Prosecutorial Policy Analysis

AI-Powered Measurement of Criminal Justice Reform in California

Dvir Yogev

February 2026

Center for Law & Justice, UC Berkeley School of Law

The Measurement Gap

District Attorneys are among the **most powerful actors** in the criminal justice system, yet we lack systematic measurement of:

- How their policies **vary across jurisdictions**
- How policies **change over time**
- Whether stated policy intent **affects outcomes**

Existing research relies on case outcomes or campaign rhetoric—neither captures the *stated policy intent* that guides line prosecutors daily.

What's Missing

No systematic, comparable measurement of internal DA policy documents exists **anywhere**.

What We Built

The first large-scale, AI-coded dataset of internal DA policy documents—infrastructure for causal inference about prosecutorial ideology.

Data: ACLU NorCal Public Records Archive

2,665 internal DA policy documents obtained by ACLU of Northern California via the California Racial Justice Act.
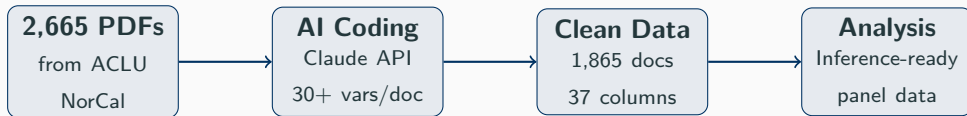
- **41 of 58** California counties represented
- Charging directives, sentencing memos, diversion protocols, racial equity initiatives, bail reform orders
- Documents span 2015–2024

These are *internal operating documents*—not press releases or campaign materials. They reveal what prosecutors actually tell their staff to do.

Why This Is Rare

Internal policy memos are almost never publicly available. The ACLU's PRA effort created a **unique research opportunity** that may not persist indefinitely.

What We Built: Research Pipeline

2,665 PDFs	→	AI Coding	→	Clean Data	→	Analysis
from ACLU		Claude API		1,865 docs		Inference-ready
NorCal		30+ vars/doc		37 columns		panel data

Dimension Coded	What's Measured
Ideological Orientation	7-point scale: clearly progressive → clearly traditional
Extensive Margin	Impact on *who enters* the system (charging, diversion)
Intensive Margin	Impact on *how severely* people are treated (sentencing)
Specific Policies	Diversion, bail reform, enhancements, three strikes, racial justice
Administrative Context	New policy vs. continuation, DA administration

Total API cost: ~$80   ·   Processing time: ~2 hours

Face Validity: The Pipeline Recovers Known Patterns

Gascón Transformation

LA County ideology score **tripled** under Gascón

Cohen's $d = 0.75$, $p < 0.001$

Geographic Clustering

- **Progressive:** Sacramento $+78\%$, Yolo $+56\%$, San Diego $+50\%$
- **Traditional:** Stanislaus -34%, Placer -21%
- Bay Area variation: Santa Clara $+0.84$ vs Alameda -0.15

2020 Racial Justice Surge

Racial justice emphasis jumped **+30pp** in one year ($12\% \rightarrow 42\%$), tracking the post-George Floyd moment precisely. Documents with high racial justice emphasis are **4.6×** more likely to be progressive ($\chi^2 = 421$, $p < 0.001$).

Why This Matters

Pipeline confirms patterns any expert would expect—evidence of validity, not artifact.

Descriptive Analysis: Theory-Grounded Patterns

Extensive > Intensive Margin

Recent reforms disproportionately emphasize **who enters** the system (**33.9%** extensive lenient) over **how severely** people are treated (**22.6%** intensive lenient). Political economy logic: diversion/declination less visible to voters than sentencing leniency—safer reforms for DAs facing reelection.

Close Elections → Progressive Policy

Elections with margins $\leq$15pp produce **+31.2pp** more progressive policies ($p = 0.010$). Continuous: $r = -0.50$ between margin and ideology ($p = 0.009$). Provides credible first stage for instrumental variables designs.

Novel contributions: The extensive-over-intensive pattern has not been documented at this scale. The election-ideology link provides the first stage for causal designs.

In the Service of Causal Inference: Policy Shocks

Disruptions Detected

- **9** significant policy disruptions (2020–23)
- SF 2020 (Boudin): highest score (0.572)
- **347** novel reform adoptions tracked

Progressive Surge (2019–2022)

- Progressive docs: $18\% \rightarrow 40\% \rightarrow 56\%$
- Linear trend: $+0.062$/yr ($p = 0.003$, $R^2 = 0.52$)

Why Funders Should Care

- Each disruption $=$ potential "natural experiment"
- Staggered adoption $\rightarrow$ modern DiD/event study designs
- Combined with outcome data $\rightarrow$ credible causal estimates

Jail Data Pilot: First Steps Toward Causal Evidence

Finding	Estimate	Significance	Survives Controls?
Ideology ↔ Jail Pop Rate	$r = -0.222$	$p = 0.009$	**Yes** (year-demeaned)
Ideology ↔ Jail Admissions	$r = -0.221$	$p = 0.009$	**Yes** (year-demeaned)
Progressive vs Traditional	-68.5/100k	$d = -0.81$	**Yes**
LA Pretrial (Gascón DiD)	-32.1/100k	Pre-trend $p = 0.90$	**Cleanest**

What Holds Up

Associations **survive year controls**.
Pretrial detention shows cleanest causal
signal (pre-trend $p = 0.90$).

What the Pilot Exposes

Naive pre/post **confounded by COVID**.
Parallel trends violated ($p = 0.021$). TWFE
non-significant ($N = 137$).

What This Infrastructure Can Test

DA Policy (Coded)	Measurable Outcome	Data Required
Bail/pretrial reform	Pretrial detention, FTA, pretrial crime	Vera (done); CA DOJ; CJARS
Diversion programs	Case dismissals, recidivism (1–3 yr)	UniCourt; CJARS recidivism
"Decline to prosecute" memos	Filing rates: drugs, property, QOL	CA DOJ OpenJustice (free)
Enhancement/3-strikes reform	Sentence lengths, enhancement filings	CA Sentencing Comm.; CJARS
Racial equity directives	B/W disparities: charging, sentencing	CA DOJ by-race; CJARS
Extensive margin (who enters system)	System contact per capita; jail admissions	CA DOJ filings; Vera (done)

The policy variation exists. What's needed: outcome linkage (Extension 2) and geographic scale (Extension 1) to power DiD, RDD, event study, and synthetic control designs.

Extension 1: Expanding Policy Coverage

Within California

- Complete remaining **17 counties** via PRA
- **Pre-2015 historical** documents for longer pre-periods
- **200-doc human validation** for publication credibility

100 Largest DA Offices Nationwide

- ~65% of US population covered
- **30–50 close elections** for RDD (vs ~5 in CA)
- Tests external validity across **different legal cultures**

Scope	Scale	Enables	Time	Cost
Complete CA + historical	~500 docs	Full panel; longer pre-periods	4–6 mo	~$25–35k
100 largest US DA offices	5–10k docs	Powered RDD; national benchmarking	12–18 mo	~$100–150k
Human validation (200 docs)	200 dual-coded	IRR; publication credibility	3 mo	~$20k

Extension 2: Merging with Outcome Data

Administrative Data (Available Now)

Source	Records
CA DOJ OpenJustice	County-year arrest, filing, incarceration
CA Sentencing Comm.	Enhancement filings, 3-strikes usage
Vera (done)	Quarterly jail pop by county

Case-Level Data (Higher Return)

Source	Records
UniCourt / PACER	Case dispositions, plea bargains
CJARS	Arrest $\rightarrow$ parole, 24 states

CJARS (Census Bureau + U. Michigan)

Harmonized individual-level criminal justice records across 24 states. Tracks people from arrest through parole. Enables testing whether a progressive DA's diversion memo actually reduces the probability an arrested individual proceeds to prosecution. Accessible via FSRDC (3–12 month application).

Value for Policymakers

- **DA accountability scorecards**—comparable measures of how each office's stated policies align with its goals

- **Rhetoric vs. practice**—does a "decline low-level drugs" memo actually reduce drug prosecutions?

- **Benchmarking tools**—a DA considering bail reform can see how similar policies performed elsewhere

- **Causal evidence on the progressive prosecutor model**—answering the central question: does it work?

Related Work

Felix Owusu's AV-funded project exploits two specific internal DA memos for a causal design. Our contribution scales this logic: **systematic measurement across 41+ offices and 1,865 documents**, creating variation needed for average causal effects and heterogeneity analysis.

11

Recommended Funding Package

$200–265k over 24–36 months

The infrastructure is built. The marginal cost of each extension is low—the pipeline, coding schema, and analysis framework are ready to scale.

Why Fund This Now

Unique Advantages

- **Infrastructure is built**—pipeline, schema, analysis framework operational
- **Data window is closing**—ACLU PRA archive may not persist
- **No competitor**—no other systematic, AI-coded DA policy database exists
- **Pilot-tested**—demonstrated linkage to outcomes; identified the right methods

Policy Impact

- Answers the question **funders care about most**: do progressive reforms actually work?
- Enables **evidence-based evaluation** of DA accountability efforts
- Creates **benchmarking tools** for jurisdictions
- First **causal estimates** of prosecutorial policy effects

Let's Talk

The descriptive associations are real. The causal question is answerable.
The infrastructure is ready. We need your support to get there.

Dvir Yogev

Post-Doctoral Researcher · UC Berkeley School of Law

dyo@berkeley.edu

github.com/dyo2112/prosecutor-policies-causal-inference

Full pilot report, figures, and replication code available in repository

Backup Slides

Case-level data is the gold standard: it links policy documents directly to line-prosecutor behavior. Are "minimize enhancements" memos actually reducing enhancement filings? UniCourt and CJARS can answer that.

Policy Disruption Detection: Methodology

Top disruptions detected:

1. San Francisco 2020 (Boudin): 0.572
2. LA County 2021 (Gascón): 0.549
3. Sacramento 2022: 0.412

Anticipated Objections

"Policy documents $\neq$ practice. How do you know what's written actually happens?"

That's exactly why Extension 2 (outcome linkage) is Priority 1. We test whether "diversion-positive" memos predict more diversions. We also plan stakeholder interviews to validate. But note: even the stated-policy variation is *itself* a novel descriptive contribution.

"You only have California. Isn't this just a Blue State story?"

That's the point of Extension 1. CA provides internal variation (progressive Bay Area vs. traditional Central Valley), but external validity requires national expansion. We target the 100 largest DA offices to include red/purple jurisdictions.

"AI coding isn't reliable enough for research."

The face validity checks are strong ($d = 0.75$ Gascón, geographic clustering). Phase 1 includes a 200-doc human validation study to establish gold-standard IRR. The $80 cost and 2-hour runtime make iteration cheap.