

# The Paper Prisons Racial Justice Act Data Tool\*

Colleen V. Chien\*

William A. Sundstrom\*

Yabo Du\*

Akhil Raj\*

Bennett Cyphers\*

Rayna Saron\*

## ABSTRACT

The California Racial Justice Act provides a novel basis for challenging racial disparities in charging, conviction, and sentencing, even in the absence of explicit intent to discriminate. However, the lack of accessible data demonstrating a “significant difference” in outcomes for “similarly situated” defendants across racial groups has hindered the Act’s implementation.

---

DOI: <https://doi.org/10.15779/Z387W67709>

Copyright © 2024 Regents of the University of California.

\* The tool described in this paper is available for public use at [www.paperprisons.org/rja](http://www.paperprisons.org/rja).

\* Colleen V. Chien is Professor of Law and Co-Director of the Berkeley Center for Law and Technology at Berkeley Law School and the founder of the Paper Prisons Initiative. She is a member of the Paper Prisons Initiative. The authors thank the California Department of Justice for granting access to its data, the editors at the *Berkeley Journal of Criminal Law* for their editing of the Article, Yangxier Su and Chau Le for bluebooking assistance, and W. David Ball for insightful discussions during the development of the paper.

\* William A. Sundstrom is the Stephen and Patricia Schott Professor of Economics of Santa Clara University. He is a member of the Paper Prisons Initiative.

\* Yabo Du is a full stack developer and web expert. He is a member of the Paper Prisons Initiative.

\* Akhil Raj is a MS graduate of Santa Clara University School of Computer Science. He is a member of the Paper Prisons Initiative.

\* Bennett Cyphers is a 2L at Berkeley Law and a member of the Paper Prisons Initiative.

\* Rayna Saron is a 3L at Berkeley Law and a member of the Paper Prisons Initiative.

This paper introduces Paper Prisons Racial Justice Act data tool ([www.paperprisons.org/rja](http://www.paperprisons.org/rja)) to address this gap by providing, in its first release, comprehensive aggregate criminal offender record information (CORI) data from the California Department of Justice on racial disparities at the county, offense, and stage-of-conviction levels from 2010 to 2021. The paper describes the provenance, processing, and production of this data. The tool covers all misdemeanor and felony offenses in California and provides statistical information for five racial groups. Using this comprehensive data, we examine how courts should interpret the RJA’s requirement that comparisons be made among “similarly situated” defendants in the “same county.”

We find that for many offenses, especially in smaller counties, and for less populous racial/ethnic groups, and for later stages of prosecution, there are often insufficient cases to meet privacy-preserving minimum sample size requirements for release of the data. For example: for 16 counties with small Black populations, not a single offense has a large enough number of arrest incidents in 2019 for the data to be reportable. Moving to later in the criminal cycle, we find that no comparisons to Black defendants with respect to felony convictions are possible for over 75% of counties. The data situation is markedly worse for Native Americans. Only five counties have enough incidents to cover more than half of the incident arrests.

If the RJA is to achieve its goal of eliminating racial bias statewide, its evidentiary standards must be interpreted flexibly and allow for the aggregation of data, for example, across geographies, time periods, and offense categories when necessary to accrue adequate sample sizes. The language of the statute and early court decisions suggest such flexibility is permissible.

Introduction.....31

I. The Standards of Proof for Bringing a Racial Justice Act Claim

    Based on a Pattern of Disparity .....32

II. Data for Determining the Presence of a Significant Difference .....39

    A. California Criminal Offender Record Information (CORI)

        Data .....39

        B. Other Data Sources .....41

III. The Paper Prisons RJA Tool.....42

IV. CORI Case Study: Overcoming Small Numbers Data

    Challenges .....50

Conclusion .....52

## INTRODUCTION

The California Racial Justice Act (RJA), enacted in 2020 and further expanded in 2022 and 2023, created a new basis for actionable racial disparity even in the absence of a showing of intent to discriminate, when the “totality of the evidence demonstrates a significant difference” in charging, conviction, or sentencing across groups defined by race, ethnicity, or national origin when compared to those who are “similarly situated” and who have engaged in “similar conduct.” But a lack of access so far to the data needed to establish the requisite “significant difference” through evidence of a “pattern of disparity” has made the Act’s promised remedies exist largely on paper only.

We attempt to address this gap by providing a tool, available at [www.paperprisons.org/rja](http://www.paperprisons.org/rja), that provides aggregate Criminal Offender Records Information (CORI) data from the California Department of Justice (CalDOJ) on racial disparities at the county-, charge-, and year-levels across all recorded criminal offender activity in California from the available years (2010 to 2021 in the current version of the tool). It provides statistical information for five racial groups, and reflects records of all offenses in California statutory codes punishable by a misdemeanor or felony, including violations of the Penal Code, Health and Safety Code, and Vehicle Code. This paper describes the standard of proof required by the Act, the underlying CORI data our tool reports on, and the tool’s processing and presentation of the data.

We use our comprehensive data to address one of the most important open questions in RJA litigation—what is the appropriate comparator when evaluating historic data for the absence or presence of significant differences by race among “similarly situated” defendants? Applying the statute’s reference to “same country” and “similarly situated” defendants to the CalDOJ data, we construct analyses of county-charge-ethnic group-outcome combinations in all California counties. We find that, despite the RJA’s intent in principle to eradicate disparities across the entire state, in practice, slicing data along multiple dimensions while respecting the privacy-preserving reporting requirements of the CalDOJ quickly diminishes the number of cases available for analysis among less populous counties, for uncommon offenses, for less populous ethnic groups, and among later-stage outcomes—arguably the very populations, geographies and situations that may benefit the most from oversight.<sup>1</sup> For example, for 16 counties with small Black populations,

---

<sup>1</sup> Under our agreement with CalDOJ, to protect privacy we do not report any comparisons for which the sample contains ten or fewer incidents for either group being

not a single offense has a large enough number of arrest incidents in 2019 for the data to be reportable. Moving to later in the criminal cycle, felony convictions, we find that no comparisons to Black defendants are possible for over 75% of counties. The data situation is demonstrably worse for Native Americans - only five counties have enough incidents involving Native Americans to cover more than half of the incident arrests.

These data limitations caution against the overly rigid interpretation of “similarly situated” and support the continued application of a flexible and expansive lens that, where data or analyses are not otherwise available, combines counties, time periods, offenses, and in certain circumstances, potentially even stages of prosecution.

Part I describes the relevant standards of proof for bringing a RJA claim based on a pattern of disparity; Part II describes data for determining the presence of a significant difference, including the California criminal offender record information (CORI) data which we feature in our RJA tool; Part III introduces the Paper Prisons RJA tool; and Part IV describes our CORI and “small numbers” case study.

#### **I. THE STANDARDS OF PROOF FOR BRINGING A RACIAL JUSTICE ACT CLAIM BASED ON A PATTERN OF DISPARITY**

The California Racial Justice Act allows defendants to challenge charges, convictions, or sentences that are sought or obtained in a racially disparate manner.<sup>2</sup> The law’s stated intent is to “eliminate racial bias from California’s criminal justice system” and “ensure that race plays no role at all in seeking or obtaining convictions or in sentencing.”<sup>3</sup> It lowers the standard of proof needed to establish an actionable claim based on racial discrimination set by *McCleskey v. Kemp*, in which the Supreme Court found that despite the presence of a “discrepancy that appears to correlate with race” in death penalty cases in Georgia, the discrepancy was not actionable absent proof of a discriminatory purpose, and must instead be accepted as “an inevitable part of our criminal justice system.”<sup>4</sup>

The original version of the RJA (housed in California Penal Code Section 745) applied only to people sentenced in trial court after January 1, 2021. In September 2022, the RJA was amended by “the Racial Justice

---

compared.

<sup>2</sup> Assemb. B. No. 2542 § 3, Reg. Sess. 2019–20 (Cal. 2020), codified as CAL. PENAL CODE § 745. Throughout the RJA, racial disparities also encompass ethnicity and national origin. *Id.*

<sup>3</sup> *Id.* at § 2(i).

<sup>4</sup> 481 U.S. 279, 312 (1987).

Act for All”<sup>5</sup> (the “2022 Amendments”), which removed this limitation and set forth a gradual schedule for applicability of the amended law, including expanded retroactivity.<sup>6</sup> Following the implementation of retroactivity, Section 745 was further modified by AB 1118 (the “2023 Amendments”), which provided that post-judgment RJA claims based on the trial record may be raised on direct appeal from a conviction or sentence (including cases with judgments entered before January 1, 2021).<sup>7</sup> A right for individuals currently incarcerated to seek habeas relief has been enshrined in the statute since its original enactment,<sup>8</sup> though this right has yet to be litigated to a published decision.

Several types of conduct comprise actionable racial disparity under the Act: the exhibition of bias or animus towards the defendant by the state, a witness, or a juror;<sup>9</sup> the use of discriminatory language about or exhibition of bias or animus towards the defendant in court (unless quoting another person);<sup>10</sup> and relevant for this article, that the prosecution more frequently sought or obtained harsher charging, conviction, or sentencing outcomes against people who are of the same race as the defendant.<sup>11</sup> For such a “historical pattern” of disparity claim, the totality of the evidence must demonstrate that there is “a significant difference” in the way that individuals with the defendant’s profile are being treated as compared to others with a different race, ethnicity, or national origin.<sup>12</sup>

The law sets forth a two-part test. In the charging or conviction context, the defendant must first show that they were “charged or

---

<sup>5</sup> Assemb. B. No. 256, Reg. Sess. 2021–22 (Cal. 2022); see also *California Racial Justice Act for All Signed into Law*, ASSEMBLY MEMBER ASH KALRA (Sept. 30, 2022), <https://a25.asmdc.org/press-releases/20220930-california-racial-justice-act-all-signed-law>.

<sup>6</sup> According to the gradual schedule, eligibility expanded to people sentenced to death or facing possible immigration consequences like deportation on January 1, 2023, and people in prison, in a county jail serving a sentence for a felony conviction, or in the Division of Juvenile Justice (“DJJ”) on January 1, 2024. Assemb. B. No. 256 § 2, Reg. Sess. 2021–22 (Cal. 2022). Eligibility will expand to people no longer incarcerated, but with a felony conviction or a juvenile case that resulted in commitment to DJJ entered after 2015 on January 1, 2025, and to anyone with a felony conviction or a juvenile case that resulted in commitment on January 1, 2026. *Id.*

<sup>7</sup> CAL. PENAL CODE § 745(b).

<sup>8</sup> *Id.* (“A defendant may file a . . . petition for writ of habeas corpus . . . in a court of competent jurisdiction, alleging a violation of subdivision (a).”).

<sup>9</sup> *Id.* § 745(a)(1).

<sup>10</sup> *Id.* § 745(a)(2).

<sup>11</sup> *Id.* § 745(a)(3)–(4).

<sup>12</sup> *Id.* § 745(h)(1).

convicted of a more serious offense than defendants of other races, ethnicities, or national origins who have engaged in similar conduct and [were] similarly situated”,<sup>13</sup> and second that “the prosecution more frequently sought or obtained convictions for more serious offenses against people who share the defendant’s race, ethnicity, or national origin in the county where the convictions were sought or obtained.”<sup>14</sup> Similarly, in the context of sentencing, a defendant must first show that a “longer or more severe sentence was imposed on the defendant than was imposed on other similarly situated individuals convicted of the same offense”,<sup>15</sup> and second that “longer or more severe sentences were more frequently imposed” either when taking into account the race of the *defendant* or when taking into account the race of the *victim*.<sup>16</sup>

Following the 2022 Amendments, subsection 745(h)(1) specifies that several types of evidence can be used to prove a significant difference, including “statistical evidence, aggregate data, or non-statistical evidence,”<sup>17</sup> inviting defendants to present different types of evidence of patterns of racial or ethnic disparity in the relevant jurisdiction. In addition, “[s]tatistical significance is a factor the court may consider but is not necessary to establish a significant difference.”<sup>18</sup>

The 2022 Amendments also clarified the appropriate comparator for showings of disparity, based on the evaluation of individuals who have “engaged in *similar conduct* and are *similarly situated*.”<sup>19</sup> The phrase “engaged in similar conduct” replaced the phrase “committed similar offenses” in the amendment process.<sup>20</sup> As such, comparisons are to be made on the basis of the defendant’s underlying conduct, rather than, for example, on the basis of system-generated data such as arrests, charging, conviction, or sentencing. The possibility that such data may be tainted by prejudice is underscored by the language of the statute, which provides that, “in evaluating the totality of the evidence, the court shall consider whether systemic and institutional racial bias, racial profiling, and historical patterns of racially biased policing and prosecution may have

---

<sup>13</sup> *Id.* § 745(a)(3).

<sup>14</sup> *Id.*

<sup>15</sup> *Id.* § 745(a)(4)(A).

<sup>16</sup> *Id.* § 745(a)(4)(A)–(B).

<sup>17</sup> *Id.* § 745(h)(1).

<sup>18</sup> *Id.*

<sup>19</sup> *Id.* (emphasis added) (stating that the amendment “compar[es] individuals who have engaged in similar conduct and are similarly situated, and the prosecution cannot establish race-neutral reasons for the disparity”).

<sup>20</sup> *Id.* (amending CAL. PENAL CODE § 745(h)(1) (2021)).

contributed to, or caused differences observed in, the data or impacted the availability of data overall.”<sup>21</sup> To carry out an analysis on individuals that are “similarly situated” does not require that all individuals in the comparison group be identical, but that factors that are relevant in charging and sentencing be similar.<sup>22</sup>

The Act sets forth “escalating burdens of proof”<sup>23</sup> with respect to the relief to which petitioners are entitled. To receive “all evidence relevant to a potential [RJA] violation,”<sup>24</sup> the petitioner must provide a “showing of good cause.”<sup>25</sup> To be entitled to an evidentiary hearing, a defendant must make a prima facie showing of a violation of the law.<sup>26</sup> To obtain an ultimate disposition (leading to relief or a denial of relief) at a hearing, a defendant must prove a violation by a preponderance of the evidence.<sup>27</sup>

As of the time of this writing, just a few published opinions have addressed the applicable standard of proof that applies to statistical information used to support relief under the RJA, one in the context of a motion for discovery under Section 745(d), and the second in the context of a petition for an evidentiary hearing under Section 745(c).

In the case of *Young v. Superior Court*, petitioner Clemon Young, an African American man, was pulled over for a traffic violation and subsequently arrested for possession and sale of ecstasy. Young moved for discovery under the RJA, alleging that his race made it more likely for him to be stopped, searched, and prosecuted, and requested discovery under Section 745(d) to obtain additional records that could support his claim.<sup>28</sup> To demonstrate good cause for the request, Young presented publicly available descriptive statistics from the state that showed that Black people are more likely to be searched during the course of traffic stops than other citizens, as well as data from a particular county, combined with specific details about his particular stop.<sup>29</sup> The trial court

---

<sup>21</sup> *Id.*

<sup>22</sup> *Id.* § 745(h)(6).

<sup>23</sup> *Young v. Super. Ct. of Solano Cnty.*, 79 Cal. App. 5th 138, 160 (2022).

<sup>24</sup> CAL. PENAL CODE § 745(d).

<sup>25</sup> *Id.* (also specifying that upon showing of good cause, the court may allow the prosecution to “redact information prior to disclosure or may subject disclosure to a protective order” in order to protect privacy rights and privileges; if the statutory privilege or constitutional privacy right cannot be addressed through redaction or a protective order, the judge shall not order disclosure).

<sup>26</sup> *Id.* § 745(c).

<sup>27</sup> *Id.* § 745(c)(2).

<sup>28</sup> *Young*, 79 Cal. App. 5th at 146.

<sup>29</sup> *Id.* at 166.

denied his motion.<sup>30</sup> But in a case of first impression,<sup>31</sup> the appellate court reversed and remanded his discovery request to the trial court.<sup>32</sup> While the appellate court in *Young* found the evidence to be “unimpressive,”<sup>33</sup> it nevertheless found that it could provide a sufficient basis for discovery once combined with facts specific to the case.<sup>34</sup> As the court described, even though “the statistical proof Young puts forward does not make out a particularly strong case of racial profiling,” it “illustrates how the good cause standard works. At this stage, he need not make a strong case but only a plausible one.”<sup>35</sup>

In *Mosby v. Superior Court*,<sup>36</sup> Michael Mosby, an African American man, brought a RJA challenge alleging that the Riverside County District Attorney’s decision to seek the death penalty in his case was racially biased. Mosby was charged with the drive-by shooting of Darryl King-Divens along with a gun enhancement for discharging a firearm, and with special circumstances including multiple murders.<sup>37</sup> Following the fatal shooting, Mosby was also convicted of two additional murders and an attempted murder in Los Angeles County.<sup>38</sup>

To challenge the DA’s decision to seek the death penalty, Mosby filed a motion that included statistical evidence of bias against Black defendants in support of his case for an evidentiary hearing and relief pursuant to Section 745(c) of the RJA.<sup>39</sup> The trial court denied his motion without prejudice, ruling that more evidence was required to make a “prima facie” case under the RJA.<sup>40</sup> Subsequently, Mosby filed a second motion, adding non-statistical evidence in the form of vignettes about defendants who had committed similar crimes but for whom the death penalty was not sought,<sup>41</sup> but the motion was again rejected. Mosby

---

<sup>30</sup> *Id.* at 144.

<sup>31</sup> *Id.* at 156 (“To our knowledge, we are the first appellate court to address the discovery provision of the Racial Justice Act.”).

<sup>32</sup> *Id.* at 166. (“Having enunciated the applicable plausible justification standard, we will remand for the trial court to evaluate the issue[.]”)

<sup>33</sup> *Id.*

<sup>34</sup> *Id.*

<sup>35</sup> *Id.*

<sup>36</sup> *Mosby v. Super. Ct. of Riverside Cnty.*, No. RIF1604905 (Cal. Ct. App. filed Jan. 25, 2024).

<sup>37</sup> *Id.* at 2.

<sup>38</sup> *Id.*

<sup>39</sup> *Id.* at 3.

<sup>40</sup> *Id.*

<sup>41</sup> With respect to non-statistical evidence,

“Petitioner presented cases in which Caucasian defendants were charged with the special circumstance of multiple murders: Robert Lars Pape killed and burned three



appealed, arguing that statistical information alone was sufficient to make a *prima facie* case.<sup>42</sup> The appellate court reversed the trial court, making several holdings about the evidence required to prevail on a motion for a hearing and relief.

First, the majority held that the statute requires petitioners to make two showings to establish a *prima facie* case, “not only statistical evidence of racial disparity in the charging of the death penalty by the District Attorney but *also* evidence of nonminority defendants who were engaged in similar conduct and were similarly situated but charged with lesser offenses,” which they described as “evidence of similar conduct *and* similarly situated defendants.”<sup>43</sup> However, though Mosby presented statistical and nonstatistical evidence, the court expressly declined to rule on whether or not *only* statistical evidence could be sufficient, stating that “we need not determine based on the evidence presented whether only statistical evidence of similar conduct and similarly situated defendants would be sufficient to support a *prima facie* case,”<sup>44</sup> while the concurrence held that the statistical evidence presented in the case alone *was* sufficient.<sup>45</sup>

Second, the majority opinion found the statistical evidence presented, in the form of three studies, satisfied the first showing. The

---

people. The District Attorney did not seek the death penalty against Pape and he was serving a LWOP sentence. Jared Bischoff killed a man who was flirting with his girlfriend and then killed his girlfriend. Bischoff stabbed his girlfriend six times until she bled to death. Bischoff was scheduled for trial, but the District Attorney chose not to seek the death penalty. Petitioner also presented evidence of Caucasian young adult defendants: James Coon, who was 26 years old at the time of his offense, robbed a clerk at a store at gunpoint and then shot the clerk because he tried to take Coon’s photograph—the District Attorney did not seek the death penalty against Coon. Melissa Unger, who was 23 years old at the time, was involved in a gang murder, which involved the kidnapping and torture of a victim—she pled guilty to voluntary manslaughter. Owen Skyler Shover was accused of killing his 16-year-old girlfriend when he was 18 years old and was facing charges but the District Attorney did not seek the death penalty. Andrew Burke was 25 years old when he stabbed to death his adopted parents/grandparents and the District Attorney chose not to seek the death penalty. Petitioner was 24 years old when the alleged offense took place but he was facing the death penalty. Finally, Petitioner presented the ‘highly aggravated’ murder by a Caucasian defendant for whom the District Attorney did not seek the death penalty. Maxamillion Eagle raped and strangled a woman, throwing her body in a trash can. Eagle had a prior conviction of assault with a deadly weapon. He was sentenced to LWOP.”

*Id.* at 15–16.

<sup>42</sup> *Id.* at 4.

<sup>43</sup> *Id.* (emphasis added).

<sup>44</sup> *Id.* at 5.

<sup>45</sup> *Id.* at 38–39 (Menetrez, J., concurring).

“Omori study” examined the charging of African American defendants with special circumstances in Riverside County from January 2016 through December 2021.<sup>46</sup> The study found that African American defendants were charged with special circumstances at a rate of 64.86 per 100,000 of the adult population, significantly higher than the rates for Caucasian (5.00 per 100,000) and Hispanic (16.84 per 100,000) defendants.<sup>47</sup> In contrast to the Omori study, which made population level comparisons, the “Peterson study” compared white and Black defendants in murder cases in Riverside County from January 2007 to July 2019.<sup>48</sup> Using a logistic regression that controlled for factors including defendant’s race/ethnicity, prior criminal history, victim’s demographics, and specifics of the crime such as the use of a firearm or the presence of multiple victims, African American defendants were found to be 1.71 times more likely to be charged with a special circumstance, 9.06 times more likely to receive a death penalty notice, and 14.09 times more likely to receive a death sentence than their Caucasian counterparts.<sup>49</sup> Finally, the “Baumgartner study” considered disparities in death penalty sentencing in Riverside County since 1972, finding that African Americans made up 66% of those sentenced to death, compared to 25% for Caucasian defendants.<sup>50</sup>

The *Mosby* decision cited another case, *Finley v. Superior Court*, in which statistical evidence was offered as part of the making of a prima facie case for relief, this time in the context of a Subsection 745(a)(1) claim for RJA relief on the basis of bias or animus towards the defendant.<sup>51</sup> In *Finley*, the defendant, an African American man, presented evidence that African American persons in San Francisco were far more likely to be stopped by police than other groups, together with evidence that the officer had no logical nonracial reason for stopping and searching him.<sup>52</sup> While the trial court denied the motion, the appellate court overturned the decision, citing the relevant standard in habeas cases, and holding that in order to show a prima facie case under the RJA,

“[A defendant] must state fully and with particularity the facts on which relief is sought, and include copies of reasonably available documentary evidence supporting the claim. The court should

---

<sup>46</sup> *Id.* at 7.

<sup>47</sup> *Id.*

<sup>48</sup> *Id.* at 8–9.

<sup>49</sup> *Id.* at 9.

<sup>50</sup> *Id.* at 10.

<sup>51</sup> *Finley v. Super. Ct. of San Francisco Cnty.*, 95 Cal. App. 5th 12 (2023).

<sup>52</sup> *Id.* at 17–18.

accept the truth of the defendant’s allegations, including expert evidence and statistics, unless the allegations are conclusory, unsupported by the evidence presented in support of the claim, or demonstrably contradicted by the court’s own records.”<sup>53</sup>

These cases illustrate the range of statistical evidence that may be offered to support a Racial Justice Act claim. Geographically, county-level to statewide data; temporally, 5 years to 51 years of data; and population level descriptive to regression statistics have all been offered, and found sufficient, to support a RJA claim.

## II. DATA FOR DETERMINING THE PRESENCE OF A SIGNIFICANT DIFFERENCE

### A. California Criminal Offender Record Information (CORI) Data

California has a complex framework for managing criminal records, which are maintained at both the county and state levels. The initial creation and maintenance of criminal records fall under the purview of state and local public law enforcement agencies, including prosecutor’s offices and courts, which are legally obligated to generate and retain these records.<sup>54</sup> For each arrest, the reporting agency is required to report to the Department of Justice applicable criminal offender record information (CORI) data, which spans, in principle, the phases of a person’s contact with the criminal justice system from arrest to release.<sup>55</sup>

The Attorney General has responsibility for the security of CORI data,<sup>56</sup> and access and dissemination of CORI records is tightly regulated.<sup>57</sup> However, statistical and research releases that do not disclose

---

<sup>53</sup> *Id.* at 23.

<sup>54</sup> See Wallace Wade, *Who’s Lying Now?: How the Public Dissemination of Incomplete, Thus Half-Truthful, Criminal Record Information Regarding a Statutorily Rehabilitated Petty Offender is an Unjust Penalty and Why Laws Regarding Expungement of and Restrictions on Dissemination of Criminal Records Information in California Must Be Reformed*, 38 W. ST. U. L. REV. 1, 7–11 (2010) (providing an overview of records collection and dissemination).

<sup>55</sup> CAL. PENAL CODE § 13150.

<sup>56</sup> *Id.* § 11077 (“The Attorney General is responsible for the security of criminal offender record information”). “Criminal offender record information” is defined as “records and data compiled by criminal justice agencies for purposes of identifying criminal offenders and of maintaining as to each such offender a summary of arrests, pretrial proceedings, the nature and disposition of criminal charges, sentencing, incarceration, rehabilitation, and release.” *Id.* § 11075(a).

<sup>57</sup> See, e.g., *id.* § 11077(b), which specifies that CORI data is to be “disseminated only in situations in which it is demonstrably required for the performance of an agency’s or official’s functions.”

the identity of the subject are allowable under the Penal Code section governing criminal history information.<sup>58</sup> Public agencies or bona fide research institutions “concerned with the prevention or control of crime, the quality of criminal justice, or the custody of correction of offenders” are allowed access to CORI data as required for the performance of their duties, including conducting research,<sup>59</sup> but the resulting dissemination should be used only for research or statistical activities and not disclose the identity of the subject.<sup>60</sup>

California Penal Code Section 13125 outlines a long list of standard data elements that should be stored in CORI systems “when applicable and available,” including, with relevance to the RJA, demographic details about the defendant and information about the arrests, charges, convictions, and sentences associated with a particular arrest incident.<sup>61</sup> The California DOJ is statutorily required to maintain

---

<sup>58</sup> *Id.* § 11105(g).

<sup>59</sup> *Id.* § 13202(a).

<sup>60</sup> *Id.* §§ 11105(g), 13202(a).

<sup>61</sup> *Id.* § 13125 (specifying the following fields; **individual level fields**: Full Name, Aliases, Monikers, Race, Sex, Date of Birth, Place of Birth (state or country), Height, Weight, Hair Color, Eye Color, CII Number, FBI Number, Social Security Number, California Operator’s License Number, Fingerprint Classification Number, Henry, NCIC, and Address; **arrest level fields**: Arresting Agency, Booking Number, Date of Arrest, Offenses Charged, Statute Citations, Literal Descriptions, Police Disposition, outcomes such as Released, Cited and Released, Turned Over To another authority, or Complaint Filed; **charge level fields**: County and Court Name, Date Complaint Filed, Original Offenses Charged in a Complaint or Citation, process outcomes like Held to Answer, Certified Plea, Disposition including Not Convicted, Dismissed, Acquitted (Court Trial or Jury Trial), Convicted (Plea, Court Trial, Jury Trial), Date of Disposition, Convicted Offenses, Sentence, Sentence Enhancement Data Elements, Proceedings Suspended, and Reason Suspended; **adjudicated court charge level fields**: County, Date Complaint Filed, Type of Proceeding (Indictment, Information, Certification), Original Offenses Charged, Disposition outcomes (Not Convicted, Dismissed, Acquitted, Convicted-felony or misdemeanor), through Plea, Court Trial, Jury Trial, Date of Disposition, Convicted Offenses, Sentence, Sentence Enhancement Data Elements, Proceedings Suspended, and Reason Suspended; **corrections level fields**: Adult Probation (County, Type of Court, Offense, probation dates, Reason for Removal); Jail data for unsentenced prisoners (Offenses Charged, Name of Jail or Institution, reception and release dates, Reason for Release, Bail details, Committing Agency); County Jail data for sentenced prisoners similarly lists convicted offense, sentence details, and release information; **Division of Juvenile Justice fields**: County, Type of Court, Court Number, Division of Juvenile Justice Number, Date Received, Convicted Offense, Type of Receipt (Original Commitment, Parole Violator), Date Released, Type of Release, Custody, Supervision, Date Terminated; **Department of Corrections and Rehabilitation fields**: mirror the above with additional Department of Corrections and Rehabilitation Number, detailed release types. **Mentally Disordered Sex Offenders fields**: County, Hospital

summary criminal history information that pertains to the identification and history of a person “including their name, date of birth, physical description, dates of arrest, charges, dispositions, and sentencing information.”<sup>62</sup>

### **B. Other Data Sources**

Several other government agencies collect and store data which may be relevant to RJA claims. The Racial and Identity Profiling Act (RIPA) of 2015 requires law enforcement agencies to report data to the California Attorney General’s office on (a) all vehicle and pedestrian stops and (b) citizen complaints alleging racial and identity profiling.<sup>63</sup> The Act also requires the annual release of certain information to the public.<sup>64</sup> In addition, the California Department of Corrections and Rehabilitation maintains extensive data about incarcerated individuals and parolees, including ethnicity, offenses, and sentences. Some aggregate data is made publicly available.<sup>65</sup> Furthermore, some federal agencies aggregate crime data from state and local agencies. For example, the National Incident-Based Reporting System (NIBRS) is an FBI initiative which captures extensive details on individual crime incidents.<sup>66</sup>

Some private entities also aggregate information which includes CORI data or its equivalent. The Criminal Justice Administrative Records System (CJARS) is a partnership between the University of Michigan and the U.S. Census Bureau which collects longitudinal electronic records from criminal justice agencies around the country in order to track criminal episodes across stages of the justice system.<sup>67</sup> It collects data from publicly available sources as well as via data use agreements with

---

Number, Date Received, Date Discharged, Recommendation).

<sup>62</sup> *Id.* §§ 11105(a)(1), (a)(2)(A)–(B).

<sup>63</sup> CAL. CODE REGS. tit. 11, § 999.224 et seq. (2024).

<sup>64</sup> *Id.* § 999.228(h).

<sup>65</sup> See *Offender Data Points*, CAL. DEP’T OF CORR. & REHAB., <https://public.tableau.com/app/profile/cdcr.or/viz/OffenderDataPoints/SummaryInCustodyandParole>.

<sup>66</sup> This includes information on victims, known offenders, relationships between victims and offenders, arrestees, and property involved in crimes. See *National Incident-Based Reporting System (NIBRS)*, BUREAU OF JUST. STATS., <https://bjs.ojp.gov/national-incident-based-reporting-system-nibrs> (last visited Apr. 7, 2024). Not all law enforcement agencies are yet certified to report crimes via NIBRS: as of May 2023, the Bureau of Justice Statistics estimated that 77% of the US population was covered by NIBRS-reporting agencies, but that number was only 55% for California. *Id.*

<sup>67</sup> *Criminal Justice Administrative Records System (CJARS)*, UNIV. MICH., <https://cjars.org> (last visited Apr. 7, 2024) (not-for-profit organization directed by Keith Finlay and Mike Mueller-Smith).

government agencies, including the California DOJ.<sup>68</sup> Access to CJARS data may be granted to researchers on a case-by-case basis pursuant to an in-depth proposal and approval process.<sup>69</sup>

California law requires that local agencies, including law enforcement, district attorneys, and superior courts, collect and report CORI data.<sup>70</sup> These agencies may additionally aggregate, store, and make this data accessible in other ways. The ACLU of Northern California has engaged in a long-term project to request data and documents relevant to the RJA from county prosecutors via the California Public Records Act.<sup>71</sup> It has made the records it receives in response available to the public.<sup>72</sup> However, the data it has received varies widely by county; some offices have withheld or redacted data claiming exemption under the PRA,<sup>73</sup> cited technical limitations in their record-keeping systems,<sup>74</sup> or stated that they do not have any responsive records at all.<sup>75</sup> As a result, the data which the ACLU has received varies widely in format, quality, and completeness.

### III. THE PAPER PRISONS RJA TOOL

To increase data accessibility for RJA litigation, we have developed and published the “Paper Prisons Racial Justice Act Tool,” available at [www.paperprisons.org/rja](http://www.paperprisons.org/rja), that provides aggregate CORI data on racial disparities at the county-, charge-, and year- levels. The source for data on this site is the CORI dataset of all arrests, charges, convictions, and sentences in California, available to researchers through the California Department of Justice Automated Criminal History System

---

<sup>68</sup> *Id.*

<sup>69</sup> See CJARS Title 13 Proposal Guide - For Data Access Through the Federal Statistical Research Data Center Network, UNIV. MICH. (2022), <https://cjars.org/proposal-guide-download/>.

<sup>70</sup> CAL. PENAL CODE § 13150.

<sup>71</sup> *Documents Related to the Implementation of the Racial Justice Act*, ACLU N. CAL., <https://www.aclunc.org/documents-related-implementation-racial-justice-act> (last visited Apr. 7, 2024).

<sup>72</sup> *Id.*

<sup>73</sup> See, e.g., *Orange County D.A. Denies Public Records Access, Chicaxs Unidxs and ACLU Sue*, ACLU S. CAL. (Oct. 19, 2022), <https://www.aclusocal.org/en/press-releases/orange-county-da-denies-public-records-access-chicanxs-unidxs-and-aclu-sue>.

<sup>74</sup> See, e.g., Response from Kern County Deputy Counsel to the ACLU and BraunHagey & Borden (Mar. 29, 2022), <https://drive.google.com/file/d/1Kdoj4v82Bx3Jx4oJFqzycrjvav7r44Ig/view?usp=sharing>.

<sup>75</sup> See, e.g., Response from the Alpine County District Attorney’s Office to the ACLU of Northern California and BraunHagey & Borden (Oct. 15, 2021), [https://drive.google.com/file/d/1g8UUHkWO\\_FWfhnJMQGF2uRnB4g-nXLVh/view?usp=sharing](https://drive.google.com/file/d/1g8UUHkWO_FWfhnJMQGF2uRnB4g-nXLVh/view?usp=sharing) (producing zero responsive records).

(ACHS) under the provisions of the CalDOJ Research Data Request process. In the initial version of the tool introduced with this writing, our records were downloaded between 9/23/2021 and 9/29/2021; the data we present range from 2010 through (nearly) the first nine months of 2021.<sup>76</sup>

While the CORI dataset provides valuable comprehensive statistics, it is not without limitations. Among the known disadvantages of the CORI data are that it does not include systematic information on the conditions of the arrest (such as whether or not a weapon was present) or other aspects of the defendant's conduct which might influence the evaluation of the "similar conduct" standard under the statute. Two additional shortcomings of our database are that it does not include juvenile or out-of-state records. Errors in underlying data are due to reporting errors and/or fundamental limitations to the Automated Criminal History System set up and maintained by the California DOJ.

The CORI source data used for the tool are anonymous, with names removed and separate individuals identified only by an internal ID code. Personally identifying information, such as date of birth or social security number, has been removed. The tool reports summaries of data that have been processed to calculate counts (raw numbers) or rates for the specified user query. These summaries are never reported at the level of a particular individual.

The tool allows website visitors to view the data in various ways. Users can customize the data presented by year, county, event point, race-ethnicity, measurement (metric), and offense. The output users see depends upon the customization for each category. The event points include arrests, court actions, convictions, felony convictions, incarceration sentences, and prison sentences (discussed further below). The metrics calculate the rate at which an event occurs for selected racial-ethnic groups, relative to their representation in the population of the selected county. We explain the different options and measurements available and how they are calculated, as well as some limitations of the data, below.

### *Unit of Analysis*

The CORI data from which the tool metrics are derived records each of the events associated with a given cycle of an individual's

---

<sup>76</sup> Rates per population reported in the tool are based on total counts of annual incidents in a calendar year. 2021 rates per population are adjusted upward by 12/9 to account for the limitation of the data to the first 9 months of the year. No adjustments are made for potential lags in reporting.

involvement with the criminal legal system, where a cycle is defined as the series of events that flow from a specific initial incident for an individual. A cycle typically begins with an arrest, in which a person may be accused of one or more offenses, and then may proceed through a series of additional steps; for example, the arrest step is frequently followed by a prosecutor filing charges in court on one or more offenses. Offenses charged in court may be the same as or different from those recorded at arrest. The offenses charged in court lead to a disposition on each offense, such as a conviction, dismissal, acquittal, diversion, or other outcome. A conviction results in a sentence, which may include incarceration in jail or prison, a fine, probation, etc. The CORI data does not provide information on actual incarcerations or completed length of term served, only on sentences of incarceration.

The tool counts the number of incidents in which a particular offense has been charged. An incident is defined as a particular cycle for a particular individual. For example, if a person has at least one arrest for violating California Penal Code Section 242 (battery) in a particular cycle, one incident of arrest on Penal Code Section 242 is added to the number of arrest incidents. Multiple counts of the same offense charged in the same arrest cycle are only counted once in the tool. For example, suppose that following a particular arrest, a person ended up convicted of three counts of California Penal Code Section 242 (battery) and two counts of California Penal Code Section 148(A)(1) (obstructing/resisting arrest). For this cycle, we would count one incident of conviction for California Penal Code Section 242 and one incident of conviction for California Penal Code Section 148(A)(1). We apply the same approach at the arrest and court levels.

If the same individual is arrested and charged with a certain offense on more than one occasion (in *different* arrest cycles), each cycle will be counted separately. For that reason, the number of incidents counted in the tool is greater than the number of individuals involved.

### *Event Points*

Racial disparities can occur at each of a number of specific event points or steps in the criminal legal system. Criminal records are complex and present an array of event types that may be defined in various ways; in designing the tool we have striven to use simple definitions of key events based on unambiguous interpretation of CORI variables. The tool provides metrics at the following specific event points or decisions, as derived from the CORI data.



Event Point	Definition and Source
Arrest	Step in the CORI data for which the CORI variable STP_ORI_TYPE_DESCR takes the value “Arrest”. This identification method includes initial arrests (ARREST/DETAINED/CITED) as well as supplemental arrests.
Court action	Step for which the CORI variable STP_ORI_TYPE_DESCR takes the value “Court”.
Conviction (including misdemeanors and felonies)	Court step for which the CORI disposition variable DISP_CODE is in the range of values 2500-2799, which encompass a variety of conviction categories. Misdemeanor charges are identified by the CORI variable OFFENSE_TOC taking the value “M”, and felony charges by OFFENSE_TOC equal to “F”.
Felony conviction	Conviction on a felony charge, using the definitions above.
Prison sentence	Conviction for which the CORI variable SENT_LOC_CODE takes the value “0” or “A”, indicating a sentence to prison.
Incarceration sentence (a sentence to prison or to county jail)	Conviction for which the CORI variable SENT_LOC_CODE takes the value “0”, “A”, or “J”, indicating a sentence to prison or jail. For both prison and incarceration sentences, we do not count suspended sentences, sentences to “fine or jail,” or sentences associated with non-conviction events, such as parole violations.

### *Year*

For arrest event points, the year of the incident recorded in the tool is the minimum (first) calendar year for the cycle. For court events (all non-arrest events), the year is the maximum (last) year for the cycle. We count any event point that occurs in a cycle up to the last event. We assign years this way to take account of the fact that someone might be arrested in one year and go to court in a subsequent year.

### *Offenses*

The CORI data set includes information on all the categories of

criminal offenses that are chargeable as a misdemeanor or felony,<sup>77</sup> most of which are in the Penal Code (PC), but also which appear in a variety of additional California codes, including Vehicle (VC) and Health and Safety (HS) Codes. Offenses chargeable only as infractions, for example certain Vehicle Code violations, are excluded, as are probation violations. (e.g. PC 1203.2)

Each code subsection is treated as a distinct offense. For example, PC Section 148(a) is treated as distinct from PC Section 148(b). In situations where subsections might more appropriately be combined, users can select multiple code subsections in the tool, and access aggregate data outputs (as described more fully in the “Combining Data” section later in this article). At the same time, users should be aware that our approach of identifying offenses with code subsections lumps together some charges that come under a single subsection but have different consequences. Examples would include so-called “wobbler” offenses that are felonies by default but may be charged as misdemeanors under some conditions. Burglary (PC 459) may be charged as first-degree or second-degree burglary, but the source data do not always distinguish the degree, so these charges are combined into a single offense code. In addition, offenses with different levels of detail in the CORI dataset are also lumped together, so that “459 PC-BURGLARY” and “459 PC-BURGLARY:FIRST DEGREE” charges are aggregated in our tool.

For any given incident, the offense cited at arrest may be different from the offense charged in court or at conviction, given prosecutorial decisions and plea bargaining. We have excluded from the tool any records in the CORI data set for which a specific code section is not provided or cannot be identified.

### *County*

The “county” reported in the tool is the county of the originating agency for the specific incident recorded, whether an arrest by a local law enforcement agency or an action by a county superior court. In a very small number of cases (0.3% of individuals and 0.08% of incidents), the county is recorded as “Unknown.” These cases are included in the California totals reported in the tool but do not appear in individual county data.

---

<sup>77</sup> This version of the tool does not include probation violations.

### *Racial Classifications*

We extracted race data as recorded in the CORI dataset as the basis for calculating racial differences in patterns of arrest, charging, conviction, and sentencing. The CORI data source indicates the racial/ethnic identity of each individual with a single mutually exclusive “race” identifier. Multiple racial identities are not recorded. While white, Hispanic, Black, and American Indian populations are specifically identified in the CORI database, we aggregated several races into the “AAPI” category,<sup>78</sup> masking considerable heterogeneity in this population.<sup>79</sup> Outside of the five groups that the tool collapses the various CORI categories into, persons of “other” and “unknown” race, representing 3.0% of the total incidents 6.9% of individuals, are not included in the tool. How, specifically, racial/ethnic groups are assigned in the CORI data (whether self-identified or assigned by authorities) is not indicated in the source. Although the CORI database also includes information on defendant national origin (country of birth), this information is not reported in the current version of the tool.

We obtain county population numbers from the 5-year estimates of the Census Bureau’s American Community Survey (ACS). The racial/ethnic classifications reported in the ACS do not correspond directly to the CORI race categories. In particular, the ACS asks separate questions about racial identity and Hispanic/Latino identity, and allows individuals to belong to more than one race. The ACS categories we use are not mutually exclusive, but in our judgment are likely to correspond reasonably well to the CORI mutually exclusive categories. The following table displays the racial categories from the CORI data and the corresponding ACS categories. The percent of incidents and percent of persons is based on the CORI data populating this version of the tool (2010-2021) and the ACS 2016-2020 population estimates.

---

<sup>78</sup> The CORI categories combined into our “AAPI” group are Asian Indian, Cambodian, Chinese, Filipino, Guamanian, Hawaiian, Japanese, Korean, Laotian, Other Asian, Pacific Islander, Samoan, and Vietnamese.

<sup>79</sup> Gabriel Schwartz & Jaquelyn Jahn, *Disaggregating Asian American and Pacific Islander Risk of Fatal Police Violence*, 17 PLOS ONE (2022) <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0274745>. There is large variation in the rates at which different AAPI groups come into contact with law enforcement, including wide disparities in fatal police violence. *Id.* In an analysis of police killings between 2013-2019, annual rates of police violence were higher for Pacific Islanders than East and South Asian Americans, approaching the level suffered by Black Americans. *Id.*

CORI race	ACS race and Hispanic identity	Percent of incidents, CORI sample	Percent of persons, CORI sample	Percent of CA population <sup>80</sup>
AAPI (combination of Asian and PI groups)	Asian American or Pacific Islander race alone	3.0%	4.0%	15.2%
Black	Black or African American race alone	16.8	13.6	5.7
Hispanic	Hispanic or Latin American, any race	41.5	42.2	39.1
Native American	Native American alone or in combination with other race(s)	0.6	0.5	2.1
White	White race alone, not Hispanic	35.1	32.9	36.5
Other / Unknown <sup>81</sup>	NA	3.0	6.9	NA

### Measurements

Three different metrics can be viewed in the current version of the tool:

1. **Raw number** is a count of the actual number of incidents in the selected category defined by county, offense, year, race, and event point. For example, a query of the tool for convictions in Alameda County in 2016 for Black persons on the offense PC Section 148(A)(1) returns a raw number of fifty-eight. This means there were fifty-eight incidents involving convictions of Black persons on the charge of PC Section 148(A)(1) recorded in Alameda County in 2016. The tool counts incidents, not individuals, so fifty-eight conviction incidents might represent fewer than fifty-eight individuals, because a given individual might have been charged with the same offense in more than one cycle.

2. **Rate per population** measures the rate at which a given event

<sup>80</sup> Population of indicated American Community Survey (ACS) racial grouping as percent of total California population from 2016-2020 ACS. *American Community Survey Data*, U.S. CENSUS BUREAU, <https://www.census.gov/programs-surveys/acs/data.html> (last visited Apr. 13, 2024). Provided for illustrative purposes. Categories are not mutually exclusive and percentages do not add to 100%.

<sup>81</sup> The CORI racial categories of “Other” or “Unknown” are not an option in the tool because there is no corresponding ACS population count.

or decision occurs for a selected racial or ethnic group, relative to that group's total population in the county. Specifically, it is the raw number of offense incidents of the requested type for the requested racial or ethnic group during the requested year, per 100 individuals of that group in the county population.<sup>82</sup>

3. **Population disparity v. white** compares the rate per population of a given racial/ethnic group with that of non-Hispanic white individuals.<sup>83</sup> The racial gap can be considered the chance that a person of the given race/ethnicity experiences a certain outcome or decision, relative to the chances of a non-Hispanic white adult, given underlying populations. A “population disparity v. white” value greater than 1.0 indicates that a specific racial/ethnic group experiences a higher rate of a particular outcome or decision compared to non-Hispanic whites, considering their respective population sizes. Conversely, a value less than 1.0 suggests that the specified group is less likely to experience the outcome relative to non-Hispanic whites, given the underlying populations. A value of 1.00 means that the two groups experience the outcome at the same rate relative to population.<sup>84</sup>

It is worth noting that the statute does not specify or limit the metrics that can be used to show a “significant difference.” For example, rather than comparing the rate per population of a given racial/ethnic group with that of non-Hispanic white individuals, as the current tool does, one might compare it to the rate per population of all other racial/ethnic groups.

---

<sup>82</sup> Formula: rate per population = raw number / county population (race-specific). Population data come from the American Community Survey (ACS), which is a large national survey run by the U.S. Census Bureau. All of our population estimates for different ethnicities use the 5-year ACS sample for the combined period 2016-2020. ACS summary tables were accessed through the Census Bureau API interface (see <https://www.census.gov/programs-surveys/acs/data/data-via-api.html>), using the R package tidycensus (see <https://walker-data.com/tidycensus/>). You can download the Census data at this link: [https://docs.google.com/spreadsheets/d/1acKdr3w9NIALgfUt8nLbtSWDqEfVxyQLKuZ3r\\_pGkes/edit#gid=840124101](https://docs.google.com/spreadsheets/d/1acKdr3w9NIALgfUt8nLbtSWDqEfVxyQLKuZ3r_pGkes/edit#gid=840124101).

<sup>83</sup> Formula: population disparity v. white = rate per population (selected race) / rate per population (white).

<sup>84</sup> For example, suppose that in a particular county in a particular year, there were two incidents in which Black adults experienced felony convictions for burglary for every 100 Black individuals in the population, and one such incident per 100 non-Hispanic white individuals. Then the population disparity for the Black relative to white population is  $2/1 = 2$ .

*Combining Data*

The tool permits data to be combined across years, counties, or offenses. Combining data is particularly useful to overcome the “small N problem,” described in the section below. When looking at raw numbers, the values displayed reflect aggregate counts of incidents (e.g. adding counts from 2018 and counts from 2019). In the case of rate per population or disparity gap per population metrics, the values displayed reflect weighted averages, taking into account combined event point and combined population counts.<sup>85</sup> When an underlying data point is unavailable, due to data limitations, the tool will simply not include that data point in the calculation and will also display a message warning the user that not all selected values are reflected in the displayed values.<sup>86</sup> As such, users of the tool are encouraged to select the “View Data” button to see which values are actually included in the aggregation, and to consider the metrics individually whenever also considering them in combination.

**IV. CORI CASE STUDY: OVERCOMING SMALL NUMBERS DATA CHALLENGES**

As California legislators confirmed when they passed the 2023 RJA Amendments, or “the Racial Justice Act For All,” the intent of the statute is to eradicate racial bias throughout the California criminal justice system, including in counties large and small, among major and minor

---

<sup>85</sup> Example where rate data is aggregated: Suppose there were 350 incidents in which Hispanic individuals had been arrested in county *X* on a charge of PC 459 (burglary) during the year in question, and the Hispanic population of county *X* was 100,000. The rate per 100 population in county *X* would be  $350/(100,000/100) = 0.35$  per 100 population. Suppose that for another county, county *Y*, there were also 350 incidents in which Hispanic individuals had been arrested on a charge of PC 459 (burglary) during the year in question, but the Hispanic population of county *Y* was 350,000. The rate per 100 population in county *Y* would be  $350/(350,000/100) = 0.10$  per 100 population. To get the aggregate rate across counties *X* and *Y*, the numerators and denominators would be added, for a combined number of arrests of Hispanic individuals of 700 and a combined Hispanic population of  $350,000 + 100,000 = 450,000$ . The combined rate per 100 population in county *X* and *Y* would be  $700/(450,000/100) = 0.157$  per 100 population.

<sup>86</sup> Example where data is limited: Suppose that in a particular county in a particular year, say 2019, there were twelve incidents in which Native American adults experienced felony convictions for burglary, and in 2018, the number of equivalent incidents was N/A, due to the total being ten or fewer. The raw count metrics for 2018 and 2019 in combination would still reflect twelve incidents, and a warning message would appear. In the case where the metric is the population disparity v. white gap, the chance that a person of the given race/ethnicity experiences a certain outcome or decision, relative to the chances of a non-Hispanic white adult, given underlying populations, and the data for either white or non-white populations is insufficient, neither will be included.

offenses, and for dominant and minority ethnic groups alike.<sup>87</sup> But slicing the data along several dimensions to facilitate comparisons of similarly situated defendants cases quickly diminishes the number of cases available for analysis, especially in less populous counties, for uncommon offenses, and for less populous ethnic groups. Below we use data from the tool to illustrate the problem and propose some potential workarounds.

Small numbers create at least two challenges. First, our tool is constrained by the California Department of Justice's rigorous and privacy-respecting regulations for the sharing of criminal records (CORI) data. Our agreement with the DOJ allows us to report comparisons only for cases in which there are more than ten incidents for each race being compared. Second, the statistical precision of any estimate depends on the sample size. There is no single rule for a minimum acceptable N, but even samples well above the DOJ's threshold of ten may be noisy. The smaller the sample, the less statistical "power" available to accurately discern significant racial gaps. In the language of the RJA, one might say there is a tradeoff between "similarity" and "significance."

Sample size is a particular challenge for comparisons involving small counties, smaller-population ethnic groups, and uncommon offenses. Figures 1 through 3 illustrate the magnitude of the small-N problem along several dimensions. Figure 1 displays the percentage of arrest incidents in each county that satisfy the  $N > 10$  threshold within offenses in 2019, for the Black-white comparison. That is to say, each bar represents the fraction of arrest incidents in that county for which one could query the tool for a Black-white comparison and obtain a result other than "NA". In the most populous counties, coverage is nearly complete. For 16 small counties with small Black populations, on the other hand, not a single offense had a large enough number of arrest incidents in 2019 to meet the  $N > 10$  requirement (these are the counties with bars of height zero). In such counties, use of the tool for statistical evidence on arrests to motivate an initial RJA claim would be wholly unavailable to Black claimants, and in several more counties would be unavailable for a sizable subset of offenses.

The small-N problem is compounded for comparisons of less frequent events or smaller demographic groups. Figure 2 replaces the arrest incidents of Figure 1 with felony conviction incidents. In this case, the fraction of incidents meeting the sample threshold falls short of fifty

---

<sup>87</sup> See *California Racial Justice Act for All Signed into Law*, ASSEMBLY MEMBER ASH KALRA (Sept. 30, 2022), <https://a25.asmdc.org/press-releases/20220930-california-racial-justice-act-all-signed-law>.

percent in all but 13 counties. *No comparisons are possible for any offense in more than half the counties.* In Figure 3, Figure 1's Black-white comparison is replaced by a comparison of arrest incidents for Native American vs. white individuals. Only five counties have enough incidents involving Native Americans to cover more than half of the arrest incidents.

Even comparisons that meet the DOJ's minimum sample threshold may rely on small samples from a statistical standpoint. Increasing the sample threshold to a more statistically reliable  $N > 50$  for each race would reduce the number of comparisons still further. And these sample size challenges would only be further exacerbated by additional controls for "similarly situated" comparisons, such as disaggregating by gender and/or prior offense records. The simple reality is that access to strong statistical evidence of racial disparities is likely to be limited to large counties, large ethnic groups, and common offenses.

Given that such constraints seem contrary to the legislature's intent, we support the aggregation of data across years, counties, and/or offenses, as has been previously done (see Section I). To illustrate one example, Figure 4 repeats Figure 3 but for the combined years 2010-2021. Aggregation allows many more offenses to clear the  $N > 10$  hurdle in many more counties. The tradeoff, of course, is that one is implicitly assuming that the cases are sufficiently similarly situated across the span of more than a decade of data. Under the right circumstances, such approaches may be advisable, enabling the RJA to reach a broader set of Californians and footprint of the criminal justice system.

## CONCLUSION

The California Racial Justice Act holds considerable promise as a tool for identifying and remedying racial disparities in the criminal justice system, but its potential has been constrained by the lack of data needed to detect the presence or absence of an actionable disparity. This paper describes the Paper Prisons Racial Justice Act data tool, a new resource for accessing comprehensive criminal record offender information on criminal justice outcomes by race at the county, offense and stage-of-conviction level, and details the underlying data, from its provenance to its processing to its production, by the tool.

We take advantage of the tool's comprehensive data to address one of the most important open questions that surrounds RJA implementation—how courts should interpret the requirement that defendants be "similarly situated" and in the "same county," and what type of data should be considered sufficient. We find that for many



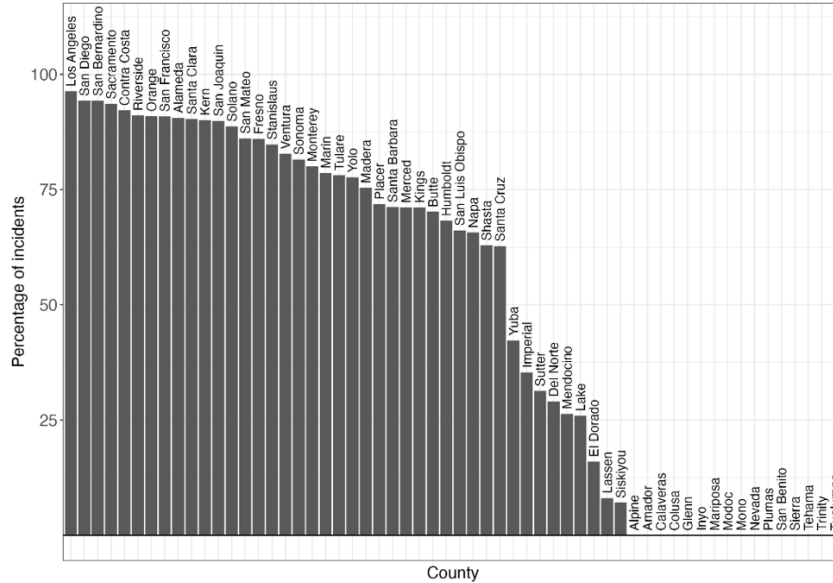
offenses, especially in smaller counties, for less populous racial/ethnic groups like Native Americans, and for the later stages of prosecution, there are often not enough cases to meet the privacy-preserving minimum sample size requirements imposed on the dissemination of individual-level criminal history data, much less to make statistically valid comparisons.

If the RJA is to have its intended impact of “eliminating racial bias” from the criminal justice system, its evidentiary standards cannot be so strict as to make it nearly impossible to demonstrate a disparity except for the most common offenses in the most populous counties and racial groups. While respecting the need for comparability, flexibility to aggregate data across geographies, time periods, and offense categories will be needed in order to accrue adequate sample sizes—and to give the RJA teeth outside California’s major urban counties. Fortunately, the language of the statute, and the cases so far, as described in Part I, reflect a level of flexibility that will likely need to be further stretched as additional cases are brought.

In addition, further data collection and research are needed at both the state to enable more surgical comparisons of “similarly situated” individuals—comparisons that take into account not just county and offense but factors like prior criminal history, strikes, and other case and defendant characteristics that we hope to incorporate into later versions of the tool. Ultimately, realization of the RJA’s vision of providing identifying and remedying racially discriminatory practices in the criminal justice system will require the sustained collaboration between state and local criminal justice agencies, researchers, and justice advocates.

**Figure 1**

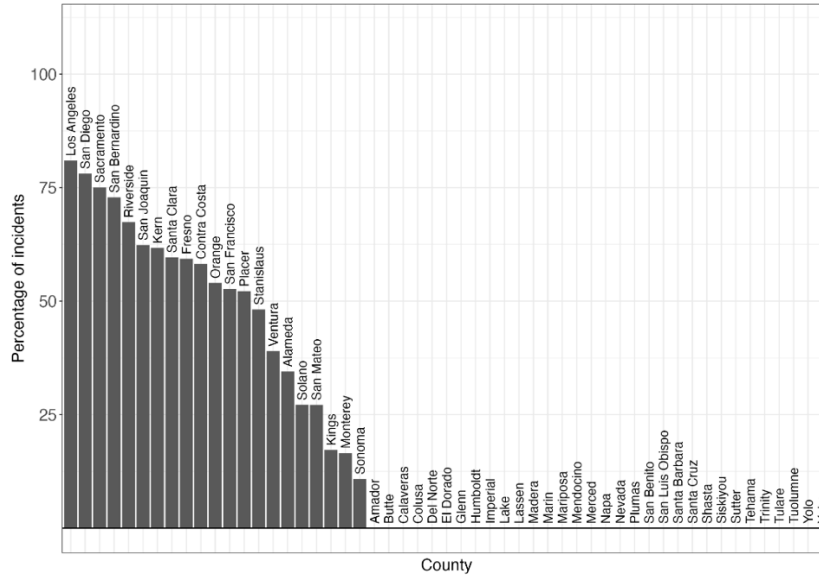
Percent of arrest incidents in 2019 with offense-specific sample size greater than 10, Black-white comparison.



Source: CORI data (see text).

**Figure 2**

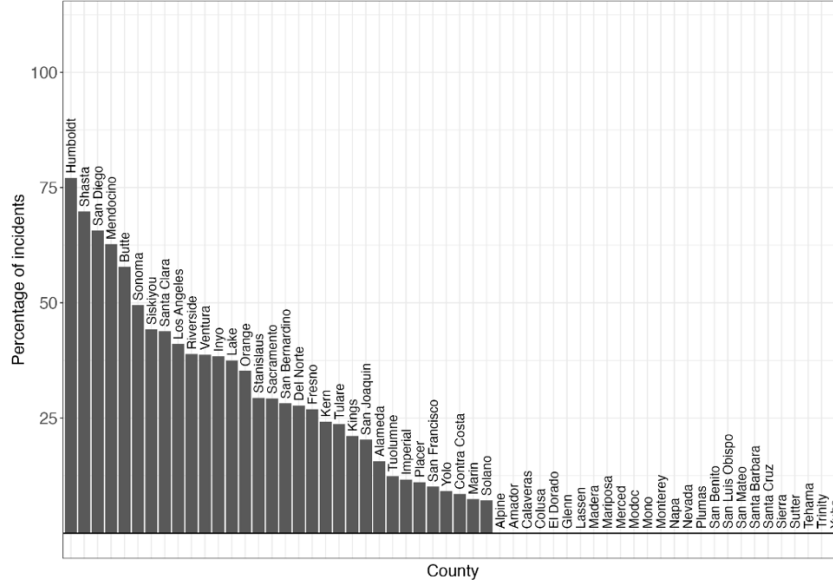
Percent of felony conviction incidents in 2019 with offense-specific sample size greater than 10, Black-white comparison.



Source: CORI data (see text).

**Figure 3**

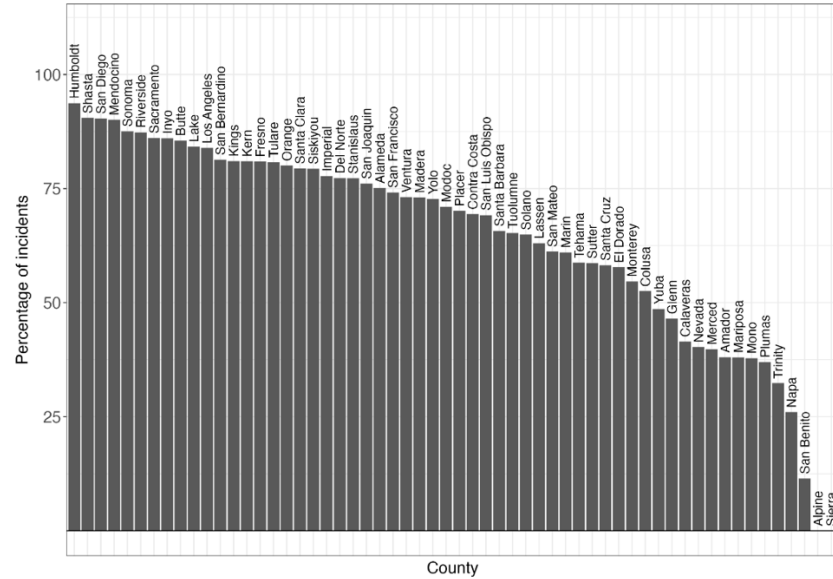
Percent of arrest incidents in 2019 with offense-specific sample size greater than 10, Native American-white comparison.



Source: CORI data (see text).

**Figure 4**

Percent of arrest incidents in 2010-2021 with offense-specific sample size greater than 10, Native American-white comparison



Source: CORI data (see text).