Chicago-Kent Law Review

Volume 98 | Issue 1 Article 7

2-20-2024

How Can You Tell If There is a Crisis? Data and Measurement Challenges in Assessing Jury Representation

Mary R. Rose

Marc A. Musick

HOW CAN YOU TELL IF THERE IS A CRISIS? DATA AND MEASUREMENT CHALLENGES IN ASSESSING JURY REPRESENTATION

MARY R. ROSE & MARC A. MUSICK*

INTRODUCTION

The premises of this special issue's theme are that society is in a time of crisis and that such a time is, in some manner, reflected in juries, perhaps to such an extent that juries are in crisis. Empirical evidence is the workhorse of any trustworthy means of arbitrating the truth of these premises. Empirical data can reveal economic, social, or political patterns that may indicate a crisis more generally. In addition, courts' providing routine data on how jury systems operate, who serves on juries, and the decisions juries reach can indicate whether or not juries are in a state of crisis. Nonetheless, one need not declare a crisis to recognize the everyday usefulness of data about courts and juries. There has long been a call for greater and higher-quality empiricism in law more generally.[1] Recently, scholars have called on courts (and fellow scholars) to identify and use empirical data that courts already possess, in order to understand courts' practices.[2] Such data may also

* Professors at University of Texas at Austin. We are extremely grateful to Bradley Silberzahn for research assistance work, Chantal Hailey for thoughtful remarks, and Nancy Marder, Sanja Kutnjak Ivković, Valerie Hans and the entire Collaborative Research Network on Lay Participation in Legal Systems within the Law and Society Association for identifying and supporting opportunities to share work on juries. All content and any errors in this paper are entirely our responsibility.

1. *See, e.g.*, Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1, 4 (2002) ("applaud[ing]" empirically based studies in law, while calling for higher-quality research); Michael Heise, *The Importance of Being Empirical* 26 PEPP. L. REV. 807, 813–815 (1999) (describing negative consequences of the "dearth" of empirically-informed legal scholarship); Oliver Wendell Holmes, *The Path of Law*, 10 Harv. L. Rev. 457, 469 (1897) ("For the rational study of the law the black-letter man may be the man of the present, but the man of the future is the man of statistics and the master of economics.").

2. *See, e.g.*, Ronald F. Wright, Kami Chavis & Gregory S. Parks, *The Jury Sunshine Project: Jury Selection Data as a Political Issue*, 2018 U. ILL L. REV. 1407, 1416–1422 (2018) (describing the limits of existing record-keeping on jury selection and the efforts required to create an dataset representing outcomes across courthouses). *See generally* Nina W. Chernoff, *No Records, No Right: Discovery and the Fair Cross-Section Guarantee*, 101 IOWA L. REV. 1719, 1749–1764 (2016) (discussing legal bases for the courts' making data available to defendants challenging jury pools).

identify patterns of practice in policing and other legal arenas.[3] In this way, we can examine juries in times of crises *and* in all other times.

A few years ago, New Jersey's Administrative Office of the Courts ("AOC") sought to use data to examine its juries, particularly who served on them, in order to assess the representativeness of their juries and patterns, if any, in who participates on juries. As one part of this review, New Jersey wished to know whether the number of peremptory challenges available to parties in criminal cases[4] contributed to jury composition, particularly with respect to race and gender. A detailed analysis and report to the AOC showed that juries commonly failed to represent their communities;[5] however, this lack of representation occurred at a much earlier phase than courtroom voir dire and use of peremptory challenges.[6] Just as other studies have shown,[7] results indicated that the juries selected in the roughly six-week study period tended, on average, to reflect the profile of the people who appeared at court to serve on the jury.[8] Although results were counterintuitive – that is, the large number of peremptory challenges was not the primary contributor to patterns of underrepresentation[9] – no finding should obscure the impressive

3. *See generally* Andrew Manuel Crespo, *Systemic Facts: Toward Institutional Awareness in Criminal Courts*, 129 HARV. L. REV 2049 (2016).

4. At the time of its study, New Jersey permitted twenty peremptory challenges to criminal defendants in the most serious felony cases and twelve to the state; other felony cases allotted ten to each side. Civil cases provided six per side. N.J. Rev. Stat. § 2B:23-13 (2014); *see also* NEW JERSEY COURTS, A GUIDE TO THE NEW JERSEY JUDICIAL CONFERENCE ON JURY SELECTION 11–12 (2021), https://www.njcourts.gov/courts/assets/supreme/judicialconference/printnolinks.pdf?c=u4y (calling New Jersey an "outlier" on peremptory challenges and providing a chart of number of peremptory challenges by state) [https://perma.cc/93SU-4SNG].

5. MARY R. ROSE, FINAL REPORT ON NEW JERSEY'S EMPIRICAL STUDY OF JURY SELECTION PRACTICES AND JURY REPRESENTATIVENESS 40 tbl.III.3 (2021), https://www.njcourts.gov/courts/assets/supreme/judicialconference/Mary_Rose_Final_Report.pdf?c=Ogy [https://perma.cc/BRN4-4LED] (showing underrepresenting in the majority of counties). The first author (Rose) wrote the New Jersey report, but as will be described below, the second author (Musick) advised on data matters to ensure any data errors were minimized.

6. *Id.* at 40 (the chart of disparities compares each county's demographic profile to those who appear for service, regardless of whether they were selected for a jury or not).

7. *See, e.g.*, Shari Seidman Diamond, Destiny Peery, Francis J. Dolan, & Emily Dolan, *Achieving Diversity on the Jury: Jury Size and the Peremptory Challenge*, 6 J. EMPIRICAL LEGAL STUD. 425, 436 (2010) ("[T]he overall impact of jury selection on the demographic composition of the jury venire was tiny."); Jacinta M. Gau, *A Jury of Whose Peers? The Impact of Selection Procedures on Racial Composition and the Prevalence of Majority-White Juries*, 39 J. CRIME & JUST. 75, 80 tbl.1 (2016) (showing similar composition of the venire and final juries); Mary R. Rose, *The Peremptory Challenge Accused of Race or Gender Discrimination? Some Data from One County*, 23 LAW & HUM. BEHAV. 695, 698 (1999) (reporting no relationship between race and the likelihood of being excused from the jury).

8. ROSE, *supra* note 5, at 49–54.

9. Instead of a generalized pattern, results pointed to concerning patterns in peremptory use in particular counties. *Id.* at 69 (summary of results stating that peremptory challenges "play a role underrepresenting [minority] groups on some juries, but only an attenuated one."). This attenuated role was likely due to the fact that attorneys failed to use the full complement of challenges granted to them, potentially blunting the effect of peremptory challenges on jury composition. *See, e.g.*, *id.* at 62 tbl.V.1 (showing the use of peremptory challenges on African Americans). The report discussed a separate harm

fact that New Jersey was willing to undertake an empirically informed examination of its own system.

Other state court systems, and the federal court system, should do likewise. Precisely because we believe courts should rigorously evaluate their jury systems using empirical data, our contribution to this symposium focuses on some of the data challenges we faced when assisting New Jersey with its study. Identifying such challenges is particularly important when, as was true for New Jersey, the court had to field a special study and generate new data to identify the race and gender of its jurors.

This article focuses on three separate issues: (1) because court systems are not typically designed to generate data for analytic purposes, researchers may have to intensively clean any data that does exist in order to make it usable and reliable; (2) efforts to comprehensively measure racial identity recognize that some people identify as multiracial, but it is unclear whether multiracial persons form a "distinctive" or "cognizable" group; and (3) groups that are smaller in number may be included in the dataset, but data analysts likely will remove them for purposes of analysis, a counterintuitive result in a study that aims to measure identity as inclusively as possible. Although there is some potential overlap across these issues (e.g., multiracial groups may be too small to analyze in some instances[10]), in general they raise separable issues. We present each of these issues because for us, they were substantial challenges, either consuming a great deal of time[11] or requiring significant judgments about how best to handle them.[12] For these reasons, these issues are worthy of discussion as court officials and scholars consider how best to use data to assist in understanding juries in the current era.

In Part I we review the law on fair cross-section claims and laws governing peremptory challenge use, and we describe what New Jersey AOC asked us to do. Part II describes the dataset we received and our efforts to clean the data. In Part III, we consider the challenges of conceptualizing multiracial persons as a "distinctive" or "cognizable" group, and we describe the choices we made when analyzing data for those who identify as

that peremptory challenges likely created: court staff had to summon far more people than needed for each jury venire. *Id.* at 87–93 (analysis and discussion of the large number of people who served on each venire, including high proportions of people who were never questioned for a case). Hence, the report did not conclude that New Jersey's unusually high number of peremptory challenges had no negative effect on the jury system. *Id.* at iii–ix (summarizing concerns about New Jersey's system of peremptory challenges).

10. *See, e.g.*, *infra*, Part III.C, Table 2 (showing that the number of multiracial persons, when considered by some categories of jury service outcomes, amounted to twenty-five or fewer individuals).

11. *Infra* Part III.B (discussing the specific reasons that data needed to be cleaned and checked for accuracy).

12. *Infra* Part III.C (discussing decisions about best way of classifying multiracial persons).

multiracial. Part IV describes the challenging issue of the best way to measure gender, given that, at present, the vast majority of people identify under the male-female binary. Part V offers some concluding thoughts.

I. REPRESENTATIVE JURIES AND ONE STATE'S ASSESSMENT

A. Standards for Fair Cross Sections and Equal Protection Requirements in Jury Selection

Under the Sixth Amendment, defendants in criminal cases are entitled to be tried by a fair cross section of the community.[13] In *Duren v. Missouri* (1979), the Supreme Court laid out a three-part prima facie test to assess whether a given amount of underrepresentation in a jury panel violates the fair cross section requirement. Under the first prong the defendant must show that the underrepresented population constitutes a "distinctive group."[14] Additionally, the defendant must show that the amount of underrepresentation is "not fair and reasonable in relation to the number of such persons in the community,"[15] and, finally, that the cause of underrepresentation is due to "systematic exclusion"[16]—that is, something inherent in the particular system used.[17] If these are proven, the government must explain the need for the practice in question, specifically whether "a significant state interest [is] manifestly and primarily advanced by those aspects of the jury-selection process . . . that result in the disproportionate exclusion of a distinctive group."[18]

The analyses New Jersey sought from us were not bound by the *Duren* standards, since their assessment was unrelated to litigation. We therefore could adopt, for example, our own determinations of what levels of underrepresentation should be cause for concern."[19] Still, as described in more detail below,[20] Duren's requirement that the underrepresentation

13. Taylor v. Louisiana, 419 U. S. 522, 528 (1975) (calling selection from a representative fair cross section of the community an "essential" component of the Sixth Amendment right to a jury). Defendants can also pursue claims under the equal protection clause, but this is generally regarded as a more challenging case to make because one must show intentional discrimination rather than racially disparate impacts of policies. *See* Nina W. Chernoff, *Wrong About the Right: How Courts Undermine the Fair Cross-Section Guarantee by Confusing It with Equal Protection*, 64 HASTINGS L.J.141, 150–165 (2012) (delineating differences between fair cross section and equal protection analyses).

14. Duren v Missouri, 439 U.S. 357, 364 (finding that women are a distinctive group, as they are "sufficiently numerous and distinct from men").

15. *Id.*

16. *Id.*

17. *Id.* at 366.

18. *Id.* at 367–68.

19. ROSE, *supra* note 5, at 10 (defining "concerning" as a 25% comparative disparity).

20. *See infra* Part III (discussing what constitutes a distinctive group in law).

involve a "distinctive group" proved an important consideration in deciding how to analyze the data.

A different set of laws governs protection against discrimination when a jury is being selected and when attorneys make decisions about how to use their limited number of peremptory challenges. In this area, court rulings or state laws indicate which groups have an equal protection interest in not being discriminated against in peremptory usage. Thus, the U.S. Supreme Court has deemed strikes based on race to be unconstitutional,[21] regardless of which side exercises the strike,[22] the race of the defendant,[23] or whether the case is civil rather than criminal.[24] Therefore, race is a "cognizable group"[25] for an equal protection analysis of peremptory strikes. The Court has also prohibited strikes based on gender[26] and stated that strikes based on Latino ethnicity would be an equal protection violation.[27] State laws may impose other restrictions on peremptory strikes.[28]

To understand in what ways final, petit juries fail to represent communities, a comprehensive study should include both issues governed by the laws described above, that is, both who appears for service at all and who is selected following voir dire in a specific case. New Jersey undertook such a study, aiming to know both to the extent to which prospective jurors who appeared at court resembled their communities as well as the effect, if any, of the large number of peremptory challenges provided to parties.

B. The New Jersey Study

At the time of the start of New Jersey's study of its jury system,[29] the state did not routinely collect demographic data about the people appearing

21. Batson v. Kentucky, 476 U.S. 79, 99 (1986) ("By requiring trial courts to be sensitive to the racially discriminatory use of peremptory challenges, our decision enforces the mandate of equal protection and furthers the ends of justice.").

22. Georgia v. McCollum, 505 U.S. 42, 59 (1992).

23. Powers v. Ohio, 499 U.S. 400, 402 (1991).

24. Edmonson v. Leesville Concrete Co., 500 U.S. 614, 616 (1991).

25. *Batson*, 476 U.S. at 96 (citing Casteneda v. Partida, 430 U.S. 482, 494 (1977)).

26. J.E.B. v. Alabama *ex rel*. T.B., 511 U.S. 127, 128–29 (1994).

27. Hernandez v. New York, 500 U.S. 352, 355 (1991).

28. *See, e.g.*, Assemb. B. 3070, 2019-20 Reg. Sess. (Cal. 2020) (protected groups include: "race, ethnicity, gender, gender identity, sexual orientation, national origin, or religious affiliation, or perceived membership in any of those groups" and offering reasons for a dismissal that are presumptively invalid); WASH. SUP. CT. GEN. R. 37. (listing presumptively invalid reasons for a peremptory strike and stating that a judge may deny use of challenge if "the court determines that an objective observer could view race or ethnicity as a factor in the use of the peremptory challenge"); ARIZ. SUP. CT. ORD. No. R-21-0020 (Aug. 30, 2021) (eliminating peremptory challenges entirely).

29. Since that time, the state's high court instructed the state to hold a conference devoted to reviewing its jury system. *See* State v. Andujar, 254 A.3d 606, 612 (N.J. 2021). An outgrowth of that conference is a new system for collecting juror demographics. *See Jury Reforms and Attorney-Conducted*

for service. New Jersey was not unique in its absence of data,[30] and the effects of not having this information are immense. In these systems, there is simply no way to know whether any practices are differentially associated with attrition from service for race or gender groups. Should a court system, or even just a party in a case, wish to know the racial, ethnic, and gender profile of people appearing for service in a given time period, it is impossible to provide a comprehensive analysis when jurisdictions do not pose demographic questions to prospective jurors, both those appearing for service and those who contact the court for exemptions or say they are disqualified. The only available information about race or gender in such systems comes from looking at the members of a specific jury pool questioned for voir dire and, perhaps, merely guessing people's racial identity. To appropriately study the demographic effects of its practices, including its large number of peremptory challenges, court administrators in New Jersey added another step to normal practices and introduced a questionnaire to jurors that captured demographic information.[31]

Data gathering for the study took place during a several-week period in September and October of 2018.[32] Fifteen counties disseminated one-page questionnaires to all persons appearing for jury service.[33] Through the questionnaire, people could voluntarily indicate their race, Latino ethnicity, and gender. Each questionnaire was affixed with a custom bar code number, linked (in a deidentified manner) to an individual juror outcome that was stored in the existing Jury Management System ("JMS") software. New Jersey thereby obtained information on respondents' race, ethnicity, and gender, as well as the outcome of their appearance. With Court authorization, these deidentified data were made available to the authors for analysis.

The outcome information in the dataset indicated whether individuals were:

- part of a "pool only," which meant that they were not called to a particular courtroom for questioning;
- selected as juror;
- excused for cause;

Voir Dire Pilot Program, N.J. CTS., https://www.njcourts.gov/attorneys/jury-reforms (last visited Feb. 2, 2023).

30. *See* PAULA HANNAFORD-AGOR & G. THOMAS MUNSTERMAN, NAT'L. CTR. FOR STATE CTS., FINAL REPORT: THIRD JUDICIAL CIRCUIT OF MICHIGAN JURY ASSESSMENT SYSTEM 26 (2006) (urging collection of data on juror demographics). According to Paula Hannaford-Agor of the National Center for State Courts, most states do not collect race or ethnicity data from jurors. E-mail from Paula Hannaford-Agor to author Mary Rose (Sept. 2, 2022, 1:19 PM) (on file with authors).

31. ROSE, *supra* note 5, at 21.

32. *Id.*

33. *Id.*

- peremptory challenged by a plaintiff's attorney in a civil case or by a prosecutor in a criminal case;
- challenged by the defendant (civil or criminal) through a peremptory;
- not reached during voir dire, meaning they were not considered for either selection or dismissal in a trial (coded in the data as "not used").[34]

These data permitted an examination of the demographic profile of all the people who appeared for service in the counties, which could be compared to each county's demographic profile. Additionally, the data revealed the racial and gender identity of those who were selected, excused, or not reached, allowing for as assessment of whether these outcomes were patterned according to race or gender.

The final dataset had 15,529 observations from fifteen counties.[35] After cleaning the data,[36] we found that data from one county had unusually high levels of missing questionnaire data,[37] rendering results uninterpretable for that area.[38] That left over 12,000 observations from fourteen counties, counties with residents that, combined, reflected 84% of the state's population.[39] Of these, 5,055 (41%) never made it to any voir dire, whereas 7,407 reflected people who underwent questioning for one or more cases.[40]

II. ENSURING THAT COLLECTED DATA ARE TRUSTWORTHY

One of our aims in the study was simply to descriptively summarize data—for example, determining the mean and standard deviation for each outcome (such as the average number of peremptory strikes used). Second, we wanted to make inferences about relationships between variables in the dataset—for example, whether race or gender are correlated with the use of peremptory strikes. To do these tasks, data must be "cleaned." Cleaning data

34. *Id.* at 22. For additional detail on the study, see Memorandum from Jessica Lewis Kelly & Lisa R. Burke to Hon. Glenn A. Grant (April 23, 2018) (on file with authors); BRIAN J. MCLAUGHLIN & LISA R. BURKE, FINAL REPORT ON THE COMBINED JURY STUDIES ON JURY REPRESENTATIVENESS AND THE IMPACT OF PEREMPTORY CHALLENGES ON THE RACIAL AND ETHNIC COMPOSITION OF PETIT JURIES 1 (2018) (on file with authors).

35. ROSE, *supra* note 5, at 21, 25.

36. *See infra* Part II.

37. ROSE, *supra* note 5, at 31 The response rate for this county was just 33% and the next lowest response rate was 71%, so this was an outlier result. With this county eliminated, the overall response rate to the AOC study was high, 84.7%. *Id.*

38. *Id.* (noting that "any analysis of the demographic composition of pools, venires, or juries would be subject to high levels of error given the levels of missing data"). As the JMS system did not hold demographic data, when questionnaires were missing, there was no way to know the demographic identities of those who served, were dismissed, not reached, or who never went to a voir dire.

39. *Id.* at 38 n.16.

40. *Id.* at 33.

involves going through the dataset one wishes to analyze and detecting, diagnosing, and editing instances of faulty data, or mistakes that can occur at all phases of data collection and analysis.[41] When datasets contain faulty data, two key properties of a study—reliability and validity[42]— are undermined. Just as one would not trust a manuscript that had not been carefully edited for errors, the aim of data cleaning is to produce a trustworthy dataset. No dataset will be perfect, but a cleaned dataset will have attempted to minimize mistakes.

A. Minimizing Mistakes in Data

Van den Broeck and colleagues have offered a helpful framework for the types of errors that can occur at each stage of a research project, including mistakes that occur in the study's design, data collection, data extraction and transformation phases, and mistakes that can occur when analyzing data.[43] Faulty data can occur in numerous situations and include: (1) observations that are missing from the dataset or that have data missing on specific questions (e.g., there are blanks where there should be answers);[44] (2) instances of excess data[45] (e.g., a subset of people have two entries in the dataset, whereas only one is expected); (3) answers may be logically impossible (e.g., an implausible outlier value, such as someone listing their age as 233[46]) or produce "strange patterns,"[47] which refer to answers that are logically inconsistent with other answers (e.g., someone says on one question that they have been arrested before but on another question says they have never had any contact with law enforcement); or (4) data analytic results

41. Jan Van den Broeck et al., *Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities*, 2 PLOS MED. 0966, 0966 Box 1 (2005).

42. *See* DEBORAH CARR ET AL., THE ART AND SCIENCE OF SOCIAL RESEARCH (2d ed. 2021) (chapter five titled "Evaluating Research," describing concepts of reliability and validity; reliability means that measures and study observations minimize random error and "noise" so that measures taken are not, for example, the product of guesses or mistakes, and validity refers to the lack of bias in measurement).

43. Van den Broeck et al., *supra* note 41, at 0967.

44. *Id.*

45. *Id.*

46. *Id.* at 0968 ("Some data points are clearly logically or biologically impossible."). As the authors note, some outliers are "true extreme[s]," meaning that the answer is correct but just highly unusual. *Id.* These can be harder to diagnose than an outlier answer that is patently false. To give another example, no newborn baby has a birthweight of seventy pounds; in that instance, in all likelihood, someone wrote "70" instead of "7" for the baby's weight, producing a faulty outlier, however, a small proportion of newborns are unusually heavy, which would mean that it's difficult to know if a value of "12" is a true outlier (an unusually heavy baby) or faulty data. *See Large for Gestational Age (LGA)*, CHILD. HOSP. OF PHILA., https://www.chop.edu/conditions-diseases/large-gestational-age-lga [https://perma.cc/5PGG-TUA5] (describing the weight for a low birthweight baby as possibly an error or a sign of malnutrition).

47. Van den Broeck et al., *supra* note 41, at 0967.

might seem "suspect"[48] (e.g., variables for height and weight do not correlate with one another as they should).

A wide variety of actions could produce mistakes like these. Missing observations ("form missing"[49]) occur because, for example, people were not approached for a study when they should have been or because they declined to participate when asked. Alternatively, how the data were managed once collected could omit people: someone entering the data could have made a mistake and left off data from a respondent; a machine that scans in data could have missed a form; or someone analyzing correct data, while in the process of transforming data or creating new datasets, may have accidentally omitted data for one or more respondents.[50] When specific answers are missing, respondents may have skipped answers on a questionnaire, or any of the data-entry/data handling issues just mentioned could be at fault.[51]

Duplicate or excess data can occur if the same person was inadvertently given two forms to fill out, or if someone entering or analyzing the data (or a machine reading the data) erroneously entered the same observation twice.[52] Two or more datasets being combined may lead to duplicate observations, particularly when the variable that links two datasets, such as a supposedly unique identification number, has mistakes in it.[53] Faulty outliers or mistakes that lead to logically inconsistent responses across variables can occur because a respondent filling out a questionnaire marked the wrong answer, because someone administering the questionnaire wrote down the wrong response, because a programmer working with the data inadvertently changed an answer, or because someone else working with the data made a mistake.[54]

All of these issues reflect the same general concern that the data contain mistakes. People who should be in the dataset are missing entirely, or they are represented in the data, but their responses are incorrect; alternatively, there are more people in the data than there should be. Ideally, datasets would contain none of these errors, but perfection may be impossible. When there is a small number of these types of mistakes, it is more likely that the mistakes represent random error in the data. However, as the amount of faulty data increases, mistakes risk being systematic rather than random

48. *Id.*
49. *Id.* at 0968 tbl.1.
50. *Id.*
51. *Id.*
52. *Id.*
53. *Id.*
54. *Id.*

occurrences, and with enough errors, the dataset may simply be too flawed to trust.

B. Concerns with the Original New Jersey Data

For several reasons, the New Jersey study was at risk for producing some amount of faulty data that needed cleaning. First, the study constituted additional work and deviations from standard practice for each county's jury staff.[55] Further, for jurors, participation in the study was voluntary.[56] Both of these factors can affect response rates for the study, with lower response rates indicating that people who should be in the study are missing. Response rates did, indeed, vary a great deal across counties,[57] likely because staff in different counties faced distinct barriers in fielding the study on any given (e.g., busy) day, or because jurors in different counties varied in their willingness to share demographic information about themselves. Because missing data can occur for a variety of reasons,[58] we were unsure whether nonresponse reflected a problem at the county level (i.e., the data were not collected) or whether something went wrong at a later stage (e.g., the data were scanned incorrectly). Given differential response rates across counties,[59] identifying the correct explanation could be consequential. If data were not collected, there was nothing further to do; however, if something went wrong with, for example, the scan of the questionnaires, data that were missing could potentially be recovered by redoing the scan.

Second, because the demographic data were an add-on, the dataset created by giving questionnaires to jurors had to be merged with the dataset routinely kept in the JMS, which stored information about the outcomes for the jurors.[60] Merging together datasets presents another opportunity for faulty errors to occur,[61] as records could be inadvertently duplicated or deleted. Duplicate values of the same identification number can be one signal of a merge problem. Problems at this stage are also evident via a pattern of missing responses on variables that suggests that the merge did not take. For

55. *See* ROSE, *supra* note 5, at 25–26 (describing the "demands [the] project placed on various types of personnel").

56. Kelly & Burke, *supra* note 34, at 9 ("The juror then completes the survey, which is still done in a voluntary and confidential way.").

57. ROSE, *supra* note 5, at 32 tbl.II.1 (showing response rates between 33% and 97.3% across counties).

58. *See* Van den Broeck et al., *supra* note 41, at 0968 tbl.1; *see also supra* Part II.A (reviewing reasons why data may be faulty).

59. ROSE, *supra* note 5, at 32 tbl.II.1.

60. *Id.* at 22.

61. Van den Broeck et. al., *supra* note 41, at 0967.

example, outcome information from the JMS may exist for a juror identification number, but the same identification number has no demographic information listed, or the reverse occurs (demographic information exists, but no associated outcome data is in the dataset). Yet missing demographic information could also arise from someone simply refusing to participate or perhaps failing to answer specific questions. Without access to the original, unmerged datasets, it is not possible to diagnose the problem. We also learned from the AOC staff that in New Jersey the same person can be sent to more than one voir dire during their jury duty period. This could create duplicate identification numbers that are accurate (they had multiple voir dire experiences, including on the same day). Again, access to the original files would be necessary to determine if the merge caused duplicate observations, or some other factor explained these.

The existence of a defined study period appeared responsible for an additional problem: implausible values. In the data, we discovered that several trials were listed as having only, say, three jurors, which is constitutionally impermissible.[62] In some cases, the data indicated that no challenges (peremptory or cause) were exercised on anyone within a voir dire— theoretically possible but highly unlikely. Finally, some people were linked with a trial number (suggesting they went to a voir dire), but there was no outcome information for them, which should not occur in a working JMS system. Once these problems were identified, the AOC staff were able to investigate them. They explained to us that sometimes jury selection commenced before the study began, creating missing data for a subset of people associated with the trial (e.g., those who had already been excused before the study began). Alternatively, the study ended before jury selection was complete, which could explain why some trials were listed as having, for example, only three jurors selected, since jury selection was not complete when data collection stopped.

With the study period problem identified, we looked closely at dates for each county across all cases and removed trials that appeared to have been affected by the study-period boundary. Still, we came across other cases within the study period that had suspicious but not impossible values, such as trials with no peremptory challenges exercised. As with the other issues described above,[63] we were concerned that perhaps the merge process, or

62. Ballew v. Georgia, 435 U.S. 223, 243–45 (1978) (holding that a criminal trial before a five-member jury violated the Sixth and Fourteenth Amendments).

63. *See* Van den Broeck et al., *supra* note 41, at 0968 tbl.1; *see also* discussion, *supra* Part II.A (reviewing reasons why data may be faulty).

some earlier phase (e.g., scanning the data) had introduced an error, and it was simply not possible to know just by looking at the merged dataset provided to us.

One final matter stemmed from the basic fact that different people within the New Jersey court system handled different facets of the study. As noted, the AOC staff designed the study, jury staff implemented it, and still another unit created the datasets. This diffusion of roles and expertise over the elements of the study is to be expected when an additional study is added to the existing and already-demanding work of courts. Yet this situation also made us mindful that helping to diagnose the source of any potential errors we found – to check original questionnaires, answer questions about data collection processes for a specific time period, to rescan data, or to recreate merged datasets – was a lot to ask of busy professionals in different departments.

In the end, we recommended to the AOC staff that the best course of action was for us to work with the original scanned data files, and to essentially recreate the merged dataset ourselves.[64] With this approach, we could not catch scanning or other data-entry errors, since it was not feasible to receive and re-enter over 15,000 original questionnaires. But we would have better control over the merge process and could observe which original datasets had which observations and then confirm that the final count of people in the dataset reflected the inputted datasets. Working with the raw files was labor-intensive and required us to substantially increase our estimate of the number of hours necessary to complete the review of the data.

C. The Value of Streamlined, In-House Data Collection

Two implications flow from this experience. The first is that calls for greater access to and use of court data need to take seriously the steps and necessary skills required to ensure that the data used are trustworthy.[65] Having data exist within a courthouse is not the same thing as being able to use those data. As the above recitation of issues indicates, we received data under what may be the best of circumstances. The AOC was a willing party who cooperatively shared their data, those sharing the data were also the ones

64. ROSE, *supra* note 5, at 24.

65. Crespo, *supra* note 3, at 2104–06. In his call for greater use of data that courts hold, Crespo raises issue of both competency and cost associated with making these data more available; he suggests partnering with social scientists with particular subject matter expertise as a way to enhance competency, and he suggests benefits that may outweigh costs of becoming more fluent with data. *Id.* at 1204–08. We support such partnerships and also believe that there are substantial benefits associated with greater empiricism in courts. We would simply note that training people to give not just any data, but clean, readily usable data would be a massive undertaking in most areas.

who designed and fielded the study and therefore knew the intended procedures, and this same group was highly invested in working with us to answer questions or research how to get additional information from others. Even with all these advantages, we faced challenges in making the data clean and therefore trustworthy, with the best course of action requiring an intensive recreation of a final dataset.

More importantly, however, our experience demonstrates why it is crucial that courts routinely collect, at the very least, the race, ethnicity, and gender profiles of its jury pools. Without this information being embedded into its jury management system, New Jersey had to reconfigure procedures to create a new database in order to assess its own practices. This reconfiguration not only required additional work from staff in different units, but it also increased the likelihood that the new steps could inadvertently create missing data or duplicate entries. If a court simply wishes to know if their jury system includes a diverse cross-section of the community (or not), the process of making those data usable and reliable should be simpler for the court and for data analysts.

Following the state supreme court ruling in State v. Andujar,[66] New Jersey will no longer have to create ad hoc procedures for system-assessments. A more defined set of in-house personnel (e.g., JMS managers or other data staff in counties or in the statewide office) will, over time, become experts in the new unified system, allowing for a more centralized source of knowledge about data collection and data processing. The latter should lead to a streamlined process for answering questions that arise when data analysts clean or examine a dataset. Other states that fail to ask for juror demographics should follow New Jersey's lead and create processes that make empirical assessments of jury systems—or other court practices—easier to undertake and therefore more helpful in addressing any concerns that exist.

Even when data are cleaned, other challenges remain. We next discuss two issues having to do with how to capture complexity when measuring people's social identities and the implications of having only a small number of people in a single category.

66. State v. Andujar, 254 A.3d 606, 612 (N.J. 2021) (initiating the Judicial Conference on Jury Selection to examining the jury selection process); *see also Jury Reforms and Attorney-Conducted Voir Dire Pilot Program*, N.J. CTS., https://www.njcourts.gov/attorneys/jury-reforms (last visited Feb. 2, 2023) (showing the outcomes of the Judicial Conference on Jury Selection).

III. MULTIRACIAL PERSONS AS MEMBERS OF A "DISTINCTIVE" GROUP

To gather information on the racial, ethnic, and gender identity of all the jurors who appeared at court during the study period, New Jersey used a form developed by the National Center for State Courts to ask these questions.[67] This form comprehensively measures identities and follows best practices by drawing on how the Census Bureau defines racial categories. The text of the race question appears in Table 1. As indicated, the question on the form listed the categories recognized by the Census Bureau and provided the Census-recognized definitions and terminology for each.[68]

Table 1. Measurement of Racial Categories on New Jersey's Study Questionnaire[69]

RACE: (check all applicable categories)

- **White or Caucasian**: A person having origins in any of the original peoples of Europe, North Africa or the Middle East.
- **African-American or Black**: A person having origins in any of the black racial groups of Africa.
- **American Indian or Alaskan Native**: A person having origins in any of the original peoples of North America and who maintains cultural identification through tribal affiliation or community recognition.
- **Asian**: A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent. These areas include, for example, China, India, Japan, and Korea.
- **Native Hawaiian or Pacific Islander**: A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. It

67. *See* Kelly & Burke, *supra* note 34, at 3 (indicating decision to use National Center for State Courts survey to measure racial categories).

68. *About the Topic of Race*, U.S. CENSUS BUREAU, https://www.census.gov/topics/population/race/about.html (Mar. 1, 2022) [https://perma.cc/8Y5C-M6NA]. Of note, Census forms that households receive do not list category definitions, but instead provide country-origin examples, particularly for the Asian category. U.S. CENSUS BUREAU, INFORMATIONAL COPY OF U.S. CENSUS FORM (2010), https://www.census.gov/content/dam/Census/programs-surveys/decennial/technical-documentation/questionnaires/2010questionnaire.pdf [https://perma.cc/AWZ7-KNAC] [hereinafter 2010 CENSUS FORM]. The 2020 form used similar categories but asked all respondents to select a race and also to write in their country origin. U.S. CENSUS BUREAU, INFORMATIONAL COPY OF U.S. CENSUS FORM (2020), https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/questionnaires-and-instructions/questionnaires/2020-informational-questionnaire.pdf [https://perma.cc/V2ZY-QA6Q] [hereinafter 2020 CENSUS FORM]. Of course, scholarly criticisms of Census Bureau's terminology and practices in measuring race and ethnicity still occur. *See, e.g.*, Hephzibah V Stramic-Pawl, Brandon A. Jackson & Steve Garner, *Race Counts: Racial and Ethnic Data in the U.S. Census and the Implications for Tracking Inequality*, 4 SOCIO. RACE & ETHNICITY 1 (2018) (explaining the history of political considerations in Census practice and reviewing areas of controversy in the modern era).

69. Kelly & Burke, *supra* note 34, at 16 (presenting juror questionnaire).

includes people who indicate their race as "Native Hawaiian," "Guamanian or Chamorro," "Samoan," and "Other Pacific Islander."

- **Other**: A person having origins other than those listed above.
- **Multi-Racial**: A person having parents of different races.

As the Census Bureau does,[70] New Jersey permitted people to mark more than one category indicating in the instructions to "check all applicable categories." This practice recognizes that people's identities may not fit neatly into one box. Recent Census analyses reveal that the number of people who consider themselves multiracial has burgeoned in recent years. Specifically in 2010, the proportion selecting more than one category on the Census form was 2.9%.[71] By 2020, that same figure was 10.2%, more than a three-fold increase.[72] In New Jersey, with a study done in 2018, the juror profile was closer to the first estimate: 3.3% of people either selected "multiracial" for their identity or checked off more than one answer for the race question.[73]

For the multiracial group, New Jersey's approach permitted people to validly describe themselves, not constraining their responses to a single category; further, people could simply self-report as "multiracial." However, for a study of jury representativeness, New Jersey's approach, particularly permitting people to mark only the broad category of "multiracial," presented a key complexity because jury representativeness jurisprudence centers on the representation of "distinctive" or "cognizable" groups. We begin by reviewing the meaning of these terms.

A. The Meaning of "Distinctive" or "Cognizable" Under the Law

The first prong of the *Duren* standard, which governs fair-cross-section cases, asks whether the allegedly underrepresented community constitutes a "distinctive" group.[74] The Supreme Court has given ambiguous guidance about what makes a group "distinctive."[75] In *Lockhart v. McCree*, the Court

70. 2020 CENSUS FORM, *supra* note 68.

71. Nicholas Jones et al., *2020 Census Illuminates Racial and Ethnic Composition of the Country*, U.S. CENSUS BUREAU (August 12, 2021), https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html [https://perma.cc/4NJM-6V2V].

72. *Id.*

73. *See infra* note 112 and accompanying text (describing who is included in the multiracial category).

74. Duren v. Missouri, 439 U.S. 357, 364 (1979).

75. *See* Lockhart v. McCree, 476 U.S. 162, 175 (1986) ("We have never attempted to precisely define the term 'distinctive group,' and we do not undertake to do so today.").

indicated that distinctiveness must be assessed in light of the purposes of a law.[76] For fair-cross-section claims under the Sixth Amendment, those purposes are:

> (1) "guard[ing] against the exercise of arbitrary power" and ensuring that the "common sense judgment of the community" will act as "a hedge against the overzealous or mistaken prosecutor;" (2) preserving "public confidence in the fairness of the criminal justice system;" and (3) implementing our belief that "sharing in the administration of justice is a phase of civic responsibility."[77]

In *Lockhart*, the Court held that those who were "non-death-qualified," and therefore excused from capital cases, did not constitute a distinctive group.[78] By way of contrast, the Court affirmed that categories such as race, gender, and ethnicity have been recognized as "distinctive" because exclusions would undermine the above aims. Exclusions based on racial, ethnic, or gender categories would be unrelated to a juror's ability to serve, and therefore the composition of juries could be "arbitrarily skewed in such a way as to deny criminal defendants the benefit of the common-sense judgment of the community."[79] Exclusions would also appear unfair, undermining public confidence, because they would be based on an "immutable characteristic."[80] Finally, exclusions would improperly deprive people in "often historically disadvantaged groups of their right as citizens to serve on juries in criminal cases."[81]

In this analysis, the notion of an "immutable characteristic" offers one type of boundary to the category "distinctive." By contrasting race, ethnicity, and gender with attitudes about the death penalty, Justice Rehnquist hints that attitudes could change.[82] Race and gender are treated as a constant in someone's life.[83] Further, the Court's passing reference to being "historically

76. *Id.*

77. *Id.* at 174–75 (quoting Taylor v. Louisiana, 419 U.S. 522, 530–31 (1975)).

78. *Id.* at 173, 177.

79. *Id.* at 175.

80. *Id.*

81. *Id.*

82. *See id.* at 176 (noting that people dismissed from capital cases because they are morally opposed to capital punishment "are singled out for exclusion in capital cases on the basis of an attribute that is within the individual's control").

83. Conceivably "immutability" may be challenged even in some instances for race or ethnicity since "passing" as a member of a different race or ethnicity has a long and fraught history in the United States, at least for those whose phenotype permits them to pass. *See, e.g.*, Randall Kennedy, *Racial Passing*, 62 OHIO ST. L.J. 1145. 1147–1156 (2001) (providing historical examples). Further, thirty-five years after *Lockhart v. McCree*, it is far more common to discuss sex and gender as categories that may be changed and for gender identity to viewed as fluid. *See, e.g.*, Jessica A. Clarke, *They, Them, and Theirs*, 132 HARV L. REV. 894, 904–910 (2019) (describing types of nonbinary identities); Lisa M. Diamond, *Gender Fluidity and Nonbinary Gender Identities in Children and Adolescents*, 14 CHILD DEV. PERSPS.

disadvantaged"[84] suggests that this concept has a role in making a group distinctive.[85] However, there are many aspects of the Court's description of distinctiveness that are open to interpretation. It remains unclear which types of exclusions generate arbitrary skew in jury composition and undermine "common-sense judgment of the community." The Court also fails to address which exclusions might seem unfair to the public and how to concretely operationalize "historical disadvantage." Although these seem like issues that should be determined, in part, by empirical and historical support, the Court opinion in *Lockhart v. McCree* did not greet empirical data with metaphorical open arms.[86]

Other Supreme Court decisions offer even less concrete language. In *Taylor*, the Court noted that the underrepresentation of women meant that the defendant's jury panel lacked "a flavor, a distinct quality,"[87] a concept Justice Rehnquist derided as "transcendental" in his dissent in *Duren*.[88] In an equal protection case, the Court held that "substantial and identifiable"[89] groups could not be excluded from jury service. "Substantial" suggests that distinctiveness may be related to a group's proportional size in the community, although precise proportions have not been developed in jury representation cases based on the fair-cross-section rule.[90] This is noteworthy because groups that encompass a small proportion of the community are structurally disadvantaged in jury representation analyses not because they fail to be "distinctive," but because it is difficult to show that omission or underrepresentation in jury pools is "not fair and reasonable," the second prong of *Duren*.[91] When a community is, say, less than about three percent of an area, even samples of 1,100 people—which is far larger than a typical

110, 111 (2020) ("[Y]ounger individuals are more likely than youth of previous time periods to adopt nonbinary rather than binary gender identities.").

84. *Lockhart*, 476 U.S. at 175.

85. It bears mentioning that excluding people on the basis of an attitude (not being death-qualified) will also disproportionately disqualify people who would be viewed as "historically disadvantaged," particularly Black people, as Justice Marshall noted in his dissent. *Id.* at 201 (Marshall, J., dissenting) ("[T]he evidence suggests that death qualification will disproportionately affect the representation of blacks on capital juries.").

86. *See, e.g.*, Samuel R. Gross & Phoebe C. Ellsworth, *Social Science and the Evolving Standards of Death Penalty Law*, *in* BEYOND COMMON SENSE: PSYCHOLOGICAL SCIENCE IN THE COURTROOM 253 (E. Borgida & S. T. Fiske eds., 2008) ("[T]he Court was faced with a substantial, consistent, and highly persuasive body of research that pointed to a conclusion opposite of the one the majority wanted to reach.").

87. Taylor v. Louisiana, 419 U.S. 522, 532 (1975).

88. Duren v. Missouri, 439 U.S. 357, 371 n.* (1979).

89. Peters v. Kiff, 407 U.S. 493, 503 (1972).

90. Berghuis v. Smith, 559 U.S. 314, 330 n.4 (2010) (explaining that the Court need not reach the question of whether to require proof that the disparity in representation must exceed a ten percent absolute disparity).

91. *Duren*, 439 U.S. at 364.

jury pool for a single case[92]—might reliably contain few or no members of that group.[93]

With little guidance from the U.S. Supreme Court, lower courts have developed the concepts of what makes a group distinctive (or cognizable) through specific cases. In *United States v. Test*, the Tenth Circuit offered three criteria for what makes a group distinctive: "(1) the presence of some quality or attribute which 'defines and limits' the group; (2) a cohesiveness of 'attitudes or ideas or experience' which distinguishes the group from the general social milieu; and (3) a 'community of interest' which may not be represented by other segments of society."[94] In *Test*, Chicanos and Blacks were deemed distinctive groups,[95] whereas, the Tenth Circuit applied the above criteria and determined that people under forty years old were not distinctive.[96]

Likewise, a California case resembles the criteria from *Test* but makes the attitudinal link to service on a jury particularly clear. *Rubio v. Superior Court* offers two criteria for a group to be "'cognizable' for purposes of the representative cross-section rule."[97] The first is what is known as the "common thread" approach.[98] A common thread is a shared perspective a person gains as a result being a member of a group, a characteristic that "impart[s] to its possessors a common social or psychological outlook on human events."[99] Second,

> [t]he party seeking to prove a violation of the representative cross-section rule must also show that no other members of the community are capable of adequately representing the perspective of the group assertedly excluded. This is so because the goal of the cross-section rule is to enhance the likelihood that the jury will be representative of significant community attitudes, not of groups per se.[100]

92. Mary R. Rose, Raul S. Casarez & Carmen M. Gutierrez, *Jury Pool Underrepresentation in the Modern Era: Evidence from Federal Courts*, 15 J. EMPIRICAL LEGAL STUD. 378, 389 (2018) (finding that clerks report a typical jury pool size to be in the range of forty to sixty depending on the case type and jurisdiction); ROSE, *supra* note 5, at 89 tbl.VIII.1 (depicting panel sizes in New Jersey).

93. The margin of error for a sample of approximately 1,100 people is plus or minus three percentage points. *See, e.g.*, DEP'T OF STAT., PENN STATE EBERLY COLL. OF SCI., *Relationship between Sample Size and Margin of Error*, *in* STAT 100: STATISTICAL CONCEPTS AND REASONING 2.3 tbl.2.1, https://online.stat.psu.edu/stat100/lesson/2/2.3 [https://perma.cc/563J-SXVE].

94. United States v. Test, 550 F.2d 577, 591 (10th Cir. 1976).

95. *Id.* at 585–86 (discussing precedential opinions for recognizing these groups as distinctive).

96. *Id.* at 591. As the opinion notes, other issues with age include the unclear natural demarcations that would create "groups." For example, it is unclear whether the correct group would be ages 21–29, 30–39, or 21–39. *Id.* at 591–93.

97. Rubio v. Super. Ct. of San Joaquin Cnty., 24 Cal. 3d 93, 98 (1979).

98. *See* Ford v. Seabold, 841 F.2d 677, 681–82 (6th Cir. 1988) (citing Barber v. Ponte, 772 F.2d 982 (1st Cir. 1985) (en banc); Willis v. Zant, 720 F.2d 1212 (11th Cir. 1983)).

99. *Rubio*, 24 Cal. 3d at 98.

100. *Id.*

In *Rubio*, the state supreme court held that those with a felony conviction and resident aliens met the first criteria of the test since their conviction for a felony, and subsequent deprivation of liberty, gave them a common social or psychological outlook, and resident aliens had the experience of being excluded from political processes.[101] However, both groups failed the second part, because, the court reasoned, someone jailed for a misdemeanor (who is jury-eligible) could adequately represent the felon-group on juries, and naturalized citizens share the experiences of resident aliens.[102] In other words, a group excluded from juries is not distinctive when their attitudes could adequately be represented by others who have had similar life experiences and therefore viewpoints.

In sum, this sampling of opinions of what constitutes a distinctive (or cognizable) group indicates that distinctiveness is not always easy to determine. Groups based on race, gender, and ethnicity seemingly always qualify as "distinctive," both because they largely fit with the notion of immutability and because non-White minority groups and women have each faced a common history of discrimination that would, presumably, shape their perspectives. Under rules outlined in *Test* and in *Rubio*, these groups cannot be represented by other segments of the community, since each group's history and experience of discrimination will differ. Groups that make up small proportions of a community may be distinctive, but these groups may face difficulties showing that their underrepresentation is not fair and reasonable.

B. The Distinctiveness of Multiracial Persons

In the questionnaire jurors responded to for the New Jersey study, people were "multiracial" if they selected this option for the race question by itself, or if individuals selected more than one racial category as their racial identity. However, based on criteria reviewed in the last section, it is not clear that under current caselaw courts would regard the category "multiracial" as cognizable or distinctive. If multiracial people were regarded as a distinctive group, their numbers would make it exceedingly difficult to prove a fair-cross-section violation.

101. *Id.* at 98–99.

102. *Id.* at 99–100. For lengthy critique of these conclusions, see the dissent by Justice Torbiner. *Id.* at 105–17 (Torbiner, J., dissenting). Note that California removed its jury felon-exclusion policy in 2019. James Binnall, *Truly Diverse Juries Must Include Citizens with Prior Criminal Convictions*, UNIV. OF CAL. PRESS: UC PRESS BLOG (NOV. 17, 2021), https://www.ucpress.edu/blog/57671/truly-diverse-juries-must-include-citizens-with-prior-criminal-convictions/ [https://perma.cc/WK98-BE9V].

To be sure, there are reasons to consider the broader group of multiracial people to be distinctive. The Census Bureau recognizes multiracial persons as a racial group,[103] which means that this is a race-based group with an immutable identity. Further, this group has faced discrimination. Multiracial persons have the shared experience of having others struggle to identify them and potentially even challenging their identity.[104] Scholars also note a shared history grounded in America's suspicion of—and historical laws against— the concept of "race-mixing."[105] At minimum, they share the common thread of identifying as having more than one race in their ancestry.[106]

At the same time, if distinctiveness requires that group members share a common perspective, a "cohesiveness of 'attitudes or ideas or experience' which distinguishes the group from the general social milieu,"[107] then the multiracial group is not distinctive because each form of being multiracial is not necessarily similar to another. As Mitchell writes in supporting a "Class of One" approach for Equal Protection claims, "creating a single classification for all multiracial people ignores the vast differences among multiracial individuals. One should hardly argue that the experiences of a Black-White, mixed-race individual would mirror the experiences of an Asian-White, mixed-race individual."[108]

Thus, although a common thread potentially exists at the aggregate level of the label "multiracial," it is easy to imagine a prosecutor or judge opposing a fair-cross-section claim by indicating that such persons cannot reasonably be considered as a single group. An alternative is to make the sub-type of "multiracial" identity its own cognizable group; that is, people

103. This is evident in the Census Bureau permitting people to select more than one race on their census form. *See supra* note 68. Further, the Census Bureau produces reports on multiracial persons. Jones et al., *supra* note 71 (describing changes in the prevalence of this group between 2010 and 2020).

104. Several scholars have considered how to conceptualize multiracial persons for questions of discrimination claims, and each mentions the commonality of a "What are you?" question directed at multiracial persons. *See, e.g.*, Nancy Leong, *Judicial Erasure of Mixed-Race Discrimination*, 59 AM. U. L. REV. 469, 477 (2010) (citing PEARL FUYO GASKINS, WHAT ARE YOU?: VOICES OF MIXED-RACE YOUNG PEOPLE 5 (1999)); Desiree D. Mitchell, *A Class of One: Multiracial Individuals Under Equal Protection*, 88 U. CHI. L. REV. 237, 256 (2021) ("Perhaps the question most often heard by multiracial individuals is, 'What are you?'"); Destiny Peery, *(Re)defining Race: Addressing the Consequences of the Law's Failure to Define Race*, 38 CARDOZO L. REV. 1817, 1819 (2017) (describing a person "often asked, 'What are you?'").

105. Leong, *supra* note 104, at 483–91; Mitchell, *supra* note 104, at 242–49; Peery, *supra* note 104, at 1835–41.

106. Mitchell, *supra* note 104, at 252 ("[W]hile there is no racial 'cohesion' among multiracial individuals as a group, they are distinguishable [for Equal Protection purposes] in that they identify as more than one race.").

107. United States v. Test, 550 F.2d 577, 591 (10th Cir. 1976).

108. Mitchell, *supra* note 104, at 261.

who are Black and White, or White and Asian would be the underrepresented group. However, although the multiracial category has grown considerably in a decade, the total population of multiracial persons in the United States is around 10% of the population.[109] This means that any subgroup is a fraction of that population; for example, Black Americans were 12.4% of the population in 2020, and when the multiracial group of Blacks who are also at least one other race are added in, that total is 14.2% of the population, meaning that latter multiracial sub-group is just 1.8% of the population.[110] Thus, a focus on a particular subtype of multiracial persons will likely run into the problem of underrepresentation being attributed to the normal process of variations arising when taking samples (a sampling error).[111]

C. Determining Categories for Analysis in the New Jersey Data

To conduct empirical research, debates about how to treat the multiracial category could not remain theoretical; the issue had to be settled in order to move forward with the analyses. One challenge stemmed from people opting to self-identify only as "multiracial," which, as noted, is unlikely to be deemed a distinctive group on its own for purposes of fair-cross-section analysis. Respondents appeared drawn to the multiracial option: more people marked this undifferentiated group than reported their more specific subtype. Specifically, of the 12,070 people who reported an answer to the race question, 278 (2.3%) chose either "multiracial" alone or multiracial in combination with just one other race (American Indian, Asian, Black, White or Other-Race), which rendered their particular multiracial subgroup impossible to identify.[112] Only another 122 people (1.0%) selected two or more specific race combinations. Most common among this latter group, nineteen people selected Black in combination with White or White and Other, and thirteen selected Asian in combination with White or White and Native Hawaiian/Pacific Islander.

The above pattern elucidates the challenges. First, maximally, the multiracial group as a whole constituted 3.3% of the sample (400 people), and it is smaller if one decides to treat any given combination (e.g., the group of people who selected White and Black for race) as a distinctive group. Until the current population ages, such that the group of people who identify as multiracial age into jury-eligibility— which is eighteen to seventy-five in

109. *See* Jones et al., *supra* note 71, fig.2.

110. Jones et al., *supra* note 71.

111. *See supra* note 93 and accompanying text.

112. Specifically, 217 people selected multiracial alone and 61 chose multiracial with one other group.

New Jersey[113] — the presence of multiracial persons in jury pools will continue to be small, making it easy to dismiss underrepresentation as mere sampling error.

Further, the numbers become even more miniscule as the outcome data are examined in more detail, one of New Jersey's aims with the study. Table 2 provides outcome information on the multiracial group and for those who selected just one racial category. Of the 400 people who identified as multiracial, 208 made it to jury selection questioning of any kind, and 64 of these persons were "not used"— neither seated, nor struck by the judge or one of the parties following voir dire. Just twenty-one experienced a peremptory challenge from either the prosecutor, plaintiff, or defense, and twenty-five served as seated jurors. These numbers are too small to be amenable to reliable statistical testing, although interestingly, in the New Jersey data, the prima facie pattern in Table 2 suggests that statistical testing would reveal very little. There is a remarkable similarity in the distribution of outcomes between multiracial and single-race persons, as the proportions across categories for both groups are close to identical.

Table 2: Outcomes for Multiracial and Single-Race Persons (12,070 people)

Outcome	Multiracial	Single-Race
No voir dire	192 (48.0%)	5,643 (48.4%)
Voir dire, but not used	64 (16.0%)	2,042 (17.5%)
Challenge for Cause	98 (24.5%)	2,776 (23.8%)
Peremptory: Prosecutor/Plaintiff	11 (2.8%)	236 (2.0%)
Peremptory: Defendant	10 (2.5%)	298 (2.6%)
Seated as a Juror	25 (6.3%)	675 (5.9%)
Total	400	11,670

113. N.J. STAT. ANN. § 2B:20-1 (West 2022).

Note: An additional 1,407 people did not provide information on the race question.

Although small numbers may undermine the ability to show a fair-cross-section violation, New Jersey officials were also interested in patterns, if any, in use of the peremptory—an area of law that governs even small numbers of people affected. The need to understand how peremptory challenges are exercised raises still another complication about how to measure and categorize multiracial persons. Significantly, use of peremptory challenges depends not only on how a juror might self-identify but also on how someone else (an attorney) sees that individual and encodes his or her race. These two concepts are not necessarily the same.

In studies of how people decode other people's racial identity, scholars have noted that it is common for observers to treat multiracial persons according to the principle of hypodescent. Under the hypodescent principle, people who identify with more than one race are (a) categorized by others into a single group, and (b) that resulting categorization is likely to skew toward the racial group that enjoys less status in society.[114] As examples, someone who is both White and Asian is likely to be categorized in single-race terms as Asian rather than White, whereas someone who is both Black and Asian is likely to be perceived by others as Black.[115] Scholars have shown that this monoracial categorization is particularly likely when people are making quick judgments of others,[116] although the principle has been consciously embedded in the legal regimes and racist social practices of our nation's past.[117] Anecdotally, it is clear that several famous multiracial people are often reduced to a single racial identity, as when former President

114. Peery, *supra* note 104, at 1834 ("[T]he race of a multiracial child is assigned to that of the socially-subordinate parent.").

115. On status measures such as income, Asian-Americans have higher status in the United States than do Black Americans. According to Census Bureau data, people who self-report that they are Asian have the highest median annual income ($94,903), far in excess of the overall median ($67,521) and more than twice the figure for Black Americans ($45,870). EMILY A. SHRIDER ET AL., U.S. CENSUS BUREAU, INCOME AND POVERTY IN THE UNITED STATES: 2020, at 5 fig.2 (2021), https://www.census.gov/content/dam/Census/library/publications/2021/demo/p60-273.pdf [https://perma.cc/X7UA-8HLE].

116. Destiny Peery & Galen V. Bodenhausen, *Black + White = Black: Hypodescent in Reflexive Categorization of Racially Ambiguous Faces*, 19 PSYCH. SCI. 973 (2008) (reporting results of a study showing people were more likely to follow the principle when instructed to make fast judgments of faces); *see also* Arnold K. Ho et al., *Evidence for Hypodescent and Racial Hierarchy in the Categorization and Perception of Biracial Individuals*, 100 J. PERSONALITY & SOC. PSYCH. 492 (2011) (reporting five studies the confirm the principle of hypodescent for both people described as Black and White and as White and Asian).

117. Peery, *supra* note 104, at 1834 (noting that the hypodescent principle in social perception "resemble[s] the old legal rules that made racial determinations dependent on relatively small traces of non-White blood and tied them to the status of an individual's parents").

Barack Obama is labeled as the "first African American to hold the office,"[118] and Tiger Woods as the first African American golfer to reach a particular milestone.[119]

The fact that other people likely perceive multiracial people as members of a monoracial category—and likely associate them with the group most likely to be subject to discrimination—means that separating multiracial people from the categories that make up their identities could mask patterns in how others treat them. The additional generic multiracial category on the New Jersey questionnaire thus interfered with the ability to test the idea that attorneys may have treated some multiracial jurors as monoracial because prospective jurors who chose "multiracial" could not otherwise be identified by the groups that make up their identity. The principle of hypodescent could be tested among the small subset of multiracial persons who listed the specific racial categories that made up their identities, but this would require categorizing them as something other than multiracial.

For all of the reasons just reviewed—because multiracial persons as a whole group, or as subtypes, are presently small in number (even if their numbers are likely to grow in the coming years); because it is not certain that the multiracial group will be treated as "distinctive" when considered as a whole; and because the principle of hypodescent strongly suggests that multiracial people are likely to be treated as monoracial—we approached the New Jersey data with the aim of reducing the multiracial category by placing people into the monoracial groups where we could.

Because scholars have carefully argued that redefining multiracial persons into monoracial groups amounts to "erasure,"[120] it is imperative to note that, as a data analytic strategy in this context, the re-coding is temporary. Categorizing multiracial individuals into monoracial categories for some analyses did not mean they "lost" the possibility of also being considered as multiracial for purposes of other analyses, such as those presented in Table 2. But for the main questions of the study—Who appears at court? Who goes to voir dire? What happens to them after voir dire?—we aimed to test the hypodescent principle, prioritizing African American

118. *Barack Obama*, WHITE HOUSE HIST. ASS'N, https://www.whitehousehistory.org/bios/barack-obama [https://perma.cc/5AQL-55QC].

119. Bob Denney, *Timeline of African-American Achievements in Golf*, PGA OF AM., https://www.pga.com/story/timeline-of-african-american-achievements-in-golf [https://perma.cc/5NRL-GWBB].

120. *See* Leong, *supra* note 104 at 521–539 (critiquing efforts to treat multiracial persons in single-race terms and reviewing academic studies and court opinions that fail to distinguish between discrimination against multiracial persons as multiracial and discrimination against one or the groups that make up a person's multiracial identity); Mitchell, *supra* note 104, at 240 ("This Comment . . . [asserts] that courts should recognize multiracial plaintiffs as just that— multiracial.").

identity, then Asian identity, and then White identity. We prioritized identifying African Americans because this group has a particularly fraught history with courts and law[121] and because New Jersey AOC indicated to us that they were also interested in how peremptory challenges may be used on this group.

The individuals who selected the "multiracial" category alone could not be recoded into any other category, except for being grouped with the "Some Other Race" category because their numbers were too small to consider by themselves. For the other multiracial persons who selected just two categories of identity, we followed these recoding rules: if someone selected "Black" as a racial identity, alone or in combination with one other race, they were coded into the "Black" category. Next, those who self-identified as "Asian" were coded into the "Asian" category.[122] Finally, those who were selected "White" were placed into the "White" category (unless they had also selected either "Black" or "Asian").[123] All remaining persons were coded as "Some Other Race," including the people who marked more than two categories. Because the numbers for each multiracial subtype were small, this recoding did not change the composition of the monoracial identity categories very much. Black respondents went from 12.6% to 12.9% after recoding; Asian respondents went from 9.0% to 9.2%, and the proportion of White respondents also increased by less than one percent (0.55%). Because of the inclusion of people who marked only multiracial, the "Other Race" group increased the most from 5.4% to 7.6%.

121. *See, e.g.*, Monica C. Bell, *Police Reform and the Dismantling of Legal Estrangement*, 126 YALE L.J. 2054 (2017) (describing theoretical and empirical bases for why Black Americans believe that the legal system and legal actors fail them); David S. Kirk & Andrew V. Papachristos, *Cultural Mechanisms and the Persistence of Neighborhood Violence*, 116 AM. J. SOC. 1190 (2011) (providing data that links neighborhood homicide rates to cynicism about the legitimacy of law and trust in police); Bill McCarthy, et al., *Neighborhood Climates of Legal Cynicism and Complaints about Abuse of Police Power*, 58 CRIMINOLOGY 510 (2020) (reviewing studies of cynicism about the law to attitudes toward police).

122. This excludes any individuals who were both Black and Asian, who were coded into the "Black" category.

123. Some Latinos also identify as White. We coded race two ways and checked if results changed substantially, with White/Hispanics coded as "White" and with this same group coded as "Other," another common identity selected by Latinos. Steven Hitlin et al., *Measuring Latinos: Racial vs. Ethnic Classification and Self-Understandings*, 86 SOC. FORCES, 587, 602-03 (2007) (describing large proportions of Latinos choosing "Other" instead of "White" when given the option to do so). The Biden administration recently proposed moving "Hispanic" into a category within the Census Bureau's question about race. *See, e.g.*, Mike Schneider, *Biden Administration Proposes New Race and Ethnicity Categories for Federal Surveys*, PBS (Jan. 26, 2023, 4:36 PM), https://www.pbs.org/newshour/politics/biden-administration-proposes-new-race-and-ethnicity-categories-for-federal-surveys [https://perma.cc/D6XC-PQDJ]. Treating Latinos as "Some Other Race" in for the New Jersey data did not change results substantially, although the White group was 65.6% of the sample instead of 70.3%.

D. The Need to Measure Race Without a Separate 'Multiracial' Category

Our experience taught us that, as thoughtful as New Jersey had been in using the measures recommended by the National Center for State Courts, including "multiracial" as a separate category people could select had the ironic effect of making it harder to understand multiracial persons in these jury pools. Notably, Census forms do not offer a separate, non-descript multiracial category that people can choose.[124] Instead, the single-race subgroups are listed, and people can select more than one category in order to describe themselves.[125] As Peery has argued, different areas of law may call for viewing multiracial persons through different lenses and approaches.[126] With the Census approach, those who select more than one group for their race can be grouped together as part of the aggregate group "multiracial" when the question of interest concerns that broader group. Yet, by reporting on the specifics of their identities, people can also be recoded in other ways if analysts wish to, for example, see if attorneys treat multiracial persons who have some Black ancestry as "Black." New Jersey has opted to take the Census approach as they move forward with the goal of routinely measuring the demographic profile of people who appear at court.[127]

More radically, courts might take seriously the conceptual distinction between race as measured by self-identification and race as perceived by others, opting to measure both. Consistent with the Census Bureau approach, people could report on how they self-identify and pick the category or categories that best describe them. Additionally, because interactions at the courthouse could be shaped by other people's assumptions about jurors' identities, courts wishing to fully understand the effects of race or ethnicity in the process might to consider asking an additional question about which racial category or categories the juror thinks others identify them as.[128]

124. *See* 2010 CENSUS FORM, *supra* note 68; 2020 CENSUS FORM, *supra* note 68.

125. *See* 2010 CENSUS FORM, *supra* note 68; 2020 CENSUS FORM, *supra* note 68.

126. *E.g.*, Peery, *supra* note 104, at 1875 ("In the context of Title VII discrimination law, for example, race may be best defined in terms of how a plaintiff is perceived based on their appearance and other racial cues, as perception determines whether someone is likely to be subjected to discrimination or not.").

127. GLENN A. GRANT, N.J. CTS., COLLECTION OF VOLUNTARY JUROR DEMOGRAPHIC INFORMATION–INITIAL IMPLEMENTATION IN BERGEN, CAMDEN, AND MIDDLESEX COUNTIES, https://www.njcourts.gov/sites/default/files/notices/2022/07/n220715a.pdf [https://perma.cc/B5T2-5GUJ].

128. *See, e.g.*, Nancy López et al., *What's Your "Street Race"? Leveraging Multidimensional Measures of Race and Intersectionality for Examining Physical and Mental Health Status Among Latinxs*, 4 SOCIO. RACE & ETHNICITY 49, 53 (2018) (giving examples of different ways to ask how other people would classify one's own race).

IV. RECOGNIZING NONBINARY GENDER IDENTITY

New Jersey law recognizes nonbinary as a gender category.[129] As a result, state forms that inquire about gender must include nonbinary gender identity as one option (i.e., a category for people who feel they are neither male nor female), and a nonbinary category will appear on the new demographic questionnaires New Jersey will implement.[130]

Although this approach measures gender identity in a more inclusive way, the approach raises an additional complication. Given the current prevalence of nonbinary persons in the population, there are few ways of recoding the data in order to make use of data from this group.

At present, particularly among the current adult (i.e., jury-eligible) population, identifying as nonbinary is rare. Studies suggest that only about 1% of the population is nonbinary.[131] This was evident in the New Jersey data. Despite the presence of just two gender categories, some people in the New Jersey study selected both male and female as their gender. Assuming none of these mistakenly or belligerently marked both categories, we can treat this group as those who wished to be considered nonbinary. Of the over 12,000 people who filled out the questionnaire and answered the question, nine (0.0007%) would be considered nonbinary by having marked both male and female.[132] On its own, that group is too small to analyze. When so few observations exist, and when they cannot be recoded into another group (as was done for multiracial persons), the only other option is to remove these individuals from the dataset for any analysis using the gender variable.

There are not many fixes for this problem. If the true prevalence had been closer to 1%, there would have been just 120 people who were nonbinary. In the New Jersey data, about 60% went to voir dire, so this 120 people would have decreased to roughly 72 with a jury selection outcome, a number that would not be large enough to analyze independently through statistical tests. Apart from waiting for there to be more nonbinary persons in the population, (and surveys suggest growth in this population among

129. N.J. STAT. ANN. § 26:8-40.12 (West 2022) (effective Feb. 1, 2019) (people may change their birth certificate gender to "undesignated/non-binary").

130. *See* GRANT, *supra* note 127 (listing "Non-Binary or Undesignated" under gender categories).

131. Anna Brown, *About 5% of Young Adults in the U.S. Say Their Gender Is Different from Their Sex Assigned at Birth*, PEW RSCH. CTR. (June 7, 2022), https://www.pewresearch.org/fact-tank/2022/06/07/about-5-of-young-adults-in-the-u-s-say-their-gender-is-different-from-their-sex-assigned-at-birth/ [https://perma.cc/V43A-NCEC] (indicating that the majority of non-binary adults are under age thirty).

132. To be sure, this is a minimum estimate because some people who did answer the gender question may have forced themselves into one of the binary categories because they did not see their nonbinary status listed.

younger persons[133]) questionnaire designers might consider a two-question approach to measuring gender: one question would comprehensively list all relevant gender categories, and a second would return to the binary and ask people to name the category that "best" describes their gender. Alternatively, as discussed above with respect to race, people might be asked to think about how others perceive them.[134] These approaches recognize the varieties of ways that people express their gender identity but also ask respondents to consider whether they feel more akin to a male or to a female or believe that others typically perceive them as male or female. Whatever the approach, goal would be to preserve the possibility of keeping people's data in analyses.

At the same time, we acknowledge that these approaches revert to the gender binary and erase the existence of nonbinary status. The problem of exclusion— "throwing out" data from small groups—is addressed only by forcing people back into the binary, an approach that would likely engender resentment among nonbinary respondents and others who seek to have legal institutions reject the gender binary.[135] Conceivably, some may see recognition and inclusion of the nonbinary category on official forms as more important than concerns about what happens to people's data when analysts start examining a dataset to look for patterns and statistical associations.

V. CONCLUDING THOUGHTS ON DATA AND MEASUREMENT CHALLENGES

We used this symposium contribution to consider quite practical issues that arise when courts take an empirical approach to assessing jury participation. First, when studies use court-collected databases, the resulting data, like all data, will need to be cleaned. Our experience suggests that this process can be quite labor-intensive, require specific expertise, and it can depend upon good access to personnel to answer clarifying questions. Some

133. Brown, *supra* note 131.

134. *See supra* note 128 and accompanying text.

135. *See, e.g.*, Clarke, *supra* note 83, at 919 ("This argument [for recognizing a third nonbinary category for gender] stakes a claim to a type of liberty, but not in the thin sense of freedom of choice. It asserts that nonbinary people should not be forced to adopt a binary sex category that is a lie."). Scholars like Clarke do not view the problem of small numbers as reason for legal doctrines—and presumably government forms—to avoid recognizing nonbinary persons as nonbinary. *Id.* at 952–953 ("The existence of complicated racial identities does not preclude the enforcement of legal doctrines that depend on statistical underrepresentation of minority groups. Neither should the existence of complicated gender identities be a barrier to collection of information on sex or gender identity.").

have called for courts to share data with outside analysts.[136] However, staff may not feel they have time to clarify issues, especially when there is no collaborative relationship like the one that existed between the authors and New Jersey. Court staff are also ill-prepared to provide detailed written information to explain the properties of the data—called "metadata"[137]—that scholars expect when they share data with each other through repositories, such as the Interuniversity Consortium on Political and Social Research (ICPSR).[138] Creating metadata documentation is a labor-intensive process,[139] and also requires a great deal of knowledge about technical aspects of data.[140] Without this information, an analyst has no way of knowing what measures are in the data and what each measure means. Not only does that make the data unusable, but the analyst also has no way of cleaning the data to ensure that the dataset does not contain serious errors. In other words, without sufficient information, court data may be useless. This reality undermines the value of having greater transparency in court practice.

Our experience likewise points to complexities in measuring identities that capture the ways that people self-identify their race or gender. Rich and inclusive measures may clash with the way that the law constructs distinctiveness and being cognizable. Further, any group faces barriers to being recognized as an underrepresented part of a "fair cross section" if that group constitutes only a small portion of the population. As a measurement issue, those designing future questionnaire items should consider ways to maximize flexibility in measurement. For example, while specific categories can be combined, broad categories ("multiracial") cannot be decomposed to be more specific and are therefore potentially less useful. By using more specific categories, data analysts can thoughtfully consider whether it is better to combine one subgroup that has small numbers with data from other groups in order to avoid removing small groups entirely. Questionnaire

136. *See* Crespo, *supra* note 3, at 2106 (urging courts to make data on judicial and court practices more transparent, even if courts have not formed partnerships with people who have social science expertise).

137. *See* ICPSR, GUIDE TO SOCIAL SCIENCE DATA PREPARATION AND ARCHIVING: BEST PRACTICE THROUGHOUT THE DATA LIFE CYCLE 34 (6th ed. n.d.), https://www.icpsr.umich.edu/files/deposit/dataprep.pdf [https://perma.cc/9EGW-SSTZ] [hereinafter DATA PREPARATION GUIDE].

138. ICPSR is among the oldest data repositories, existing for sixty years. *ICPSR: The Founding and Early Years*, ICPSR, https://www.icpsr.umich.edu/web/pages/about/history/early-years.html [https://perma.cc/N5QL-3NZR].

139. *See generally* DATA PREPARATION GUIDE, *supra* note 137 (providing a detailed, fifty-eight-page guide to describe the process of archiving data to make it available to others).

140. *See, e.g.*, *id.* at 33–36 (describing specific best practices for a metadata file, which includes information about sampling, weighting, handling of missing values, and technical information to include for each variable).

designers might also consider ways to measure not only self-identity, but also people's experiences with others' perceptions of their identities.

We describe these issues to bring attention to those designing questionnaires, specifically those who are inexperienced with such considerations. The recommendations offered are not the only possible ways to measure race,[141] and there are other design issues to consider.[142] In our estimation, greater collaboration between courts and social scientists with the appropriate training allows for courts to design their materials more carefully and opens opportunities for the public to have greater insight into how courts work.

Courts that aim to understand levels of diversity in the pools of people reporting for jury duty, and among the people who actually serve, deserve credit because even an imperfect effort is preferable to the current state of affairs in many areas where no data exists at all. To assess its own system, New Jersey undertook a multicounty data collection effort by introducing a juror demographics questionnaire into a system where one did not exist. Because we were tasked with working with the resulting data and assisting New Jersey in understanding its practices, we are able to offer some lessons about the challenge of cleaning data and issues that arose when attempting to accurately measure the identities of all people. However, these lessons in no way detract from our belief that any court system seeking to understand, and perhaps even critically analyze its practices, should be lauded.

141. In Professor Rose's report, she suggested that people might be encouraged to select the single category that "best describes" their racial identity. ROSE, *supra* note 5, at 103. The Census-based approach discussed in this paper is, however, a better and more flexible way to capture both multiracial identities and the components of that identity.

142. When studies ask about both Hispanic/Latino identity and race, the Hispanic question should precede the race question because when Hispanic/Latino identity follows the race question, people have an increased tendency to skip the item. *Id.* at 96–98.