration). Cross-jurisdiction dispersion in predicted prosecution probabilities is greatest in borderline BAC ranges and narrows as BAC increases, consistent with informal thresholds and routinized coping under capacity constraints (Lipsky, 1980; Feeley, 1979; Kohler-Hausmann, 2018).

H3 (Coupling of baseline leniency and responsiveness). Baseline leniency and severity responsiveness are positively associated across jurisdictions, such that jurisdictions more prosecution-prone at baseline also react more sharply to increases in BAC—consistent with organizational accounts in which local risk management and managerial incentives generate correlated decision rules (Mou, 2022; Frederick and Stemen, 2012; Bowman, 2024).

These hypotheses align the paper's core conceptual distinction—baseline leniency versus severity responsiveness—with observable implications that can be tested using a multilevel charging model and BAC-indexed severity profiles.

Methods

Data and analytical sample

We assembled a nationwide corpus of procuratorial case documents on the offence of dangerous driving (drunk driving). These documents are produced by procuratorial authorities

and record whether the state initiates a formal public prosecution or issues a decision not to prosecute, alongside structured case attributes. The unit of analysis is the case. Jurisdiction is defined at the basic level (county/district), using the basic-level procuratorial authority recorded in the document—an operationalization that corresponds closely to the most local unit of justice governance in routine criminal processing.

The full corpus contains 1,836,317 documents. For multilevel estimation, we construct a jurisdiction-balanced analytical sample to ensure broad geographic coverage while keeping computation tractable. The main models are estimated on 470,693 cases drawn from 2,694 basic-level jurisdictions. Table 1 reports descriptive statistics for all variables used in the main analyses. Where covariates have missing or implausible values, we use simple, transparent preprocessing rules (described briefly below and fully documented in the Supplementary Information) to retain cases for estimation.

Measures. The outcome variable is public prosecution, coded as 1 if the procuratorate initiated public prosecution and 0 if the case resulted in non-prosecution.

The key explanatory variable is blood alcohol concentration (BAC) measured in mg/100 ml. BAC is a continuous and legally salient indicator of intoxication severity in drunk-driving enforcement. For estimation, BAC is standardized (z-score) so that coefficients are interpretable as changes associated with a one–standard deviation increase in BAC.

We adjust for a set of defendant-, incident-, and procedure-related covariates commonly recorded in the documents. These include age and sex; indicators for prior criminal record and prior DUI-related record; incident characteristics such as traffic accident, injury, and hit-and-run; legally relevant driving-condition indicators such as unlicensed driving and speeding/overloading/overcapacity; and post-event/procedural indicators such as plea leniency, forgiveness/understanding, compensation, and presence of defense counsel. Year fixed effects are included to absorb nationwide time-varying shifts in enforcement intensity and policy environments.

To make the model intercept interpretable as an “average case,” binary covariates are mean-centered (so the intercept corresponds to the sample-average value of each indicator rather than an all-zero profile), and age is standardized. For transparency and to maximize sample retention, missing binary indicators are coded as 0 when the document does not record the relevant circumstance, and implausible age values are treated as missing and imputed to the sample median before standardization.

Table 1. Descriptive statistics of study variables (analytical sample)

Variable	Mean	SD	Min	Max
Public prosecution (1 = prosecute)	0.869	0.338	0	1

BAC (mg/100 ml)	158.313	65.929	80	990
Age (years)	46.084	9.619	20	93
Male	0.975	0.155	0	1
Prior criminal record	0.118	0.323	0	1
Prior DUI-related record	0.027	0.162	0	1
Traffic accident	0.387	0.487	0	1
Injury	0.188	0.391	0	1
Hit-and-run	0.016	0.126	0	1
Unlicensed driving	0.122	0.327	0	1
Speeding/overloading/overcapacity	0.016	0.127	0	1
Plea leniency indicator	0.500	0.500	0	1
Compensation	0.082	0.274	0	1
Forgiveness/understanding	0.072	0.258	0	1
Defense counsel present	0.333	0.471	0	1

Notes: The analytical sample contains 470,693 cases from 2,694 basic-level jurisdictions (after jurisdiction-balanced sampling). For binary variables, the mean equals the proportion of cases with value 1. BAC is measured in mg/100 ml.

Analytical strategy

We use multilevel logistic regression to quantify cross-jurisdiction disparity in charging and to separate two conceptually distinct components: baseline leniency (stable differences in the general propensity to prosecute) and severity responsiveness (differences in how sharply prosecution probability rises with BAC).

Static benchmark of overall spatial disparity: ICC/VPC (and MOR) at mean BAC. We first report a single summary of how strongly prosecution outcomes cluster by local jurisdictions, following multilevel heterogeneity reporting practices commonly used in recent work on spatial disparity in Chinese criminal justice outcomes (Xiong et al., 2025). In logistic mixed models, a standard latent-variable approximation treats the level-1 residual variance as $\pi^2/3$. For a random-intercept model, the variance partition coefficient (VPC), often referred to as an intraclass correlation coefficient (ICC), is:

$$ICC \equiv \frac{\sigma_0^2}{\sigma_0^2 + \pi^2/3},$$

where $\sigma_0^2$is the between-jurisdiction variance of the random intercept (Merlo et al., 2006; Austin et al., 2017).

Our preferred specification includes both random intercepts and random slopes on BAC, so the between-jurisdiction variance in the linear predictor varies with BAC. Nevertheless, a useful static benchmark is obtained at $BAC_z = 0$(the mean BAC after standardization), where the between-jurisdiction variance reduces to the random-intercept variance σ_0^2. We therefore report the ICC evaluated at $BAC_z = 0$as a headline measure of baseline geographic clustering, and then use the random-slope results and BAC-indexed dispersion figures to

describe how heterogeneity changes with severity.

For interpretability on an odds-ratio scale, we also compute the median odds ratio (MOR) at $BAC_z = 0$:

$$MOR \equiv \exp\left(0.6745\sqrt{2\sigma_0^2}\right),$$

which can be read as the median multiplicative difference in prosecution odds between two randomly selected jurisdictions for otherwise identical "average" cases (Larsen et al., 2000; Merlo et al., 2006).

Multilevel charging model with correlated random intercepts and slopes. Let Y_{ij} indicate whether case i in jurisdiction j results in public prosecution ($Y_{ij} = 1$) rather than non-prosecution ($Y_{ij} = 0$). The main specification is:

$$\text{logit}\big(\Pr(Y_{ij} = 1)\big) = \beta_0 + \beta_1 BAC_{z,ij} + X_{ij}\gamma + \delta_{t(i)} + u_{0j} + u_{1j} BAC_{z,ij},$$

where $BAC_{z,ij}$ is standardized BAC, X_{ij} is the vector of controls, and $\delta_{t(i)}$ are year fixed effects.

The jurisdiction-specific random intercept u_{0j} captures baseline leniency (baseline prosecution propensity for otherwise average cases), while the random slope u_{1j} captures severity responsiveness (how strongly prosecution probability increases with BAC in jurisdiction j). We allow (u_{0j}, u_{1j}) to follow a bivariate normal distribution with an unrestricted covariance matrix:

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}\right),$$

so the intercept–slope correlation $\rho = \sigma_{01}/(\sigma_0\sigma_1)$ captures whether jurisdictions that are more prosecution-prone at baseline also respond more sharply to increases in BAC.

Translating model components into interpretable disparity profiles and implied thresholds. Because heterogeneity in log-odds units is difficult to interpret directly, we translate the fitted model into predicted prosecution probabilities for an "average" case. For each jurisdiction j, we compute predicted probabilities over a BAC grid while holding controls at their sample means (equivalently, setting centered indicators to zero) and evaluating year at the reference category used in estimation:

$$\hat{p}_j(b) = \text{logit}^{-1}\left(\hat{\beta}_0 + \hat{\beta}_1 b + \bar{X}\hat{\gamma} + \hat{\delta} + \hat{u}_{0j} + \hat{u}_{1j} b\right).$$

At each BAC value b, we summarize cross-jurisdiction dispersion using the distribution of

$\hat{p}_j(b)$across jurisdictions, focusing on the 5th–95th percentile spread:

$$Disparity(b) = \hat{p}_{95}(b) - \hat{p}_{05}(b).$$

Figures restrict BAC to the 1st–99th percentiles of the observed distribution for readability, while model estimation uses the full analytical sample. To avoid undue influence of small-sample jurisdictions in visualization, figures describing jurisdiction-level distributions exclude jurisdictions with fewer than 150 sampled cases.

Finally, to express severity responsiveness as an implied "threshold" quantity, we compute for each jurisdiction the implied BAC level at which an average case reaches a predicted prosecution probability of 0.90. Figure 3 presents the distribution of these implied thresholds, with the x-axis restricted to 80–350 mg/100 ml for readability.

Results

Overall cross-jurisdiction clustering: static ICC/VPC (and MOR) at mean BAC

We first quantify how strongly charging outcomes cluster by local jurisdictions. Evaluated at mean BAC ($BAC_z = 0$), the estimated static ICC/VPC on the latent (logit) scale is **0.454**, indicating that approximately **45%** of the residual variance in underlying prosecution propensity is attributable to between-jurisdiction differences rather than within-jurisdiction case-to-case variation, net of observed covariates and year fixed effects. As an optional interpretive metric, the corresponding MOR at $BAC_z = 0$is **4.84**, implying that for two otherwise identical "average" cases drawn from two randomly selected jurisdictions, the median difference in prosecution odds is nearly fivefold.

Main multilevel estimates

Table 2 reports estimates from the main multilevel logistic model in which both the jurisdiction-specific intercept and the jurisdiction-specific BAC slope vary and are allowed to be correlated, net of year fixed effects and observed case characteristics. The analytical sample includes **470,693** cases from **2,694** basic-level jurisdictions.

Consistent with BAC serving as a salient severity signal, prosecution likelihood rises steeply with BAC: a one–standard deviation increase in BAC is associated with a **2.09** increase in log-odds (**OR = 8.08**, 95% CI: 7.64–8.56). Several legally salient aggravating indicators are also positively associated with prosecution, including unlicensed driving (**OR = 7.42**, 95% CI: 6.90–7.97), traffic accidents (**OR = 4.25**, 95% CI: 4.03–4.47), and hit-and-run behavior (**OR = 3.64**, 95% CI: 2.96–4.49). Forgiveness/understanding is negatively associated with prosecution (**OR = 0.42**, 95% CI: 0.39–0.45), whereas compensation is not statistically distinguishable from zero in this specification.

The key results lie in the random effects. There is substantial cross-jurisdiction

heterogeneity in baseline prosecution propensity (**Var(intercept) = 2.736**) and in severity responsiveness measured by the BAC slope (**Var(slope) = 1.425**). Importantly, the intercept–slope correlation is positive (**Corr = 0.340**), implying that jurisdictions that are more prosecution-prone in baseline terms also tend to respond more sharply to increases in BAC. This pattern is consistent with geographically heterogeneous "severity-to-charging" mappings rather than a single, severity-invariant notion of local strictness.

Table 2. Multilevel logistic model of public prosecution decisions[1]

Panel A. Fixed effects (year fixed effects omitted)

Predictor	Coef. (log-odds)	SE	Odds ratio	95% CI (OR)
Intercept	1.734***	0.037	5.66	[5.27, 6.09]
BAC (standardized)	2.090***	0.029	8.08	[7.64, 8.56]
Age (standardized)	-0.061***	0.007	0.94	[0.93, 0.95]
Male (centered)	0.339***	0.038	1.40	[1.30, 1.51]
Prior criminal record (centered)	1.771***	0.034	5.88	[5.50, 6.28]
Prior DUI-related record (centered)	1.032***	0.067	2.81	[2.46, 3.20]
Traffic accident (centered)	1.447***	0.026	4.25	[4.03, 4.47]
Injury (centered)	0.319***	0.033	1.38	[1.29, 1.47]
Hit-and-run (centered)	1.293***	0.107	3.64	[2.96, 4.49]
Unlicensed driving (centered)	2.004***	0.037	7.42	[6.90, 7.97]
Speeding/overloading/overcapacity (centered)	0.224***	0.068	1.25	[1.10, 1.43]
Plea leniency indicator (centered)	2.225***	0.019	9.25	[8.91, 9.60]
Compensation (centered)	-0.041	0.043	0.96	[0.88, 1.04]
Forgiveness/understanding (centered)	-0.871***	0.042	0.42	[0.39, 0.45]
Defense counsel present (centered)	2.298***	0.020	9.95	[9.57, 10.35]

Panel B. Random effects and model fit

Quantity	Value
Var(random intercept)	2.736
Var(random BAC slope)	1.425
Corr(intercept, slope)	0.340
Cases (N)	470,693
Jurisdictions (J)	2,694
logLik	-82,529.45
AIC	165,118.89

[1] Model includes year fixed effects (omitted from table). BAC and age are standardized (mean 0, SD 1). Binary covariates are mean-centered. Random intercept and random slope on BAC are estimated at the basic-level jurisdiction (county/district) level.

Notes:

(1) Year fixed effects are included (reference year = 2021) but omitted for space.

(2) "Centered" binary variables are mean-centered; coefficients retain the same 0→1 contrast magnitude, but the intercept corresponds to an "average case" (covariates at their means).

*(3) Significance: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.*

Disparity is concentrated near the margin

Figure 1 visualizes how cross-jurisdiction heterogeneity changes across the BAC distribution. Even after conditioning on observed defendant characteristics, case circumstances, procedural indicators, and year fixed effects, predicted prosecution probabilities for an "average" case exhibit substantial dispersion across jurisdictions. Crucially, this heterogeneity is not constant: it is largest in borderline cases at relatively low BAC levels and shrinks as BAC increases, consistent with the idea that local discretion is most consequential near implicit decision thresholds.

At the point of maximum dispersion (BAC ≈ 97.19 mg/100 ml), the predicted prosecution probability ranges from approximately 0.056 (5th percentile jurisdiction) to 0.909 (95th percentile jurisdiction), a gap of 0.853 in absolute probability for otherwise similar cases. As BAC rises, dispersion declines steadily, reflecting convergence toward near-certain prosecution among high-severity cases.

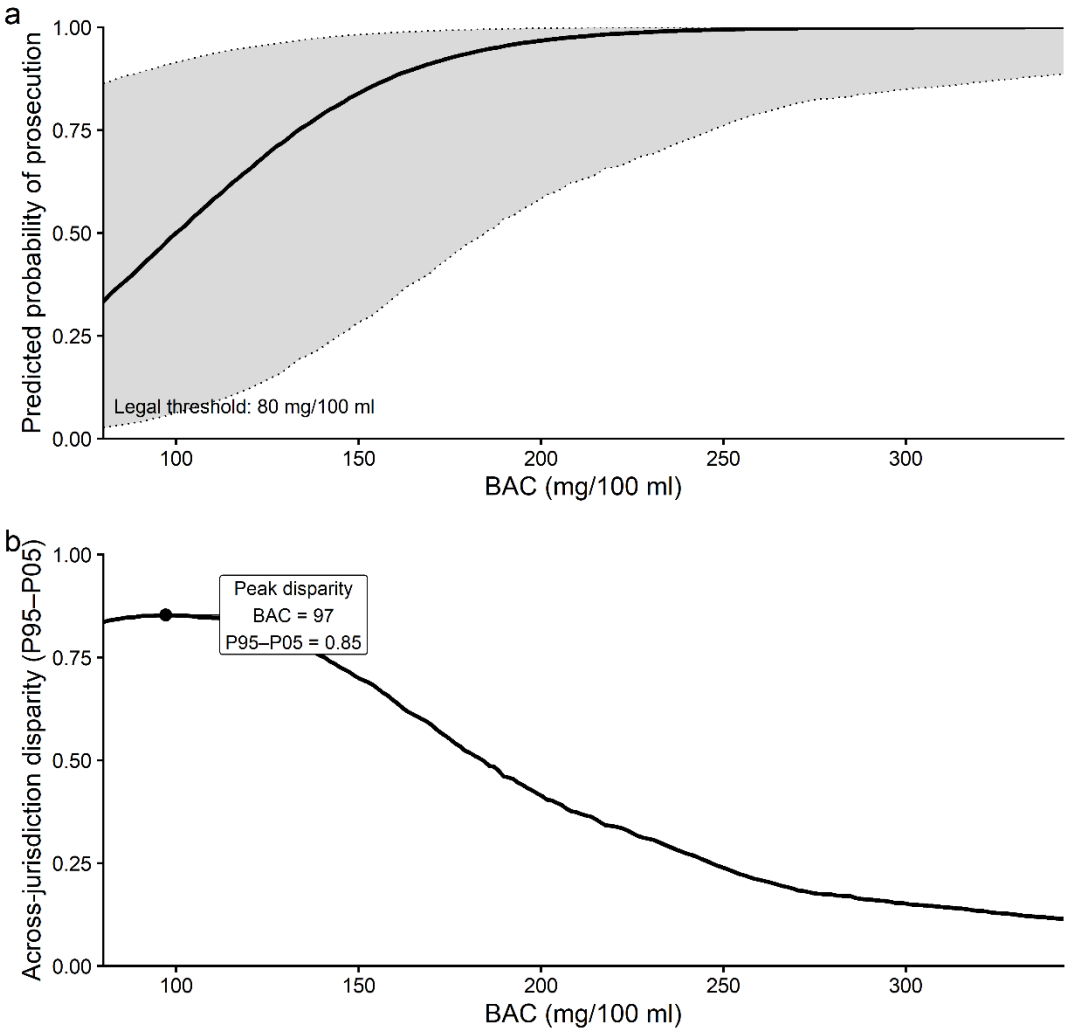


Fig. 1 | Cross-jurisdiction dispersion in prosecution probabilities varies sharply over BAC.[2]

Baseline leniency and severity responsiveness are positively coupled

Figure 2 plots jurisdiction-specific random intercepts against random slopes on BAC. The positive association indicates that baseline prosecution propensity and severity responsiveness tend to move together: jurisdictions that are more prosecution-prone at baseline also tend to increase prosecution probability more sharply as BAC rises. This coupling is consistent with the model-implied positive

[2] **a**, Predicted probability of public prosecution as a function of blood alcohol concentration (BAC, mg/100 ml) for an "average" case, showing the cross-jurisdiction distribution (5th–95th percentiles; shaded band) and the median (solid line). Predictions are generated from the multilevel logistic model with covariates and year fixed effects (Model m3_ctrl), with covariates held at their sample means (centered controls set to zero) and year evaluated at the reference category. **b**, Cross-jurisdiction disparity measured as the 5th–95th percentile spread (P95–P05) in predicted prosecution probabilities at each BAC value. For readability, BAC is restricted to the 1st–99th percentiles of the observed distribution; jurisdictions with fewer than 150 sampled cases are excluded from the figure.

intercept–slope correlation (**ρ ≈ 0.34**), reinforcing the interpretation that local "strictness" is multidimensional rather than reducible to a single parameter.

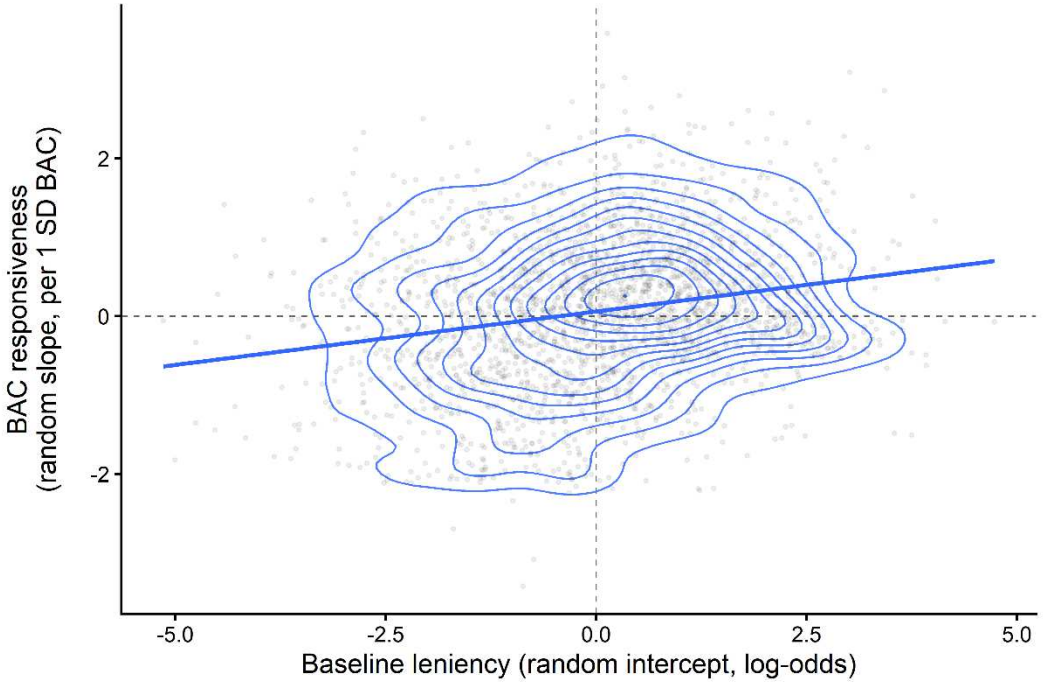


Figure 2 | Baseline leniency and severity responsiveness across jurisdictions.

Each point represents a basic-level jurisdiction (county/district). The x-axis shows the jurisdiction-specific random intercept (baseline propensity to prosecute), and the y-axis shows the jurisdiction-specific random slope on standardized BAC (responsiveness to case severity). Contour lines summarize the density of jurisdictions. The fitted line is from a linear regression of slope on intercept. The multilevel model estimates a positive intercept–slope correlation (**ρ ≈ 0.34**), indicating that jurisdictions that are more prosecution-prone at baseline also tend to increase prosecution probability more sharply as BAC rises.

Using the fitted multilevel model, we recover an implied BAC threshold for each jurisdiction—the BAC level at which an average case reaches a predicted prosecution probability of 0.90. The distribution of these implied thresholds is wide. In the analytical sample of stable jurisdictions (N ≥ 150 per jurisdiction), the median implied 0.90-threshold is approximately 166 mg/100 ml, with an interquartile range of roughly 138–204 mg/100 ml. The dispersion extends further into the tails: the 5th to 95th percentile range spans approximately 94–337 mg/100 ml. This broad threshold heterogeneity complements the random-slope evidence: jurisdictions differ not only in baseline propensity to prosecute, but also in the implied severity threshold at which prosecution becomes near-certain.

Implied local prosecution thresholds vary widely

To translate the random-intercept/random-slope structure into a threshold quantity, we

recover an implied BAC threshold for each jurisdiction—the BAC level at which an "average" case reaches a predicted prosecution probability of 0.90 (holding covariates at their centered means and using the model's year reference category). Figure 3 shows that implied thresholds are widely dispersed. The median implied 0.90-threshold is approximately **165.59 mg/100 ml**, with an interquartile range of roughly **138.39–203.83 mg/100 ml**. The tails are substantial: the 5th to 95th percentile range spans approximately **94.41–337.32 mg/100 ml**.

For readability, Figure 3 restricts the x-axis to **80–350 mg/100 ml**, which contains the bulk of the implied-threshold distribution; **7.72%** of jurisdictions fall outside this displayed range.

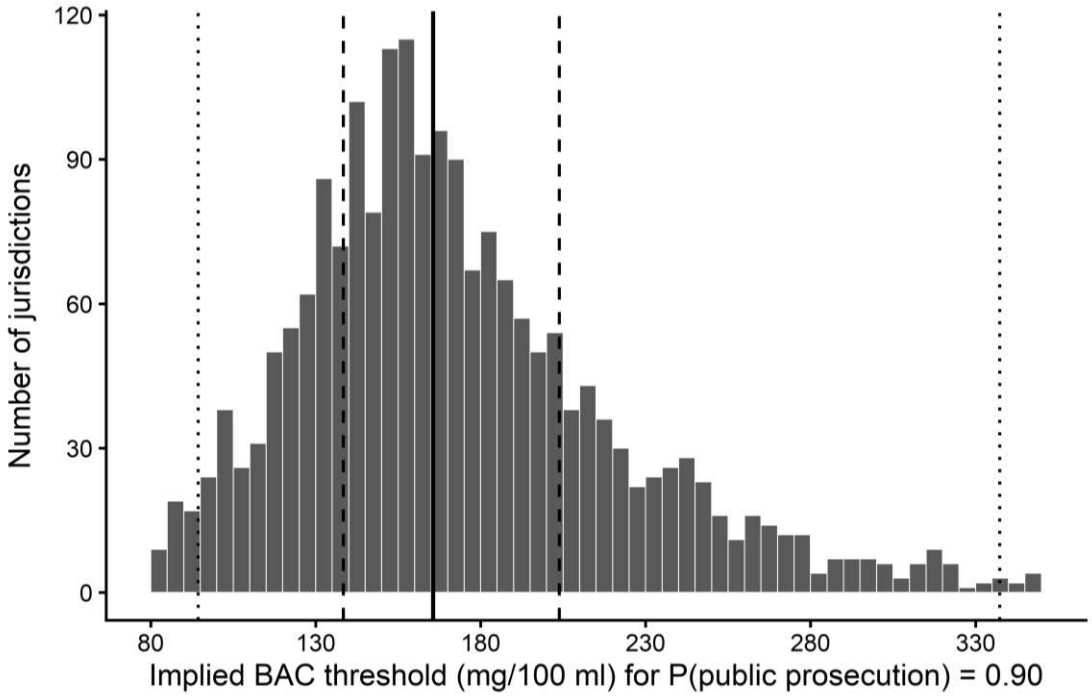


Figure 3. Distribution of implied prosecution thresholds across jurisdictions.

Histogram of the implied BAC threshold at which an "average" case reaches a predicted probability of public prosecution of 0.90, computed from the multilevel logit model with jurisdiction-specific random intercepts and random slopes on BAC (controls held at their centered means; year fixed effects with 2021 as the reference category). Vertical lines denote the median (solid), interquartile range (dashed), and 5th/95th percentiles (dotted). For readability, the x-axis is restricted to 80–350 mg/100 ml; 7.72% of jurisdictions have implied thresholds outside this window.

Discussion and conclusion

This study brings the prosecutorial charging/non-prosecution stage into the empirical conversation about "like cases treated alike." Charging is widely recognized as the criminal

process's principal gateway: it determines whether a person enters the adjudicative pipeline, the legal framing under which courts see the case, and the bargaining space that follows, while remaining difficult to supervise through ordinary judicial review in routine practice (Vorenberg, 1981; Davis, 2007; Sklansky, 2016). Yet charging has also been comparatively difficult to study at scale because systematic, linkable prosecution-stage records are often unavailable, leaving a persistent empirical "black box" relative to downstream outcomes such as sentencing (Miller and Wright, 2008; Wright et al., 2022). Using a nationwide corpus of procuratorial documents in China's drunk-driving dangerous-driving cases and a multilevel random-intercept/random-slope design, we make that upstream discretion empirically visible and decompose geographic disparity into two conceptually distinct components: baseline leniency (jurisdiction-specific intercepts) and severity responsiveness (jurisdiction-specific slopes on BAC). The central finding is that both dimensions are substantively important and jointly imply that "which places are stricter" is conditional on case severity rather than severity-invariant.

Even after conditioning on observed case circumstances, procedural indicators, and year fixed effects, prosecution outcomes remain strongly structured by local jurisdiction. The static latent-scale ICC/VPC evaluated at mean BAC indicates substantial between-jurisdiction clustering in the underlying propensity to prosecute, consistent with the broader multilevel literature's view that spatial heterogeneity can remain large even under formally uniform legal rules (Snijders and Bosker, 2012; Xiong et al., 2025). On an interpretive scale, the corresponding MOR at mean BAC implies that two otherwise identical average cases drawn from two randomly selected jurisdictions can differ markedly in prosecution odds—a magnitude that is difficult to reconcile with a purely technical view of geographic variation as minor noise (Merlo et al., 2006). This is particularly salient in the drunk-driving domain because it is a high-volume offence category that has repeatedly been framed by Chinese authorities as requiring standardized handling after the offence's criminalization (NPC Standing Committee, 2011; Supreme People's Court et al., 2013; Supreme People's Court et al., 2023).

The most policy-relevant pattern is not merely that jurisdictions differ, but that disparity is sharply severity-dependent. Figure 1 shows that cross-jurisdiction dispersion in predicted prosecution probability peaks among borderline cases at relatively low BAC and then contracts as BAC rises and prosecution approaches near-certainty in high-severity cases. This "largest divergence near the margin" profile aligns closely with theoretical accounts that emphasize thresholding and routinization under capacity constraints. In the street-level bureaucracy tradition, frontline legal actors facing heavy caseloads and bounded information develop coping strategies—categorization, routinization, and informal cutpoints—that effectively become "policy in action," with discretion most consequential precisely where

cases are contestable (Lipsky, 1980). Socio-legal research on local legal institutions similarly emphasizes stable local routines and shared working understandings that can produce patterned geographic divergence even under common formal law (Eisenstein et al., 1988). In other words, the evidence is consistent with the view that what varies geographically is not only "overall toughness," but the administration of practical thresholds—exactly the kind of variation that equal-treatment ideals find hardest to discipline because it operates inside routine screening rather than through overtly reviewable adjudication (Vorenberg, 1981; Sklansky, 2016).

A second contribution is to show that spatial disparity in charging is multi-dimensional. The random-intercept variance indicates substantial baseline differences in prosecution propensity, while the random-slope variance indicates that jurisdictions also differ meaningfully in how strongly they translate BAC-defined severity into prosecution. Moreover, the positive intercept–slope correlation (Figure 2) implies that these are not independent axes: jurisdictions that are more prosecution-prone at baseline also tend to escalate more sharply with increasing BAC. This finding matters for both theory and governance. Theoretically, it supports the methodological and substantive move—developed most prominently in sentencing and court-community research—toward treating contextual disparity as multi-parameter, where not only baseline levels but also the operation of key predictors varies across places (Ulmer and Johnson, 2004; Johnson, 2006). From a governance perspective, the coupling implies that jurisdictional "strictness" is not well summarized by a single league-table ranking. A jurisdiction may appear relatively lenient at one part of the BAC distribution yet relatively strict elsewhere because its severity gradient is steeper. This conditionality echoes the broader institutional argument that when discretion is constrained at later stages, distributive leverage can shift upstream toward prosecutors, increasing the importance of understanding not only levels of prosecution but also how prosecutors administer gradients and thresholds of seriousness (Stuntz, 2001; Starr and Rehavi, 2013).

Figure 3 translates the random-intercept/random-slope structure into an intuitive threshold quantity: the implied BAC at which an average case reaches a predicted 0.90 probability of prosecution. This translation is substantively informative because it maps abstract random effects into a governance parameter that closely matches how practitioners and policymakers often conceptualize routine case handling: the severity point at which prosecution becomes effectively "nearly certain." The wide dispersion of implied thresholds indicates that some jurisdictions reach near-certain prosecution at BAC levels close to the legal criminalization boundary, while others reach it only at much higher BAC, even after adjusting for observable case circumstances and procedural indicators. In the Chinese institutional context, this matters because the criminal process formally authorizes

discretionary exits for minor cases and, in recent policy guidance, explicitly seeks to clarify the handling of lower-severity drunk-driving cases (NPC Standing Committee, 2018; Supreme People's Court et al., 2023). Our results suggest that, in practice, local charging authorities continue to differ substantially in how they operationalize seriousness calibration in precisely the region where discretion is most action-consequential.

These findings carry an implication for standardization that follows directly from the paper's baseline-versus-responsiveness framework. If the largest geographic divergence is concentrated at the margin, then reforms that primarily target average rates of prosecution or non-prosecution will be poorly aligned with where inequity is generated. Instead, the most leverage for reducing unequal treatment lies in the seriousness-conditioned administration of borderline cases: clarifying, auditing, and supervising the application of "minor circumstance" exits and seriousness bands in the low-to-moderate BAC range, where small shifts in implicit thresholds produce large shifts in who enters the criminal process. This also suggests that monitoring should be distributional rather than purely average-based. Because baseline leniency and responsiveness are coupled, and because strictness is severity-conditional, a monitoring approach that reports only overall prosecution rates risks obscuring the most consequential disparity—differences in how jurisdictions translate increments of BAC into charging decisions near key severity regions.

Several limitations should frame interpretation and suggest avenues for future research. First, although BAC is an unusually salient and continuous severity signal, prosecutors plausibly respond to additional case features not fully observed in the documents (e.g., evidentiary strength, negotiation dynamics, or local policing and referral practices). To the extent these unobservables vary geographically, part of the estimated jurisdiction effects may reflect differences in case mix or upstream enforcement rather than charging standards alone. Second, while the jurisdictional unit corresponds to the local basic-level charging authority recorded in the procuratorial documents, observed heterogeneity may also incorporate local institutional ecology more broadly, including policing, detention practices, and defense availability. Research designs that link referral-stage information, exploit policy discontinuities, or incorporate richer local institutional covariates could better separate these channels. Third, the main multilevel models are estimated on a stratified analytical sample for tractability and balanced jurisdictional representation. This design is appropriate for estimating variance components and for characterizing how heterogeneity changes over severity, but future work could assess sensitivity to alternative sampling intensities or implement scalable estimation approaches for very large hierarchical generalized linear models.

Taken together, the evidence suggests that local discretion in prosecutorial charging persists—and does so in a patterned way—in a domain where formal severity measurement is

unusually clear and where national authorities have repeatedly signaled a desire for standardized handling (NPC Standing Committee, 2011; Supreme People's Court et al., 2013; Supreme People's Court et al., 2023). The main theoretical takeaway is that spatial inequality at the charging stage is best understood as a two-parameter decision function rather than a single "toughness" dimension. Jurisdictions differ both in baseline propensity and in severity responsiveness, and the resulting disparity is concentrated at the margin where discretionary exits are feasible. If the rule-of-law aspiration of consistent treatment is to be advanced at the charging stage, the results imply that the key governance problem is threshold administration: aligning how local prosecutors translate legally salient seriousness signals into charging decisions, especially in borderline cases where discretion is most consequential.

References

Albonetti, C. A. Prosecutorial discretion: The effects of uncertainty. **Law Soc. Rev.** 21, 291–313 (1987).

Bowman, N. A., Lowrey-Kinberg, B. & Gould, L. *A prosecutorial decision-making model: case characteristics, district context, and defendant race and ethnicity.* **Law Soc. Rev.** 58, 452–480 (2024). https://doi.org/10.1017/lsr.2024.24

Davis, A. J. *Arbitrary Justice: The Power of the American Prosecutor* (Oxford University Press, 2007).

Eisenstein, J., Flemming, R. B. & Nardulli, P. F. *The Contours of Justice: Communities and Their Courts* (Scott, Foresman, 1988).

Feeley, M. M. *The Process Is the Punishment: Handling Cases in a Lower Criminal Court* (Russell Sage Foundation, 1979).

Flemming, R. B., Nardulli, P. F. & Eisenstein, J. *The Craft of Justice: Politics and Work in Criminal Court Communities* (University of Pennsylvania Press, 1992).

Franklin, T. W. *The local influence on prosecutorial dismissals: a multilevel analysis of case- and county-level factors.* **J. Crim. Justice** 38, 494–501 (2010).

Frederick, B. & Stemen, D. *The Anatomy of Discretion: An Analysis of Prosecutorial Decision-Making in New York County, New York* (Vera Institute of Justice, 2012).

Garrett, B. L., Crozier, L., Gifford, C. D., Grodensky, A., Quigley-McBride, A. & Teitcher, M. *Open prosecution.* **Stan. L. Rev.** 75, 1365–1430 (2023).

He, X. *Plea leniency and prosecution-centredness in China's criminal process*. **China Q.** 255, 659–683 (2023). https://doi.org/10.1017/S0305741023000854

Johnson, B. D. *The multilevel context of criminal sentencing: integrating judge- and county-level influences.* **Criminology** 44, 259–298 (2006). https://doi.org/10.1111/j.1745-9125.2006.00047.x

Johnson, B. D., Ulmer, J. T. & Kramer, J. H. *The social context of guidelines circumvention: the case of federal district courts.* **Criminology** 46, 737–783 (2008). https://doi.org/10.1111/j.1745-9125.2008.00117.x

Kohler-Hausmann, I. *Misdemeanorland: Criminal Courts and Social Control in an Age of Broken Windows Policing* (Princeton University Press, 2018).

Kutateladze, B. L., Andiloro, N. R., Johnson, B. D. & Spohn, C. *Cumulative disadvantage: examining racial and ethnic disparity in prosecution and sentencing.* **Criminology** 52, 514–551 (2014). https://doi.org/10.1111/1745-9125.12047

Langer, M. & Sklansky, D. A. (eds.) *Prosecutors and Democracy: A Cross-National Study* (Cambridge University Press, 2017).

Li, E. *Haste makes waste: why China's new plea leniency system is doomed to fail.* **Asian J. Comp. Law** 17, 76–105 (2022). https://doi.org/10.1017/asjcl.2022.2

Lipsky, M. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services* (Russell Sage Foundation, 1980).

Miller, M. K. & Wright, R. F. *The black box.* **Iowa L. Rev.** 94, 125–196 (2008).

Mou, Y. T*he constructed truth: the making of police dossiers in China.* **Soc. Leg. Stud.** 26, 69–88 (2017). https://doi.org/10.1177/0964663916664303

Mou, Y. *Accountable for a lifetime: reforms in prosecutorial accountability in China.* **China Law Soc. Rev.** 6, 36–78 (2022). https://doi.org/10.1163/25427466-07010003

Natapoff, A. *Punishment Without Crime: How Our Massive Misdemeanor System Traps the Innocent and Makes America More Unequal* (Basic Books, 2018).

Nardulli, P. F., Eisenstein, J. & Flemming, R. B. *The Tenor of Justice: Criminal Courts and the Guilty Plea Process* (University of Illinois Press, 1988).

Sklansky, D. A. *The nature and function of prosecutorial power.* **J. Crim. Law Criminol.** 106, 473 (2016).

Sklansky, D. A. *The problems with prosecutors.* **Annu. Rev. Criminol.** 1, 451–469 (2018). https://doi.org/10.1146/annurev-criminol-032317-092226

Starr, S. B. & Rehavi, M. M. *Mandatory sentencing and racial disparity: assessing the role of prosecutors and the effects of Booker.* **Yale Law J.** 123, 2–80 (2013).

Stuntz, W. J. *The pathological politics of criminal law.* **Mich. L. Rev.** 100, 505 (2001).

Ulmer, J. T. *Recent developments and new directions in sentencing research.* **Justice Q.** 29, 1–40 (2012). https://doi.org/10.1080/07418825.2011.624115

Ulmer, J. T. & Johnson, B. D. *Sentencing in context: a multilevel analysis.* **Criminology** 42, 137–178 (2004). https://doi.org/10.1111/j.1745-9125.2004.tb00516.x

Vorenberg, J. *Decent restraint of prosecutorial power.* **Harv. L. Rev.** 94, 1521–1573 (1981).

Wright, R. F., Baughman, S. B. & Robertson, C. T. I*nside the black box of prosecutor discretion.* **UC Davis L. Rev.** 55, 2133–2207 (2022).

Wu, Y. & Lin, L. *Constrained power expansion: China's procuratorial reforms within and beyond criminal justice.* **Mod. China** 50, 371–401 (2024).

Xiong, M., Xia, Y. & Yu, X. *Sentencing equilibrium in rape cases: a legal and political explanation of jurisdictional uniformity in China.* **Humanit. Soc. Sci. Commun.** 12, 59 (2025). https://doi.org/10.1057/s41599-025-04368-z

Chinese statutes and policy documents

National People's Congress. *Criminal Law of the People's Republic of China* (中华人民共和国刑法) (1997; as amended).

Standing Committee of the National People's Congress. *Criminal Law Amendment (VIII) of the People's Republic of China* (中华人民共和国刑法修正案（八）) (2011).

Standing Committee of the National People's Congress. *Criminal Procedure Law of the People's Republic of China* (中华人民共和国刑事诉讼法) (2018 Amendment).

Supreme People's Court, Supreme People's Procuratorate & Ministry of Public Security. *Opinion on Several Issues Concerning the Application of Law in Handling Criminal Cases of Driving Motor Vehicles While Intoxicated* (关于办理醉酒驾驶机动车刑事案件适用法律若干问题的意见；法发〔2013〕15 号) (2013).

Supreme People's Court, Supreme People's Procuratorate, Ministry of Public Security & Ministry of Justice. *Opinions on Handling Criminal Cases of Dangerous Driving While Intoxicated* (关于办理醉酒危险驾驶刑事案件的意见) (2023).

Data availability. The data underlying this study are derived from publicly accessible Chinese procuratorial case documents. The compiled research dataset contains large-scale case-level records and sensitive personal information, which prevents unrestricted public release.

For the purposes of peer review, the full raw data, cleaned analytical datasets, variable definitions, and all statistical code and scripts required to reproduce the results have been made available to the editors and reviewers.

Access to the data may be granted upon reasonable request to the corresponding author, subject to applicable legal, ethical, and data protection requirements.

Code availability. The statistical analysis was conducted in R. Replication code sufficient to reproduce the models and figures can be made available upon reasonable request.

Competing interests. The authors declare no competing interests.

Ethical Approval & Informed Consent. This article does not contain any studies with human participants performed by any of the authors.