Received: 2 December 2022 | Revised: 21 June 2023 | Accepted: 21 July 2023

DOI: 10.1111/1745-9125.12359

CRIMINOLOGY

ORIGINAL ARTICLE

The problem with criminal records: Discrepancies between state reports and private-sector background checks

Sarah Lageson[1] | **Robert Stewart**[2]

[1]School of Criminal Justice, Rutgers University, Newark, USA

[2]University of Maryland, College Park, USA

Correspondence
Sarah Lageson, School of Criminal Justice, Rutgers University, Newark, 123 Washington Street, Newark, NJ 07102-3094, USA.
Email: sarah.lageson@rutgers.edu

Drs. Lageson and Stewart have noted equal authorship on this article.

Replication Data and Code
Replication data and code are provided at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/BDMWJN.

The authors thank Katherine Bright, Kateryna Kaplun, Tyler Keller, Anat Kimchi, Cassandra Philippon, Juan Sandoval, Elizabeth Webster, and Ryan Whyte, who were instrumental in helping with data collection for this study. The authors would also like to thank Sarah Tahamont, Robert Apel, Valerio Baćak, and Jordan Brensinger for comments on earlier drafts.

Funding Information
Funding was provided by the U.S.

Abstract

Criminal records are routinely used by employers and other institutional decision-makers who rely on their presumed fidelity to evaluate applicants. We analyze criminal records for a sample of 101 people, comparing official state reports, two sources of private-sector background checks (one regulated and one unregulated by federal law), and qualitative interviews. Based on our analysis, private-sector background checks are laden with false-positive and false-negative errors: 60 percent and 50 percent of participants had at least one false-positive error on their regulated and unregulated background checks, and nearly all (90 percent and 92 percent of participants, respectively) had at least one false-negative error. We define specific problems with private-sector criminal records: mismatched data that create false negatives, missing case dispositions that create incomplete and misleading criminal records, and incorrect data that create false positives. Accompanying qualitative interviews show how errors in background checks limit access to social opportunities ranging from employment to education to housing and violate basic principles of fairness in the legal system.

Department of Justice, Office of Justice Programs, National Institute of Justice, under Award 2017-IJ-CX-0036.

KEYWORDS
background checks, criminal records, data, technology

1 | INTRODUCTION

In a recent case in U.S. District Court, *Henderson v. The Source for Public Data* (2021), a group of plaintiffs claimed they had been denied employment due to inaccurate criminal records reported through the course of routine background checks. According to one plaintiff, his background check report included criminal record data about a different but similarly named person with a felony conviction. Another plaintiff alleged that his report included a set of criminal convictions that had been dismissed. As a result of these inaccuracies, the plaintiffs claimed they wrongly faced 20 employment discrimination as employers were making decisions using bad data.

The *Henderson* (2021) case sheds light on broader questions about the efficacy and accuracy of criminal record data in a context in which 94 percent of employers conduct background checks (HR Research Institute, 2020), criminal records are considered public data (Bushway & Kalra, 2021; Sandoval & Lageson, 2022), and a multibillion-dollar industry collects, aggregates, and re-sells criminal record information (Lageson, 2020). Criminal background checks are ubiquitous across a broad set of institutions, including employment, housing, volunteering, dating, and education (Blumstein & Nakamura, 2009; McElhattan, 2019; Stoll & Bushway, 2008). One national study found 58.4 percent of respondents (and 71.1 percent of those who applied for a job within the past year) were asked whether they had a criminal record at some point during their last job search (Denver et al., 2018). The prevalence of inquiries implies that employers find the information contained in criminal records useful when evaluating applicants. Amidst the widespread assumption that criminal records are an effective tool for identifying risk or safety, we lack an understanding of their reliability or accuracy (Bushway et al., 2007; Bushway & Kalra, 2021).

In response, we analyze a set of criminal records for 101 people in one state and compare official state records, two sources of private-sector criminal record data, and first-person qualitative accounts. By comparing multiple sources of criminal record information for the same person—and by comparing these accounts to their lived experiences—we identify a key set of accuracy problems with private-sector criminal records. In what follows, we describe the types of criminal record information available to the public and review theoretical and empirical literature related to broader social and legal forces, as well as the organizational structures and decision-making, that contribute to data error.

2 | CRIMINAL RECORDS AND BACKGROUND CHECKS

"Criminal record" is a broad term that typically refers to a collection of information derived from police departments (who arrest and book individuals into jail), criminal courts (that process charges, adjudicate guilt or innocence, and assign a final case disposition), and corrections agencies (that incarcerate or supervise people who have been convicted). Ostensibly, criminal records reflect criminal justice processing outcomes after criminal offense events, but can take many forms. The term can be used to describe governmental records compiled into state-level reports (colloquially known as "RAP sheets"), which vary in format between jurisdictions. It can also be used to describe criminal history data drawn from government sources, which are

compiled in the private sector. In most jurisdictions, criminal records are considered public information (Sandoval & Lageson, 2022) and many criminal justice agencies have increasingly made digital criminal record data available on their websites or for purchase (Lageson, 2020). These public records are copied, scraped, and aggregated by third parties and matched to other sources of public and consumer data (SEARCH, 2001). Consequently, a person may have hundreds—if not thousands—of versions of their "criminal record," patched together from both state and private-sector sources (National Consumer Law Center [NCLC], 2019).

Criminal records can have an outsized role in people's lives, conditioning access to opportunities and resources across a wide range of areas (Miller & Stuart, 2017; Uggen & Blahnik, 2014; Uggen & Stewart, 2015). The relationship between criminal records and employment access has been studied extensively during the last two decades, including how serious and less serious records affect employer callbacks (Pager, 2003; Uggen et al., 2014), whether credentials and reentry programs mitigate criminal record effects (Bushway & Apel, 2012; Denver & DeWitt, 2022; Lindsay, 2022), and how employers view applicants with records (Kurlychek et al., 2019; Lageson et al., 2015). Criminal record effects have also been studied in other contexts, including housing, education, and social relationships (Evans, 2016; Evans & Blount-Hill, 2020; Leasure, 2019; Stewart & Uggen, 2020). These lines of research have shown that across contexts, decision-makers use criminal records from a variety of public and private sources even when they find this information difficult to understand or interpret or lack institutional mandates about how to use such information (Furst & Evans, 2017; Lageson et al., 2015).

Local and national movements have addressed the deleterious effects of having a criminal record, but even these policies depend in part on complete and accurate criminal history information in public- and private-sector settings. For instance, expungement and record-sealing policies, such as "Clean Slate," limit the reach of official criminal records after some point in time (Love & Poggenklass, 2023). "Dirty data," however, have impeded efforts to scale and automate expungement effectively (Chien, 2020). The proliferation of records into the private sector also poses difficulties. Policies only seal records under government custody; they have limited, if any, jurisdiction over copies of criminal record data already in the public sphere, online, or in private-sector databases. If a background check erroneously reports a recently sealed record, the effectiveness and fairness of these policies comes into question.

Yet, employers often require an applicant to submit to a criminal background check produced by a private-sector consumer reporting agency before rendering a final employment decision. Even in jurisdictions with "Ban the Box" or "Fair Chance" hiring policies that prohibit asking about criminal records on an initial application, employers can and do conduct background checks in later stages of the hiring process (Community Legal Services of Philadelphia, 2018; Maurer, 2018; Raphael, 2021). Although recent research has shown that employer and human resources (HR) concerns about liability are incongruent with recent court decisions regarding negligent hiring lawsuits (McElhattan, 2022a), the steadfast assumptions that criminal background checks are valuable sources of information have persisted, and the background check industry has continued to thrive.

2.1 | Regulated and Unregulated Background Checks

The growth of the private sector in supplying criminal record information has dramatically changed the landscape of background checking. The widespread availability of bulk criminal court data through purchasing, scraping, or other means has allowed private companies to offer employers, landlords, and the general public nearly unfettered access to criminal history

information, even in jurisdictions where state law has prohibited broad public access to centralized governmental criminal record repositories (Bushway et al., 2007; Jacobs, 2015). In the digital age, the growth of data brokers as middlemen has facilitated the mass disclosure of bulk criminal court data and correctional data, thereby expanding the reach and crystallizing the stigma of criminal records.

In the current technological and policy climate, a broad variety of criminal record sources are available. Approximately three categories of criminal records exist. "Official" records are produced by state or federal agencies and include a comprehensive account of an individual's criminal history, including arrests, charges, and convictions, within the respective agency's jurisdiction. These records are often compiled using fingerprints or a specific state ID. "Regulated" criminal background checks are produced by companies called "consumer reporting agencies" (CRAs) that are required to comply with federal regulations for background checking under the Fair Credit Reporting Act of 1970 (FCRA) for employment and housing determinations. "Unregulated" criminal background checks are produced by data aggregators that compile public records and publish them on the internet for a subscription fee, also known as "people search" services (Sherman, 2021). These aggregators affirmatively opt out of FCRA regulations and offer more informal reports for the public.

Industry research has indicated that, of the 93 percent of employers who report conducting a background check for job applicants, most prefer to use private providers of criminal history information (rather than official governmental sources). Specifically, only 19 percent of private employers report ever using a fingerprint check, as opposed to 89 percent of firms who regularly use state/county-level databases and 84 percent who use national databases. (Professional Background Screening Association [PBSA], 2021). The PBSA has also indicated that the perceived superiority of fingerprint background checks as the "gold standard" is a "misnomer" because these governmental reports "have significant limitations and are not recommended as a single source for a background check for employment, volunteer or tenant screening purposes" (PBSA, 2021, p. 8). Although 83 percent of firms in the PBSA's survey report quality and accuracy as a top concern, they also cite speed (58 percent) and cost (54 percent) as important in choosing a provider. Given that most state-run checks place limits on the information employers can view, are costly, and often require written consent and fingerprinting of the subject (Sandoval & Lageson, 2022), not surprisingly, most companies prefer to use the private sector to obtain criminal record information.

2.1.1 | Regulated background checks

Although the FCRA was originally designed for regulating financial information used for lending decisions, FCRA-compliant background checks also report criminal history information for employment and housing determinations (Kiviat, 2017). The FCRA regulates the processes and procedures of CRAs that produce criminal background checks (and credit checks), which must comply with its various provisions. The regulation stipulates that CRAs "shall follow reasonable procedures to assure maximum possible accuracy of the information concerning the individual about whom the report relates" (15 U.S. Code § 1681e(b)). CRAs draw on various data sources, often including proprietary databases of bulk criminal record data, third-party data purveyors, and in some cases "court runners" who obtain criminal record information directly from courthouses (Jacobs & Crepet, 2008). CRAs require consent from the subject of the criminal history search

(15 U.S. Code § 1681a(o)(5)(A)). If the subject of a search finds inaccurate information reported in their background check, they can file a dispute with the CRA challenging the information (15 U.S Code § 1681g(c)). The employer that required the background check, however, is not obligated to wait until the dispute is resolved to make a hiring decision and may move on to other applicants (U.S. Government Accountability Office, 2015, p. 20).

The promise of "regulated" background checks has been questioned by experts and courts alike as subjects of background checks regularly complain about incorrect information appearing on their reports, while having limited legal remedies (NCLC, 2019). In the era of big data, this issue has grown as most background check information is sourced from data brokers who act as intermediaries, aggregating public records and providing access to extensive data sets (NCLC, 2019). As these data brokers are considered data "furnishers" rather than CRAs, they face limited liability for producing inaccurate data (15 U.S Code § 1681s-2(a)). In addition to the limitations of the FCRA, a multitude of unregulated sources now exist, enabling users to access criminal record information easily and more affordably through unregulated data aggregation and dissemination processes.

2.1.2 | Unregulated background checks

The FCRA regulates the use of private background checks in scenarios in which an adverse hiring or housing outcome may result, but a diverse array of state policies governs the compilation and disclosure of governmental criminal legal system information. Many U.S. states regulate access to their repository of compiled criminal histories (what we call in this article the "State Report"), but police arrest logs, correction facility rosters, and other types of criminal justice system information are considered public record (Sandoval & Lageson, 2022). This availability is especially true for criminal court records, which are presumptively open under state and federal common law. Before the digital age, these records were difficult to access. In this paper-based world, as the Supreme Court of the United States described in *United States Department of Justice v. Reporters Committee for Freedom of the Press* (1989), criminal records existed in "practical obscurity" and therefore did not pose a major privacy or reputational threat (File, 2017).

Digitization has dramatically changed this landscape. Data aggregators have purchased, scraped, and aggregated millions of records newly digitized and publicly released by criminal justice agencies across all branches of government. Redistributing records deemed public by government is legal, ushering in a lucrative industry of criminal record databases, mugshot websites, and "people search" services that provide subscription-based reports on people's home addresses, phone numbers, social media accounts, real estate transactions, and criminal histories. These websites all affirmatively operate outside the purview of the FCRA and often post disclaimers regarding proper uses of their reports—namely, not to be used in housing or employment contexts (Lageson, 2020). Of course, many employers and landlords conduct routine internet searches of candidates and may thus inadvertently learn about a criminal history and simply choose not to consider the applicant, even in lieu of an "official" background check (Cooley & Parks-Yancy, 2016; Kluemper, 2013).

2.2 | Theory and Research in Criminal Record Data Quality

As early as 2006, the U.S. Department of Justice warned that "much more needs to be done to achieve uniformity in the improvement of record quality and completeness" (United States Office

of the Attorney General, 2006, p. 126). As access and uses of criminal records have expanded, concerns have persisted regarding serious data quality problems in criminal justice databases (Laudon, 1986; Logan & Ferguson, 2016; National Employment Law Project, 2014). Common errors include incorrect social security numbers, inaccurate names, incorrect or "illogical" birth dates, mismatched identities, and the reporting of sealed or expunged records (NCLC, 2019). Yet, private companies that use criminal record data often "disclaim responsibility of the accuracy and completeness of the reports they provide and refuse to correct detected errors" (Logan & Ferguson, 2016, p. 562). We next theorize how organizational structure, incentives, and function creep of criminal record data create and contribute to error.

2.2.1 | Organizational incentives and disincentives for data accuracy

Organizational theory helps explain why data error occurs and persists by emphasizing how ambiguity, incentives, and repeated decisions shape organizational outputs (Shapira, 1996), such as the administrative agency data that become criminal background reports. The structure of both the criminal legal system and the companies that use legal system data influence possibilities for data error, particularly due to decentralization of recordkeeping and sometimes conflicting legal and policy guidance. Similar data quality concerns exist in other institutional domains, such as in medical records (Weiskopf et al., 2013), but quality control efforts and public policy has better attended to areas outside the criminal legal system (see, e.g., Pine & Bossen, 2020), in particular, because common incentive structures, such as financial and legal liabilities, are less present in the criminal records context.

First, the decentralized structure of the criminal legal system across jurisdictions (such as counties and states) and institutions (such as police, courts, and prisons) means data are similarly siloed and fragmented in their creation (Bureau of Justice Statistics [BJS], 2001). Agencies primarily generate and maintain data to fulfill their own internal operational needs, rather than for external uses (Lynch, 2018). Moreover, diverse and often conflicting public policies govern data collection, retention, and sharing across different sectors of the criminal justice system (Sandoval & Lageson, 2022). Although agencies operate under varying organizational and policy contexts, the data they produce are often treated as having equal quality and fidelity by end users when aggregated into person-level criminal record dossiers, with consumers rarely differentiating between the sources or contexts of the information.

Second, incentive systems create more data problems in both governmental and private-sector organizations. The structure of the private market tends to shift liability from party to party: CRAs rely on data brokers, and data brokers rely on various state agencies to produce raw data. Data brokers consequently lack strong incentives for accuracy, instead passing responsibility for verifying information to the CRA or the subjects of records themselves. The industry especially benefits from public records, which are low cost and available through webscraping or freedom of information requests (Wayne, 2012), thus, operating outside state policies for the central criminal record repository.

State workers tasked with collecting and organizing criminal history data receive broad, qualified immunity for any errors they create (Lageson, 2023). Data brokers and agencies that furnish source data to CRAs must comply with some accuracy provisions of the FCRA, but they are largely protected from lawsuits (National Consumer Law Center, 2023; see, for instance, *Sprague v. Salisbury Bank & Tr. Co*., 2020). For a plaintiff to recover against a CRA, they must overcome newly stringent requirements under a recent Supreme Court decision that held a person must be materially harmed by a data error before pursuing a claim—and that the

existence of the error itself does not constitute such a concrete harm (*TransUnion LLC v. Ramirez*, 2021).

Third, people harmed by error are disincentivized to pursue remedy. Because criminal record data exist in so many domains, the subject of a report is faced with myriad time-consuming approaches: They can dispute the record directly with the CRA that produced the report, they can file a complaint with the Consumer Financial Protection Bureau, or they can contest their state record directly with the central criminal records repository—all efforts that constitute "data work" (Brensinger, in press) and may be unlikely given the documented "systems avoidance" of people with criminal histories (Brayne, 2014). The FCRA itself prompted one federal judge to describe the law as "imposing such burdensome procedural requirements that no layperson could possibly be expected to comply" (*Burrell v. DFS Serv., L.L.C.*, 2010).

Finally, criminal record data can be characterized by function creep: "the expansion of a system or technology beyond its original purposes" (Koops, 2021, p. 29), which tends to embed and bury data error. Function creep occurs when administrative criminal record data on individuals are collected for specific internal purposes and uses (Lynch, 2018) but "have migrated to other [purposes and uses] that extend and intensify surveillance and invasion of privacy beyond what was originally understood and considered socially, ethically and legally acceptable" (Dahl & Sætnan, 2009, p. 84). Criminal record function creep is rooted in the institutionalization of background checking across domains (Corda, 2016) and the wide availability of criminal records as a public data source (Bushway & Kalra, 2021; Lageson, 2020). Background checking, however, has continued to receive broad public, political, and legislative support (McElhattan, 2019, 2022b; Thacher, 2008), and the public continues to seek out other people's criminal records (one study shows 15 percent of U.S. adults report searching online in the past year to see whether a person has a criminal record [Lageson et al., 2019]) even though these data were originally collected for governmental purposes and do not easily translate into these other uses.

These incentive systems and criminal record function creep directly create data integrity problems on the ground level. For instance, when private-sector companies purchase or otherwise acquire administrative data, those data are generally available only at the case level and do not include *unique* person-level identifying information, often referred to as "ground truth" identifiers. Without access to ground truth identifiers, such as biometric data or even consistently present state ID numbers or case numbers, third parties must rely on matching algorithms using combinations of alternative data points (Tahamont et al., 2021; Tahamont et al., 2023). These algorithms use person-level information (e.g., name, alias, birthdate, state) of varying quality and completeness to link together and aggregate criminal record data for their customers (United States Government Accountability Office, 2015). The imprecision of these matching methods, however, increases the likelihood of inaccurate records, especially false-positive errors in which spurious criminal record events are erroneously associated with an individual.

2.2.2 | Empirical research on criminal record data quality

Empirical research into the quality of private-sector criminal records has noted that "the high number of false negatives and the poor quality of the private records raise some troubling questions about the use of private background screening firms" (Bushway et al., 2007, p. 176). One analysis of state records in New York showed an 80 percent error rate (Center for Community Alternatives, 2015), and a federal analysis showed that a system used to screen employees reported incorrect information 42 percent of the time (Appelbaum, 2015). In 2020, 191,000 complaints

regarding inaccurate background checks were filed with the Consumer Financial Protection Bureau (2021).

In a motivating study for this project, Bushway et al. (2007) compared FBI records with private-sector criminal background checks for 120 probationers in Virginia. The study revealed a significant false-negative problem, with the FBI reporting criminal histories for only 72 percent of the sample, and the private background check company reporting criminal histories for 47 percent of the sample even though all 120 people in the sample were previously incarcerated. Only 42 of the 120 people in the sample were found in both the FBI and private-sector databases. The authors concluded that "we do not have a great deal of confidence that the information provided by either source represents an accurate description," particularly due to the high number of false negatives (Bushway et al., 2007, p. 191). The authors further recommended that "future research should focus on identifying the prevalence of false positives" (p. 188), which we use to motivate our project.

Notably, this study was conducted in the early ages of digitization and aggregation of criminal record data. The private company that provided criminal history information to the researchers created their background reports by directly requesting data from a single county criminal court (and perhaps used a "court runner" or a staff member employed by the company that directly interfaced with court clerks to obtain records for a single person; see Bushway et al., 2007, p. 188–189). Since their study was published, major shifts in data collection and aggregation have changed the scope of criminal record data available to private companies. Most companies now rely on subscriptions to data aggregators that continually scrape or purchase bulk data from courts, correctional agencies, and police departments (NCLC, 2019). Although background check companies used to contact courts directly, they can now run a subjects' name through dozens, if not hundreds, of databases to locate criminal record information.

Our study also departs slightly from that of Bushway et al. (2007) by including qualitative interviews in response to recent calls to improve and validate quantitative data sources (Bjerre-Nielsen & Glavind, 2022). These calls encourage using qualitative data to better understand large-scale data sets on human behavior (Grigoropoulou & Small, 2022) and to concretely establish known identifiers for merging administrative data, such as social security numbers or biometric indicators like fingerprints (Tahamont et al., 2023). Although known identifiers (or "ground truth" identifiers) can be difficult or costly to obtain, even a modest investment in ground truth training data can lead to dramatic increases in merged data quality (Tahamont et al., 2023). Unfortunately, criminal records often lack consistent known identifiers that span across data sources because data are created in siloed governmental agencies and then aggregated without careful attention to detail (Borakove et al., 2015; SEARCH, 2001). Thus, qualitative data can help create ground truth.

Assessing the reliability of criminal records is important for several reasons. Criminal records are openly and routinely used to justify discriminatory hiring and employment practices (Kurlychek et al., 2019; Lindsay, 2022; Pager, 2003). The negative mark of a criminal record is well documented in social science literature (Uggen & Stewart, 2015). Criminal records increase social inequality as they are assigned disproportionately to racial minorities, particularly at the preconviction stages of arrest and initial charges (Kurlychek & Johnson, 2019), well before a person is found guilty of a crime, is still legally innocent, and may have their charges dropped or dismissed due to a lack of evidence. Yet, these nonconviction records routinely appear in internet searches and background checks (Lageson, 2020). Racial disparities are also correlated with poor criminal record data at the state level (McElhattan, 2021). Yet, even amidst the salience of

criminal background checks on a broad set of life outcomes, we lack a robust understanding of the reliability and accuracy of criminal record data.

Accordingly, this study asks the following: 1) How accurate are criminal records across public and private platforms? 2) How often do publicly and privately sourced criminal records match? 3) What are the experiences of criminal record subjects who must navigate different versions of their criminal records?

3 | METHODOLOGY

We use a multimethodological approach to investigate criminal record accuracy. We compared criminal record reports for 101 New Jersey participants,[1] all of whom had at least one criminal charge in their history, from three sources: 1) the official state record (hereafter, "State Report"); 2) a background check report from a consumer reporting agency ("Background Check"); and 3) a report generated from an open-source, public records data aggregator ("Data Aggregator Report"), also known as a "people search" service. We then asked respondents to review their criminal records in a qualitative interview.

3.1 | Participant Recruitment and Data Sources Overview

We recruited participants through partnerships with legal aid and workforce development centers in New Jersey. Once enrolled, participants participated in an intake interview with the research team and completed a state "personal records request" form, which the research team submitted to the New Jersey State Police to obtain a confidential copy of their official state criminal record (the State Report). In addition to the request form, the New Jersey State Police requires subjects to provide fingerprints to *IdentoGo*, a private company that contracts with New Jersey State Police for biometric matching to their State Bureau of Identification's (SBI's) central repository of criminal history data. The fingerprints and report fees (totaling $40.66 for each participant at the time of data collection) were paid for by the research study.

State Reports contain all arrests, criminal charges, and case dispositions recorded in the state linked to the record subject's name and fingerprints. Data points produced by law enforcement, criminal courts, and corrections are collected by the NJ State Police's SBI, which "functions as the central repository for the receiving, verifying, coding, processing, and dissemination of all criminal history record information used by criminal justice agencies for criminal justice purposes and noncriminal justice agencies for licensing/employment purposes" (New Jersey State Police, 2022). When an individual's fingerprints are first entered into the Automated Fingerprint Identification System (e.g., after an arrest, during the booking process, or after a criminal conviction), the SBI assigns a unique SBI number (NJ Rev Stat § 53:1-15). The SBI aggregates case-level criminal record data to the individual level using the SBI number. In relational terms, the SBI number, which is based on biometric data (here, fingerprints), serves as the primary key linking the various criminal record data points together across the state sources within the repository. These data are compiled into the State Report primarily for end users with unrestricted access, such as government actors or the report subject themselves. Third-party users (such as employers and volunteer

[1] A total of 104 people participated in the study; three participants were dropped from this analysis because they had only arrest records, with no corresponding criminal charges filed.

organizations) can obtain copies of the State Report with various levels of arrest-only and nonconviction records redacted per the permission level of the user and with the consent of the record subject (N.J.Admin Code § 13:59-1.1).

Next, we ordered private criminal history reports for each participant through a FCRA-compliant background check company (the "Background Check") and a popular unregulated, non-FCRA–compliant people search website (the "Data Aggregator Report"). The Background Check Company contracts with different data vendors that each conduct criminal record searches at a different level (using county, state, or national databases). The average cost for each Background Check was $49.32. The Data Aggregator Company, on the other hand, charged a flat monthly subscription fee of $35.12. With a subscription, Data Aggregator users can conduct unlimited searches that, in addition to criminal record information, also include residential, family, social media, e-mail, and government licensure information.

The Background Check Company provided us with the full report it compiled based on its subscriptions with various data vendors. Note that if the report were requested by a landlord or employer to assess an applicant (and potentially trigger an "adverse" outcome), the Background Check Company is mandated to remove arrests that did not lead to a conviction from more than 7 years ago but can report convictions from any time in a person's life. The reports we received from the Company were left intact, showing all arrests and nonconviction records as far back as their data revealed incidents. The reports we analyzed thus constitute the source data that innumerable private background check companies use to develop reports furnished to employers and landlords. The degree to which individual companies scrutinize and verify the underlying private-source data varies widely by their respective internal company policies. Because we focus on the *source* data rather than on their interpretation of the data, our approach allows for a more fulsome comparison with the state RAP sheet that, similarly, included the state's official version of a person's entire criminal history.

The Data Aggregator Company provides reports for nonadverse action scenarios, such as curiosity, screening friends, dates, or family members, or conducting investigations. The Data Aggregator Company we used for the study affirmatively opts out of the FCRA regulatory scheme, while selling background checks to consumers. The Company simply warns users against using the report for scenarios that trigger FCRA compliance, such as employment or housing. Many companies follow this business model, such as InstantCheckmate, Global Background Screening, BeenVerified, Intelius, and Garbo (the nonprofit background screening organization formerly integrated into various online dating platforms operated by Match Group, including Match.com, Tinder, and Stir).

Participants received copies of their state criminal records in the mail after being fingerprinted at IdentoGo, submitting their paperwork, and paying the required fee to the New Jersey State Police. Each participant brought their State Report to a "record review" interview with the first author. The record review interview protocol was based in part on a standardized medical clinical interview used to document the health history of a patient, wherein qualitative interview questions grounded in administrative data are used to contextualize and connect the lived experience of the individual to their record on file. In our interviews, we asked our participants to describe the context surrounding each arrest and criminal case on their record. We also asked whether they had discovered their criminal record in other electronic sources (such as third-party background checks and publicly available websites) and what, if any, remedies they planned to pursue. If the participant was interested in seeking assistance with criminal record expungement, their case was referred directly to one of two legal aid partners. Participants received compensation of $40 for their participation in the study.

We included this qualitative component of the study for several reasons. Relying on comparisons of administrative criminal record data does not shed light on the subjective experience of having a criminal record, navigating public and private sources of data, and attending to the resulting harms that can occur from mismatching or incomplete data. We conducted the interviews at the same time we received the State Report and before analyzing data from the private sources. This approach allowed us to focus on the State Reports as the baseline, enriched by the additional context provided by the interviewee who guided us through their record. Interviewees highlighted events or details they believed were inaccurate or misrepresentative and identified when probation violations, pretrial interventions, or plea bargains helped create the criminal record event. We flagged these instances in our fieldnotes.

In our analysis, we frequently returned to the interview fieldnotes and transcripts to understand discrepancies we uncovered in the quantitative matching analysis. This approach helped explain key patterns of data inaccuracy as we followed up on events or details flagged by our interviewees. Thus, our use of the interviews was interactive and back and forth with the quantitative analysis. Space constraints prevent us from fully presenting interview results alongside the matching analysis, but we use illustrative quotations and case examples here to help describe the main findings.

3.2 | Conceptualizing Accuracy to Match and Compare Records

Criminal record data were collected from each of the three sources for each participant, but the sources varied in how they identified relevant records for our participants. Records for the official State Reports were compiled by the New Jersey State Police after an applicant submitted biometric data (e.g., fingerprints), name, social security number, and birthdate. The Background Check Company required the full name (first, middle, last), possible aliases, date of birth, and social security number (which the Company uses to identify address history and additional aliases). For the Data Aggregator, the research team searched the web interface using all parameters the Aggregator made available, consisting of first name, middle initial, last name, age, and state of residence.

To compare charge events across sources, criminal charge and disposition data from each of the three criminal record report sources were compiled into three respective data sets. The available data points were inconsistent, however, across the three report sources. Each source provided a different combination of charge/case identifying data points (date types, identifying case number types, charge level) with varying degrees of missingness. Thus, confidently matching individual, specific charges across sources presented obstacles. We also found considerable discrepancies in the specific terms, word order, and use of abbreviations across the three criminal record sources used to describe the same charges and dispositions within the same participant.

We took a reductive approach in our data preparation to maximize comparability. First, we limited our analysis to adult criminal records. Only nine participants in our sample had at least one juvenile charge (totaling 19 juvenile charges in the sample), and none of those charges or cases were reported by the private-sector sources (indeed, the private-sector sources found no criminal record information for five of those participants, despite them all having adult records). We therefore removed those cases from the data to provide the cleanest possible comparative analysis.

Second, we standardized charge descriptions and dispositions following the BJS's strategy for collecting and processing diverse multistate criminal history data for statistical and research analysis (Durose et al., 2019). Charge descriptions in all three sources were converted from New

Jersey offense descriptions to BJS National Corrections Reporting Program (NCRP) offense codes using the National Archive of Criminal Justice Data (NACJD) crosswalk that includes BJS's national recode of state and federal offenses (NACJD, 2022). BJS codes are broader categories of offenses, such as "Drug Trafficking," "Drug Possession," "Burglary," "Forgery/Fraud," "Aggravated Assault," and "Larceny/Theft." For example, "cds w/i distribute," "cds/analog - distribute on/near school property/bus," and "manufacture/distribute cds" were all recoded as "Drug Trafficking." Charge dispositions were also recoded into one of three broader categories: "convicted" (e.g., "guilty," "plea to accusation," "defn plead to lesser"), "dismissed" (e.g., "continued for dismissal," "acquitted," "pre-trial intervention"), or "unknown" (e.g., "pending," "arraigned", "dispo unavailable").[2]

Accuracy, however, is a moving target. Our primary motivation for this project was to evaluate the degree to which private-sector data sources represent the official version of a person's criminal record through a comparative approach; in doing so, we made the methodological choice to treat the State Report as the "ground truth." We recognize that state records also contain errors and omissions, so we bolstered our evaluation of the state record with the qualitative interviews. Nonetheless, our matching framework treats the State Report as the baseline. This choice is also grounded in real-world application: The state record is used by the FBI to conduct national background checks, is considered more robust because it relies on biometric identifiers and social security numbers, and in practice, is considered the baseline source of a person's criminal record in the state. If, for instance, a New Jersey resident wishes to contest their State Report, they must obtain the relevant court or probation disposition paperwork and provide it to the state police to have the State Report updated. In this sense, court, corrections, and police data are considered piecemeal bits of supporting data to the compiled criminal history report. U.S. Supreme Court precedent in *DOJ v. Reporters Comm. for Free Press* (1989) further establishes the RAP sheet as a primary source of a criminal record because it is organized around a person's name, as opposed to being organized around the structure of the criminal justice agencies that supply the baseline information.

If an interviewee contested the information in their State Report (which was common), we did not alter the quantitative state data to match their account, but instead we tracked these comments through the accompanying qualitative data. Because participants received their report in the mail, they had the opportunity to look over the information before our interview; however, most participants expressed that this was their first time seeing the information and that they found it difficult to understand. In response, the initial part of the record review interview consisted of walking the participant through the composite parts of the report and soliciting contextual information about each arrest and criminal charge.

The resulting analysis is a comparison of how private-sector criminal records from a regulated source (the Background Check) and an unregulated source (the Data Aggregator Report) compare with the official government-produced criminal history report (the State Report). We contextualize these findings with data from participant interviews to draw out the consequences and experiences of having errors in criminal records.

[2] The Background Check Company provided three raw data extracts for each participant (county, state, national). For ease of comparison, we merged the three extracts into one data set before merging with the other two data sources. In cases of conflicting dispositions, we coded the merged disposition in the following priority: "conviction," "dismissal," "unknown" (e.g., if the state extract reported a "conviction" and the national extract reported a "dismissal" for the same charge, the charge in the merged Background Check data set was coded as "conviction").

TABLE 1 Types of criminal record data error.

Error Type	Description	Possible Contributing Factors
False negative	Private-sector background check fails to report incident found in official government record	Data entry error in source data; private-sector database outdated
False positive	Private-sector background check reports incident that does not appear in official government record ("orphan" records)	Matching error; background check company drawing data from alternative sources (e.g., law enforcement arrest data); case has been expunged or otherwise sealed
Missing, misleading and inaccurate information	Missing or incorrectly reported case dispositions (e.g., conviction rather than dismissal); incorrect type of charge reported (e.g., felony rather than misdemeanor)	Private sector using outdated version of criminal record data; background check company not in compliance with regulations

TABLE 2 Charges and dispositions by source.

Charges/Dispositions	State Report	Background Check Company	Data Aggregator Report
All Charges			
Charges	1,323	1,123	973
Dispositions			
Convictions	401	350	632
Dismissal/Acquittal/Diversion	821	669	117
Disposition unknown	101	104	224
Charges by Participants			
Participants with charges	101	77	56
Participant dispositions			
Convictions	92	68	47
Dismissal/Acquittal/Diversion	89	62	30
Disposition unknown	28	22	38

Note: The top section displays counts of charges and charge dispositions by criminal record source for all charges in the sample. The bottom section displays counts of sample participants with at least one reported charge and respective disposition type by the three criminal record sources.

4 | RESULTS

Governmental error (i.e., missing case information or final dispositions) and private-sector error (i.e., when imperfect techniques of data aggregation, matching, or sourcing) led to false negatives and false positives in consumer background check reports. In what follows, we describe how these errors manifest: as mismatched data that create false negatives in third-party records, as missing case dispositions that create incomplete and misleading criminal records, and as incorrectly reported criminal records that lead to false positives in third-party reports. Table 1 displays a summary of the error types we identified in our data.

4.1 | Missing Data and False Negatives

The first column of the top section of table 2 shows incident-level charges and the distribution of possible dispositional outcomes: conviction, dismissal, or unknown/missing. The bottom section

shows the distribution of charges and possible distributions across sample participants. According to the government-produced State Reports, all study participants had been arrested and charged with at least one crime, but only 92 were convicted of at least one charge (that is, nine participants were charged with at least one crime but never convicted). The State Reports included a combined total of 1,323 criminal charges, 30 percent of which (401) resulted in a criminal conviction and 62 percent of which (821) were dismissed. Table 2 also includes cases with unknown case dispositions.

Columns 2 and 3 of table 2 present criminal record data from the private-sector sources. Several key differences emerge. First, as displayed in the bottom panel, not all participants were located by private-sector data sources. Although all participants had at least one criminal charge, the Background Check Company—which produces federally regulated background checks for employers and landlords—found criminal charges for only 77 people (1,123 total criminal charges across the sample), 68 of whom were reported to have at least one conviction. The Background Check Company reported no criminal record information (charges or convictions) for almost one quarter of our sample, creating a false-negative scenario in which a criminal background check purchased by an employer or a landlord reports a clean criminal background when the person has a record, according to the state repository. The Data Aggregator report located criminal records for an even smaller proportion of our sample (56 participants with a total of 973 charges). Despite only finding criminal records for slightly more than half of our participants, however, the Data Aggregator Report noted 632 total convictions, or 58 percent *more* overall convictions than did the State Reports ($n = 401$ convictions), the official criminal record of the state.

Table 3 summarizes the proportion of private-sector data error at both the incident level (that is, individual criminal charges) and the person level. Of the 1,323 total charges included in the State Reports (our denominator), the regulated Background Check Company failed to report 672 of these charges (approximately half of the State Report total). The unregulated Data Aggregator Report missed even more (1,065, or 80 percent). Private-sector error was widespread among the participants in our sample (bottom panel of table 3): Ninety percent of participants had at least one false-negative error on their background check, and 92 percent had at least one false-negative error on their Data Aggregator Report.

Second, each source painted a much different picture of a person's criminal record. Figure 1 depicts the types of information for each participant for each source. The left column represents the official State Report, the center column represents the Background Check (regulated) report, and the right column represents the Data Aggregator Report (unregulated). Participants #1–101 are indicated by the *y*-axis. For each source, convictions, dismissals, and unknown case dispositions are reported, respectively, by dark-shaded, light-shaded, and gray colors. If we assume data accuracy across sources (as users of criminal records might), the distribution of criminal charges and corresponding case outcomes should appear similar across data sources. The visualization shows, however, that error is not systematically assigned by source or extensiveness of the record.

4.1.1 | Missing case dispositions

Thus far we have described mismatching data that creates false-negative problems, when a person appears to not have a criminal record when requested from a private vendor, even though a record

TABLE 3 Proportion of errors by private-sector source.

Charges/Participants	Background Check Company	Data Aggregator Report	Combined Private Sector[a]
Charges			
False negatives			
State Report charges (denominator)	1,323	1,323	1,323
Charges not reported by private sector reported on State Report	672	1,065	613
Charges not reported by private sector reported on State Report (%)	50.8%	80.5%	46.3%
False positives (i.e., "orphan" records)			
Private-sector charges (denominator)	1,123	973	1,776
Charges reported by private sector not reported on State Report	472	715	1,066
Charges reported by private sector not reported on State Report (%)	42.0%	73.5%	60.0%
Participants			
False negatives[b]			
Participants (denominator)	101	101	101
Participants with at least one false negative	91	93	95
Participants with at least one false negative (%)	90.1%	92.1%	94.1%
False positives (i.e., "orphan" records)			
Participants (denominator)	101	101	101
Participants with at least one false positive	61	50	77
Participants with at least one false positive (%)	60.4%	49.5%	76.2%
Any error (false negatives or false positives)			
Participants (denominator)	101	101	101
Participants with at least one error (false positive or false negative)	97	101	101
Participants with at least one error (false negative or false positive) (%)	96.0%	100.0%	100.0%

[a]If both reports were combined to create a more comprehensive record, how would that combined report compare with the State Report? Some charges were reported by both private-sector sources; therefore, this is not a sum of the previous two columns but takes into account where the sources overlap.

[b]Participants for whom a private-sector report found no criminal record data are counted as false negatives.

exists at the governmental level. We now turn to the specific problem of missing dispositions in both governmental and private-sector sources.

An unknown or missing case disposition can result from data entry error, from data loss as paper governmental records were digitized, or through poor-quality data matching techniques.

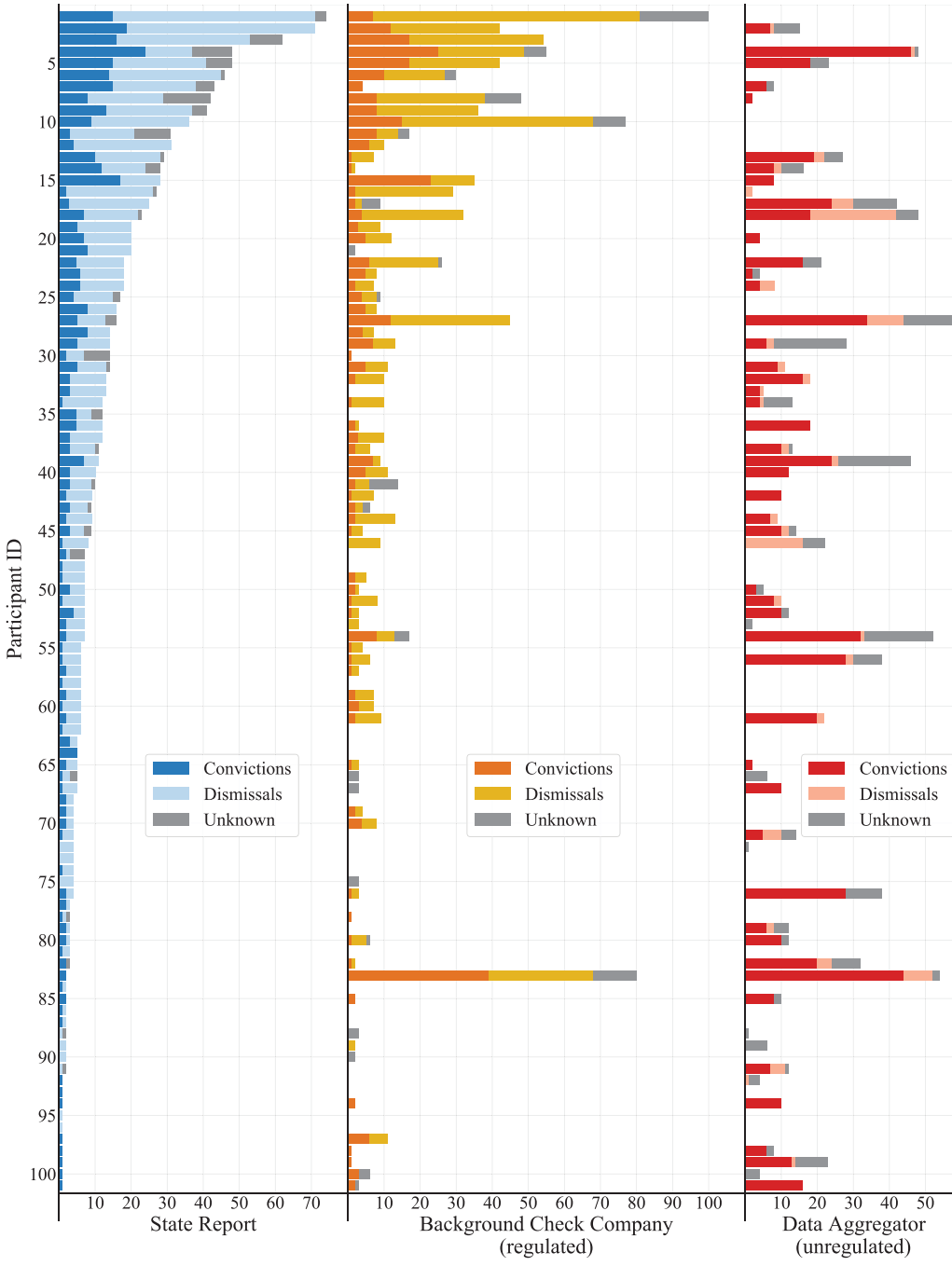


FIGURE 1 Participant charges and dispositions across data sources. [Color figure can be viewed at wileyonlinelibrary.com]

A missing case disposition can cause significant problems for the subject of a background check as it implies that a case is still open or pending. Missing case dispositions also prevent an individual from expunging or sealing their criminal record unless they successfully track down the original court docket paperwork and present it to the state office responsible for maintaining the

central repository (in New Jersey, the state police). As shown in table 2, all three data sources listed charges with unknown or missing disposition data. A disposition was coded as missing or unknown if it contained no information (e.g., blank or empty) or if it did not indicate a clear outcome for the charge. Examples included "disposition unavailable," "pending court," "refer to grand jury," "case pending," "sent to prosecutor," and more.

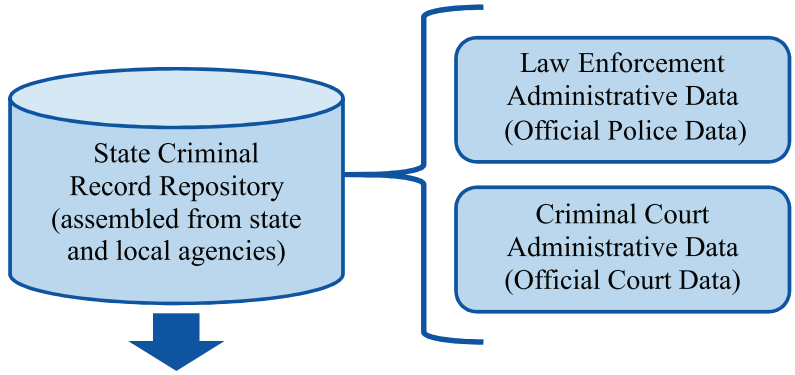
State Reports contained a significant number of missing dispositions (bottom panel of table 2): A total of 28 study participants had at least one charge with a missing disposition on their State Report, and 7.6 percent of their 1,323 charges were missing final case dispositions. Nearly three quarters of the State Report's missing dispositions were for cases filed more than 10 years before data collection, with the oldest missing data from 1978. Analyses of state and federal criminal record repositories have identified significant data quality issues for almost four decades (Laudon, 1986); therefore, this result is not a new phenomenon but an underappreciated and persistent one.

Private-sector records, too, contained missing dispositions. In total, 9.3 percent of charges reported by the (regulated) Background Check Company were missing disposition information, and 22 of the 77 participants who had a record in the Background Check Company database had at least one missing disposition. The (unregulated) Data Aggregator Report contained a much larger proportion of missing disposition data, with more than 1 in 5 (or 23 percent of) charges missing dispositions. Of the 56 people for whom the Data Aggregator Report found records, 38 (or approximately two thirds) had at least one missing disposition.

Missing data can linger for years with pernicious consequences. One of our participants (#76 in figure 1) was midway through the process to expunge a 1994 drug conviction, decided to withdraw his petition after his state record showed a "pending court" disposition from an arrest in 1981 due to criminal charges never being filed, but the police department failing to update their records.[3] He feared an open warrant would result in his arrest should he present his expungement petition to the court. Although he was legally eligible for an expungement, the missing data on his record led him to drop out of the process.

Another participant (#83 in figure 1) described his unsuccessful effort to get his record expunged in 2012. He had low-level cocaine possession charges in 1986 and 1991, both of which led to convictions and probation sentences. His State Report also includes two arrests for cocaine possession in 1987 and 1990 that appear as separate incidents, both of which are listed as "pending court." When we spoke to him, he identified these two arrests not as separate but instead as probation violations stemming from the 1986 case (which we confirmed by verifying all three events have the same case number). New Jersey's expungement process, however, requires applicants to present disposition paperwork for any incomplete records in the court file. He initially went to a local legal aid organization for assistance in clearing up his record but could not afford the $100 fee. He tried to start the expungement process himself, but the "pending court" status became a roadblock because the original records had been destroyed following the court's retention policy. He decided to try and explain the situation to the local prosecutor, but he told us, "The prosecutor said I'm stuck with it for the rest of my life." The prosecutor did, however, have advice for him: "She said there's nothing we can do, but when the governor leaves I should ask for a pardon." He did not receive a pardon, and 10 years later, he has still not successfully had his record expunged due to missing data, although he is legally eligible.

[3] Because the court never filed charges related to this arrest, these missing charges do not appear in our data set as a criminal history event for matching purposes.



	Official State Report	Regulated Background Check	Unregulated Data Aggregator
Participant	Reported Charge(s)	Reported Charge(s)	Reported Charge(s)
075	Forgery/Fraud	Forgery/Fraud	[No data found]
	Contempt of Court	Contempt of Court	[No data found]
	Contempt of Court	[No data found]	[No data found]
	Trespass	[No data found]	[No data found]
		Burglary	
100	Auto Theft	Auto Theft	Auto Theft
		Auto Theft	Auto Theft
		Auto Theft	Auto Theft
		Auto Theft	Auto Theft
		Auto Theft	
		Auto Theft	
095	Assault on Police	[No data found]	[No data found]
092	Drug Possession	[No data found]	[No data found]
			Obstruction of LE
			Obstruction of LE
			Obstruction of LE
			Assault on Police

FIGURE 2 Examples of criminal record data across three sources. [Color figure can be viewed at wileyonlinelibrary.com]

4.2 | False Positives: "Orphan" Records

We next identify charges that appear in private-sector data but lack a corresponding match in official state data. We refer to these false positives as "orphan" records a concept used in database management to refer to a record that appears in a secondary source or table but does not appear in the primary source (Babad & Hoffer, 1984). Figure 2 illustrates the relationship between the official state data and the two private criminal record sources, with four participants from the sample as examples. The criminal record data contained in the State Report is aggregated from local police departments and criminal courts and reflects the most current version of the official state criminal record data for each participant. Orphan records (false positives) are indicated by red or orange text for private-sector–reported criminal charges that have no corresponding charge in the State

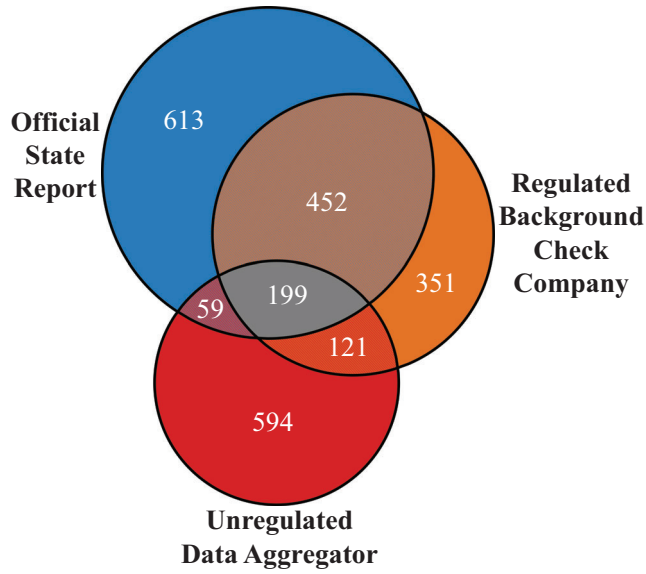


FIGURE 3 Overlap of charge coverage between three criminal record sources. [Color figure can be viewed at wileyonlinelibrary.com]

Report. "[No data found]" indicates that the private-sector report found no information on the respective State Report charge and is thus a false negative.

Orphan data likely results from substandard data matching and management techniques as private companies aggregate, compile, and store criminal record data. The Venn diagram (figure 3) shows the overlap of the charges reported by the three data sources for all our participants. Each circle represents charges reported by the three respective criminal record data sources. The reported number is the *N* of criminal charges that appear in each source (or in overlapping sources). Fields that do not intersect with the official State Report region indicate orphan records. The upset plot (figure 4) provides an additional depiction of the degree to which each source matched the other two sources. Filled circles indicate charges that appear in each respective row's source. The bars indicate the relative distribution of charges that appear in the various record source combinations. The last three columns represent orphan records.

The matching analysis shows that only 8 percent of our participants' criminal charges were reported by all three sources, and only 30 percent accurately matched at least two sources (that is, the first three bars). Furthermore, our matching analysis yielded a troubling quantity of orphan records reported by private-sector sources. Of the 2,389 matched criminal charges, 45 percent (or 1,066) were orphan records not reported in the state record (see figure 3 and columns 5–7 in figure 4). On average, 35 percent of the charges listed on each of our participants' comprehensive criminal record (or the combination of all three records) were orphan records (median = 28 percent). The types of crimes reported in orphan data were not trivial but contained serious offenses, which could have damaging impacts on people's lives. Table 4 describes the types of orphan records that appeared in private background checks.

Qualitative data help illustrate how orphan records impact lives. Participant #101, who appears in the bottom row of figure 1, was charged and convicted with one count of felony-level promoting prostitution. His regulated Background Check reported somewhat similar results, showing two charges and convictions of promoting prostitution that appear under the same case number but were filed on different dates. The Background Check also includes an orphan charge for terroristic threats with an unknown disposition. His unregulated Data Aggregator Report

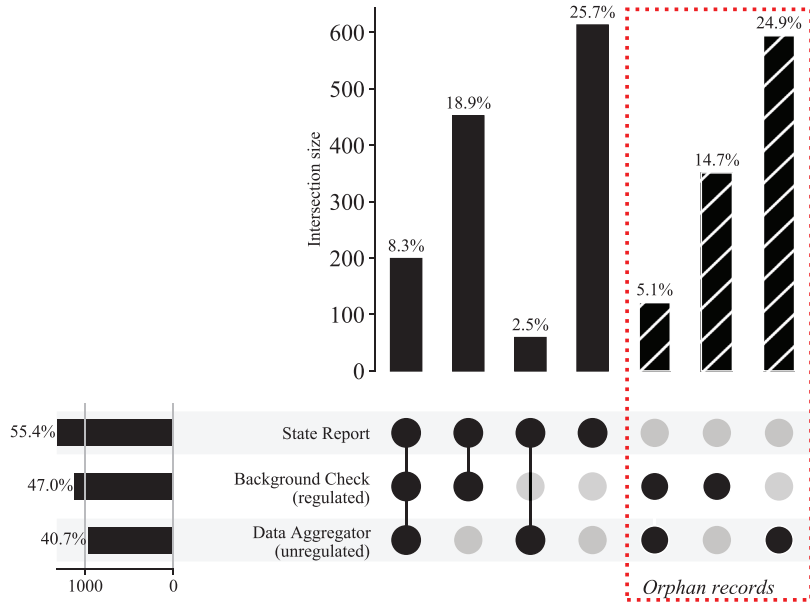


FIGURE 4 Distribution of sample charges across three data sources. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 4 Frequency of offense type of orphan charges.

Offense Type[a]	Frequency ($N = 1{,}066$)
Drugs (trafficking, possession)	340
Forgery/Fraud/Bribery/Vice	122
Weapons offenses	113
Public Order/Misc	113
Assault (aggravated, on police, simple)	99
Property/Theft (burglary, auto theft, receiving stolen property)	95
Armed robbery	75
Criminal process (obstruction, contempt, escape)	48
Family offenses (cruelty, neglect)	30
Other violent (sexual abuse, kidnapping)	17
Murder/Manslaughter	14

[a]Orphan charges were collapsed into comprehensive charge categories for display purposes.

results, however, show ten separate charges for promoting prostitution and six charges for aggravated assault. Taken across sources, participant #101 has one official state charge resulting in a conviction but 17 orphan charges. Although some Data Aggregator Report results could be duplicate entries erroneously reported as separate criminal charges, the service does not consistently provide enough information (such as case numbers and relevant dates) to specify whether these stem from an isolated, single event or reflect multiple, separate incidents that occurred over an extended period of time.

Participant #97 was arrested in 1990 for sexual assault but was charged and convicted with misdemeanor simple assault. This information is all that appears in his State Report. His regulated Background Check shows 11 charges, including seven counts of drug trafficking, two counts of drug possession, one count of receiving stolen property, and one count of felony conspiracy. His unregulated Data Aggregator Report shows nothing, reporting instead that he has no criminal history. This participant also told us he had been working for the same company for nearly three decades. His company adopted a new HR policy that required background checks on new and *current* employees, and he was fired shortly thereafter when they found his 1990 incident. He found a new job with a contractor in the aerospace industry, but he was let go because his new background check included a conviction for an assault on a police officer in another state. After closer inspection, he realized the erroneous assault was for another person with the same name, birthdate, and a match on three of the last four digits of his social security number.

Participant #83, who was referenced as having two official charges but missing dispositions, presents an even more extreme case: Ninety-eight percent of the charges reported by the three data sources for this person were orphan records. Recall that participant #83's official State Report listed only two drug felony charges (cocaine possession in 1986 and cocaine sales in 1991) that both resulted in convictions. Although the two convictions were also noted by the regulated Background Check Company and the unregulated Data Aggregator Report, the similarities end there. The Background Check reported *78 additional charges* filed from 1986 to 2009: 37 that led to convictions, 29 dismissals, and 9 with missing dispositions. The convictions included armed robbery (5 separate charges), burglary (5), receiving stolen property (7), forgery/fraud (2), weapons offenses (6), and arson (1). According to the Background Check, this participant was also charged with murder in 1991, although it was reported as dismissed. Similarly, the unregulated Data Aggregator Report included 54 charges in addition to the two on the State Report (44 convictions, 8 dismissals, and 2 missing disposition), although a different mixture of charges and dispositions than in the Background Check. Like the Background Check Company, the Data Aggregator Report noted convictions for armed robbery (2) and weapons offenses (4) but also for drug possession (9), drug sales or trafficking (11), racketeering (2), resisting arrest (2), and child neglect/endangering welfare of a child (8).

Extensive review of this participant's records obtained from the county sheriff, local police, and municipal and superior court systems; several e-mail exchanges with representatives from those offices; and consultations with numerous legal aid attorneys revealed the source of the data error: Our research participant had an alias tied to his name in the state court records database. Another defendant in the state used the same alias at a different point in time. The same name was thus entered into both persons' state court identity files. When state court data were aggregated in the release or sale to the Background Check Company and Data Aggregator Report, their respective matching algorithms presumably linked *all* criminal record events associated with the other defendant's alias to anyone *else* connected to that alias, including participant #83. When we spoke to the participant, he explained he had been unable to obtain employment due to the nature of what appears on his background check and had been working to remedy these data errors for a decade. Not only did he make repeated requests with the local prosecutor, but he also sought help from the local police, the municipal court office, and the superior court office for disposition paperwork that he could then provide background check companies to dispute the charges. To date he has been unsuccessful.

4.3 | Misleading and Inaccurate Records that Undermine Public Policy and Individual Rights

Criminal records also contain misleading information that can characterize a person as a present danger or threat, or they can be easily misinterpreted to exaggerate the seriousness of the criminal offense. For many study participants, this type of misleading information involved the disclosure of criminal charges that were filed but were later dismissed. Considering that "overcharging" and "charge stacking" are common, documented tactics prosecutors use to induce defendants to take a plea bargain for a lesser charge (Husak, 2023), these observations become problematic.

Criminal charges are considered public record; therefore, many private-sector background checks report dismissed charges collected from court records, even if state policy or the FCRA bars disclosure of such records after a set time period for regulated background checks and State Reports. For the 101 sample participants, the regulated Background Check Company and the unregulated Data Aggregator Report noted 669 and 117 charges (respectively) that were dismissed, acquitted, or otherwise closed without a conviction. Although the State Report requested by an employer or volunteer organization (with the consent of the record subject) would not include these charges on an official background check, data aggregators are legally allowed to report this information, undermining the goals of the state's criminal record disclosure policy to limit the potential harms of nonconviction records.

Even when initial criminal charges were accurately represented by private-sector sources, the final case dispositions were routinely reported incorrectly. In cases when the charges in the State Report matched the charges in the regulated Background Check report (N = 651), 34 percent of the case *dispositions* did not match (e.g., the State Report listed the charge as dismissed, whereas the Background Check reported the charge as resulting in a conviction). When both the State Report and the unregulated Data Aggregator Report noted matching criminal *charges* (N = 258), 56 percent had conflicting case *dispositions*. In total, 145 dismissals for 51 people erroneously showed up as convictions in either of the private sources. Put differently, half of the study sample had at least one dismissal—that is, where they were not legally guilty of the charge—which showed up as a guilty conviction in a private-sector report. Such data error not only misrepresents reality but also violates basic principles of due process and fairness, with lasting consequences.

For example, participant #61 described being routinely denied housing and employment due to the disclosure of serious criminal charges that were later dismissed. In 2010, she was arrested in a pharmacy parking lot after attempting to fill a forged prescription and getting into a minor car accident with the responding police officer. The prosecutor charged her with assault on a police officer and attempted murder because of the fender bender, alongside the prescription forgery charge. She was convicted only for forgery and resisting arrest, and the judge dismissed the charges related to the car accident. Yet, the regulated Background Check reports two counts of attempted murder and assault on a police officer and the unregulated Data Aggregator Report erroneously notes 20 felony convictions for obstruction, forgery, and aggravated assault and two unknown dismissed records. Although the participant has completed treatment and had no arrests in more than a decade, evidence of a purported attempt to murder and harm police officers shows up on background checks and on internet websites. "I take responsibility," she said in her interview. "I should not have been there, but for it to say attempted murder is just crazy; it just wasn't like that."

5 | DISCUSSION

The problem with criminal records is that they regularly contain errors, are easily misinterpreted, and are thus unreliable tools for evaluating individuals. For the 101 participants in our sample, 60 percent and 50 percent had at least one false-positive error on their regulated and unregulated background checks, respectively, and nearly every participant (90 percent and 92 percent, respectively) had at least one false-negative error (see table 2). The extent of mismatch suggests that a person's record can be so disparate as to be basically unrecognizable across sources. Beyond just being fragmented, erroneous, and variable, contemporary criminal record information can paint a starkly inaccurate portrait of a person. As figure 1 explains, one must not even have an extensive record to be subject to error; rather, any person with any type of criminal record is subject to extensive errors.

Although this study does not measure error in background checks for those with *no* history of criminal legal involvement, these error rates remain significant to a large proportion of the U.S. population, affecting the estimated one in three American adults who have some type of criminal record, and the estimated 31 million U.S. residents who have a background check conducted for employment reasons alone each year (Brame et al., 2012; Denver et al., 2018; Greenspan & DeBacco, 2014).

Inaccurate criminal records carry serious consequences (Horn, 2019). Criminal records are used to deny people employment, housing, and social and institutional opportunities. Although the dissemination of accurate conviction records may be justified on public policy grounds (Bushway & Kalra, 2021), the widespread dissemination and perpetuation of inaccurate criminal records undermines these goals. As such, these practices violate basic principles and policies of the criminal justice system, including the presumption of innocence and constitutionally guaranteed rights of due process. More practically, the spirit and promise of criminal legal diversion programs that offer people an opportunity to avoid the mark of a criminal record if they complete a probationary period are undermined by inaccurate records. Similarly, criminal record expungements—in which old records are sealed from public view after completing a period without any further criminal legal involvement—are rendered moot when these sealed records remain in private databases.

Organizational structures, conflicting public policies, and financial incentives likely play a key role in criminal record data accuracy problems. The concentration of harms among vulnerable and marginalized groups, alongside the significant barriers to remedy such harms, means data accuracy may have long lasting consequences for both individuals and broader social stratification. These contextual factors deserve more careful scrutiny amidst widespread and persistent background checking in America.

5.1 | Data Error in an Automated and Algorithmic World

The data inaccuracies in the regulated Background Check Company and the unregulated Data Aggregator Report also raise important questions about whether the private sector should report criminal histories at all and, more broadly, questions the efficacy of using criminal records for a widely expanding set of institutional and economic purposes if the data are so unreliable. Our study thus points to the problem of "function creep" as records created for the criminal legal system are used across nonlegal settings with consequences for American inequality. Administrative

criminal record data is commonly used as a purportedly reliable indicator of behavior, with uses spanning social science research, predictive modeling, and risk assessment and sentencing tools. As such, many of these enterprises are built on a shaky foundation.

The widespread reliance on probabilistic matching algorithms in the private sector exacerbates the issue of data inaccuracy. Probabilistic algorithms attempt to match records based on the likelihood of a match, often using data points such as names, birthdates, and addresses, rather than definitive confirmation using unique identifiers, such as biometric or unique identification numbers. Thus, this approach is inherently prone to create errors in background checks, as well as bias in statistical analyses. Individuals with similar names or demographic information may be incorrectly, and unfairly, matched with another person's criminal record, leading to significant consequences. Researchers and users of administrative data must be attentive to the sources of their data, how those data were created and maintained, and the intended purpose(s) for collecting those data (Lynch, 2018). In addition, administrative data users should be cautious about the quality of their data sources, recognize the limitations of probabilistic matching algorithms, and consider alternative methods or additional safeguards to minimize errors and biases in criminal records, such as incorporating qualitative data to evaluate whether quantitative data sources are measuring what they appear to be measuring (Bjerre-Nielsen & Glavind, 2022; Grigoropoulou & Small, 2022).

As criminal record data become increasingly ubiquitous and used for purposes for which they were not originally collected, such as applicant assessments, predictive modeling, and risk analysis tools, problems in source data can be inadvertently amplified, leading to flawed assessments and unfair outcomes. Moreover, the opacity of many algorithmic systems makes it difficult for individuals to challenge the accuracy or fairness of decisions based on their criminal records (Harron et al., 2017). Thus, researchers and other users of administrative data should prioritize transparency, accountability, and the development of robust methods to assess the validity and fairness of the algorithms that underpin criminal records and related decision-making processes. This approach would help ensure that the use of criminal records data in various contexts is not only accurate but also ethically and legally sound.

5.2 | Implications, Limitations, and Next Steps

The data accuracy issues we describe raise questions about how useful background checking is for its purported purposes. For instance, both private-sector sources reported "clean" backgrounds for significant portions of our sample. Given the institutionalization of criminal background checking in the United States and expectations put onto hiring managers to conduct such due diligence, false negatives provide a misleading sense of employer protection from liability or workplace safety. In turn, our qualitative accounts show how people's understanding of their own records may diverge from both state and private-sector versions, which can be consequential for those individuals (see also Myrick, 2013). For instance, if an employer asks an applicant about what appears on a background check, but that report relies on inaccurate data, the employer may believe the applicant is lying about their criminal record because the underlying report is inaccurate (see Roberts, 2015).

Some areas of harm also could be concentrated in particular groups. In 2021, the Consumer Financial Protection Bureau warned against using name-only matching techniques for background screening due to the heightened likelihood for mismatch "to be greater among Hispanic, Black, and Asian individuals because there is less surname diversity in those populations than

among the non-Hispanic white population," citing recent U.S. Census data (Chopra, 2021, para. 7). Our analysis also shows that error is more concentrated among older records, likely due to the transition from paper to digital records, which points to a potential harm for the aging population. This issue may be a consideration for elder law and estate planning. The use of faulty criminal record data may also lead to increasing racial inequities as states with more racially punitive legal systems also maintain the faultiest criminal record systems (McElhattan, 2021). Future research should closely examine these differences, in particular, for juvenile records, older records, and other specific groups.

The findings and methodological approach also have implications for criminology. First, this study responds to recent calls for "small data analysis" (Bjerre-Nielsen & Glavind, 2022; Grigoropoulou & Small, 2022) to promote better understanding of big data analytics and to include qualitative field data into quantitative analyses. Results show the importance of context in understanding administrative data, and the study empirically demonstrates the value of qualitative data in big data applications. Our study suggests that scientists can better understand social phenomena by going "into the field" and becoming closer to the production of data used in secondary analyses. Researchers may, for example, consider developing data validity statements to accompany their quantitative studies. Writ large, this process will involve recognizing that governmental data not originally produced or structured for research purposes can still be used for such applications after seeking and establishing context and ground truth and elevating the lived experience of data subjects.

This study has limitations. By emphasizing the research value of qualitative interviews, the research is limited by the number of participants and the geographic scope of one state. In addition, we did not measure error as it relates to people with no criminal history; instead, study participants had at least one criminal charge (but not necessarily a criminal conviction) in their past. Participants were also not systematically sampled from the broader state population and were instead recruited from partnerships with legal aid organizations and workforce centers that advertised our study and assisted in recruiting eligible participants. Next steps for expansion and application of this research would involve taking the lessons learned from this small data analysis to identify broader patterns of missing and mismatching data across all state criminal justice data, in data shared with federal agencies, and in audits of private-sector criminal record data repositories.

This area is clearly ripe for policy reform. Policy responses at the state level might include greater privacy protection for record subjects (particularly for nonconviction records), creating legal requirements for accuracy in private-sector data, and creating the ability for record subjects to check their own records for mistakes at no cost in both public and private sectors (see Lageson, 2023). States can also create consumer data and privacy laws that go beyond the FCRA, such as forbidding CRAs from reporting arrest records (United States Bureau of Consumer Financial Protection, 2022). States could also better centralize and regulate the disclosure of bulk criminal history information through a single agency contact, such as the criminal record repository or a unified state court system, with data-sharing agreements with vendors that include accuracy and data refreshing provisions. This approach would allow states to directly offer criminal history checks for registered employers and landlords from a single source, and with a limited scope of information. Thus, employers would rely less on the private sector and shift the market toward generating state resources that could be reinvested in better data quality measures.

Several broader implications follow from this research. Inaccurate and misrepresentative criminal records produced by the private sector harm both people with records and people who rely

on records to make decisions. Criminal records used for employment, education, and housing contexts may rely on incorrect information, leading to illegal discrimination or failure to comply with public safety regulations. Continued public disclosures of old, incorrect, expunged, or misrepresentative criminal records also undermines due process and public safety goals. Inaccurate labeling may increase recidivism and prevent fair access to safe and stable housing, employment, and family life—all of which have been shown to reduce crime (Crutchfield & Pitchford, 1997; Duwe & Henry-Nickie, 2021; Kirk et al., 2018; Uggen, 1999; Warr, 1998). Data error and its attendant consequences should be better framed as a legal system harm.

ORCID

Sarah Lageson https://orcid.org/0000-0002-4108-4365
Robert Stewart https://orcid.org/0000-0002-6326-915X

REFERENCES

Appelbaum, B. (2015, February 28). Out of trouble, but criminal records keep men out of work. *The New York Times*. https://www.nytimes.com/2015/03/01/business/out-of-trouble-but-criminal-records-keep-men-out-of-work.html

Babad, Y. M., & Hoffer, J. A. (1984). Even no data has a value. *Communications of the ACM*, *27*(8), 748–756.

Bjerre-Nielsen, A., & Glavind, K. L. (2022). Ethnographic data in the age of big data: How to compare and combine. *Big Data & Society*, *9*(1), 20539517211069893. Advance online publication. https://journals.sagepub.com/doi/10.1177/20539517211069893

Blumstein, A., & Nakamura, K. (2009). Redemption in the presence of widespread criminal background checks. *Criminology*, *47*(2), 327–359.

Borakove, M. E., Wosje, R., Cruz, F., Wickman, A., Dibble, T., & Harbus, C. (2015). From Silo to System: What Makes a Criminal Justice System Operate Like a System? *The Justice Management Institute*, https://ncfsc-web.squiz.cloud/__data/assets/pdf_file/0024/52449/From-Silo-to-System-30-APR-2015_FINAL.pdf

Brame, R., Turner, M. G., Paternoster, R., & Bushway, S. D. (2012). Cumulative prevalence of arrest from ages 8 to 23 in a national sample. *Pediatrics*, *129*(1), 21–27. https://doi.org/10.1542/peds.2010-3710

Brayne, S. (2014). Surveillance and system avoidance: Criminal justice contact and institutional attachment. *American Sociological Review*, *79*(3), 367–391.

Brensinger, J. (In press). Personal data work: Managing personal information in everyday life. *American Sociological Review*.

Bureau of Justice Statistics. (2001). *Report of the National Task Force on Privacy, Technology, and Criminal Justice Information*. https://bjs.ojp.gov/library/publications/report-national-task-force-privacy-technology-and-criminal-justice-information

Burrell v. DFS Serv., L.L.C., 753 F. Supp. 2d 438, p. 444 (D.N.J. 2010).

Bushway, S. D., & Apel, R. (2012). A signaling perspective on employment-based reentry programming: Training completion as a desistance signal. *Criminology & Public Policy*, *11*(1), 21–50. https://doi.org/10.1111/j.1745-9133.2012.00786.x

Bushway, S. D., Biggs, S., Taxman, F., Thanner, M., & Van Brakle, M. (2007). Private providers of criminal history records: Do you get what you pay for? In S. D. Bushway, M. A. Stoll, & D. F. Weiman (Eds.), *Barriers to reentry? The labor market for released prisoners in post-industrial America* (pp. 174–200). Russell Sage Foundation.

Bushway, S. D., & Kalra, N. (2021). A policy review of employers' open access to conviction records. *Annual Review of Criminology*, *4*(1), 165–189. https://doi.org/10.1146/annurev-criminol-061020-021622

Center for Community Alternatives. (2015). *Boxed out: Criminal history screening and college application attrition*. https://www.communityalternatives.org/wp-content/uploads/2019/11/boxed-out.pdf

Chien, C. (2020). America's paper prisons: The second chance gap. *Michigan Law Review*, *119*, 95.

Chopra, R. (2021). *Statement regarding the advisory opinion to curb false identity matching*. Consumer Financial Protection Bureau. https://www.consumerfinance.gov/about-us/newsroom/statement-regarding-the-advisory-opinion-to-curb-false-identity-matching/#note1

Community Legal Services of Philadelphia. (2018, February 9). Reporting expunged or sealed cases in commercial background checks violates the Fair Credit Reporting Act. https://clsphila.org/criminal-records/fcra-and-expungements/

Cooley, D., & Parks-Yancy, R. (2016). Impact of traditional and internet/social media screening mechanisms on employers' perceptions of job applicants. *The Journal of Social Media in Society*, *5*(3), 151–186. https://thejsms.org/index.php/JSMS/article/view/180

Corda, A. (2016). More justice and less harm: Reinventing access to criminal history records. *Howard Law Journal*, *60*, 1–60.

Crutchfield, R. D., & Pitchford, S. R. (1997). Work and crime: The effects of labor stratification. *Social Forces*, *76*, 93–118.

Dahl, J. Y., & Sætnan, A. R. (2009). "It all happened so slowly"—On controlling function creep in forensic DNA databases. *International Journal of Law, Crime and Justice*, *37*(3), 83–103. https://doi.org/10.1016/j.ijlcj.2009.04.002

Denver, M., & DeWitt, S. E. (2022). "[It's] what you do after the mistake that counts": Positive employment credentials, criminal record stigma, and potential pathways of mediation. *Criminology*, *61*(1), 5–39. https://doi.org/10.1111/1745-9125.12319

Denver, M., Pickett, J. T., & Bushway, S. D. (2018). Criminal records and employment: A survey of experiences and attitudes in the United States. *Justice Quarterly*, *35*(4), 584–613. https://doi.org/10.1080/07418825.2017.1340502

DOJ v. Reporters Comm. for Free Press, 489 U.S. 749 (1989).

Durose, M. R., Cooper, A. D., & Snyder, H. N. (2019). *Collecting and processing multistate criminal-history data for statistical analysis* (Technical report NCJ 250750). Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.

Duwe, G., & Henry-Nickie, M. (2021). *A better path forward for criminal justice: Training and employment for correctional populations*. Brookings. https://www.brookings.edu/research/a-better-path-forward-for-criminal-justice-training-and-employment-for-correctional-populations/

Evans, D. N. (2016). The effect of criminal convictions on real estate agent decisions in New York City. *Journal of Crime and Justice*, *39*(3), 363–379. https://doi.org/10.1080/0735648X.2016.1166068

Evans, D. N., & Blount-Hill, K. L. (2020). Swipe right? Experimental analyses of app-based dating in the age of criminal stigma. *Corrections*, 1–23.

File, P. C. (2017). A history of practical obscurity: clarifying and contemplating the twentieth century roots of a digital age concept of privacy. *University of Baltimore School of Law Journal of Media Law & Ethics*, *6*, 4–21.

Fair Credit Reporting Act. 15 U.S. Code § 1681 et seq. (1970).

Furst, R. T., & Evans, D. N. (2017). Renting apartments to felons: Variations in real estate agent decisions due to stigma. *Deviant Behavior*, *38*(6), 698–708. https://doi.org/10.1080/01639625.2016.1197635

Greenspan, O. M., & DeBacco, D. A. (2014). *Survey of state criminal history information systems, 2012* (NCJ 244563). Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.

Grigoropoulou, N., & Small, M. L. (2022). The data revolution in social science needs qualitative research. *Nature Human Behaviour*, *6*(7), 904–906.

Harron, K., Dibben, C., Boyd, J., Hjern, A., Azimaee, M., Barreto, M. L., & Goldstein, H. (2017). Challenges in administrative data linkage for research. *Big Data & Society*, *4*(2), 2053951717745678. https://doi.org/10.1177/2053951717745678

Henderson v. Source for Public Data. 540 F. Supp. 3d 539 (Dist. Court 2021).

Horn, A. E. (2019). Wrongful collateral consequences. *George Washington Law Review*, *87*(2), 315–372.

HR Research Institute. (2020). *Background screening: Trends and uses in today's global economy*. https://pubs.thepbsa.org/pub.cfm?id=459B8AB7-0CEA-625E-0911-A4A089DE5118

Husak, D. (2023). Six Questions About Overcriminalization. *Annual Review of Criminology*, *6*, 265–284.

Jacobs, J., & Crepet, T. (2008). The expanding scope, use, and availability of criminal records. *New York University Journal of Legislation and Public Policy*, *11*(2), 177–214.

Jacobs, J. B. (2015). *The eternal criminal record*. Harvard University Press.

Kirk, D. S., Barnes, G. C., Hyatt, J. M., & Kearley, B. W. (2018). The impact of residential change and housing stability on recidivism: pilot results from the Maryland Opportunities through Vouchers Experiment (MOVE). *Journal of Experimental Criminology*, *14*, 213–226.

Kiviat, B. (2017). The art of deciding with data: Evidence from how employers translate credit reports into hiring decisions. *Socio-Economic Review*, *17*(2), 283–309. https://doi.org/10.1093/ser/mwx030

Kluemper, D. H. (2013). Social network screening: Pitfalls, possibilities, and parallels in employment selection. In *Social media in human resources management* (Vol. 12, pp. 1–21). Emerald Group Publishing Limited. https://www.emerald.com/insight/content/doi/10.1108/S1877-6361(2013)0000012005/full/html

Koops, B.-J. (2021). The concept of function creep. *Law, Innovation and Technology*, *13*(1), 29–56. https://doi.org/10.1080/17579961.2021.1898299

Kurlychek, M. C., Bushway, S., & Denver, M. (2019). Defensible decisions: Balancing employer and prospective employee rights in an era of criminal background checks. *Equality, Diversity and Inclusion: An International Journal*, *38*(5), 529–546. https://doi.org/10.1108/EDI-09-2018-0176

Kurlychek, M. C., & Johnson, B. D. (2019). Cumulative disadvantage in the American criminal justice system. *Annual Review of Criminology*, *2*, 291–319.

Lageson, S. E. (2020). *Digital punishment: Privacy, stigma & the harms of data-driven criminal justice*. Oxford University Press.

Lageson, S. E. (2023). Criminally bad data: Inaccurate Criminal Records, Data Brokers, and Algorithmic Injustice. *University of Illinois Law Review*, *5*, 1771–1810.

Lageson, S. E., Denver, M., & Pickett, J. T. (2019). Privatizing criminal stigma: Experience, intergroup contact, and public views about publicizing arrest records. *Punishment & Society*, *21*(3), 315–341.

Lageson, S. E., Vuolo, M., & Uggen, C. (2015). Legal ambiguity in managerial assessments of criminal records. *Law & Social Inquiry*, *40*(1), 175–204.

Laudon, K. C. (1986). Data quality and due process in large interorganizational record systems. *Communications of the ACM*, *29*(1), 4–11.

Leasure, P. (2019). Securing private housing with a criminal record. *Journal of Offender Rehabilitation*, *58*(1), 30–49. https://doi.org/10.1080/10509674.2018.1549182

Lindsay, S. L. (2022). Damned if you do, damned if you don't: How formerly incarcerated men navigate the labor market with prison credentials. *Criminology*, *60*(3), 455–479. https://doi.org/10.1111/1745-9125.12307

Logan, W. A., & Ferguson, A. G. (2016). Policing criminal justice data. *Minnesota Law Review*, *101*, 541–616.

Love, M., & Poggenklass, R. (2023). *The frontiers of dignity: Clean slate and other criminal record reforms.* Collateral Consequences Resource Center. https://ccresourcecenter.org/wp-content/uploads/2023/01/22-report.1.9.22.rev6_.pdf

Lynch, J. (2018). Not even our own facts: Criminology in the era of big data. *Criminology*, *56*(3), 437–454.

Maurer, R. (2018) "Ban the box" turns 20: What employers need to know. *SHRM*. https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/ban-the-box-turns-20-what-employers-need-to-know.aspx

McElhattan, D. (2019). "$40 to make sure": Background check laws and the endogenous construction of criminal risk. In A. Sarat (Ed.), *After imprisonment: Special issue* (1st ed.). Emerald Publishing.

McElhattan, D. (2021). Punitive ambiguity: State-level criminal record data quality in the era of widespread background screening. *Punishment & Society*, *24*(3), 367–386. https://doi.org/10.1177/1462474521989502

McElhattan, D. (2022a). The exception as the rule: Negligent hiring liability, structured uncertainty, and the rise of criminal background checks in the United States. *Law & Social Inquiry*, *47*(1), 132–161.

McElhattan, D. (2022b). The proliferation of criminal background check laws in the United States. *American Journal of Sociology*, *127*(4), 1037–1093.

Miller, R. J., & Stuart, F. (2017). Carceral citizenship: Race, rights and responsibility in the age of mass supervision. *Theoretical Criminology*, *21*(4), 532–548. https://doi.org/10.1177/1362480617731203

Myrick, A. (2013). Facing your criminal record: Expungement and the collateral problem of wrongfully represented self. *Law & Society Review*, *47*(1), 73–104.

National Archive of Criminal Justice Data. (2022). *National Corrections Reporting Program resource guide*. https://www.icpsr.umich.edu/web/pages/NACJD/guides/ncrp.html

National Consumer Law Center. (2019). Broken records redux: How errors by criminal background check companies continue to harm consumers seeking jobs and housing. https://www.nclc.org/wp-content/uploads/2022/09/report-broken-records-redux.pdf

National Consumer Law Center. (2023). Advocates support CFPB commitment to closing accountability loopholes for data brokers selling dossiers on Americans in secret. NCLC Press release, August 15, 2023.

https://www.nclc.org/advocates-support-cfpb-commitment-to-closing-accountability-loopholes-for-data-brokers-selling-dossiers-on-americans-in-secret/

National Employment Law Project. (2014). The "Wild West" of Employment Background Checks. https://www.nelp.org/wp-content/uploads/2015/03/Wild-West-Employment-Background-Checks-Reform-Agenda.pdf

New Jersey State Police. (2022). *Identification & information technology section*. https://nj.gov/njsp/division/administration/identification-information-technology.shtml

Pager, D. (2003). The mark of a criminal record. *American Journal of Sociology*, *108*(5), 937–975. https://doi.org/10.1086/374403

Pine, K. H., & Bossen, C. (2020). Good organizational reasons for better medical records: The data work of clinical documentation integrity specialists. *Big Data & Society*, *7*(2), 2053951720965616. Advance online publication. https://doi.org/10.1177/2053951720965616

Professional Background Screening Association. (2021). *Background screening: Trends in the U.S. and abroad*. https://pubs.thepbsa.org/pub.cfm?id=FB36B937-C9D5-A941-7720-4047386F38B0

Raphael, S. (2021). The intended and unintended consequences of ban the box. *Annual Review of Criminology*, *4*(1), 191–207. https://doi.org/10.1146/annurev-criminol-061020-022137

Roberts, J. (2015). Expunging America's rap sheet in the information age. *Wisconsin Law Review*, *2*, 321.

Sandoval, J. R., & Lageson, S. E. (2022). Patchwork disclosure: Divergent public access and personal privacy across criminal record disclosure policy in the United States. *Law & Policy*, *44*(3), 255–277. https://doi.org/10.1111/lapo.12193

SEARCH. (2001). *Report of the National Task Force on Privacy, Technology, and Criminal Justice Information* (NCJ 187669). Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.

Shapira, Z. (1996). Introduction and overview. In Z. Shapira (Ed.), *Organizational Decision Making* (Cambridge Series on Judgment and Decision Making, pp. 3–8). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511584169.003

Sherman, J. (2021). Data brokers and sensitive data on U.S. individuals: Threats to American civil rights, national security, and democracy. *Duke Sanford Cyber Policy Program*, 9. https://techpolicy.sanford.duke.edu/wp-content/uploads/sites/4/2021/08/Data-Brokers-and-Sensitive-Data-on-US-Individuals-Sherman-2021.pdf

Sprague v. Salisbury Bank & Tr. Co., 969 F.3d 95 (2d Cir. 2020).

Stewart, R., & Uggen, C. (2020). Criminal records and college admissions: A modified experimental audit. *Criminology*, *58*(1), 156–188. https://doi.org/10.1111/1745-9125.12229

Stoll, M. A., & Bushway, S. D. (2008). The effect of criminal background checks on hiring ex-offenders criminal background checks. *Criminology & Public Policy*, *7*(3), 371–404.

Tahamont, S., Jelveh, Z., Chalfin, A., Yan, S., & Hansen, B. (2021). Dude, where's my treatment effect? Errors in administrative data linking and the destruction of statistical power in randomized experiments. *Journal of Quantitative Criminology*, *37*(3), 715–749.

Tahamont, S., Jelveh, Z., McNeill, M., Yan, S., Chalfin, A., & Hansen, B. (2023). No ground truth? No problem: Improving administrative data linking using active learning and a little bit of guile. *PLOS ONE*, *18*(4), e0283811.

Thacher, D. (2008). The rise of criminal background screening in rental housing. *Law & Social Inquiry*, *33*(1), 5–30.

TransUnion LLC v. Ramirez. 141 S. Ct. 2190, 2197 (2021).

Uggen, C. (1999). Ex-offenders and the conformist alternative: A job quality model of work and crime. *Social Problems*, *46*, 127–151.

Uggen, C., & Blahnik, L. (2014). *The increasing stickiness of public labels*. Taylor & Francis.

Uggen, C., & Stewart, R. (2015). Piling on: Collateral consequences and community supervision. *Minnesota Law Review*, *99*, 1871–1910.

Uggen, C., Vuolo, M., Lageson, S. E., Ruhland, E., & Whitham, H. K. (2014). The edge of stigma: An experimental audit of the effects of low-level criminal records on employment. *Criminology*, *52*(4), 627–654.

United States Bureau of Consumer Financial Protection. (2021). *Consumer response annual report: January 1–December 31, 2020*. https://files.consumerfinance.gov/f/documents/cfpb_2020-consumer-response-annual-report_03-2021.pdf

United States Bureau of Consumer Financial Protection. (2022, July 11). *The Fair Credit Reporting Act's limited preemption of state laws* (Interpretive rule 87 FR 41042).

United States Government Accountability Office. (2015, February). Criminal history records: Additional actions could enhance the completeness of records used for employment-related background checks (GAO-15-62 Report to Congressional Requesters). https://www.gao.gov/assets/gao-15-162.pdf

United States Office of the Attorney General. (2006). The Attorney General's report on criminal history background checks. https://bjs.ojp.gov/sites/g/files/xyckuh236/files/media/document/ag_bgchecks_report.pdf

United States Department of Justice v. Reporters Committee for Freedom of the Press. 489 U.S. 749 (1989).

Warr, M. (1998). Life-course transitions and desistance from crime. *Criminology*, *36*(2), 183–216.

Wayne, L. D. (2012). The data-broker threat: Proposing federal legislation to protect post-expungement privacy. *The Journal of Criminal Law and Criminology (1973-)*, *102*(1), 253–282.

Weiskopf, N. G., Rusanov, A., & Weng, C. (2013). Sick patients have more data: the non-random completeness of electronic health records. In *AMIA annual symposium proceedings* (Vol. 2013, p. 1472). American Medical Informatics Association.

How to cite this article: Lageson, S., & Stewart, R. (2024). The problem with criminal records: Discrepancies between state reports and private-sector background checks. *Criminology*, *62,* 5–34. https://doi.org/10.1111/1745-9125.12359

AUTHOR BIOGRAPHIES

Sarah Lageson is an associate professor at the School of Criminal Justice, Rutgers University—Newark. Her research examines technology, law, and society.

Robert Stewart is an assistant professor of criminology and criminal justice at the University of Maryland—College Park. His research focuses on the social, political, and collateral consequences of criminal legal involvement.