The Wrongful Conviction Law Review

Volume 5, Issue 1

Editorial Board

Volume 5, Issue 1 **Cited as (2024) 5:1 Wrongful Conv L Rev**

Table of Contents

Introduction

Broadening the Community, Strengthening the Movement: An Introduction to the Innocence Project "Just Data" 2024 Scholarship Issue 1-7
Vanessa Meterko, Jaime S. Henderson, Ngozi Ndulue, Tebah Browne, Valena Beety

Articles

Wrongful Convictions with Prison Sentences in Spain: Exoneree Characteristics, Crime Types, and Contributing Factors 8-26
Nuria Sánchez, Guadalupe Blanco-Velasco, Linda M. Geven, Jaume Masip, Antonio L. Manzanero

Wrongful Conviction in Texas: 'Sex Assaults', False Guilty Pleas, Stranger Rape with Misidentification, and Drug Offenses 27-41
Matthew Barry Johnson, Janquel D. Acevedo

"Not Scientific" to Whom? Laypeople Misjudge Manner of Death Determinations as Scientific and Definitive 42-58
Jeff Kukucka, Oyinlola Famulegun

In Pursuit of Innocence: A Study of Race and Ethnicity Differences in Time-to-Exoneration 59-79
Virginia Elizabeth Braden

A Computational Decision-Tree Approach to Inform Post-Conviction Intake Decisions 80-102
Kalina Kostyszyn, Carl J. Wiedemann, Rosa M. Bermejo, Amie Paige, Kristen W. Kalb-DellaRatta, Susan E. Brennan

Innocence Discovery Lab - Harnessing Large Language Models to Surface Data Buried in Wrongful Conviction Case Documents 103-125
Ayyub Ibrahim, Huy Dao, Tarak Shah

When Pleas Precede Evidence: Using Bayesian Analyses to Establish the Importance of a *Reasonable* Standard for Evidence Prior to Plea Offers 126-145
Miko M. Wilford, Joseph E. Gonzale, Annmarie Khairalla

Broadening the Community, Strengthening the Movement: An Introduction to the Innocence Project "Just Data" 2024 Scholarship Issue

Vanessa Meterko
Research Manager
Innocence Project

Jaime S. Henderson
Director of Data Science and Research
Innocence Project

Ngozi Ndulue
Special Advisor on Race and Wrongful Convictions
Innocence Project

Tebah Browne
Forensic Science Policy Specialist
Innocence Project

Valena Beety
Robert H. McKinney Professor of Law
Indiana University Bloomington

This special issue of *The Wrongful Conviction Law Review* features a collection of seven research papers presented at the Innocence Project's 2023 Just Data: Advancing the Innocence Movement conference. This annual event, organized by the Innocence Project and partners in the larger Innocence Network, brings together a diverse group of scholars and advocates, including those who have been wrongfully convicted themselves, to share the latest social science, discuss emerging issues, and make connections to inform our collective work and promote new applied research. Social science research, often conducted quietly and behind the scenes, is critical to data-driven litigation and policy work in courthouses and statehouses around the world. It informs social work and public education about wrongful convictions. Rigorous, robust research is an essential tool in the mission to free innocent people, prevent future wrongful convictions, and create fair, compassionate, and equitable systems of justice for everyone.

This year's scholarship builds directly on previous years (e.g., Kostyszyn and colleagues [this issue] citing Berube et al., 2023). Several broad themes emerged in this collection: place-based research, with studies uncovering and exploring wrongful convictions in Spain and across Texas; the ever-present role of race and ethnicity at points throughout the criminal legal system – from expert testimonies to exoneration trajectories; the unique power of interdisciplinary research in leveraging data to improve case evaluation and accountability in the system; and the persistent problems in the plea-bargaining process.

To complement the novel research on these topics, Just Data: Advancing the Innocence Movement 2023 featured a keynote address on race and bias by Stanford University Professor and

Innocence Project Board member Dr. Jennifer Eberhardt, and insights from Innocence Network representative, Valena Beety, as well as three exonerees working in the field: podcaster, writer, filmmaker, and founder of House of Renewed Hope, Christopher Scott; lawyer and consultant, Chris Ochoa; and Innocence Project Reentry Coach, Rodney Roberts. Additionally, leading thinkers Cierra Robson (Ida B. Wells Just Data Lab) and Mitha Nandagopalan (Innocence Project) joined moderator Ngozi Ndulue (Innocence Project) for a conversation about technology, how new frontiers (e.g., artificial intelligence) collide with race and racism, and implications for innocence work and for society. Together, this is how our knowledge grows and a movement advances.

To begin, two studies explore how wrongful convictions are influenced by location, showcasing place-based research in both the United States and abroad. These articles focus on the contrasting examples of wrongful convictions in Spain and in Texas.

Dr. Nuria Sánchez presented as lead author on *Wrongful Convictions with Prison Sentences in Spain: Exoneree Characteristics, Crime Types and Contributing Factors*, on behalf of her co-authors Guadalupe Blanco-Velasco of Ontario, Canada; Linda Geven, of Leiden, Netherlands; Jaume Masip of Salamanca, Spain; and Antonio L. Manzenro, of Madrid, Spain. This international team of researchers coded decisions made by the Spanish Supreme Court from 1996 to 2022, identifying 89 wrongful convictions.

Their research revealed that 92% of wrongly convicted people in Spain were male, and the majority had a prior conviction. In great contrast to the United States, 85% of the sentences were less than 5 years. Professional misconduct was the leading contributing factor for these wrongful convictions, followed by the same factors common in the United States: faulty forensic science, eyewitness misidentification, false confessions, and false testimony. This timely piece is being published just as the European Registry of Exonerations (EUREX) is launched, compiling data specific to European exonerations.

Dr. Matthew Barry Johnson presented his research on wrongful convictions in Texas, co-authored with Janquel D. Acevedo. They analyzed data from the National Registry of Exonerations (NRE) and found that the state of Texas has the most sexual assault wrongful convictions. In over half of the Texas cases, exculpatory DNA ultimately led to the exoneration, and false guilty pleas were a major contributing factor.

Their research also documented a prior drug analysis problem in Houston, where people were wrongly convicted of drug offenses after legal substances were misidentified and then misrepresented as illegal controlled substances. In over 97% of these cases, innocent defendants pled guilty. This article crucially emphasizes how, under pressure, a guilty plea can become the seemingly best action – even when the defendant is innocent of any crime. (Later in this issue Dr. Miko Wilford and colleagues present an in-depth analysis of the risk of false guilty pleas.)

Both of these papers document the importance of national and international registries, such as the U.S. NRE and the newly launched EUREX. These tools provide researchers with stepping stones to learn what leads to wrongful convictions, and strategies for changing our systems.

Christopher Scott, a leader in challenging wrongful convictions, commented on both papers. Mr. Scott was himself wrongly convicted in Dallas, Texas, and as an exoneree he works to free other wrongly convicted people. Scott shared the impact of race and wrongful convictions, where a majority of exonerees in Dallas County – like himself – were wrongly convicted by an all white jury. He believes the lack of sufficient compensation for jurors leads to fewer people of color serving on juries. Scott's real life experience parallels Dr. Johnson's research in the field: he was wrongly convicted by an eyewitness misidentification, and research demonstrates the accuracy issues with cross-racial identifications as well as the importance of double-blind presentation in line ups like the one used against Scott. Prosecutors failed to disclose all the evidence in Scott's case. Now, research on discovery has pushed for change and greater disclosure, including the Michael Morton Act in Texas named after another exoneree. Scott's insights share how the practical research of Dr. Johnson and Dr. Sánchez not only proves these issues occur, but can help change laws and assist innocent people fighting their own cases.

People of color are more likely to be wrongly convicted (Gross et al., 2022), and numerous factors such as jury composition and eyewitness misidentification can adversely impact case outcomes as evidenced by the experience of Christopher Scott and the work of Dr. Johnson and Janquel D. Acevedo. The structural inequities of the criminal legal system and wrongful convictions are well documented. Two studies in this issue add to the already impressive accumulation of research on disparate treatment and outcomes.

In an attempt to investigate more nuanced biases at play in the trial process, Dr. Jeff Kukucka and Oyinlola Famulegan explored varying the ethnicity and socioeconomic status of a woman on trial, the perceived certainty of medical examiner testimony on determinations of death, and whether these variables influenced case verdicts. The study revealed incongruence between what jurors determined to be scientific and what experts deemed scientific regarding determinations of death. More specifically, jurors believed that determinations of death were scientific evidence and if the death was determined accidental, it was not as convincing if the woman on trial was Latina and of low socioeconomic status as compared to an affluent white woman.

Along with evidence that legal and extralegal factors influence verdicts in a disparate manner (Kukucka & Famulegan [this issue]), scholars have found they also affect the time to exoneration for innocent people who have been wrongly convicted. The first of its kind using survival analysis, Dr. Virginia Braden's research revealed significant differences in time to exoneration across race/ethnicity. Analyzing a national sample of exonerees, the study found that Black people experienced significantly longer times to exoneration than their white and Hispanic/Latino/a/x counterparts. Further, factors such as inadequate legal defense and age resulted in longer times to exoneration for Hispanic/Latino/a/x people compared to white people. Where identifiable variables are responsible for the markedly longer times to exoneration for innocent Black people (e.g., inadequate defense, official misconduct, etc.) these disparities can be addressed by transparent changes in policy and practice.

The Just Data: Advancing the Innocence Movement 2023 conference further explored the role of race in wrongful convictions by examining the ways that new technology is fueling racially discriminatory investigation and outcomes in the criminal legal system. In "Digital Dilemmas:

Exploring the Intersection of Technology, Race, and Wrongful Conviction," panelists Cierra Robson, the Associate Director of the Ida B. Wells Just Data Lab at Princeton University, and Mitha Nandagopalan, Staff Attorney in the Innocence Project's Strategic Litigation Department, explained how technology shapes the criminal legal system. Robson discussed how technology that relies on historical patterns to predict future crime and target enforcement continues to entrench racial bias that has underlied the criminal legal system since its inception. She also shared how today's use of technology is actively reshaping policing, including the net widening effect of "proactive policing" that draws more people into law enforcement's orbit based on predictive algorithms. Robson observed that all of these practices "have implications for wrongful conviction primarily because many of them put more people in the process of police activity in ways that are not substantiated by an actual threat, but instead a risk of a threat."

Innocence Project Staff Attorney, Mitha Nandagopalan, discussed the way that technologies like facial recognition, ShotSpotter, and automated license plate readers could bring entire neighborhoods under surveillance, placing people on law enforcement's radar who would not otherwise have police contact. Nandagopalan noted that "algorithms, data, and technology reflect…the priorities, preferences, and often the prejudices of the people who generate them." Using ShotSpotter as an example, they emphasized the need to examine the accuracy of new technology and whether there is evidence that it is actually making people safer.

Both panelists saw both promise and challenges in research about how new technologies are transforming policing. Robson praised the excellent qualitative research being done in the field, and discussed her research on lawyers' use of risk assessments to advocate for their clients. Nandagopalan noted the importance of examining the intersection between human decision making and emerging technology, recognizing the transparency challenges with accessing policing-related data. They also underlined the importance of ensuring that any research is accountable to the communities affected by the practices being studied. Without including affected communities in research planning and execution, "the research itself can run the risk of just reinforcing the same inequities that the technology itself is furthering and … become a tool for that larger trend and historical pattern of racial bias."

Turning to the day-to-day work of addressing wrongful convictions, innocence organizations receive hundreds of intake applications yearly and face difficulties processing all of them. With the advancement of technology, algorithms or artificial intelligence (AI) systems may be used to facilitate and possibly accelerate the process. In the "Data to Deliverance: Leveraging Research to Inform Post-conviction Work" session of the conference, two studies were presented that demonstrated how algorithms can be utilized to aid the intake process of innocence organizations.

Kalina Kostyszyn and colleagues introduced a technique that uses decision trees to assist the intake process. They examined 3,284 exoneration cases to determine patterns or features associated with successful cases. This was done using two methods: the Berube et al. (2023) latent class analysis (LCA) method and decision trees. Using the LCA method, they found that a four-class model provided the best statistical fit, corroborating Berube et al. (2023) while using a larger data set. The four predicted classes are as follows: intentional errors, witness mistakes, investigative corruption, and failures to investigate. Next, decision trees were used to further

analyze the four classes. Using the "Six Canonical Factors," decision trees were used to predict classifications from the LCA methods (6-factor model). Additional trends within the four classes were examined by looking beyond the canonical factors (extended model). Several trees were generated to examine the data. The accuracy of each branch and the accuracy and precision of the entire tree were determined. Each model was run twice, using 10-fold cross-validation and 75/25 split. The 6-factor runs displayed the strongest trends.

Ayyub Ibrahim and colleagues discussed the Innocence Discovery Initiative, a collaborative effort between the Innocence Project New Orleans, Public Data Works, and the Human Rights Data Analysis Group, that provides an advanced methodology for reviewing potential wrongful conviction case files. The initiative consists of the following five stages:

1. The compilation of a CSV index of metadata
2. The evaluation of filenames to identify relevant case files
3. The utilization of FastAI to train an image classifier
4. The extraction of information from the files identified in the second phase
5. Cross-referencing of the extracted officer names and titles with the Louisiana Law Enforcement Accountability Database

Ultimately, this methodology can be used to assist in the identification of potential wrongful convictions.

Finally, Dr. Miko Wilford, Dr. Joseph E. Gonzales, and Dr. Annmarie Khairalla introduced their analysis of the plea-bargaining system in the United States. They noted that in most jurisdictions, prosecutors do not have to establish a reasonable basis for guilt before offering plea deals. In theory and practice, this means that the State could have very little concrete evidence against a person at the time they are offered a plea deal. Given the pressures (e.g., pretrial detention) and consequences (e.g., the "trial penalty"), there is a real risk of innocent people pleading guilty during this process. As a recent report from the American Bar Association's Plea Bargain Task Force articulates, "the state may induce the defendant to plead guilty with incentives that make it irrational for even an innocent person to turn down the deal" (p. 15). Indeed, nearly a quarter (839 of 3,466) of the wrongful convictions captured in the NRE database to date involved false guilty pleas.

During this year's conference, Rodney Roberts offered a description highlighting the dilemma that innocent people face based on his personal experience. He shared that the choice between taking your chances at trial and facing a life sentence or falsely pleading guilty to guarantee freedom in a few years feels like a choice "between lucifer and satan." Chris Ochoa described being threatened with the death penalty if he went to trial and the way in which the stress of his pre-trial detention impacted his mother's health. When she had a stroke, Mr. Ochoa made "the hardest decision I have ever had to make in my life; I had to plead guilty to something I know I didn't do…there was no way out."

Recognizing that everyone is legally innocent until proven guilty, Wilford and colleagues took a statistical approach to investigating the risk of wrongful conviction via guilty plea by those who are factually innocent and factually guilty. They used Bayesian analyses, which they note are

uniquely suited to illuminate the impact of prior probability of guilt on the informativeness of a particular outcome (i.e., a guilty plea). As the authors expected, results revealed that when plea offers are accepted at lower prior probabilities of guilt, the probability that a plea is actually false is significantly higher than when prior probabilities of guilt are higher. Thus, demonstrating the risks of plea offers that precede concrete evidence, the research team concluded that until prosecutors are 85% confident in the accused's guilt, one in 20 guilty pleas will be false. They conclude by offering policy recommendations for open file discovery, which allows everyone involved to fully evaluate the evidence in a criminal case and, hopefully, avoid the impossible choices that Rodney Roberts and Chris Ochoa were forced to make.

This collection of studies underscores the pertinence of ongoing research in the realm of wrongful convictions. The articles are a staunch reminder that wrongful convictions are a global issue, structural racism profoundly affects people of color who are wrongly convicted, we must be mindful of and act against the implicit and explicit power of pleas for vulnerable people ensnared in the legal system, and we must deliberately monitor the use of technology and statistics to expand the number of people that can be assisted while carefully balancing the biases that can impact those seeking help.

Although consistent themes have been identified in known innocence cases, there is still action to be taken and much to learn from social science scholars. The authors in this issue have provided numerous recommendations for policy and future studies. Additionally, innocence practitioners have identified numerous research questions and unaddressed topics that can be found on the Innocence Project's website (i.e., A Call to Action, Innocence Project, n.d.) and implore the research community to continue endeavors that will continue to influence policy and practice. We encourage researchers to contact the Innocence Project and the Innocence Network for guidance and collaboration on work that will continue to support the efforts of the innocence movement and make meaningful progress to eradicate wrongful convictions.

References

American Bar Association Criminal Justice Section. (2023.) 2023 Plea Bargain Task Force Report. https://www.americanbar.org/content/dam/aba/publications/criminaljustice/plea-bargain-tf-report.pdf

Berube, R., Wilford, M. M., Redlich A. D., & Wang, Y. (2023). Identifying patterns across the six canonical factors underlying wrongful convictions. The Wrongful Conviction Law Review, 3(3), 166-195. https://doi.org/10.29173/wclawr82

Gross, S. R., Possley, M. Otterbourg, K., Stephens, K., Weinstock Paredes, J., & O'Brien, B. (2022). Race and wrongful convictions in the United States. National Registry of Exonerations (NRE). https://www.law.umich.edu/special/exoneration/Documents/Race%20Report%20Preview.pdf

Innocence Project (n.d.). A Call to Action: Applied Research Questions. https://innocenceproject.org/wp-content/uploads/2023/06/A-Call-to-Action.pdf

Innocence Project. (2019, April 1). Breakthrough Discovery Law Passes, Repealing State's "Blindfold Law," Leading to Greater Transparency and Fewer Wrongful Convictions. https://innocenceproject.org/breakthrough-discovery-law-passes-repealing-states-blindfold-law-leading-to-greater-transparency-and-fewer-wrongful-convictions/

National Association of Criminal Defense Lawyers (NACDL). (2018.) The Trial Penalty: The Sixth Amendment Right to Trial on the Verge of Extinction and How to Save It. https://www.nacdl.org/getattachment/95b7f0f5-90df-4f9f-9115-520b3f58036a/the-trial-penalty-the-sixth-amendment-right-to-trial-on-the-verge-of-extinction-and-how-to-save-it.pdf

National Registry of Exonerations (NRE). (n.d.). Exoneration detail list. https://www.law.umich.edu/special/exoneration/Pages/detaillist.aspx?View=%7BFAF6EDDB-5A68-4F8F-8A52-2C61F5BF9EA7%7D&FilterField1=Group&FilterValue1=P

Wrongful Convictions with Prison Sentences in Spain: Exoneree Characteristics, Crime Types, and Contributing Factors

Nuria Sánchez
Universidad de Salamanca
Spain

Guadalupe Blanco-Velasco
Ontario Tech University
Canada

Linda M. Geven
Leiden University
Netherlands

Jaume Masip
Universidad de Salamanca
Spain

Antonio L. Manzanero
Universidad Complutense de Madrid
Spain

Researchers worldwide have extensively explored the factors contributing to wrongful convictions and the characteristics of individuals affected by these miscarriages of justice for over a century. Despite these global efforts, limited research has been conducted on this issue in Spain. This study seeks to address this gap. We trained coders to assess available review judgments issued by the Spanish Supreme Court from 1996 to 2022. We identified 88 cases of individuals wrongly sentenced to deprivation of liberty. Our findings indicated that 92% of those wrongfully convicted were male, with the majority having a prior criminal record. Most exonerations involved minor crimes, and 85% of individuals were sentenced to less than 4.5 years of deprivation of liberty. Professional misconduct emerged as the primary contributing factor, followed by the misapplication of forensic science, misidentifications, false testimonies, and false confessions. This project sheds light on wrongful convictions in Spain, emphasizing the need for comprehensive measures to address this issue. The current results have practical implications for justice professionals, policymakers, and legal practitioners. It is crucial to educate professionals in the judicial system on the causes of judicial errors, the biases that may influence them, and best practices to improve processes and reduce the occurrence of wrongful convictions.

Keywords: Wrongful convictions, Miscarriage of justice, Judicial errors, Prison, Exonerations.

I. Introduction
  A. Characteristics of Individuals who Have Been Wrongfully Convicted
  B. Types of Crimes with a Higher Incidence of Judicial Errors
  C. Factors Contributing to Wrongful Convictions
    a. Perjury or False Accusation
    b. Professional Misconduct
    c. Eyewitness Misidentification by Victims and Witnesses
    d. Misapplication of Forensic Science
    e. False Confessions
  D. Current Study
II. Method
  A. Sample
  B. Coding Procedure and Reliability
III. Results
  A. Characteristics of Wrongfully Convicted People
  B. Characteristics of Wrongful Conviction Cases
  C. Factors Contributing to Wrongful Convictions
IV. Discussion
  A. Rate of Wrongful Convictions

B. Gender
C. Citizenship
D. Prior Criminal Records
E. Crime Type
F. Contributing Factors
G. Limitations
V. References

I Introduction

From 1989 until November 2023, the United States National Registry of Exonerations (2023a) has documented over 3,400 instances of individuals being exonerated due to wrongful convictions for crimes they did not commit. In 2023, the European Registry of Exonerations (EUREX, 2023) was established with the primary objective of compiling data specific to European exonerations and analyzing the causes and consequences of wrongful convictions in European countries. Notably, in Spain, the General Council of the Judiciary initiated the recording of judicial errors in 2010. According to the 2022 annual report (General Council of the Judiciary, 2023), 33 cases have already been officially acknowledged.

Wrongful convictions have important personal, social, and institutional consequences. On a personal level, exonerees may face significant social, economic, health, and safety consequences (Schönteich, 2014). These individuals not only lose their freedom, but also face job loss and deterioration in their social relationships (Rogers-Degeer, 2023). Kukucka *et al*. (2022) concluded that 50% of the exonerees they studied reported clinically significant symptoms of posttraumatic stress disorder and/or depression. Additionally, after their exoneration, wrongfully convicted individuals also have difficulties finding a job (Clow, 2017) or accessing housing (Kukucka *et al*., 2021).

At the social level, exonerees' relatives often become secondary victims of wrongful convictions (Jenkins, 2013). Furthermore, the conviction of an innocent person often implies that the truly guilty individual remains free and may commit additional crimes. Altogether, public awareness of miscarriages of justice diminishes citizens' trust in the justice system (Norris & Mullinix, 2020). Moreover, once an innocent person is exonerated, that person has the right to receive economic compensation, which involves a significant expense for the state (Ortiz-Pradillo, 2023). Over a 10-year period, the Spanish Justice Administration spent close to three million euros to compensate individuals who had been victims of improper pretrial detention (Sánchez *et al*., 2017).

For over a century, researchers worldwide have conducted studies on the contributing factors of wrongful convictions and the characteristics of individuals who have been wrongfully convicted (Gould & Leo, 2010). This phenomenon has been examined in various countries, including Australia (Dioso-Villa, 2015), England and Wales (Helm, 2022), Italy (Lupária, 2020), the Netherlands (Brants, 2012), and the United States (Gross *et al.*, 2005; West & Meterko, 2015). Nevertheless, in Spain, research on this topic has been limited (Sánchez *et al*., 2017; Ortiz-Pradillo, 2023). The objectives of these international studies often include describing the characteristics of wrongfully convicted individuals, identifying the types of crimes in which errors are more prevalent, and determining the factors contributing to judicial mistakes.

A. Characteristics of Individuals Who Have Been Wrongfully Convicted

The National Registry of Exonerations (2023a) reported that 92% of the individuals who were exonerated were male. Consistent findings by various researchers (Dioso-Villa, 2015; Duce, 2015; Gross *et al.*, 2005) indicate that females typically constitute no more than 15% of exonerees. This gender distribution aligns with the broader demographics of the incarcerated population in the respective countries (Duce, 2015), suggesting an absence of gender bias in wrongful convictions.

In the United States, there is a notable overrepresentation of individuals from ethnic minorities among those wrongfully convicted (Gross *et al*., 2005). The majority of exonerees are Black, comprising 53%, while White exonerees constitute only 33% (National Registry of Exonerations, 2023a). Similarly, in other countries such as Australia, indigenous individuals make

up 14% of exonerees (Dioso-Villa, 2015). This racial bias is particularly pronounced in sexual assault cases, where the primary evidence often rests on the testimony of a White victim accusing a Black perpetrator. In these instances, misidentifications are likely, as same-race faces are recognized with greater accuracy than cross-race faces (e.g., Katzman & Kovera, 2023; Meissner & Brigham, 2001; Smith & Hattery, 2011).

Prior meta-analytical research has also indicated that the probability of a guilty verdict is heightened when jurors possess information about an individual's criminal history (Devine & Caughlin, 2014). Similar findings were observed by Gould *et al.* (2014) in their analysis of 460 cases of wrongful convictions and accusations. They concluded that individuals with a criminal record face a higher likelihood of being wrongfully convicted compared to those without such a background. Gould *et al*. (2014) proposed two key explanations for this bias: (a) the practice of law enforcement officers of showing pictures of individuals with prior criminal records to victims or eyewitnesses, which increases the likelihood of these individuals being selected in a photo array; and (b) the potential for police and judges to give greater attention to individuals with criminal records, possibly leading to the oversight of exculpatory evidence, especially if the victim has previously identified them as the perpetrator.

B. Types of Crimes with a Higher Incidence of Judicial Errors

There is an ongoing debate on whether wrongful convictions occur more frequently for certain types of offenses than for others (Gould & Leo, 2010). For instance, are serious crimes more prone to errors than minor offenses? Gross *et al.* (2005) conducted an examination of 340 exonerations in the United States spanning from 1989 to 2003. Their findings indicate that 95% of these exonerations were linked to rape and murder cases, which constitute only 2% of the total convictions for serious crimes in the country (Gross, 2008). However, the actual frequency of judicial errors remains unknown, and it is possible that the identified cases represent only a fraction of the total occurrences (Leo & Gould, 2009). Minor offenses could experience a higher incidence of wrongful convictions, but limited resources allocated to their investigation (e.g., Burrows *et al*., 2005; Hunt *et al.*, 2019), the lower penalty severity, and fewer requests to re-open cases may lead to overlooking them (Gould & Leo, 2010). Additionally, the criteria for reopening a case often hinge on the availability of analyses of biological evidence, such as DNA testing. Such evidence is much more commonly found in cases of rape or murder compared to minor offenses like theft or threats (Garrett, 2017). Consequently, cases of minor offenses not only face resource constraints, but also may lack sufficient evidence to warrant reopening (Gould & Leo, 2010).

Given the significance and prevalence of judicial errors, particularly in serious offenses, numerous studies and reports have focused exclusively on cases involving imprisonment or the death penalty (Harmon & Lofquist, 2005). Some of these reports have specifically examined the duration wrongfully convicted individuals spend incarcerated in the United States. The Innocence Project (2023b) has calculated that exonerees, on average, spend 16 years in prison, while the National Registry of Exonerations (2021) has reported an average of nine years. In Australia, the average prison time for exonerees is lower at 4.5 years (Dioso-Villa, 2015).

C. Factors Contributing to Wrongful Convictions

Wrongful convictions can stem from various factors, with the primary contributing factors encompassing perjury or false accusations, mistaken identifications by victims and witnesses, misapplication of forensic science, false confessions, and professional misconduct (Innocence Project, 2023b; National Registry of Exonerations, 2023b). While a singular factor may account for errors in a particular case, more frequently, it is a combination of factors that leads to a wrongful conviction (Berube, *et al*. 2022; Yaroshefsky & Schaefer, 2014).

a. Perjury or False Accusation

According to the National Registry of Exonerations (2023b), perjury or false accusation accounts for the highest percentage of wrongful convictions, being present in 64% of the 3,421 cases registered to date in the United States. This factor refers to both jailhouse informants as well as witnesses and victims who provide false testimony (Gross *et al*., 2005). Jailhouse informants may be incentivized to falsely accuse an innocent person by providing incriminating information in exchange for benefits in their criminal trial (Natapoff, 2006). Victims, too, may have specific interests, such as financial compensation or harming the innocent person they are accusing.

However, discerning whether perjury exists requires a crucial distinction between deliberate lying by victims or witnesses and honest mistakes (e.g., Masip *et al*., 2004; Sporer, 2008). For instance, if a victim of a sexual assault makes a misidentification but genuinely believes she/he is pointing to the actual offender, this would be considered an error, not perjury.

b. Professional Misconduct

Professional misconduct is the second most prevalent factor for wrongful convictions in the United States. It explains, at least in part, 60% of these convictions (National Registry of Exonerations, 2023b). Legal practitioners acting improperly, whether deliberately or unintentionally (Gould & Leo, 2010), might contribute to the occurrence of judicial errors. While deliberate misconduct is infrequent, unintentional inappropriate behaviors arguably may happen more often. Given the inherent uncertainty in criminal investigations and the need for professionals to make multiple decisions during the investigation process, various biases and subjective influences can impact their decision-making. Ask and Fahsing (2019) examined the reasoning processes involved in criminal investigations. They concluded that these influences may play a role in several phases of the criminal investigation process: during the collection and processing of information, during hypothesis testing, and during the structuring and description of the criminal event. For instance, over a crime investigation, "tunnel vision" might become highly relevant. This happens when different actors in the system develop an initial belief (suspicion), cling to it, and solely seek to confirm that belief (Godsey, 2017). Consequently, professionals such as crime investigators may not gather information to test multiple plausible hypotheses. Instead, they only collect information that confirms their pre-existing beliefs, and reject any information creating cognitive dissonance (Findley & Scott, 2006). Tunnel vision can contribute to wrongful convictions as it hinders the objectivity of criminal investigations.

Inadequate legal representation can also fall under the category of professional malpractice (Huff, 2004). This extends beyond poor preparation or professional incompetence among attorneys. At times, individuals facing charges may be unable to afford a private attorney, and the excessive workload for public defenders, the lack of control systems, or insufficient incentives may lead to some public defenders not effectively performing their tasks (Duce, 2013).

c. Eyewitness Misidentification by Victims and Witnesses

Identifying an offender poses a challenge for human memory. According to the National Registry of Exonerations of the United States (2023b), misidentifications are the third most important factor contributing to wrongful convictions—ranking just below perjury and professional misconduct. These misidentifications, at least partially, account for 27% of exonerations.

The identification of a culprit depends on factors related to the incident, the victim or witness, and system or process variables (Wells, 1978). Regarding incident-related factors, circumstances such as insufficient lighting can result in challenges in perceiving critical information, consequently interfering with identification procedures. Concerning victim or witness variables, factors like the cross-race effect (Meissner & Brigham, 2001), personal prejudices, and stereotypes might influence person identification or bias the perception and interpretation of events. Finally, concerning system or process variables, issues such as the selection of lineup fillers, the prelineup instructions provided to witnesses, whether double-blind procedures are employed, and whether repeated identification attempts are conducted with the same witness and suspect, among others, may have an impact on (mis)identifications (Wells *et al*., 2020).

d. Misapplication of Forensic Science

False or misleading forensic evidence is the fourth most prevalent factor of wrongful convictions, explaining, at least in part, 25% of cases (National Registry of Exonerations, 2023b). Misapplication of forensic science involves the use of either invalid or unvalidated forensic methods, misleading expert testimony, or forensic experts' misconduct (see, e.g., Scott *et al*, 2014).

Forensic techniques, including fingerprint analysis, hair microscopy, and tissue analysis, require analysts to undertake pattern comparison and interpret results. It is essential to recognize that analysts, being human, are fallible. Despite the integration of technology in forensic analyses, both sample collection and the decision-making process are contingent upon human involvement.

The presence of prior expectations and biases has the potential to impede the impartiality of procedures and decisions (e.g., Dror, 2020; Dror *et al*., 2021; Herrero, 2021).

Several books and scientific papers have documented real cases where this factor played a role (e.g., Dror, *et al*., 2006; Morgan, 2023). One of these cases is the investigation of the March 11, 2004, train bombings in Madrid. Initially, the Spanish police were unable to find a match for a fingerprint they had collected from the crime scene. Consequently, they sought assistance from law-enforcement agencies from other countries through Interpol (EFE, 2006). FBI experts and an independent examiner in the United States initially matched the print to Brandon Mayfield, a Muslim-American convert living in Oregon. However, when examiners from the Spanish police compared the latent print from the crime scene with Mayfield's fingerprint, they determined the outcome was inconclusive. Later on, the Spanish police reported that the latent print belonged to a different individual. The FBI eventually acknowledged their error, conducting an analysis to identify the shortcomings in the Mayfield examination process (Stacey, 2004). Mayfield was released two weeks after his arrest (EFE, 2006). Dror *et al*. (2006) cited this case to illustrate how irrelevant and misleading contextual information may negatively impact experts' decisions.

e. False Confessions

False confessions refer to rich and elaborate self-incriminating statements made by an individual admitting to a crime they did not actually commit (Gudjonsson, 2018). According to the National Registry of Exonerations (2023b), this factor, at least partially, accounts for 13% of wrongful convictions in the United States.

Suspects may falsely confess for various reasons (e.g., Kassin, 2022; Kassin *et al.*, 2010). Inappropriate interview techniques employed by the police, involving psychological tactics such as coercion, threats, false promises, and maximization and minimization tactics (see Kassin *et al*., 2010), can compel innocent suspects to confess as a means of escaping an aversive situation (coerced-compliant false confession) or even lead them to believe they actually committed the crime (coerced-internalized false confession; see Kassin, 2022; Kassin *et al*., 2010). Additionally, certain dispositional risk factors, such as vulnerability (e.g., being a child or a person with an intellectual disability) or exhibiting high levels of suggestibility or compliance, also contribute to the phenomenon of false confessions (Gudjonsson, 2018; Kassin *et al.*, 2010). False or coerced confessions pose a significant danger to the system, as legal practitioners frequently presume the confession's veracity (e.g., Kassin *et al*., 2005). Such a belief may lead practitioners to deem independent corroboration unnecessary or to overlook contradictory evidence. Furthermore, research has shown that confessions have a greater impact on jurors than other types of evidence (Kassin & Neumann, 1997).

In a study conducted in Spain, over 80% of the 89 Civil Guard and 126 National Police investigators surveyed reported having obtained at least one false confession from a suspect. Approximately 20% of respondents indicated that at least one of those false confessions occurred under coercion (Schell-Leugers *et al.*, 2023).[1]

A separate, less common type of false confession is voluntary. Voluntary false confessions may occur without external pressure, and they may result from psychopathology, or from a rational process to obtain a benefit (e.g., public notoriety) or to protect someone else (Kassin, 2022). Aebi and Campistol (2013) analyzed 1394 Spanish news articles about voluntary false confessions motivated by intangible benefits. The authors categorized the cases into two separate groups: Social-topic-based confessions, which were aimed at promoting changes in criminal law (i.e., on topics such as euthanasia and abortion); and confessions aimed at protecting an individual or group. In the latter category, the authors reported cases involving individuals within the Roman ethnicity who falsely confessed to protect relatives, young people confessing to protect relatives or friends, and terrorists falsely confessing to acting alone to protect their criminal organization.

D. Current Study

This research is centered on wrongful conviction cases in Spain, providing a comprehensive description of both the exonerees and the cases involved. Specifically, we

[1] However, this study also revealed that coercive techniques are employed only rarely by the Spanish police (see Schell-Leugers *et al*., 2023).

examined exonerees' gender, citizenship, and criminal records, as well as the type of crimes and the duration of the prison sentences. We also examined contributing factors to these judicial errors, including perjury or false accusation, professional misconduct, eyewitness misidentification, misapplication of forensic science, and false confessions.

The criteria employed to determine which cases to include in the study adhere to the legal definition of the term "exonerated" as outlined by Findley (2010). Specifically, we included cases where an individual was wrongfully convicted in a final judgment and, subsequently, a later judgment by the Spanish Supreme Court annulled the initial judgment based on new evidence revealing a factual error committed in the first judgment.

The Spanish judicial system permits the reopening of a closed case only if new facts or circumstances about the case emerge. This refers to information that was unknown at the time of the conviction (Blanco-Velasco *et al*., 2023). In such situations, the convicted individual can file a "review appeal" to the Supreme Court—a process that is regulated by the Spanish Criminal Procedure Law (articles 954 to 961).

The law specifies certain requirements for a review appeal. Grounds for submission include false testimony, a convicting judgment based on documents subsequently declared to be false, a confession obtained through violence or coercion, judicial misconduct by judges or magistrates, or new facts or evidence that were unknown at the time of the initial conviction. Additionally, since 2015, if the European Court of Human Rights issues a judgment reporting a violation of those human rights outlined in the European Convention for the Protection of Human Rights and Fundamental Freedoms and its Protocols (e.g., the right to a fair trial), the convicted individual can file a review appeal.

II Method

A. Sample

The judgments considered for this study were sourced from the Spanish Supreme Court, and pertained to cases in which a review appeal was either considered or admitted. Data were collected from two legal databases, namely Aranzadi and Cendoj. The time frame for the selected judgments ranged from May 24th, 1996 (the inception date of the current Spanish Criminal Code) to December 31st, 2022.

The search term "Recurso de revisión" ("Review appeal") was used both in the Aranzadi and Cendoj databases. In the Aranzadi database, the term was used to explore the criminal jurisprudence thesaurus. For the Cendoj database, the search was filtered by the type of entity, specifically selecting "Tribunal Supremo" ("Supreme Court") and "Sala de lo penal" ("Criminal Division"). Given that review appeals are regulated by the Criminal Procedure Law, this law was chosen as the Legislation domain to specifically search for judgments citing this regulation. After eliminating duplicate judgments, 447 review judgments were initially retained for analysis. However, as detailed below, the focus was narrowed to cases involving a deprivation of liberty, resulting in 88 judgments.

To enhance the depth of information for each case, alongside the review judgments, an analysis of previous judgments for each case was undertaken whenever available (e.g., first instance, appeals).

B. Coding Procedure and Reliability

A manual was developed encompassing pertinent coding variables, drawing on insights from prior literature (see Table 1). These variables can be categorized into three broad clusters:

1) *Characteristics of the wrongfully convicted individual*: Gender, citizenship, and whether the exoneree had a prior criminal or police record.

2) *Characteristics of the case*: Type of offense for which the judgment was overturned, the imposed penalty, and the duration of the deprivation of liberty.

3) *Factors contributing to judicial errors*: False testimonies by victims or eyewitnesses, misconduct of legal professionals (including lawyers, judges, prosecutors, and the police), misidentifications by victims or eyewitnesses, misapplication of forensic science, false confessions, and other causes not fitting in the aforementioned categories.

Seven coders (two researchers and five research assistants) attended two training sessions conducted before coding different subsets of judgments. Out of the total 447 review judgments, 302 (67.56%) were independently assessed by two of these coders. The remaining 145 judgments (32.44%) were coded by one of the pre-trained researchers only.

The coding process unfolded across six rounds, facilitating the refinement and combination of coding criteria to minimize discrepancies. By avoiding coding a large number of judgments in a single round, the risk of a coder misunderstanding specific categories and introducing systematic errors was mitigated. Furthermore, coding in multiple rounds provided the opportunity to correct misinterpretations before completing the coding of all statements. Following each round, coders engaged in discussions to address discrepancies and reach a consensus.

A central aspect in this research was whether the petitioner claimed innocence or not. Following Gould and Leo (2010), and considering the Spanish legislation, a distinction must be made between factual innocence and procedural errors. Procedural errors do not necessarily imply the person's innocence. For instance, in Spain, an individual convicted twice for the same criminal event may file a criminal review appeal with the Supreme Court to vacate one of the judgments. In such cases, a procedural error occurred, yet the ground for the appeal is not factual innocence. Coders were tasked with determining, for each judgment, whether the petitioner claimed innocence, did not claim innocence, or if this information was not included in the judgment. Intercoder reliability for this variable was high (Krippendorff's Alpha averaged across all rounds, and weighted by the number of cases coded in each round, was .91). Only cases where the petitioner claimed innocence were retained for further consideration in this study.

The second crucial variable was whether the petitioner had received a sentence involving deprivation of liberty (such as a prison sentence, an arrest, being confined to a [closed] forensic facility, or any other penalty involving deprivation of liberty). Consequently, coders determined the presence or absence of each of a number of penalty types included in the Spanish legislation.[2] Thereafter, we grouped the assessments from each coder into three dichotomous variables to determine whether the punishment entailed deprivation of liberty (yes/no variable), whether the punishment did *not* involve a deprivation of liberty (e.g., community service or fines; yes/no variable), or whether the judgment did not provide enough information to assess this aspect (yes/no variable). The mean weighted (across rounds) Krippendorff's Alphas for these three variables were .90, .88, and .81, respectively. In total, after coders resolved their discrepancies and reached a consensus, there were 88 cases of individuals wrongfully convicted to deprivation of liberty. These were the cases subjected to our analysis.

As shown in Table 1, Krippendorff's Alpha was ≥ .80 (indicating near-perfect agreement; see Hughes, 2021; Landis & Koch, 1977) for all variables except *crimes against freedom* and *criminal record*. The relatively smaller frequencies for these two latter variables may explain their lower reliability (De Swert, 2012). Nevertheless, agreement for these variables was substantial (Hughes, 2021; Landis & Koch, 1977).

[2] Categories of coded penalties were prison, major prison, minor prison, major arrest, minor arrest, weekend arrest, deprivation of liberty (to categorize any type of deprivation of liberty not specified in the judgment), confinement to a (closed) forensic facility, community service, fines, court costs, ancillary penalties, other type of penalty, and the judgment does not mention the penalty. Some of these penalties (specifically, major prison, minor prison, major arrest, and minor arrest) are no longer included in the current Spanish Criminal Code, but in the previous one. However, note that we analyzed revision judgments. These judgments could resolve a case of a person convicted under the previous Criminal Code.

Table 1. Average Reliability Weighted by the Number of Cases Coded in Each Coding Round

Variable	Krippendorff's Alpha
Characteristics of the individual	
Gender (Male/Female/Other)	.99
Citizenship [a]	.87
Criminal record (Yes/No/Not indicated)	.72
Characteristics of the case [b]	
Duration of deprivation of liberty	.94
Minor offenses	.82
Homicide	1.00
Abortion	1.00
Crimes of assault	1.00
Injuries to the fetus	1.00
Genetic manipulation	1.00
Crimes against freedom	.75
Torture	.99
Human trafficking	.93
Sex crimes	.99
Relief omission	1.00
Crimes against privacy	1.00
Crimes against honor	1.00
Crimes against domestic relations	1.00
Crimes against property	.97
Crimes of illegal financing of political parties	1.00
Crimes against the public treasury and social security	1.00
Crimes against worker's rights	.93
Crimes against the rights of foreign citizens	1.00
Offenses related to land use, urban planning, and the environment	.95
Crimes against public safety	.98
Forgery	1.00
Crimes against the Public Administration	.95
Crimes against the Justice Administration	1.00
Crimes against the Spanish Constitution	1.00
Crimes against public order	.91
Crimes of treason, against the peace or independence of the State, and related to National Defense	1.00
Crimes against the international community	1.00
No information available on crime type	1.00
Contributing factors [c]	
False testimony of victims and witnesses	.96
Legal professionals' misconduct	.80
Misidentifications	.89
Misapplication of forensic science	.81
False confessions	.86
Other causes	.84
No information available about the cause	.92

[a] The coders identified the specific country of citizenship for each appellant using a drop-down menu. This menu also provided options such as "Not Indicated" (to be selected when the judgment included no information about citizenship) and "Foreign" (to be selected when the sentence mentioned the appellant's foreign status without indicating the specific country). Additionally, the coders identified cases of dual citizenship (i.e., Spanish and other). Due to data protection concerns, we refrain from disclosing individual citizenship details and re-coded this variable into four categories: *Spanish*, *Foreign* (encompassing both the "Foreign" category and individuals with citizenship other than Spanish), *Dual Citizenship*, and *Not Indicated*. [b] All variables under "Characteristics of the case" were continuous. *Duration of deprivation of liberty* was coded in terms of the number of days. The rest of variables related to crime type; since a single case could involve more than one crime attributed to the wrongfully convicted person, these variables were coded indicating the number of crimes leading to wrongful conviction(s) in each case. [c] Each of the contributing factors was coded as present or absent.

III Results

A. Characteristics of Wrongfully Convicted People

Most (92.05%) of the 88 individuals wrongfully convicted to deprivation of liberty in Spain were males. Only 7.95% of exonerees were female. The citizenship of the petitioner was not specified in 59.09% of the cases. Among the cases with available information, 52.78% of petitioners were Spanish, 44.44% were foreigners, and one person had Spanish citizenship but was born abroad. In 36.36% of the cases (32 individuals), no information was provided regarding whether the person had criminal or police records. In the remaining cases, 55.36% of wrongfully convicted individuals had a criminal or police record, while 44.64% did not. The record status could be either sealed or active at the time of the conviction.

B. Characteristics of Wrongful Conviction Cases

The various forms of deprivation of liberty encompassed a range from imprisonment to house arrest. In a negligible proportion of judgments (1.14%), no information was provided on the type of offense for which the appellant had been convicted. As indicated in Table 2, the substantial majority of appellants (70.45%) had been wrongfully convicted for a single offense. Conversely, only 25 out of the 88 individuals in our sample (28.41%) had been unjustly convicted for two or more offenses.

Table 2. Number of Offenses for Which the Petitioners Were Wrongfully Convicted

Number of crimes	Frequency	Percentage
One	62	70.45
Two	15	17.05
Three	4	4.55
Four	4	4.55
Six	2	2.27
No information available about the crime(s)	1	1.14

Table 3 shows that the predominant offense type was property crime, followed by public safety offenses (e.g., drug trafficking or road safety crimes) and crimes of assault. The percentages for the remaining offenses were all below 10%. There were no cases involving the offenses that are listed in Table 1 but not in Table 3.

Table 3. Frequencies and Percentages for the Most Prevalent Types of Crimes

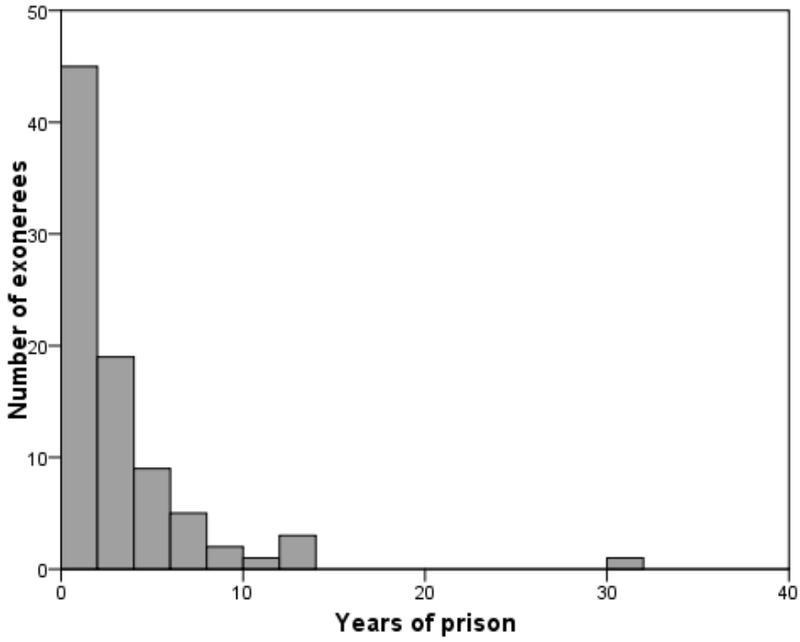
Type of crime	Frequency (one or more crimes of each type)	Percentage
Crimes against property	40	45.45
Crimes against public safety	17	19.32
Crimes of assault	15	17.05
Minor offenses	8	9.09
Sex crimes	7	7.95
Forgery	7	7.95
Crimes against freedom	5	5.68
Crimes against the Justice Administration	5	5.68
Crimes against public order	4	4.55
Homicides	3	3.41
Offenses related to land use, urban planning, and the environment	3	3.41
Torture	1	1.14
Crimes against domestic relations	1	1.14
Crimes against the rights of foreign citizens	1	1.14
Not reported	1	1.14

Note. The sum of the total percentage is greater than 100% because one single person could be convicted for more than one crime.

Information on the duration of imprisonment was unavailable for three individuals in the judgments. For the remaining cases, prison terms spanned from 3 days to 30 years. Despite the

substantial range between these extremes, the majority (85%) of individuals received sentences of less than 4.5 years (Mean = 3 years; Median = 1.5 years). Figure 1 provides a visual representation of the dispersion in this variable within the sample.

Figure 1. Distribution of the Number of Exonerees Based on Sentence Duration



C. Factors Contributing to Wrongful Convictions

As shown in Table 4, professional misconduct was the most prevalent factor contributing to wrongful convictions, accounting for, at least in part, 63.64% of the coded cases. The second most prevalent factor was misapplication of forensic science, present in approximately one of every four cases. Misidentifications and false testimonies of victims and witnesses ranked as the third and fourth most frequent causes, respectively.

Table 4. Frequencies and Percentages of Factors that Contributed to Wrongful Convictions

Contributing factors	Frequency	Percentage
Misconduct of legal professionals (lawyers, judges, prosecutors, the police…)	56	63.64
Misapplication of forensic science	24	27.27
Misidentifications	15	17.05
False testimony	14	15.91
Other causes	12	13.64
False confessions	8	9.09
Not reported	4	4.55

Note. The sum of the total percentage exceeds 100% because wrongful convictions could be caused by more than one factor.

IV Discussion

Leo and Gould (2009) noted that wrongful convictions have seldom been scrutinized using a systematic and rigorous scientific methodology. Indeed, numerous published scholarly documents have either consolidated instances of miscarriages of justice or discussed the legal aspects contributing to them (Leo, 2005). Conversely, Leo (2017) advocates for the advancement of the innocence movement through systematic and rigorous research.

Additionally, LaPorte (2018) found inconsistencies in how the Innocence Project and the National Registry of Exonerations classified the factors contributing to some specific wrongful conviction cases. This discrepancy can be attributed, at least in part, to each organization employing different coding criteria or different definitions. LaPorte's report led both organizations

to work together towards a consensual definition of "False or Misleading Forensic Evidence", as well as to recode their cases to eliminate discrepancies (see Cole *et al*., 2022).

In accordance with Leo's (2017) recommendations, and to mitigate potential coding discrepancies such as those identified by LaPorte (2018), we elaborated a comprehensive coding scheme, and used it to systematically examine wrongful convictions in Spain. We focused on the rate of wrongful convictions, the characteristics of exonerees, the crime types for which most wrongful convictions occurred, and the contributing factors.

A. Rate of Wrongful Convictions

At least 88 individuals have been wrongfully convicted to custodial sentences in Spain under the current Criminal Code. Over the same research period, there were 3,044 exonerations documented in the United States (National Registry of Exonerations, 2023a). Due to significant differences in population size between the two countries,[3] directly comparing absolute numbers would be misleading. Hence, we computed the ratio of wrongful convictions per 100,000 population in each country. The rate is lower for Spain (0.18) than it is for the United States (0.91).

B. Gender

In 2022, 92% of individuals incarcerated in Spain were males (Ministry of the Interior[4], 2023). The proportion of wrongfully convicted males that we identified also stands at 92%. Therefore, in line with previous research conducted in other countries (Dioso-Villa, 2015; Duce, 2015; Gross *et al*., 2005), there is no gender bias evident in Spain.

C. Citizenship

In Spain, 30.1% of prison inmates are foreigners (Ministry of the Interior, 2023).[5] Information about the citizenship of the wrongfully convicted individuals was included in only 40% of the analyzed judgments. However, according to these judgments, one half of these individuals were foreign citizens. These data are consistent with the notion that minorities may be at a higher risk for wrongful convictions. For instance, research conducted in the United States has found an overrepresentation of ethnic minorities among wrongfully convicted persons (Gross *et al*., 2005).[6]

D. Prior Criminal Records

Gould *et al*. (2014) observed that individuals with a criminal history are more susceptible to be wrongfully convicted than those with no prior records. In our research, information on the criminal records of the petitioners was available for 64% of the cases. Among these cases, the majority of wrongfully convicted individuals (55%) had a prior criminal record. This finding aligns with previous research.

We compared the rate of wrongfully convicted individuals with prior records with the recidivism rates of the Spanish prison population. According to a report from the General Secretariat of Penitentiary Institutions (2017), out of 688 inmates for whom relevant data were available, 62.50% had prior convictions, and the rest did not. Consequently, the rate of wrongfully

[3] The United States has a population of approximately 336 million (United States Census Bureau, 2023), while Spain has a population of approximately 48.5 million (Statistics National Institute, 2023).

[4] The roles of the Spanish Ministry of the Interior differ from those of the U.S. Department of Interior. Rather, they are very similar to those of the UK Home Office (see online: https://transparencia.gob.es/transparencia/en/transparencia_Home/index/PublicidadActiva/OrganizacionYEmpleo/Funciones/Funciones-MINT.html; in English. https://en.wikipedia.org/wiki/Ministry_of_the_Interior_(Spain)).

[5] The Ministry of the Interior (2023) report compiles the population of Spanish and foreign inmates in Spain from 1996 to 2022 (the same years we considered for this study). The average percentage of foreign individuals in Spanish prisons over these 27 years has been 28.15%.

[6] We were unable to examine ethnicity because Spanish judgments do not indicate the defendant's ethnicity (presumably because the Spanish population is more ethnically homogeneous than the US population; e.g., Infoplease Staff, 2023).

incarcerated innocents with prior criminal records is lower than the rate of the Spanish inmate population at large with such records.

E. Crime Type

While the majority of exonerations in the United States have taken place for serious crimes, particularly homicide, there is an ongoing debate about whether judicial errors are more prevalent for serious or minor crimes (Gould & Leo, 2010). In this study, the relatively short duration of custodial sentences and the observation that 82% of exonerations involved crimes such as theft, offenses against public safety, or assault, suggest that, in Spain, the exoneration cases did not primarily involve extremely serious offenses.

To examine whether certain types of crimes are more prone to judicial errors, we compared the percentages in Table 3 with data from the prison population in Spain (Ministry of the Interior, 2023). The Ministry of the Interior (2023) records the crimes for which inmates are serving their sentences in prison, while our dataset includes individuals wrongfully deprived of liberty for *at least* one offense. It is important to note that a single individual could be convicted for more than one crime, including offenses with a sentence other than deprivation of liberty. To illustrate, if a person had been wrongfully convicted for homicide (resulting in a prison sentence) and theft (resulting in a fine), we counted one homicide and one theft in the types of crime. Consequently, there may be an overrepresentation of minor offenses in our dataset.

To address this concern, we excluded from our dataset the one person whose judgment did not report the crime and those individuals who had been penalized for more than one offense (28%), focusing solely on the remaining 70%. Figure 2 illustrates that the most overrepresented crimes in the study sample (compared to the prison population) were offenses against public safety (with a difference of 6%), and forgery and assault crimes (both with a difference of 5%). As shown in Table 5, *Z* tests for comparison of proportions indicated that forgery is the sole crime whose difference in proportions was significant ($Z = -2.06$, $p = .042$, Cohen's $h = 0.32$). Once again, these findings suggest that judicial errors in Spain are more likely to occur for minor offenses.

The two most underrepresented crimes among wrongfully convicted individuals were homicides and sex crimes (present 4% less often in our sample than in the prison population at large. However, none of these differences were significant according to Z tests for the comparison of proportions (see Table 5).[7] This is unexpected, considering that homicides and sex crimes are the most common crime types in wrongful conviction cases in the United States (Gross *et al*., 2005) and Australia (Dioso-Villa, 2015). The reason for the underrepresentation of these offenses in our sample remains unclear.

[7] These data must be analyzed with caution, since the prison population report (Ministry of Interior, 2023) treats offenses and minor offenses related to gender-based violence as an independent type of crime. The report indicates there were 4,782 cases of gender-based violence, representing 10.52% of the total incarcerated population. Redistributing the offenses categorized as gender-based violence into the Figure 2 categories might reduce the overrepresentation of innocent individuals in some of those categories. Conversely, if the percentage of homicides increases by adding homicides currently included in the gender-based violence category, the difference between the inmate population and wrongfully convicted individuals would increase for homicides.

Figure 20. Percentages of Individuals Wrongfully Convicted for a Single Offense, and of the Total Incarcerated Population in Spain, for Each Crime Type

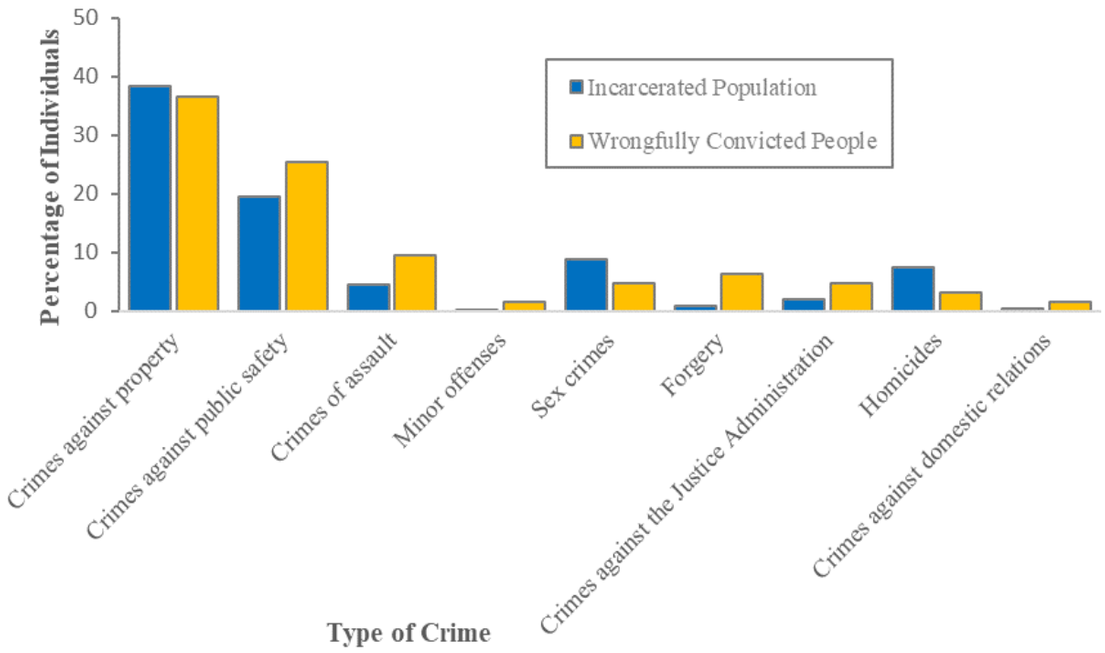


Table 5. Z Tests for the Comparison of Proportions of the Wrongfully Convicted Individuals, and of all the Incarcerated Population in Spain, for Each Crime Type

Crime type	Z	p	Cohen's h
Crimes against property	0.41	.797	0.06
Crimes against public safety	-1.08	.312	0.15
Crimes of assault	-1.45	.157	0.21
Minor offenses	-1.23	.223	0.21
Sex crimes	1.12	.251	0.16
Forgery	-2.06	.042	0.32
Crimes against the Justice Administration	-1.09	.289	0.16
Homicides	1.30	.187	0.19
Crimes against domestic relations	-0.90	.371	0.14

F. Contributing Factors

The National Registry of Exonerations (2023b) identified false testimonies and professional misconduct as contributing, at least in part, to over 60% of wrongful convictions in the United States. In Spain, professional malpractice is the predominant factor, comparable to the United States. However, false testimonies are ranked third, alongside mistaken identifications, explaining only 16% of cases. Misapplication of forensic science is the second-most prevalent factor, present in 27% of cases.

Our results on factors contributing to wrongful convictions in Spain have practical implications for justice professionals, policymakers, and legal practitioners:

> First, the number of cases reopened in Spain through review appeals that we were able to retrieve is certainly small (88 cases). This can indicate that, in general, Spanish legal professionals perform their duties well—but note it is also possible that the small number of cases is a consequence of the requirements for bringing a review appeal forward being very narrow. Either way, our data revealed that *professional malpractice* is the main factor behind wrongful convictions in Spain, which suggests there is room for improvement. Indeed, the most effective approach to reducing the number of wrongful convictions in the country may be enhancing the performance of legal practitioners. It is crucial to educate justice professionals on the causes of judicial errors, the biases that may influence them, and best practices to improve processes and reduce the occurrence of wrongful convictions. These topics should be included in all undergraduate and postgraduate law programs, and should be

part of the curriculum for the exams to become a judge, prosecutor, lawyer, or law enforcement officer. Additionally, these topics should also be incorporated into ongoing professional development courses, workshops, or seminars available for justice professionals. In addition, the police and judges should verify the identity of suspects, ensure suspects are aware of the charges against them, and properly verify their alibis; lawyers should ensure that their defendants truly understand the charges they face, and so on. For instance, in some cases, in being caught, the real perpetrator pretended to be someone else (i.e., the innocent person who would later be convicted), either just verbally or by using that person's personal identification document (or a photocopy of it). Had the perpetrator's identity been verified using fingerprints from the national identity document database, some of these judicial errors might have been avoided. Errors can also be prevented by comparing the fingerprints of the person who was detained on the day of the crime with those of (a) the person attending the trial, and (b) the individual entering prison. In some other instances, the appellant was incarcerated, detained, or in hospital at the time of the crime. A thorough investigation would have revealed this circumstance. Investigation judges should have verified these individuals' alibi, and the defense lawyers should have sought this evidence and should have presented it in court.

Second, regarding the *misapplication of forensic science*, Cooley and Oberfield (2007) propose the establishment of external and independent oversight of crime laboratories, along with the implementation of a certification system to verify professionals' qualifications and thus reduce forensic misconduct. Forensic scientists should only employ those methods endorsed by the scientific community and should comprehensively elucidate the limitations of those procedures and the implications of their findings within their reports. Consequently, scientific reports should be drafted with precision and without ambiguity to facilitate comprehension by legal professionals (LaPorte, 2018). In addition, a number of measures that forensic experts can take to minimize the risk of cognitive biases are briefly listed by Dror (2020).

Third, regarding *eyewitness misidentification*, an American Psychology-Law Society (Division 41 of the American Psychological Association) committee has crafted a set of recommendations to improve eyewitness identification procedures (Wells *et al*., 2020). Those guidelines include some pre-lineup interview recommendations, detail the process for selecting lineup fillers, suggest some pre-lineup instructions to be given to witnesses, recommend using a double-blind procedure, suggest confidence judgments should be collected at the time of the identification, and emphasize the significance of video recording the entire process. Additionally, they highlight the need to have evidence-based suspicions before conducting a lineup, and indicate that repeated identifications with the same suspect, as well as showups, should be avoided.

Fourth, regarding *false testimony*, a comprehensive criminal investigation encompassing the scrutiny of multiple hypotheses and the pursuit of external evidence could potentially contribute to mitigating miscarriages of justice.

Finally, related to *false confessions*, it is essential for the police to employ science-based, legal, and ethical interview methods. Numerous studies indicate that coercive interrogations, whose sole objective is to elicit confessions, yield less reliable and precise information than non-coercive investigative interviews, and increase the likelihood of false confessions (e.g., Meissner *et al.*, 2014; Vrij *et al.*, 2017). Recently, an international committee of experts in the fields of interviewing, law enforcement, criminal investigations, national security, military, intelligence, psychology, criminology, and human rights from around the world have drafted the "Méndez principles" on effective interviewing for investigations and information gathering (Méndez, 2021). The Méndez Principles are to replace coercive, confession-oriented interrogations with rapport-based investigative interviews instructed by science, law, and ethics (see Méndez, 2021).

G. Limitations

We exclusively examined the information presented in judgments. Conversely, other studies have also drawn on supplementary sources such as media reports and statements from victims and perpetrators (Dioso-Villa, 2015; Gross *et al*., 2005). Our approach has both positive and negative aspects. On the positive side, the information within judgments is typically well-substantiated. However, on the negative side, this information can sometimes be incomplete. For example, we were unable to ascertain the citizenship of 59% of petitioners. Additionally, factors

not explicitly mentioned in the sentences might have contributed to the judicial errors. While our research captures the minimum number of causes present in Spanish cases, we cannot guarantee that other factors did not also contribute. Despite this limitation, our research is crucial as it sheds light on and serves as a starting point for understanding wrongful convictions in Spain.

V References

Aebi, M. F., & Campistol, C. (2013). "Voluntary" false confessions as a source of wrongful convictions: The case of Spain. In C. R. Huff & M. Killians (Eds.), *Wrongful Convictions and Miscarriages of Justice: Causes and Remedies in North America and European Criminal Justice Systems.* Routledge. Online: https://doi.org/10.4324/9780203597286

Ask, K., & Fahsing, I. (2019). Investigative decision making. In R. Bull & I. Blandón-Gitlin (Eds.), *The Routledge International Handbook of Legal and Investigative Psychology* (pp. 84-101). Routledge. Online: https://doi.org/10.4324/9780429326530

Berube, R., Wilford, M. M., Redlich, A. D., & Wang, Y. Identifying patterns across the six canonical factors underlying wrongful convictions. (2022) Wrongful Conviction Law Review, 3(3), 166-195. Online: https://doi.org/10.29173/wclawr82

Blanco-Velasco, G., Sánchez, N., & Geven, L. M. (2023). Miscarriages of justice in Spain. In J. Robins (Ed.), *Murder, Wrongful Conviction and The Law: An International Comparative Analysis* (pp. 122-135). Routledge. Online: https://doi.org/10.4324/9781003251484

Brants, C. Wrongful convictions and inquisitorial process: The case of the Netherlands. (2012) University of Cincinnati Law Review, 80, 1069-1114. Online: https://scholarship.law.uc.edu/uclr/vol80/iss4/2

Burrows, J., Hopkins, M., Hubbard, R., Robinson, A., Speed, M., & Tilley, N. (2005). *Understanding The Attrition Process in Volume Crime Investigations*. Home Office. Online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment data/file/115828/hors295.pdf

Clow, K. A. (2017). Does The "Wrongful" Part of Wrongful Conviction Make A Difference In The Job Market? In R. Ricciardelli & A. M. F. Peters (Eds.), *After Prison: Navigating Employment and Reintegration* (pp. 243-257). Wilfred Laurier University Press.

Cole, S. A., Meterko, V., Chu, S., Cooper, G., Paredes, J. W., Possley, M., & Otterbourg, K. (2022). *The Contribution of Forensic and Expert Evidence to DNA Exoneration Cases: An Interim Report.* National Registry of Exonerations and Innocence Project. Online: https://n2t.net/ark:/88112/x2g890

Cooley, C. M., & Oberfield, G. S. Increasing forensic evidence's reliability and minimizing wrongful convictions: Applying Daubert isn't the only problem. (2007) Tulsa Law Review, 43(2), 285-380. Online: https://digitalcommons.law.utulsa.edu/tlr/vol43/iss2/6

De Swert, K. (2012). *Calculating Inter-Coder Reliability in Media Content Analysis Using Krippendorff's Alpha*. Online: https://www.polcomm.org/wp-content/uploads/ICR01022012.pdf

Devine, D. J., & Caughlin, D. E. Do they matter? A meta-analytic investigation of individual characteristics and guilt judgments. (2014) Psychology, Public Policy, and Law, 20(2), 109–134. Online: https://doi.org/10.1037/law0000006

Dioso-Villa, R. A repository of wrongful convictions in Australia: First steps toward estimating prevalence and causal contributing factors. (2015) Flinders Law Journal, 17(2), 163-202. Online: https://ssrn.com/abstract=3009055

Dror, I. E. Cognitive and human factors in expert decision making: Six Fallacies and The Eight Sources of Bias. (2020) Analytical Chemistry, 92(12), 7998-8004. Online: https://dx.doi.org/10.1021/acs.analchem.0c00704

Dror, I. E., Charlton, D., Perón, A. E. Contextual information renders experts vulnerable to making erroneous identifications. (2006) Forensic Science International, 156(1), 74-78. Online: https://doi.org/10.1016/j.forsciint.2005.10.017

Dror, I. E., Melinek, J., Arden, J. L., Kukucka, J., Hawkins, S., Carter, J., & Atherton, D. S. Cognitive bias in forensic pathology decisions. (2021) Journal of Forensic Sciences, 66(5), 1751-1757. Online: https://doi.org/10.1111/1556-4029.14697

Duce, M. ¿Debiéramos preocuparnos de la condena de inocentes en Chile?: Antecedentes comparados y locales para el debate [Should we be concerned about the conviction of innocent people in Chile?: Comparative and local background for the debate]. (2013) Ius et Praxis, 19(1), 77-138. Online: https://doi.org/10.4067/S0718-00122013000100004

Duce, M. The conviction of innocents in Chile: An empirical approximation from the results of wrongful conviction claims admitted by the Supreme Court during the 2007-2013 period. (2015) Política Criminal, 10(19), 159-191. Online: https://doi.org/10.4067/S0718-33992015000100006

EFE. (2006, March 30). Un policía denunció presiones del FBI para inculpar a un abogado estadounidense en el 11-M. *El Mundo*. Online: https://www.elmundo.es/elmundo/2006/03/30/espana/1143730221.html

EUREX. (2023). The European Registry of Exonerations. Online: https://www.registryofexonerations.eu/

Findley, K. A. Defining innocence. (2010) Albany Law Review, 74(3), 1157-1208. Online: https://doi.org/10.2139/ssrn.1815304

Findley, K. A., & Scott, M. S. The multiple dimensions of tunnel vision in criminal cases. (2006) Wisconsin Law Review, 2, 291-398. Online: https://ssrn.com/abstract=911240

Garrett, B. L. (2017). Convicting the innocent redux. In D. S. Medwed (Ed.), *Wrongful Convictions and The DNA Revolution: Twenty-Five Years Of Freeing The Innocent* (pp. 40-56). Cambridge University Press. Online: https://doi.org/10.1017/9781316417119.004

General Council of the Judiciary. (2023). *La justicia dato a dato 2023: Estadística judicial* [Justice datum by datum 2023: Judicial statistics]. Online: https://www.poderjudicial.es/stfls/ESTADISTICA/FICHEROS/JusticaDatoaDato/Justicia%20Dato%20a%20Dato%202022.pdf

General Secretariat of Penitentiary Institutions. (2017). *Documentos penitenciarios 16. La estancia en prisión: consecuencias y reincidencia* [Prison documents 16. Imprisonment: Consequences and recidivism]. Online: https://www.interior.gob.es/opencms/pdf/archivos-y-documentacion/documentacion-ypublicaciones/publicaciones-descargables/instituciones-penitenciarias/La_estancia_en_prision_126170566_web.pdf

Godsey, M. (2017). *Blind injustice*. University of California Press.

Gould, J. B., Carrano, J., Leo, R. A., & Hail-Jares, K. Predicting erroneous convictions. (2014) Iowa Law Review, 99(2), 471-522.

Gould, J. B., & Leo, R. A. One hundred years later: Wrongful convictions after a century of research. (2010) The Journal of Criminal Law & Criminology, 100(3), 825-868. Online: https://scholarlycommons.law.northwestern.edu/jclc/vol100/iss3/7

Gross, S. R. Convicting the innocent. (2008) Annual Review of Law and Social Science, 4, 173-192. Online: https://doi.org/10.1146/annurev.lawsocsci.4.110707.172300

Gross, S. R., Jacoby, K., Matheson, D. J., Montgomery, N., & Patil, S. Exonerations in the United States 1989 through 2003. (2005) Journal of Criminal Law & Criminology, 95(2), 523-560. Online: https://www.jstor.org/stable/3491344

Gudjonsson, G. H. (2018). *The Psychology of False Confessions: Forty Years of Science and Practice*. Wiley. Online: https://doi.org/10.1002/9781119315636

Harmon, T. R., & Lofquist, W. S. Too late for luck: A comparison of post-Furman exonerations and executions of the innocent. (2005) Crime & Delinquency, 51(4), 498-520. Online: https://doi.org/10.1177/0011128705275977

Helm, R. K. Wrongful conviction in England and Wales: An assessment of successful appeals and key contributors. (2022) Wrongful Conviction Law Review, 3(3), 196-217. Online: https://doi.org/10.29173/wclawr79

Herrero, C. Los conocimientos de la psicología más allá de la prueba pericial [Knowledge of Psychology Beyond Expert Evidence]. *Quaestio Facti.* (2021) Revista Internacional sobre Razonamiento Probatorio, *2*, 363-408. Online: https://doi.org/10.33115/udg_bib/qf.i2.22533

Huff, C. R. Wrongful convictions: The American experience. (2004) Canadian Journal of Criminology and Criminal Justice, 46(2), 107-120. Online: https://doi.org/10.3138/cjccj.46.2.107

Hughes, J. Krippendorffsalpha: An R package for measuring agreement using Krippendorff's Alpha Coefficient. (2021) The R Journal,13(1), 413-425. Online: https://doi.org/10.32614/RJ-2021-046

Hunt, P. E., Saunders, J., & Kilmer, B. Estimates of law enforcement costs by crime type for benefit-cost analyses. (2019) Journal of Benefit Cost Analysis, 10, 95-123. Online: http://doi.org/10.1017/bca.2018.19

Infoplease Staff (2023, July 27). *Ethnicity and Race by Country*. Infoplease. Online: https://www.infoplease.com/world/social-statistics/ethnicity-and-race-countries

Innocence Project. (2023a). *DNA Exonerations in the United States (1989-2020).* Innocence Project. Online: https://innocenceproject.org/dna-exonerations-in-the-united-states/

Innocence Project. (2023b). *Explore the Numbers: Innocence Project's Impact*. Innocence Project. Online: https://innocenceproject.org/exonerations-data/

Jenkins, S. Secondary victims and the trauma of wrongful conviction: Families and children's perspectives on imprisonment, release and adjustment. (2013) Australian & New Zealand Journal of Criminology, 46, 119-137. Online: https://doi.org/10.1177/0004865812470384

Kassin, S. M. (2022). *Duped: Why Innocent People Confess—And Why We Believe Their Confessions.* Prometheus Books.

Kassin, S. M., Drizin, S. A., Grisso, T., Gudjonsson, G. H., Leo, R. A., & Redlich, A. D. Police-induced confessions: Risk factors and recommendations. (2010) Law and Human Behavior, 34(1), 3-38. Online: https://doi.org/10.1007/s10979-009-9188-6

Kassin, S. M., Meissner, C. A., & Norwick, R. J. "I'd know a false confession if I saw one": A comparative study of college students and police investigators. (2005) Law and Human Behavior, *29*(2), 211–227. Online: https://doi.org/10.1007/s10979-005-2416-9

Kassin, S. M., & Neumann, K. On the power of confession evidence: An experimental test of the fundamental difference hypothesis. (1997) Law and Human Behavior, 21(5), 469–484. Online: https://doi.org/10.1023/A:1024871622490

Katzman, J., & Kovera, M. B. Causes of racial disparities in wrongful convictions based on mistaken identifications: Own-race bias and differences in evidence-based suspicion. (2023) Law and Human Behavior, 47(1), 23-35. Online: https://doi.org/10.1037/lhb0000503

Kukucka, J., Clow, K. A., Horodysky, A. M., Deegan, K., & Gayleard, N. M. Do exonerees face housing discrimination? An email-based field experiment and content analysis. (2021) Psychology, Public Policy, and Law, 27(4), 570-580. Online: https://doi.org/10.1037/law0000323

Kukucka, J., Horodyski, A. M., & Dardis, C. M.. The Exoneree Health and Life Experiences (ExHaLE) study: Trauma exposure and mental health among wrongly convicted individuals. (2022) Psychology, Public Policy, and Law, 28(3), 387–399. Online: https://doi.org/10.1037/law0000358

Landis, J. R., & Koch, G. G. The measurement of observer agreement for categorical data. (1977) Biometrics, 33(1), 159-174. Online: https://doi.org/10.2307/252 9310

LaPorte, G. Wrongful convictions and DNA exonerations: Understanding the role of forensic science. (2018) National Institute of Justice Journal, 279, 1-16. Online: https://nij.gov/journals/279/Pages/wrongfulconvictions-and-dna-exonerations.aspx.

Leo, R. A. Rethinking the study of miscarriages of justice: Developing criminology of wrongful conviction. (2005) Journal of Contemporary Criminal Justice, 21(3), 201-223. Online: https://doi.org/10.1177/1043986205277477

Leo, R. A. The criminology of wrongful conviction: Decade later. (2017) Journal of Contemporary Criminal Justice, 33(1), 82-106. Online: https://doi.org/10.1177/1043986216673013

Leo, R. A., & Gould, J. B. Studying wrongful convictions: Learning from Social Science. (2009) Ohio State Journal of Criminal Law, 7(7), 7-30. Online: https://ssrn.com/abstract=1531903

Lupária, L. Unveiling wrongful convictions between the U.S. and Italy: Cross-learning from each other's mistakes. (2020) Wrongful Conviction Law Review, 1(1), 101-123. Online: https://doi.org/10.29173/wclawr12

Masip, J., Garrido, E., & Herrero, C. Defining deception. (2004) Anales de Psicología, 20(1), 147-171. Online: https://revistas.um.es/analesps/article/view/27631/26801

Meissner, C. A., & Brigham, J. C. Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. (2001) Psychology, Public Policy, and Law, 7(1), 3-35. Online: https://doi.org/10.1037/1076-8971.7.1.3

Meissner, C. A., Redlich, A. D., Michael, S. W., Evans, J. R., Camilletti, C. R., Bhatt, S., & Brandon, S. Accusatorial and information-gathering interrogation methods and their effects on true and false confessions: A meta-analytic review. (2014) Journal of Experimental Criminology, 10, 459-486. Online: https://doi.org/10.1007/s11292-014-9207-6

Méndez, J. E. (2021). *Principles On Effective Interviewing for Investigations and Information Gathering (The Méndez Principles)*. Online: https://www.apt.ch/en/resources/publications/new-principles-effective-interviewing-investigations-and-information

Ministry of the Interior (2023). *Anuario estadístico del ministerio del Interior: 2022* [Statistical yearbook of the Ministry of the Interior: 2022]. Online: https://www.interior.gob.es/opencms/pdf/archivos-y-documentacion/documentacion-ypublicaciones/anuarios-y-estadisticas/ultimo-anuario-estadistico/Anuario_estadistico_2022_126150729.pdf

Morgan, J. (2023). *Wrongful Convictions and Forensic Science Errors: Case Studies and Root Causes.* CRC Press. Online: https://doi.org/10.4324/9781003202578

Natapoff, A. Beyond unreliable: How snitches contribute to wrongful convictions. (2006) Golden Gate University Law Review, 37(1), 107-120. Online: https://digitalcommons.law.ggu.edu/ggulrev/vol37/iss1/5

National Registry of Exonerations. (2021). *25000 Years Lost to Wrongful Convictions*. Online: https://www.law.umich.edu/special/exoneration/Documents/25000%20Years.pdf

National Registry of Exonerations. (2023a). *The National Registry of Exonerations*. Online: https://www.law.umich.edu/special/exoneration/Pages/Exonerations-in-the-United-States-Map.aspx

National Registry of Exonerations. (2023b). *% Exonerations by Contributing Factor*. Online: https://www.law.umich.edu/special/exoneration/Pages/ExonerationsContribFactorsByCrime.aspx

Norris, R. J., & Mullinix, K. J. Framing innocence: An experimental test of the effects of wrongful convictions on public opinion. (2020) Journal of Experimental Criminology, 16, 311-334. Online: https://doi.org/10.1007/s11292-019-09360-7

Ortiz-Pradillo, J. C. (2023). Compensation for wrongful convictions in Spain. In W. Jasinski & K. Kremens (Eds.), *Compensation for Wrongful Convictions: A Comparative Perspective* (pp. 70-86). Routledge. Online: https://doi.org/10.4324/9781003229414-5

Rogers-Degeer, A. (2023). *The Social Reintegration of Exonerees: A Systematic Literature Review*. [Doctoral dissertation, University of Arizona Global Campus]. ProQuest Dissertations and Theses Global. Online: https://www.proquest.com/docview/2810842110

Sánchez, N., Sobral, J., & Seijo, D. Judicial error in preventive prison: People in prison who will never be convicted. (2017) Revista Iberoamericana de Psicología y Salud, *8*(1), 36-43. Online: https://doi.org/10.23923/j.rips.2017.08.004

Schell-Leugers, J. M., Masip, J., González, J. L., Vanderhallen, M., & Kassin, S. M. Police interviewing in Spain: A self-report survey of police practices and beliefs. (2023) Anuario de Psicología Jurídica, 33(1), 27-40. Online: https://doi.org/10.5093/apj2022a4

Schönteich, M. (2014). *Presumption of Guilt: The Global Overuse of Pretrial Detention.* Open Society Justice Initiative.

Scott, M. T., Manzanero, A. L., Muñoz, J. M., & Köhnken, G. Admisibilidad en contextos forenses de indicadores clínicos para la detección del abuso sexual infantil [Admissibility of clinical indicators for the detection of child sexual abuse in forensic contexts]. (2014) Anuario de Psicología Jurídica, 24, 57-63. Online: https://doi.org/10.1016/j.apj.2014.08.001

Smith, E., & Hattery, A. J. Race, wrongful conviction & exoneration. (2011) Journal of African American Studies, 15, 74-94. Online: https://doi.org/10.1007/s12111010-9130-5

Sporer, S. L. Lessons from the origins of eyewitness testimony research in Europe. (2008) 737-757. Online: https://doi.org/10.1002/acp.1479 Applied Cognitive Psychology, 22(6),

Stacey, R. B. A report on the erroneous fingerprint individualization in the Madrid train bombing case. (2004) Journal of Forensic Identification, 54(6), 706-718. Online: https://www.proquest.com/docview/194800751

Statistics National Institute. Estadística continua de población (ECP) a 1 de octubre de 2023 [Continuous population statistics as of October 1, 2023]. (2023) Notas de Prensa. Online: https://www.ine.es/daco/daco42/ecp/ecp0323.pdf

United States Census Bureau. (2023). *U.S. and World Population Clock.* United States Census Bureau. Retrieved December 10, 2023, from Online: *https://www.census.gov/popclock/*

Vrij, A., Meissner, C. A., Fisher, R. P., Kassin, S. M., Morgan, C. A., & Keinman, S. M. Psychological perspectives on interrogation. (2017) Perspectives on Psychological Science, 12(6), 927-955. Online: https://www.doi.org/10.1177/1745691617706515

Wells, G. L. Applied eyewitness-testimony research: System variables and estimator variables. (1978) Journal of Personality and Social Psychology, 36(12), 1546–1557. Online: https://doi.org/10.1037/0022-3514.36.12.1546

Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. (2020) Law and Human Behavior, 44(1), 3–36. Online: https://doi.org/10.1037/lhb0000359

West, E., & Meterko, V. DNA exonerations 1989-2014: Review of data and findings from the first 25 years. (2015) Albany Law Review, 79(3), 717-795. Online: https://ssrn.com/abstract=2986970

Yaroshefsky, E., & Schaefer, L. (2014). Defense Lawyering and Wrongful Conviction. In A. D. Redlich, J. A. Acker, R. J. Norris, & C.L. Bonventre, (Eds.), *Examining Wrongful Convictions: Stepping Back, Moving Forward* (pp. 123-140). Carolina Academic Press.

Wrongful Conviction in Texas: 'Sex Assaults', False Guilty Pleas, Stranger Rape with Misidentification, and Drug Offenses[1]

Matthew Barry Johnson
Department of Psychology, CUNY John Jay College of Criminal Justice
New York, NY
U.S.A.

Janquel D. Acevedo
School of Psychology, The University of Queensland
Brisbane
Australia

Citing IP data, Johnson (2021), reported that sexual assault/rape was the most common offense associated with exoneration in the US. Also, stranger rape accounted for 72% of the entire IP database. To further examine the role of sexual assault, the current study examined all exonerations in Texas, the US state with the most sexual assault exonerations. Using NRE data, descriptive analyses, and reclassifying sexual assaults, we find drug offenses are the most common crime type associated with exonerations in Texas but sexual assault/rape accounts for a significant portion of Texas exonerations. Contrary to a common assumption, we also find that exculpatory DNA does not explain the substantial proportion of sexual assaults among exonerations. We also examine the role of stranger rape misidentification, youthful complainant recantations (perjury/false allegations) and false guilty pleas in the NRE Texas database. Finally, we discuss other patterns within the Texas exonerations and policy implications.

I. Introduction
  A. Why Texas?
  B. Current Study
II. Methods
III. Findings
IV. Case Illustrations
  A. False Guilty Pleas
    1. Blackshire & Johnson
    2. Dahn Clary
  B. Youthful Complainant Recantation
    1. San Antonio Four
    2. Tony Hall

[1] This paper is dedicated to the memory and honor of ***Timothy Cole*** and ***Cameron Todd Willingham***. An initial version of this report was presented by the author at the Austin, Texas Criminal Defense Lawyers Association Meeting on February 11, 2022. The authors thank Attorney Betty Blackwell for inviting and arranging the address. A second version was presented at the Innocence Project, "Just Data: Advancing the Innocence Movement" virtual conference - November 9, 2023. Luisianna Tejada provided valuable research assistance.

C. Stranger Rape, Misidentification
1. Michale Phillips
2. Ricardo Rachell
D. Drug Offenses
1. Meghan Alegria
V. Discussion

I Introduction

What can exonerations tell us about wrongful conviction? Quite a bit since the record of each exoneration documents an erroneous criminal investigation and prosecution. That is, the errors of omission reflected in the failure to identify the actual offender (if there was one), coupled with the prosecution of an innocent person (an error of commission).[2] In this way, each exoneration illustrates what went awry in the process. However, caution is warranted because what went wrong in any given prosecution may have also gone awry in other prosecutions which resulted in dismissals or acquittals.[3] Further caution is warranted because we do not know if exonerations are representative of wrongful convictions. It may very well be that the wrongful conviction processes for the exonerated are different, in fundamental ways, from the wrongfully convicted who have not been exonerated. So, we begin this inquiry with curious interest, mixed with caution.

According to Johnson,[4] in the US, sexual assaults predominate among confirmed wrongful convictions. In addition, Johnson reported that disaggregation among known wrongful convictions in sexual assault finds these convictions are disproportionately concentrated among stranger rape investigations, as opposed to the more common acquaintance sexual assaults. Seventy-two percent of the Innocence Project (IP) exoneration list are stranger rape cases. However, earlier, it was asserted the association of wrongful conviction to rape is an artifact of DNA testable samples being common in sexual assaults. Neufield & Scheck wrote in the forward of Connors et al:

> Since there does not seem to be anything inherent in sexual assault cases that would make eyewitnesses more prone to mistakes than in robberies or other serious crimes where the crucial proof is eyewitness identification, it naturally follows that the rate of mistaken identifications and convictions is similar to DNA exoneration cases.[5]

This early explanation for the prevalence of sexual assaults among exonerations is limited in several ways. It suggests the misidentification outcome is tied to 'eye-witness errors' rather than the broader criminal investigation (elaborated below). Secondly, it does not account for the

---

[2] Matthew Barry Johnson, *Wrongful Conviction in Sexual Assault: Stranger Rape, Acquaintance Rape, and Intra-familial Child Sexual Assaults* (New York: Oxford University Press, 2021).

[3] Jon B. Gould, *et al*, "Predicting Erroneous Convictions: A Social Science Approach to Miscarriages of Justice" (2012) 99 Iowa Law Rev, 471, online: <https://ssrn.com/abstract=2231777>

[4] Johnson, *supra* note 2.

[5] Edward Connors *et al*, *Convicted by Juries, Exonerated By Science: Case Studies In The Use Of DNA Evidence To Establish Innocence After Trial* (Alexandria: DIANE Publishing, 1996) at xxxi

frequency of wrongful conviction in sexual assault stemming from false confessions (such as the Central Park Five, the Norfolk Four, Christopher Ochoa, the Beatrice Six, Jeffrey Deskovic, Byron Halsey and others).[6] The significance of these wrongful convictions in sexual assault stemming from false confessions is often overlooked because these cases are typically rape/murders and classified as 'murders'.

We suggest wrongful conviction in sexual assault is not linked to misidentification as 'encapsulated' error, but rather to practices and dynamics of the crime investigation [7] which likely apply to wrongful conviction in sexual assault associated with false confessions as well. That is, there are offense specific obstacles to accurate identification of suspects, coupled with offense specific biases and incentives among those who conduct the investigation, identification, and prosecution that account for the disproportionate concentration of wrongful convictions among (stranger) sexual assaults. The presence of DNA tells us how the cases were exonerated, not how they were wrongfully convicted.

There have been additional questions surrounding wrongful conviction and sexual assault. Considerable attention has been focused on wrongful convictions secondary to false guilty pleas.[8] Johnson & Cunningham [9] reported 17 cases of innocent defendants who pled guilty to rape charges. Another identified contributor to wrongful conviction in sexual assault has been allegations from youthful complainants that were found to be unreliable. Johnson [10] described a number of cases (Brian Banks, Jarrett Adams, Gary Dotson, Edgar Coker) in which a young person made a fabricated sexual assault allegation, to a private confidant, which resulted in unanticipated criminal prosecution and resulting conviction of an innocent party.

---

[6] Johnson, *supra* note 2

[7] Johnson, *supra* note 2; see also Margaret B. Kovera "The role of suspect development practices in eyewitness identification accuracy and racial disparities in wrongful conviction" (2023) 18:1 Soc Issues Policy Rev, 1–23, online: <https://doi.org/10.1111/sipr.12102>; Jacqueline Katzman & Margaret B. Kovera, "Potential Causes of Racial Disparities in Wrongful Convictions Based on Mistaken Identifications: Own-Race Bias and Differences in Evidence-Based Suspicion" (2023). 47:1 Law & Hum Behav 23, online: <https://doi.org/10.1037/lhb0000503>.

[8] Allison D. Redlich *et al*, "The Psychology of Defendant Plea Decision Making" (2017) 72:4 Am Psychol, 339–352, online: https://doi.org/10.1037/a0040436; Melanie B. Fessinger & Margaret Bull Kovera "An Offer You Cannot Refuse: Plea Offer Size Affects Innocent but Not Guilty Defendants' Perceptions of Voluntariness" (2023) 47:6 Law & Hum Behav, 619, online: <https://doi.org/10.1037/lhb0000548>; Tina M. Zottoli, *et al*, "Plea discounts, time pressures, and false-guilty pleas in youth and adults who pleaded guilty to felonies in New York City" (2016) 22:3 Psychol Public Pol L 250, online: <https://doi.org/10.1037/law0000095>.

[9] Matthew B. Johnson & Sydney Cunningham, (2015) Why Innocent Defendants Plead Guilty to Rape Charges. *The Crime Report*, June 30, 2015. online: <http://www.thecrimereport.org/viewpoints/2015-06-why-innocent-defendants-plead-guilty-to-rape-charges>.

[10] Johnson, *supra* note 2

A. Why Texas?

According to NRE data, as of 2022, the three states in the United States with the most exonerations are Illinois, Texas, and New York. However, relative to Illinois and New York, Texas has the most exonerations stemming from sexual assault allegations (combining NRE classification of cases with Sexual Assault and Child Sex Abuse). Thus, summarizing and disaggregating the NRE data on Texas sexual assault exonerations may contribute vital insights into wrongful convictions stemming from sexual assault allegations.

B. Current Study

The current study explores five questions:

1. Do sexual assault/rape prosecutions predominate among the Texas exonerations?
2. Is DNA evidence the principal means of exoneration among the Texas sexual assault/rape exonerations?
3. Is eye-witness misidentification the principal erroneous evidence among the Texas sexual assault/rape wrongful exonerations?
4. Do stranger rape misidentification cases predominate among Texas sexual assault/rape exonerations?
5. Do false guilty pleas predominate among Texas sexual assault/rape, and other types of exoneration cases?

II Methods

With the above questions in mind, we examined the record of wrongful conviction in the US State of Texas, as revealed by data published on the NRE. This data source was searched electronically to gain a descriptive account of exonerations in the State. As of March 18, 2022, there are a total 401 confirmed wrongful convictions in Texas reported on the database. The Texas cases were found by using the "Detailed View" option in "Browse Cases" tab on the NRE website. Then filtering the data by left-clicking "ST" (State) and selecting "TX" (Texas) for Texas state cases and "F-TX" for Texas federal cases, of which there were five**.** We report on the 396 Texas state prosecutions.

In our classification, we combined the NRE 'Sexual assault' cases with other prosecutions where sexual assault allegations were part of the offense/investigation (such as rape/murders) to designate the category 'Sexual Assault/Rape.' The Sexual Assault/Rape category includes cases where there was indeed a sexual assault but an innocent person was convicted (such as Timothy Cole), cases that involved false rape charges (i.e. San Antonio Four) which led to the conviction of innocent person(s), as well as cases where the formal charges did not include a sexual assault offense but sexual assault was included in the report of the crime. For instance, Anthony Massingill and Cornelius Dupree were charged with armed robbery in Dallas County, Texas in 1979. It was reported the female robbery victim was also sexually assaulted by both defendants but, according to the NRE, the prosecutors did not include sexual assault charges because it would not have

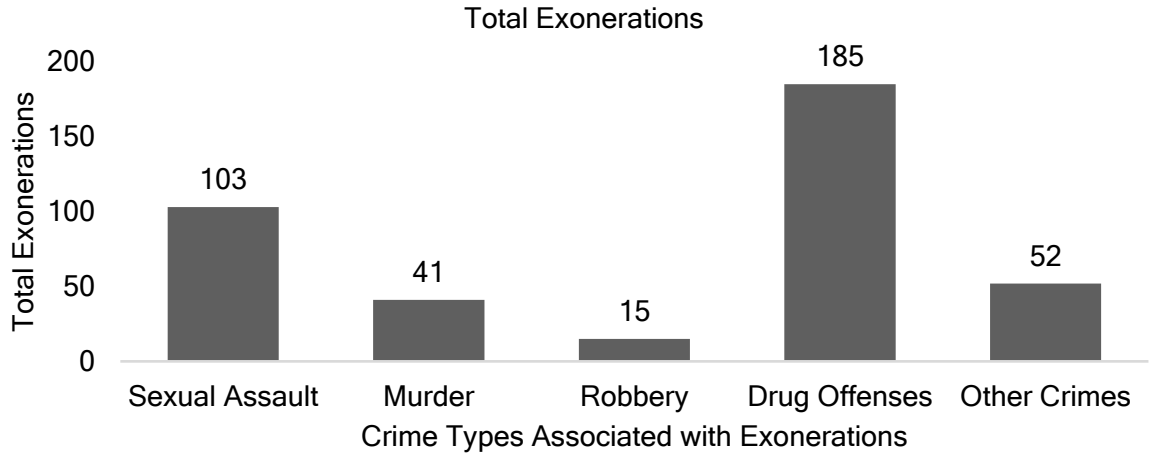
resulted in a longer sentence.[11] We designate the resulting wrongful conviction of Massingill and Dupree as 'Sexual Assault/Rape' since it involved sexual assault allegations though not formally charged as such. We compare the frequency of wrongful convictions in 'Sexual Assault/Rape' to the number in 'Murder' (w/o sexual assault), 'Child abuse', 'Drug offenses', 'Robbery', and the 'Other' category.

Lastly, we ran all descriptive analyses on Stata/SE 17.0. In addition, we provide case illustrations describing prototypical cases of the common themes within the exonerations, such as false guilty pleas, stranger rape misidentification, and youthful complainant recantations.

III Findings

First, as presented in Figure 1, Sexual Assault/Rape exonerations account for a significant portion of Texas exonerations (26.0%, n=103), but Drug offenses (46.7%, n=185) are the most common crime type associated with wrongful conviction in Texas. Additional crime types with substantial proportions among the Texas exonerations are Murder (10.3%, n=41), Robbery (3.7%, n=15), and 'Other Crimes' (13.1, n=52).

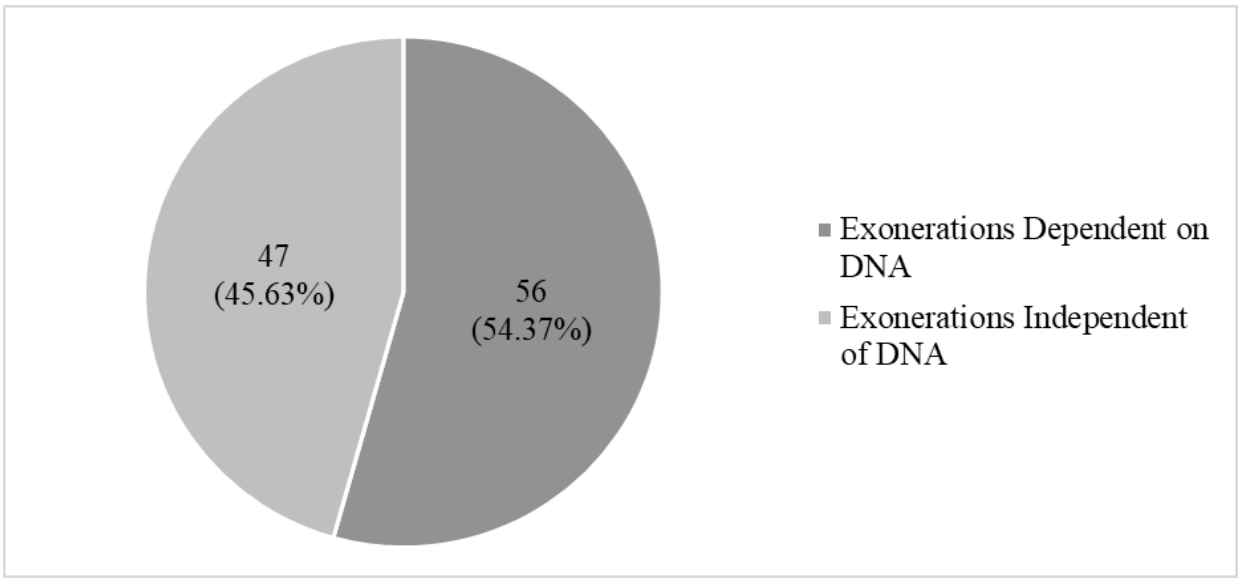
Figure 1. Frequency of Texas Exonerations (N=396) Associated with Crime Type



Second, Figure 2 indicates a substantial portion (46%, n=47) of the n=103 sexual assault/rape exonerations were independent of DNA evidence. This finding suggests prevalence of sexual assaults among exonerations is not (solely) an artifact of testable DNA.

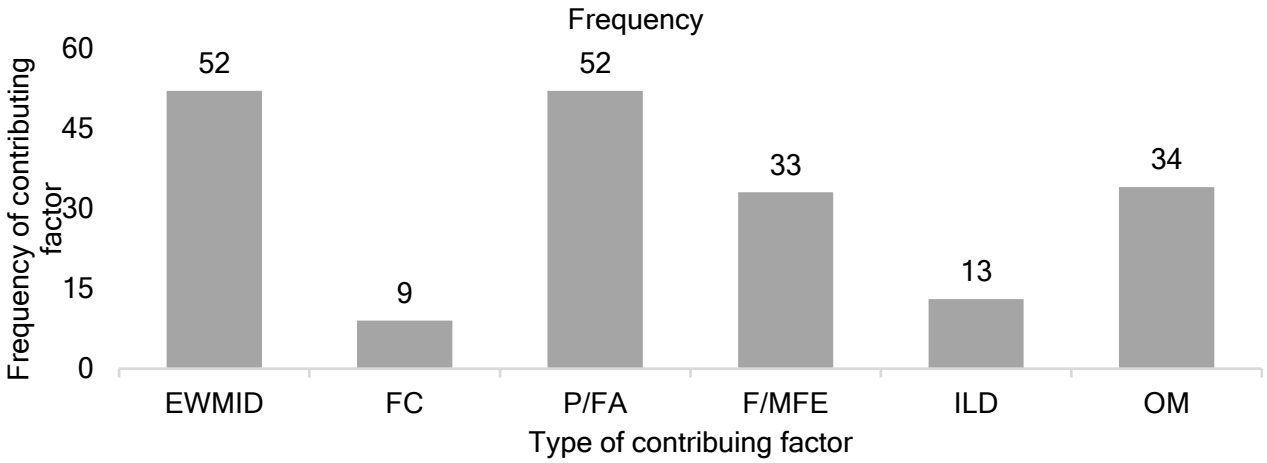
[11] Maurice Possley *"Anthony Massingill"* (2014) National Registry of Exonerations, online: <https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=4534>

Figure 2. Frequency of Sexual Assault/Rape Exoneration (n=103) Independent of DNA Evidence, 47 (45.6%)



Third, Figure 3 illustrates eye-witness mis-identification was a major contributor to the Texas sexual assault/rape exonerations (50.49%, n=52), though perjury/false allegation was an equal contributor (50.49%, n=52). Other contributing features were false confessions (8.7%, n=9), false/misleading forensic evidence (32.0%, n=33), inadequate legal defense (12.6%, n=13), and official misconduct (33.0%, n=34). The total percentages equal more than 100 because most exonerations had several contributing factors.

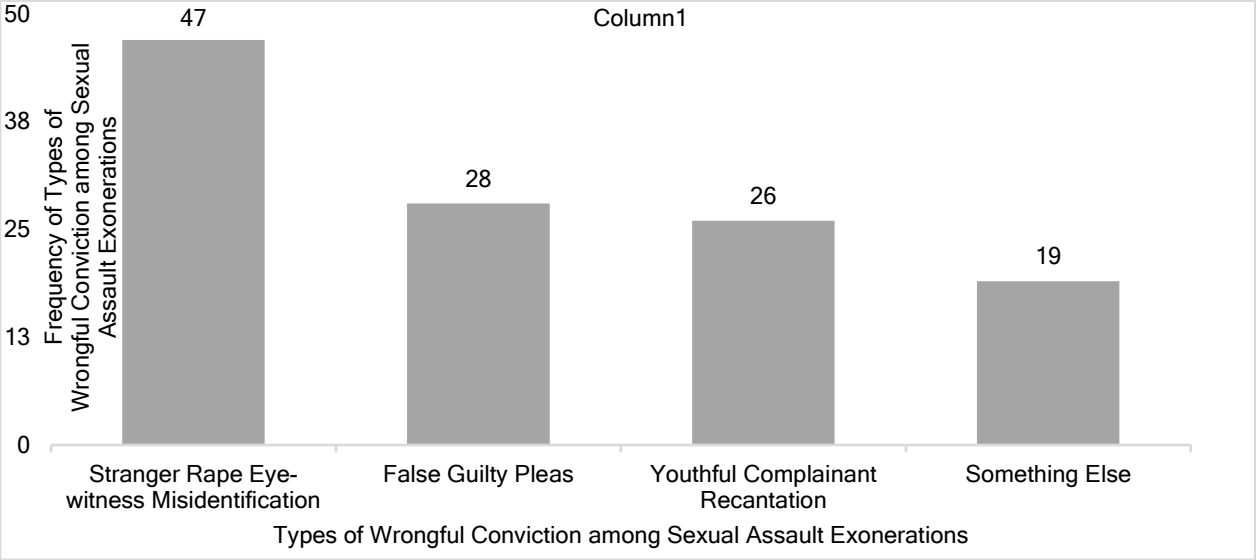
Figure 3. Frequency of Wrongful Conviction Contributing Factors Among the n=103 Sexual Assault/Rape Exonerations



Note. EWMID = eye-witness misidentification; FC = false confession; P/FA = perjury or false accusations; F/MFE = false or misleading forensic evidence; ILD = inadequate legal defense; and OM = official misconduct. Total percentage is > 100 since each case typically has multiple (non-exclusive) contributors.

Fourth, among the 103 sexual assault/rape exonerations, stranger rape eye-witness misidentification was a prominent contributor (45.6%, n=47), accounting for almost half of the cases. False guilty pleas (27.1%, n=28) and youthful complainant recantations (25.2%, n=26) (a particular type of perjury/false accusation) were also significant contributors. Eighteen-point four percent (n=19) of the sexual assault exonerations were attributable to various other sources of error. In addition to false guilty pleas being a major contributor to wrongful conviction in sexual assault/rape, false guilty pleas occurred in virtually all wrongful convictions in drug offenses. Below, we summarize select cases to further illustrate the findings.

Figure 4. Frequency of Stranger Rape Eye-Witness Misidentification, False Guilty Pleas, and Youthful Complainant Recantations Among the n=103 Texas Sexual Assault/Rape Exonerations



Note. There are 17 (16.50%) exoneration cases that have more than one type. For instance, there are youthful complainant recantation exonerations that have false guilty pleas as well.

IV Case Illustrations

A. False Guilty Pleas

Johnson & Cunningham[12] presented 17 innocent defendants who plead guilty to rape charges. They noted five non-exclusive factors (1- death penalty threats; 2 - severe penalties, other than death, combined with vulnerabilities; 3 - false confessions that led to false guilty pleas; 4 - the wrongful conviction of co-defendants that led to false guilty pleas; and 5 - false rape charges) that contributed to the outcome. Below we present two cases of innocent defendants who pled to sexual assault/rape charges, among the 27 noted on the NRE list from Texas.

[12] Johnson & Cunningham *supra* note 9.

1. Blackshire & Johnson

James Blackshire and Antrone Johnson, Black males, aged, 18 and 17 respectively, were accused, in Dallas County, in March of 1994, of sexual assault of a 13-year-old student at the school.[13] In February of 1995, they both entered guilty pleas with "deferred adjudication", meaning if they remained arrest free for 10 years, the conviction would be vacated. They were ordered to pay fines, a monthly sex offender registration fee, and to attend sex abuse therapy. In August, Blackshire was arrested for failing to pay fees and not attending therapy. He pled again to a 40-year sentence. Ten days later, Johnson was accused in a case, similar to the original one. His probation was revoked, and he was sentenced to life. He thereafter pled guilty to the second charge. Johnson's family hired new counsel and, with the assistance of the Dallas Conviction Integrity Unit (CIU), exculpatory evidence, in the separate prosecutor's files, was identified which led to Johnson and Blackshire's exoneration. Specifically, in the first case, the prosecutor's note indicated the day prior to the initial plea, the complainant reported there was no sexual assault, the defendant had taken her in the bathroom, she did not want to do it, so they stayed in there and pretended, before leaving the bathroom. In addition, the prosecutor's file included reports by school personnel indicating the complainant was not reliable and regarded as a "great liar". This evidence was not shared with the defense. Review of the file in the second case against Johnson, which was handled by a different prosecutor, revealed the complainant reported she engaged in sex, in the school, with three other students on the relevant date. This report also was not shared with the defense. The District Attorney joined the defense motion to set aside the convictions and sentences, which was affirmed by the Texas Court of Criminal Appeals. Thus, Blackshire was exonerated for the first charge and the probation violation. Johnson was exonerated for the first charge and the probation violation, though his guilty plea to the 2nd charge was undisturbed. Thus, DNA did not contribute to the exonerations of the sexual assault charges.

2. Dahn Clary

Dahn Clary, Jr., a 41-year-old white male, was arrested on March 13, 1997 and charged with the aggravated sexual assault of his best friend's 11-year-old son, who reported multiple instances of abuse in 1996.[14] Clary pled guilty to aggravated assault on February 28, 1998, receiving 'deferred adjudication' with the condition he complete a sex offender therapy program and remain arrest-free for ten years. According to Clary, his attorney advised him a trial would likely result in a life sentence. However, Clary's deferred adjudication was revoked six years later due to his failure to attend therapy sessions and comply with the program, leading to a prison sentence. The 'victim' later recanted his claim of abuse, admitting to his mother and signing a sworn statement that he fabricated the sexual abuse accusation due to resentment toward Clary's lack of involvement in his life. Subsequently, Clary filed a habeas corpus petition, granting a writ of habeas corpus. Finally, on June 13, 2016, Clary's charges were dismissed. This illustration of false guilty pleas also illustrates a youthful complainant recantation and a sexual assault/rape exoneration, independent of DNA evidence.

---

[13] Maurice Possley, *James Blackshire,* National Registry of Exonerations, online: <https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=3934>.

[14] Possley, Maurice, *"Dahn Clary, Jr."* National Registry of Exonerations, online: <https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=4923>

B. Youthful Complainant Recantation

Johnson [15] discussed the matter of sexual accusations that are later recanted. Recantation may be associated with an initial false allegation but there are circumstances where actual victims recant for a variety of reasons such as pressure from defendants or others, dissatisfaction with the prosecution process, or desire to protect the defendant from legal consequences. Whether a recantation reflects a false accusation requires assessment of the specific context and situation. The series of exonerations from Texas include 26 cases where recantations from youthful complainants were deemed credible and/or corroborated, and thus noted in the exoneration. Two examples are provided.

1. San Antonio Four

The San Antonio Four (Elizabeth Ramirez, Kristie Mayhugh, Cassandra Rivera, and Anna Vasquez), Hispanic females, were 'out' lesbians, and charged with sexually molesting Elizabeth Ramirez's two nieces.[16] The prosecution considered Ramirez to be the 'ring-leader'. Ramirez was convicted first, in 1997. The other three defendants refused plea offers and were convicted together in 1998. The convictions were based on false/misleading forensic evidence (F/MFE) regarding alleged injury to one child's hymen as well as 'victim' testimony. Post-conviction investigation in 2010 found the younger victim recanted, explaining she was coerced to make the false accusation by her father who had been romantically rejected by Ramirez. After additional medical consultation, the state's original medical expert reversed her earlier trial testimony acknowledging it was mistaken, ultimately resulting in an exoneration independent of DNA. The defendants were released in 2013. In 2016, they were officially exonerated by the Texas Court of Criminal Appeals and awarded compensation. An overview of the wrongful convictions and subsequent exonerations is provided in the documentary series, "Southwest of Salem".[17]

2. Tony Hall

Twenty-Five-year-old, white male, Tony Hall was charged with fondling a 7-year-old child whom he was baby-sitting in 1992.[18] Hall denied the charges, passed a polygraph exam, rejected a plea offer, and proceeded to a bench trial. Hall was convicted of aggravated sexual assault and sentenced to 15 years in prison. While imprisoned, he endured physical and sexual assaults. Hall was routinely denied parole because he refused to admit guilt. Hall served his entire sentence, was released in 2008, and required to register as a sexual offender. Two years after his release, Hall had an encounter with this accuser, who was a young adult. The accuser reported a vague memory of the trial and no awareness that Hall had been imprisoned. The accuser reported a clear memory

---

[15] Johnson, *supra* note 2.

[16] Maurice Possley, *"Elizabeth Ramirez"* (2016) National Registry of Exonerations, online: <https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=5048>

[17] Deborah Esquenazi, "*Southwest of Salem: The Story of the San Antonio Four*" (2016) Exoneration Detail List, National Registry of Exonerations, online: <https://www.law.umich.edu/special/exoneration/Pages/detaillist.aspx>

[18] Maurice Possley, *"Tony Hall"* (2012) National Registry of Exonerations, online: <https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=4025>

his mother had insisted he accuse Hall and physically beat him to comply. The accuser provided an affidavit. With the assistance of counsel, Hall obtained testimony from the accuser's aunt (the mother's sister) that she observed the mother physically beat the child to force him to make the allegations against Hall (who was a distant relative). Though the mother denied this, the State District Judge set aside the conviction and it was affirmed by the Texas Criminal Court of Appeals.

As described in Johnson[19], and illustrated above, youthful complainant recantations (YCRs) are a particular type of Perjury/False Allegation. A review of cases suggests different patterns. For instance, a youthful complainant may fabricate an assault and make a private disclosure without intending there will be legal consequences. However, the party who receives the disclosure, directs and/or pressures the complainant to make an official report which leads to criminal investigation and erroneous prosecution, as in the Gary Dotson case and the Brian Banks case. There are other cases, such as the Dahn Clary case, where youthful complainants fabricate the charges stemming from their own hostility toward the accused. There is also a third type where youths become pawns in disputes among adults and are persuaded/coerced to make false allegations, as in the San Antonio Four and Tony Hall cases. Regardless of how the false allegation emerged, complaints from youths generate sympathy, the youths are usually naïve to the consequences of false accusations, and youths are especially vulnerable to coercion.

C. Stranger Rape, Misidentification

As noted above, Johnson[20] reported wrongful convictions in sexual assaults were concentrated among stranger rape prosecutions which comprise 72% of all IP exonerations. In addition, Johnson described an *incapacitated victim-false confession path* to wrongful conviction as well as a distinct *capable victim-misidentification path*. In the capable victim, misidentification path, the rape victim/witness misidentifies an innocent person as the assailant. As described by Johnson, the erroneous identification is commonly associated with biased identification procedures, cross racial identification challenge, [21] and other factors (elaborated in the discussion). There are 47 prosecutions with this characteristic among the series of 103 sexual assault/rape exonerations from Texas. Since the NRE does not routinely report the race/ethnicity of the offense victim, the frequency of cross racial misidentification in this series is not apparent. A feature that clearly demonstrates the bias in the identification process is multiple eye-witness misidentification (MEM). One case illustrating cross racial misidentification is provided below, followed by a second case with MEM.

[19] Johnson, *supra* note 2.

[20] Johnson, *supra* note 2.

[21] Laura Connelly, "Cross- racial identifications: Solutions to the "they all look alike" effect" (2015) 21:1 Mich J Race & L 25, online: <https://doi.org/10.36643/mjrl.21.1.cross-racial>; Innocence Project "*Re-evaluating Line-ups: Why Witnesses Make Mistakes and How to Reduce the Chance of a Misidentification",* An Innocence Project Report (2009) online: <https://innocenceproject.org/reevaluating-lineups-why-witnesses-make-mistakes-and-how-to-reduce-the-chance-of-a-misidentification/>

1. Michale Phillips

In September of 1990, a 16-year-old white teen was raped in a Dallas motel room.[22] According to the victim, in the struggle, she was able to remove the assailant's ski mask, and recognized him as (36-year-old) Michael Phillips, a Black male, and former resident of the motel. A month later she selected Phillips from a six-photo line-up. The following month, Philips accepted a guilty plea with a 12-year sentence. Philips had a prior conviction for burglary. Phillips later reported he was advised by counsel to accept the plea because the jury would surely take the word of the white victim and he would be sentenced to life. Two years after serving the 12-year sentence, Phillips pled guilty to failing to register as a sex offender. While detained, he submitted a pro se habeas motion asserting DNA testing would exonerate him. The motion was denied. In 2007 after the Dallas County CIU was established, all county sexual assault convictions where available DNA had not been tested where examined. Phillips was excluded as a contributor of the semen from the vaginal swab, and it was matched via CODIS to Lee Marvin Banks a felony offender, who admitted residing at the motel during the relevant time period though he denied committing the offense. Phillips' convictions were vacated, and he was awarded compensation.

This was a stranger rape, misidentification wrongful conviction that also involved cross-racial misidentification and a false guilty plea. Phillips was exonerated from the original rape charge as well as the violations associated with the conditions of his parole. The exoneration was dependent on DNA evidence.

2. Ricardo Rachell

On 10/20/02, an eight-year-old, Black male child, was lured with an offer to earn money by a Black male adult, riding a bicycle, in southeast Houston. The child was anally raped by the perpetrator and discovered by neighborhood adults running and crying. The child was taken home and initially reported someone had tried to kill him with a knife.[23] The child, and a six-year-old who had seen the man on the bike, described him as a Black man about 30 years old. The following day the child's mother left the house, without the child victim, and observed Ricardo Rachell in the neighborhood. She suspected the Rachell was the perpetrator. She returned home, got the child, and the child confirmed Rachell was the assailant. Rachell had a pronounced facial disfiguration from a prior gunshot injury. The police apprehended Rachell and the child again identified Rachell, seated in the rear of the police car. Later that day, the child was interviewed, the sexual nature of the assault was discovered, and a rape kit was secured. Following a police/prosecution consultation, Rachell was arrested 10/24/02 and voluntarily provided DNA samples. The DNA samples were never tested. On 11/16/02, while Rachell was in jail, another eight-year-old, Black male child was sexually assaulted in the same community, with a similar MO. Rachell was convicted 6/03/03 at trial, with the in-court identifications by the two children (MEM). Two appeals on Rachell's behalf were rejected. On 10/23/03 the Houston Police Department, Juvenile Sex Crimes Unit, behaviorally linked the 11/16/02 assault with another sexual assault on an African American boy in southeast Houston. On 4/08/04, Andrew Wayne

[22] Maurice Possley *"Michael Phillips"* (2014), National Registry of Exoneration, online: <https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=4483>

[23] Johnson, *supra* note 2.

Hawthorne pled guilty to the three assaults and was subsequently DNA matched to the evidence from the crime scene that resulted in Rachell's conviction. However, Rachell was not released until 2009.[24] He was exonerated and compensated in 2011.

This was a stranger rape misidentification with two witnesses making the mistaken identification. MEM is not uncommon, occurring in 38% of confirmed eye-witness misidentifications.[25] Confidence in the (faulty) identification increases with the repeat presentation of the suspect, even though the reliability of the identification is tainted/compromised. The Rachell prosecution also included the conviction of an innocent for crime committed by a serial sex offender, a common feature in 67 stranger rape wrongful convictions.

D. Drug Offenses

Gross (2018) commented the Texas misdemeanor and felony drug offense exonerations, "… have the same basic plot".[26] An alleged illegal substance is seized from the search of a car. A passenger or driver is charged and arrested. If the defendant has priors, the bail is prohibitively high. While in jail, the defendant is offered a plea deal which involves release the same day or within weeks. A not-guilty plea will result in months of detention prior to trial, with an uncertain outcome. Many defendants, whether they believe they are guilty or innocent, take the plea offer so they can go home. Gross explained these exonerations were concentrated in Harris County (the Houston area), the only county that conducted post-adjudication lab testing of controlled substances. It was often found the substances were not the alleged illegal drugs.

1. Meghan Alegria

Meghan Alegria, a 20-year-old, white female, was arrested 11/18/21 and charged with possession of 'PCP' (phencyclidine) in her confiscated cigarettes. On 11/21/11 she pled to possession of a controlled substance with a 3-year parole sentence. In 2014 the Harris County District Attorney learned lab tests results from cases that had been resolved were not being forwarded to the prosecuting attorneys. The sample from Alegria's case was tested January 27, 2012 and found to be negative for illegal substance. The defense attorney was informed and this eventually led to Alegria's exoneration in 2020.

V Discussion

We reviewed the NRE data on all exonerations in the State of Texas as of March 2022 (N=396), to determine if sexual assault/rape was the predominant crime type. Unexpectantly, we found Drug offenses (46.7%) were the most common offense type though sexual assault/rape prosecutions did comprise a substantial proportion (26.0%) of the exonerations. A considerable

[24] Maurice Possley, *"Ricardo Rachell"* (2012) National Registry of Exonerations, online: <https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=3559>

[25] Innocence Project, *supra* note 21.

[26] Samuel Gross, "Errors in Misdemeanor Adjudication" (2018) 98:3 *BUL Rev*, at 999-101, online: <https://repository.law.umich.edu/articles/2003>

portion (46%) of the n=103 sexual assault/rape exonerations were independent of DNA evidence, suggesting the disproportionate frequency of sexual assault/rape prosecutions among the Texas exonerations was not a mere artifact of testable DNA. The leading evidence types among the n=103 sexual assault/rape wrongful convictions were eye-witness misidentification and perjury/false accusation, each occurring in about half the cases. Many cases had multiple contributing factors (such as an eye-witness misidentification and official misconduct). Among the 103 sexual assault exonerations, stranger rape misidentification was the most common pattern, evident in 45.6% (n=47) of the cases. There were also a substantial number of false guilty pleas (n=28, 27.1%) and youthful complainant recantations (25.2%, n=26). False guilty pleas also occurred in all the wrongful convictions in drug offenses.

The loss of freedom stemming from false guilty pleas ranges across different types of crimes (drug offenses, sexual assaults, and murder). The Christopher Ochoa false guilty plea to an Austin, Texas rape and murder[27] is only one of several well-known cases where innocent defendants pled guilty to serious crimes.

In the past decade there has been increased attention to the role of false guilty pleas in wrongful convictions.[28] Several remedies to reduce the occurrence of innocents pleading guilty have been advanced such as open file discovery, enhancing 'voluntariness' by increasing attorney advisement and decreasing time pressure prior to pleas, and recognizing unique due process vulnerability among juvenile defendants. We also add, increased funding for public defense (as it relates to attorney compensation, caseload size, and available resources).

As noted above, YCRs are a particular type of Perjury/False Allegation. YCR played a significant role in many of the Texas wrongful convictions in sexual assault. Also noted above, these false accusations can arise from a variety of circumstances which warrant assessment when evaluating the evidence in a given case. These accusations pose a challenge for investigators (both defense or prosecution) because thorough assessment of the complaint could have the untoward effect of discouraging victims from reporting assaults, which is also an adverse outcome. This recognition suggests a need for special training to enable investigators to remain supportive of complainants, while assessing their complaints.[29] While adults also recant complaints, the heightened vulnerability of youth recognized in other criminal investigative contexts (i.e. false confessions) warrants attention to youthful complainant accusations.

Early wrongful conviction scholar Borchard (1932) used the term 'manufactured evidence' to refer to instances where investigators (intentionally) framed suspects to seal convictions. Johnson [30] describes how the 'manufacture of evidence' can occur inadvertently, as well as intentionally. Johnson presented a 'continuum of intentionality' in the manufacture of evidence,

---

[27] Innocence Project *"America's Guilty Plea Problem Under Scrutiny"* (2017), online: <https://innocenceproject.org/americas-guilty-plea-problem-scrutiny/>

[28] Redlich *et al*, *supra* note 8; Fessinger & Kovera, *supra* note 8; Zottoli *et al*, *supra* note 8; Tarika Daftary-Kapur & Tina M. Zottoli "A first look at the plea deal experiences of juveniles tried in adult court" (2014) 13:4 Int J Forensic Ment Health 323, online: https://doi.org/10.1080/14999013.2014.960983

[29] Johnson, *supra* note 2.

[30] Johnson, *ibid.*

ranging from spontaneous misidentifications [31] by victim/witnesses that occur independent of law enforcement identification procedures on the left, to situations where police fabricate evidence and/or provide intentionally false testimony on the right. Johnson also elaborated the 'black box' character of this manufacture of evidence, where there is an absent, or limited, record of the circumstances that led to a witness identification, or a confession, or the testimony by an informant, or the opinion of an 'expert' or analyst.

With specific regard to eye-witness misidentifications, the notion these misidentifications are "encapsulated" witness errors [32] is contrary to the essential recognition of the law enforcement investigation procedures in the misidentification. We suggest there are features of stranger sexual assaults that can make eyewitnesses more prone to error. The stress/trauma and profound personal violation of the assault is one. The police and public reaction to the offense, including incentives/rewards for an arrest and conviction, is another.[33]

When the teen sexual assault victim picked out Michael Phillips from the six-photo spread, she was merely demonstrating she could distinguish Michael Phillips from the five other photos. It did not indicate he was her assailant. However, the procedures made her identification appear reliable. This is how her identification of Phillips, as her rapist, was 'manufactured', albeit unintentionally. Similarly, the two witness misidentifications of Ricardo Rachell (MEM) were manufactured by the police investigation where the rape victim's mother, who had never seen the assailant, developed a hunch the disfigured man had attacked her son. The traumatized, 8-year-old victim, concurred with his mother's impression, and identified Rachell seated in the rear of the police car. At trial the prosecution had two eye-witnesses (the eight-year-old and the six-year- old) provide the critical in-court identifications. MEM, where multiple witnesses are making the same erroneous identification, is a clear indicator of (forensic confirmation) bias in the criminal investigation.

Since the actual offender (culprit) was absent in each of these identifications, there was considerable risk an innocent person would be (mis-)identified as the assailant. Wrongful conviction eye-witness researchers [34] have observed, "…the dangers of misidentification increase dramatically when the actual culprit is not included [in the] identification procedure". Further, this research indicates the 'base rate' (or frequency) of culprit absent line-ups determines the likelihood any particular identification is accurate. Wells & Quigley-McBride calculated from available lab studies, "…if the culprit-present base rate was 75%, then the chance that an identified suspect was innocent was 9%. However, if the culprit-present base rate was only 25%, then the chance that an identified suspect was innocent ballooned to a whopping 53%" (p. 292). The researchers explain culprit absent line-ups in real world police investigation is not at all uncommon. A study of actual line-ups conducted by the Houston Police Department estimated 65% of witnesses were shown

[31] Matthew B. Johnson & Sydney Melendez "Spontaneous Misidentification in Wrongful Rape Conviction" (2019) 37:3 *Am J of Forensic Psychol*, 5-20, online: <https://psycnet.apa.org/record/2020-50561-001>

[32] Johnson, *supra* note 2 at 84.

[33] Johnson, *supra* note 2.

[34] Gary Wells & Adele Quigley-McBride "Applying Eyewitness Identification Research to the Legal System: A Glance at Where We Have Been and Where We Could Go?" (2016) 5:3 JARMAC, at 292 online: <https://doi.org/10.1016/j.jarmac.2016.07.007>

culprit absent line-ups.[35] Therefore, educating law enforcement and legal authorities about the inherent risk of misidentification where there is a culprit absent line-up (or show-up) has emerged as a clear policy objective.

Other researchers [36] have reported data that suggest the increased rates of misidentification of Black suspects [37]-is not (primarily) a result of cross-racial identification error but rather due to law enforcement investigators' lower threshold of evidence before placing a Black, as opposed to a white, suspect in a line-up. The recognition of these base rates considerations is not novel, nor unique, to wrongful conviction connected to eye-witness misidentification. Gudjonsson noted, "… the rate of false confessions in a given population is dependent, to a certain extent, on the base rate of guilty suspects interrogated".[38] Where the base rate of guilty suspects interrogated is high, the risk of false confession will be low, and where the base rate of guilty suspects is low, the rate of false confessions will be high.

What can exonerations tell us about wrongful convictions? Single exonerations, not so much, but the aggregate data available through the NRE and IP are valuable sources in the effort to understand, prevent, and challenge wrongful convictions. Contributions from multiple sources, archival data (such as the NRE and IP), as well as controlled lab research, social and historical investigations all advance knowledge in the field.

Two notable limitations to the data presented in this report warrant mention. First, we were unable to ascertain the frequency and proportions of criminal convictions, in the State of Texas, associated with different crime types. These data would provide a more complete picture of the relevance of exonerations associated with different crime types and features. Second, we did not measure inter-rater reliability as applied to the crime type classification and features reported. We hope to have these data available for future reports.

---

[35] John T. Wixted *et al*, "Estimating the reliability of eyewitness identifications from police lineups" (2015) 113:2 PNAS 304, online: <https://doi.org/10.1073/pnas.151681411>

[36] Jacqueline Katzman & Margaret B. Kovera, "Potential Causes of Racial Disparities in Wrongful Convictions Based on Mistaken Identifications: Own-Race Bias and Differences in Evidence-Based Suspicion" (2023). 47:1 Law & Hum Behav 23, online: <https://doi.org/10.1037/lhb0000503>; Margaret B. Kovera "The role of suspect development practices in eyewitness identification accuracy and racial disparities in wrongful conviction" (2023) 18:1 Soc Issues Policy Rev, 1–23, online: <https://doi.org/10.1111/sipr.12102>

[37] Connelly, *supra* note 21.

[38] Gisli H. Gudjonsson, "*The Psychology of Interrogations and Confessions*" (West Sussex: John Wiley & Sons, 2003) at 173.

"Not Scientific" to Whom? Laypeople Misjudge Manner of Death Determinations as Scientific and Definitive

Jeff Kukucka, Oyinlola Famulegun
Department of Psychology, Towson University,
Towson, MD
U.S.A.

When someone dies unexpectedly, a medical examiner may perform an autopsy to determine how they died (i.e., manner of death). Recent studies found that cognitive bias can affect manner of death judgments, such that extraneous non-medical information may cause the same death to be judged as either a homicide or accident, which has significant legal ramifications. In response, leading medical examiners clarified that manner of death is "not scientific" and "often does not fit well in court." Yet medical examiners often testify in court, and little is known about how factfinders appraise their judgments. To address this gap, we conducted two experiments in which mock jurors read and evaluated a medical examiner's testimony at a murder trial (modeled after the real-world case of Melissa Lucio), while varying the expert's opinion (i.e., homicide or accident) and the defendant's attributes (i.e., an affluent white or underprivileged Latina woman). Overall, participants rated the medical examiner's testimony as highly scientific, credible, and convincing, and it strongly affected their verdicts and belief in the defendant's guilt, irrespective of the defendant's attributes. Moreover, participants unexpectedly rated the expert as even more credible if they ruled the death a homicide rather than an accident. Our data thus reveal a worrisome disconnect between how medical examiners characterize their judgments (i.e., as nonscientific and tentative) and how jurors appraise those judgments (i.e., as highly scientific and practically dispositive). We discuss ways to remedy this disconnect, including reforming death investigation practices to curtail bias and encourage standardization and transparency.

I. Introduction
II. Method
  A. Participants and Design
  B. Procedure
  C. Materials
    a. Narrative Summary
    b. Expert Testimony
  D. Measures
    a. Ratings of Expert Testimony
    b. Guilt Judgments
    c. Comprehension Test
III. Results
  A. Analytic Plan
  B. Verdicts
  C. Likelihood of Guilt
  D. Ratings of Expert Testimony
IV. Discussion

A. Policy Recommendations
V. References

I Introduction

Jurors tend to believe that expert judgments of forensic evidence are objective, dispositive, and virtually infallible (e.g., Crozier *et al*., 2020; Martire *et al*., 2019; Koehler, 2017). However, it is now well-established that forensic experts' judgments are susceptible to *cognitive bias*, such that extraneous influences (e.g., ordering of information, knowledge of irrelevant case facts or theories) can elicit different interpretations of the same evidence (see Kukucka & Dror, 2023). In one study, for example, firearms experts changed 28% of their opinions of the same bullets when given different information about where those bullets were found (Kerstholt *et al*., 2010). Cognitive bias has likewise been shown to affect expert judgments of DNA mixtures, fingerprints, toxicology results, bloodstain patterns, skeletal remains, and digital evidence, among others (see Kukucka & Dror, 2023). When cognitive bias prompts conflicting opinions of the same evidence, one of those opinions must be incorrect, which can produce a miscarriage of justice—and indeed, forensic science errors have now been implicated in over 850 wrongful convictions (National Registry of Exonerations, 2023), including 53% of DNA exonerations (Innocence Project, 2023).

Medicolegal judgments are likewise susceptible to cognitive bias, such that extraneous information can lead medical experts to interpret the same injury as either criminal or accidental (see Kukucka & Findley, 2023). In one study, for example, physicians and nurses more often misdiagnosed an accidental pediatric leg fracture as child abuse if led to believe that the child's parents were unmarried and had a strained relationship—information that should have no bearing on their medical judgment (Anderst *et al*., 2016). Similarly, another study found that hospital staff more often interpreted the same radiographs as evidence of child abuse if given prejudicial but medically irrelevant information about the child's parents (e.g., that they were unemployed, unmarried, immigrants, drug users, or welfare recipients; Loos *et al*., 2021).

Dror *et al*. (2021) showed that cognitive bias can also affect postmortem *manner of death* determinations (i.e., opinions as to whether a sudden unexpected death was an accident, homicide, suicide, natural, or indeterminate—which, depending on the jurisdiction, may be rendered by a coroner, forensic pathologist, or medical examiner). In their study, medical experts read a case file about a child who was found unresponsive at home, taken to a hospital, and died soon thereafter, and they were asked to opine as to how the child had died. All of the experts received the same information about the child's injuries and medical history; however, some were told that the child was white and brought to the hospital by her grandmother, while others were told that the child was Black and brought to the hospital by her mother's boyfriend. These extraneous details had a strong effect on manner of death opinions: Despite identical medical information, experts in the latter group were about 12 times more likely to rule the child's death a homicide.

Dror *et al*.'s findings sparked strong and varied reactions from the medical examiner community. Some were supportive, lauding the researchers for "call[ing] attention to the powerful impact that bias can have on [medical examiners'] diagnostic decisions" and "offer[ing] important suggestions on how to mitigate the impact of bias" (Graber, 2021; see also Duflou, 2021; Obenson,

2021). Others were extremely hostile, dismissing Dror *et al*.'s results as invalid and politically motivated, levying baseless allegations of research misconduct (all of which were thoroughly investigated and found to have no merit), demanding that the paper be retracted, and threatening legal action against the authors, journal, and publisher (see Peat, 2021).

Importantly, Dror *et al*.'s critics also took this opportunity to clarify the nature and purpose of manner of death determinations. A commentary by 86 prominent forensic pathologists from 29 U.S. states (including the then-President of the National Association of Medical Examiners) explained that "manner determination is not a 'scientific' determination" but rather "a cultural determination… for the purpose of public health statistics" for which "there is no 'right' answer [or] criteria for correctness" and which "often does not fit well in court" (Peterson *et al*., 2021a). Other medical examiners expressed similar viewpoints, including that "manner determination is a nonscientific determination for statistical purposes [such that] any individual determination is questionable" (Oliver, 2021). Indeed, medical examiners often reach different opinions of the same death even in the absence of biasing extraneous information: In one study, for example, 152 experts determined manner of death for 30 case scenarios, only four of which produced greater than 90% consensus, while 17 showed less than 70% consensus (Hanzlick *et al*., 2015).

Nevertheless, medical examiners *do* present manner determinations in courtrooms, where jurors likely consider them highly scientific just as they do other forensic expert testimony. Take for example the case of Melissa Lucio, who is the first Hispanic woman to be sentenced to death in Texas. In February 2007, Melissa's two-year-old daughter Mariah fell down a flight of stairs and suddenly died two days later. Melissa was later convicted of murdering her daughter based largely on testimony from Dr. Norma Jean Farley, who performed Mariah's autopsy and ruled her death a homicide, opining that her injuries could only have resulted from abuse.

Critically, however, two police officers had attended Mariah's autopsy and told Dr. Farley in advance that Melissa had allegedly confessed to abusing her daughter, which raises concern that Farley's determination was biased by this and/or other extraneous information (e.g., that Melissa was an unmarried, impoverished, former drug user). Accordingly, numerous independent experts have since repudiated Farley's testimony in this case; as one pathologist wrote, "the investigation into Mariah's death appears to have been significantly prejudiced, not evidence based, and without an adequate consideration of alternative issues" (Ophoven, 2022).

Unfortunately, jurors are poor at recognizing and discounting biased forensic expert testimony (Kukucka *et al*., 2020; Thompson & Scurich, 2019)—or worse, jurors may actually favor biased testimony insofar as it reinforces their own biases. Research on *stereotype congruence* has shown that the degree of correspondence between a defendant's traits and the stereotypic traits of a criminal offender affects how jurors process and appraise information (Jones & Kaplan, 2003; McKimmie *et al*., 2013), including forensic science evidence (Smalarz *et al*., 2016). As a result, jurors may value the same evidence differently depending on the defendant's characteristics; in one study, for example, jurors found a confession more compelling if the defendant fit a criminal stereotype, even if the confession was coerced (Smalarz *et al.*, 2018). Similarly, jurors may deem the same expert testimony more credible—regardless of its objective validity—when the defendant's characteristics stereotypically congruent with the expert's opinion.

We therefore designed the current studies with two aims. First, in light of pathologists' assertions that manner of death determinations are not scientific, we aimed to test whether jurors feel similarly—or whether there is a worrisome disconnect between how experts and non-experts regard manner determinations. Second, given that stereotypes and other extraneous information have been shown to affect both expert and juror decisions, we aimed to test whether jurors' evaluations of the same medical expert testimony depend on the stereotypic congruence between the expert's opinion and the defendant's traits—e.g., jurors may consider inculpatory testimony more credible if the defendant matches a criminal stereotype, and vice versa.

To address these questions, we conducted two experiments (one with college students and one with community members) that followed the same procedure and used materials modeled closely after the aforementioned case of Melissa Lucio. In both studies, participants acted as mock jurors in the trial of a mother who was accused of murdering her daughter. First, participants read background information about the case, which described the defendant as either an affluent white woman or an underprivileged Latina woman. Then, participants read the trial testimony of the medical examiner who performed the child's autopsy, which described the child's injuries and ruled the death either an accident or a homicide. Finally, participants rendered a verdict, estimated the likelihood of the defendant's guilt, and rated the degree to which the medical examiner's testimony was scientific, credible, and convincing. Our hypotheses were as follows:

> **H1:** The medical examiner's testimony will affect participants' belief in the defendant's guilt, such that they will more often judge the defendant as guilty if the medical examiner testifies that the child's death was a homicide rather than an accident.
>
> **H2:** Participants' evaluations of the medical examiner's testimony will depend on the stereotypic congruence between the manner of death determination and the defendant's characteristics. If the medical examiner rules the child's death a homicide, jurors will rate their testimony more favorably if the defendant is an underprivileged Latina woman as opposed to an affluent white woman. Conversely, if the medical examiner rules the child's death an accident, jurors will rate their testimony more favorably if the defendant is an affluent white woman as opposed to an underprivileged Latina woman.

II Method

All materials are available on the Open Science Framework (http://osf.io/jekvm).

A. Participants and Design

Study 1 included 319 college students who completed the study online for course credit. Study 2 included 251 community members who completed the study online via Prolific for a $3.00 credit. In both studies, each participant was randomly assigned to one of four cells in a 2 (Stereotype: Innocent vs. Guilty) X 2 (Opinion: Accident vs. Homicide) between-person design. We later excluded data from participants who failed a comprehension test (n = 107 in Study 1; n = 32 in Study 2), leaving samples of N = 212 and 219 for Studies 1 and 2, respectively.

Study 1 participants were mostly female (81.1%; 15.6% male, 3.3% non-binary) and white (46.7%; 27.8% Black, 8.5% Hispanic), with a mean age of 19.1 (*SD* = 2.8; range = 18 – 43). Study 2 participants were also mostly female (53.4%; 40.2% male, 4.6% non-binary) and white (64.8%; 11.0% Black, 9.6% Hispanic), with a mean age of 38.6 (*SD* = 13.1; range = 18 – 75).

B. Procedure

Both studies followed the same procedure. First, participants were asked to imagine being a juror in the murder trial of a woman who is charged with killing her two-year-old daughter, and they read a narrative summary of the child's death and the ensuing investigation (approximately 500 words). By random assignment, this narrative described the defendant as either an affluent, married, white stay-at-home mother (*Innocent Stereotype* condition) or a poor, unmarried, unemployed Hispanic mother with a history of drug use (*Guilty Stereotype* condition), but all other details of the narrative (as explained below) were identical between conditions.

Next, participants read a 12-page transcript of trial testimony from the medical examiner who performed the child's autopsy, in which she described her credentials and the child's injuries. By random assignment, the medical examiner concluded that the child died from either falling down the stairs (*Accident* condition) or being physically abused (*Homicide* condition), though her description of the child's injuries was identical between conditions.

After reading the transcript, participants rated the degree to which they found the medical examiner's testimony to be scientific, credible, and convincing. Then, they rendered a verdict in the case and estimated the likelihood that the defendant had killed her daughter. Lastly, participants completed a comprehension test and a demographic questionnaire.

C. Materials

a. Narrative Summary

By random assignment, each participant first read one of two versions of a narrative that provided background information about the case. The two versions differed in terms of how they described the defendant. In the *Innocent Stereotype* condition, the defendant was Natalie Martin—a 38-year-old woman who was born in the U.S., graduated college, and worked as a museum curator until becoming pregnant with her first child. She is now a stay-at-home mom who owns a house with her husband and two children. In the *Guilty Stereotype* condition, the defendant was Natalia Martinez—a 38-year-old woman who immigrated to the U.S, dropped out of high school, and worked as a museum janitor until entering a drug treatment program. She is now unemployed and rents an apartment with her boyfriend and two children from prior relationships.

All other details of the narrative were identical between conditions and were modeled after the case of Melissa Lucio. In both versions, the defendant was unable to wake her two-year-old daughter Olivia from a nap, so she called 911, but emergency personnel were unable to resuscitate Olivia, and she was pronounced dead at 5:30pm. Police then became suspicious of the defendant because of her body language and "unusually calm demeanor," so they interviewed her at the police station from 9:30pm until 3am, during which they accused her of abusing her daughter.

Although she denied this at first, the defendant eventually "accepted responsibility" for her daughter's death, but she later recanted that statement, claiming that police had coerced her. Nonetheless, the defendant was charged with capital murder and pleaded not guilty.

The narrative then explained that the defense is arguing that Olivia's injuries and death were due to her falling down a "steep flight of wooden stairs" two days earlier. The prosecution is not disputing that Olivia fell down the stairs, but they are arguing that her injuries and death were due to physical abuse. Lastly, the narrative explained that participants would now read and evaluate the trial testimony of the medical examiner who performed Olivia's autopsy.

b. Expert Testimony

Next, participants read a 12-page transcript of the trial testimony of the medical examiner who performed Olivia's autopsy, who we named Dr. Claudia J. Farris. To develop this transcript, we began with the original 55-page transcript of Dr. Norma Jean Farley's testimony at Melissa Lucio's 2008 murder trial, changed the names of the individuals involved and/or discussed, and removed portions that were purely procedural (e.g., discussing exhibit numbers, objections, sidebars) or redundant (e.g., repeated questions). Then, we truncated the remainder of the transcript in a way that maintained its overall content and structure: Dr. Farris explained her qualifications (~1 page), explained the duties of a forensic pathologist in general terms (~1 page), described the injuries that she observed during Olivia's autopsy (including her manner of death determination; ~7 pages), was cross-examined (~2 pages), and was briefly re-directed (~1 page).

Participants who were randomly assigned to the *Homicide* condition read this truncated version of the original transcript, which consisted of near-verbatim statements made by Dr. Farley (now Dr. Farris) that described Olivia's injuries as non-accidental (e.g., "This child was severely abused"), refuted the alternative (e.g., "The pattern of injuries from a fall generally wouldn't look like this"), and ruled her death a homicide (e.g., "The manner was homicide"). Dr. Farris then conceded on cross-examination that it is possible for a child to die from falling downstairs, and then reiterated on re-direct her conclusion that Olivia's death was the result of abuse.

Participants who were randomly assigned to the *Accident* condition read a minimally altered version of this transcript in which we changed Dr. Farris' interpretation of Olivia's injuries, but not her description of those injuries. All told, we changed only 106 of the 3,322 words (3.2%) between the two versions of the transcript. Paralleling the Homicide transcript, the Accident transcript described Olivia's injuries as accidental (e.g., "This child suffered a severe fall"), refuted the alternative (e.g., "The pattern of injuries from abuse generally wouldn't look like this"), and ruled her death an accident (e.g., "The manner was accident"). Dr. Farris then conceded on cross-examination that it is possible for a child to die from being physically abused, and then reiterated on redirect her conclusion that Olivia's death was the result of a severe fall.

D. Measures

a. Ratings of Expert Testimony

Participants rated the degree to which they felt that the medical examiner's testimony was scientific, credible, and convincing, each on a separate scale from 1 (*not at all*) to 5 (*very*).

b. Guilt Judgments

Participants rendered a verdict (i.e., guilty or not guilty) and estimated the likelihood that the defendant had in fact murdered her daughter, using a sliding scale from 0% to 100%.

c. Comprehension Test

Lastly, participants completed a five-item multiple-choice test to ensure that they read and understood the narrative and transcript. Two items pertained to details of our manipulations, including: (a) whether the defendant was previously a museum curator (*Innocent Stereotype* condition) or in a drug treatment program (*Guilty Stereotype* condition) and (b) whether the medical examiner ruled the child's death an accident (*Accident* condition) or a homicide (*Homicide* condition). Prior to analysis, we excluded data from 107 (33.5%) and 32 (12.7%) participants who answered one or both of these items incorrectly in Studies 1 and 2, respectively.

III Results

A. Analytic Plan

Data from both studies are available on the Open Science Framework (http://osf.io/jekvm). For verdicts, we performed a binary logistic regression with Stereotype (Innocent vs. Guilty), Opinion (Accident vs. Homicide), and their interaction as predictors of verdict (i.e., guilty or not guilty). For each numerical measure (i.e., likelihood of guilt and three ratings of the medical examiner's testimony), we performed both frequentist and Bayesian 2 (Stereotype) X 2 (Opinion) ANOVAs. Descriptive statistics for these numerical measures are shown in Table 1, and ANOVA results are shown in Table 2. Bayesian analyses compare the relative strength of the evidence for an effect and the evidence against an effect; Bayes factors (i.e., BF_{10} values; see Table 2) greater than 1 indicate stronger evidence of an effect, whereas BF_{10} values less than 1 indicate stronger evidence of no effect (see Quintana & Williams, 2018).

Table 1. Means (and Standard Errors) for Ratings of Guilt (0-100) and Expert Testimony (1-5)

			Stereotype		Opinion	
		Overall	Innocent	Guilty	Accident	Homicide
Likelihood of Guilt	Study 1	63.43 (1.85)	61.34 (2.22)	64.82 (2.14)	48.16_a (2.20)	78.01_b (2.16)
	Study 2	55.58 (2.39)	56.58 (2.09)	54.84 (2.02)	27.40_a (2.02)	84.02_b (2.04)
Scientific	Study 1	4.33 (.05)	4.39 (.08)	4.29 (.07)	4.42 (.07)	4.26 (.07)
	Study 2	4.65 (.04)	4.71 (.06)	4.60 (.06)	4.66 (.06)	4.66 (.06)

Credible	Study 1	4.33 (.05)	4.40 (.08)	4.27 (.07)	4.20_a (.08)	4.47_b (.07)
	Study 2	4.63 (.04)	4.60 (.06)	4.66 (.06)	4.54_a (.06)	4.72_b (.06)
Convincing	Study 1	4.13 (.06)	4.22 (.09)	4.04 (.08)	3.92_a (.09)	4.34_b (.08)
	Study 2	4.50 (.05)	4.50 (.08)	4.50 (.07)	4.42 (.08)	4.59 (.08)

Note. Means not sharing a common subscript were significantly different at $p < .05$ (see Table 2).

Table 2. Results of 2 (Stereotype) X 2 (Opinion) ANOVAs on Ratings of Guilt and Expert Testimony

		Stereotype		Opinion		Interaction	
		F	BF_{10}	F	BF_{10}	F	BF_{10}
Likelihood of Guilt	Study 1	1.27	0.26	93.90***	4.21×10^{15}	0.04	0.21
	Study 2	0.37	0.16	389.81***	1.19×10^{47}	0.46	0.24
Scientific	Study 1	1.07	0.25	2.14	0.42	0.00	0.22
	Study 2	1.62	0.32	0.00	0.15	0.16	0.22
Credible	Study 1	1.70	0.30	6.98**	4.02	1.78	0.47
	Study 2	0.52	0.21	4.80*	1.58	0.01	0.20
Convincing	Study 1	2.08	0.33	12.48***	49.45	3.60	1.00
	Study 2	0.00	0.15	2.42	0.47	0.00	0.22

Note. *$p < .05$; **$p < .01$; ***$p < .001$

B. Verdicts

Consistent with H1, a 2 (Stereotype: Innocent vs. Guilty) X 2 (Opinion: Accident vs. Homicide) factorial logistic regression on verdicts revealed significant effects of Opinion in both Study 1, Wald $\chi^2(1) = 34.75$, $p < .001$, $OR = 2.65$ [95% CI: 1.92, 3.67], and Study 2, Wald $\chi^2(1) = 88.36$, $p < .001$, $OR = 8.72$ [95% CI: 5.55, 13.70], such that participants more often judged the defendant as guilty when the medical examiner ruled the death a homicide (84.3% in Study 1; 90.8% in Study 2) as opposed to an accident (43.3% in Study 1; 12.7% in Study 2).

There was no effect of Stereotype in either Study 1, $\chi^2(1) = 0.44$, $p = .506$, $OR = 0.90$ [95% CI: 0.65, 1.24], or Study 2, Wald $\chi^2(1) = 1.76$, $p = .184$, $OR = 0.87$ [95% CI: 0.86, 2.13]. Moreover, the Stereotype X Opinion interaction was not significant in either Study 1, Wald $\chi^2(1) = 0.03$, $p = .874$, or Study 2, Wald $\chi^2(1) = 0.87$, $p = .351$, indicating that the medical examiner's testimony had an equivalent effect on verdicts regardless of the defendant's characteristics.

C. Likelihood of Guilt

Further supporting H1, a 2 (Stereotype) X 2 (Opinion) ANOVA on likelihood of guilt estimates (0-100%) revealed strong effects of Opinion in both studies (see Table 2), such that participants were more confident in the defendant's guilt when the medical examiner ruled the death a homicide as opposed to an accident (see Table 1 for descriptive statistics).

D. Ratings of Expert Testimony

Overall (i.e., collapsed across studies and conditions), participants rated the medical examiner's testimony as highly scientific (M = 4.50 out of 5, SE = .03), highly credible (M = 4.48, SE = .04), and highly convincing (M = 4.32, SE = .04). In each case, most participants selected either '5' (60.8% for scientific, 60.6% for credible, 53.4% for convincing) or '4' (29.2% for scientific, 28.5% for credible, 30.2% for convincing), which indicates that a sizeable majority of participants considered the medical examiner's testimony to be highly scientific, credible, and convincing regardless of the manner determination or the defendant's characteristics.

As shown in Table 2, there were no main effects of Stereotype on any of these ratings in either study, such that judgments of the medical examiner's testimony were uniformly positive regardless of the defendant's characteristics. However, we unexpectedly found main effects of Opinion on credibility ratings in both studies and on convincingness ratings in Study 1 (but not Study 2), such that—irrespective of the defendant—participants rated the medical examiner's testimony as more credible if they ruled the death a homicide rather than an accident.

Finally, and contrary to H2, there were no significant Stereotype X Opinion interactions on any ratings of the medical examiner's testimony in either study. Thus, participants rated the medical examiner's testimony as equally (and highly) scientific regardless of the stereotypic congruence (or lack thereof) between the examiner's manner determination and the defendant's characteristics, and participants consistently rated homicide determinations as more credible than accident determinations regardless of the defendant's characteristics.

IV Discussion

Dror *et al.* (2021) demonstrated that extraneous information can affect medical experts' judgments of whether a sudden death was a homicide or an accident—judgments that often hold legal ramifications insofar as they imply whether a crime was committed. In response, leading forensic pathologists clarified that manner of death determinations are "not scientific" (Peterson *et al*., 2021a) and that "any individual determination is questionable" (Oliver, 2021). Despite this, our data show that laypeople find manner determinations very compelling: Across two studies, 90% of mock jurors rated a medical examiner's testimony about manner of death as highly scientific (i.e., a 4 or 5 on a 5-point scale), 89% rated it as highly credible, and 84% rated it as highly convincing. Accordingly, the medical examiner's opinion was practically dispositive on trial outcomes: 88% of mock jurors judged the defendant as guilty of murder if the medical examiner ruled the death a homicide, compared to only 28% if they ruled the same death an

accident. Our data thus reveal a troubling disconnect between how non-experts regard manner determinations and how experts feel that their determinations *should* be regarded.

Contrary to our prediction, participants' evaluations of the medical examiner's testimony were equally favorable regardless of whether the defendant fit a criminal stereotype, which may reflect a ceiling effect. However, we unexpectedly found that jurors rated the testimony as more credible (in both studies) and more convincing (in Study 1 only) when the medical examiner ruled the death a homicide rather than an accident, even though the circumstances of the death were otherwise identical. The presumption of innocence—i.e., that jurors should assume the defendant to be innocent unless and until they are proven guilty—is a cornerstone of the American legal system. However, this finding suggests that jurors find otherwise-identical testimony more persuasive when it is inculpatory rather than exculpatory, which casts doubt on their ability to follow this principle. Other studies have likewise found that jurors initially estimate the probability of a defendant's guilt as around 50% (when it should be 0%; Scurich *et al*., 2016), that merely being charged with a crime invites stronger presumptions of guilt (Scurich & John, 2017), and that a sizeable minority of individuals believe it is worse to wrongly acquit a guilty person than to wrongly convict an innocent person (Garrett & Mitchell, 2022). Thus, our data raise the concerning possibility that jurors not only show unwarranted confidence in manner of death determinations in general, but also inherently trust incriminating determinations more than others.

Legal scholars have argued that expert testimony on manner of death should not be admissible, arguing that it is both unreliable and so dispositive that it usurps the factfinding role of the jury (Findley & Strang, 2022; Simon, 2019). Some medical examiners have likewise explained that manner determinations do not belong in courtrooms, noting that they are meant to inform "public health statistics… not trial results" (Peterson *et al.*, 2018b) and voicing concern over "misuse of manner determination by the courts" (Peterson *et al*., 2018a). Notably, manner determination is an American invention; in most other countries, coroners do not certify manner or testify in court (e.g., Oliver, 2014). For example, Australian death investigators do not classify "manner of death into ascribed, clearly defined categories," nor do they include it on death certificates (Phillips *et al*., 2015). In Italy, "the death certificate is not used in court as evidence of the cause and manner of death, but is used only to prove that a person is dead" (Di Vella & Campobasso, 2015). And in India, "the autopsy physician opines only on the cause of death, [whereas] manner of death is determined by the police" (Sharma & Bajpai, 2015).

Short of eliminating manner determinations altogether, how else might we discourage factfinders from giving undue weight to potentially unreliable manner of death opinions? Above all, medical examiners should make abundantly clear in court—just as they have done in writing—that manner determinations are "nonscientific," there are no "criteria for correctness," and their aim is not to find the "right answer" (Peterson *et al*., 2021a). Furthermore, they should be fully transparent about the rationale for their manner determination—i.e., the extent to which it is based on medical information within the purview of their expertise, as opposed to non-medical information gleaned from police or other sources—and if it is based solely on the latter (as in the infamous case of *Iowa v. Tyler*; see Weedn, 2021), then it should not be admissible.

However, these ostensibly simple solutions may not be as foolproof as they appear. First, even if experts' deficiencies are made clear, the mystique of expertise may overshadow those

deficiencies, as in one study where mock jurors rated forensic experts as highly skilled, competent, and convincing even if they admitted to a staggering error rate (~30%; Crozier *et al*., 2020). Second, it is well-known that people have only limited ability to introspect on their own decision-making processes (e.g., Nisbett & Wilson, 1977) and are often blind to their own biases (e.g., Pronin & Hazel, 2023), such that medical examiners may be unable to faithfully articulate the rationale for their determinations even if they are willing. Third, when experts claim to be immune to bias (Kukucka *et al*., 2017), jurors may believe them; in one study, for example, jurors equally trusted forensic examiners who denied that biasing information had affected their opinion and other examiners who never received that information (Kukucka *et al*., 2020).

When medical examiners neglect to explain the limitations and/or basis of their manner determinations, the onus falls on attorneys to draw out that information. Indeed, FRE Rule 705 states that expert witnesses need not disclose the basis for their opinion unless cross-examination requires it. Unfortunately, it appears that attorneys often fail to detect unreliable manner of death opinions. Despodova *et al*. (2020) had defense attorneys imagine representing a client charged with murder and review a case file, including an autopsy report that ruled the death a homicide. They found that attorneys rated the autopsy as equally probative and reliable regardless of whether it was patently biased (i.e., the medical examiner admitted that his knowledge of the defendant's recanted confession colored his interpretation of the decedent's injuries) or unbiased (i.e., he was unaware of the confession). Moreover, fewer than half of attorneys who read the patently biased autopsy report said that they would raise the issue of bias on cross-examination.

For attorneys who do cross-examine forensic experts, its efficacy remains unclear. Some studies found that cross-examination had little or no effect on jurors' trust in dubious testimony (e.g., Garrett *et al*., 2020; McQuiston-Surrett & Saks, 2009), while others found that it lessened trust indiscriminately (e.g., Scanlon *et al*., 2021)—neither of which is ideal. Cross-examination should ideally sensitize jurors to the quality of forensic testimony so that they discount it only when it is unreliable. Some research suggests this may be possible (e.g., Crozier *et al*., 2020; Thompson & Scurich, 2019), but more work is needed to identify specific effective approaches. While our study included a cross-examination, it was rather benign and constant across conditions, and so we cannot say whether or how it affected jurors' decision-making. Some studies have tested other ways to make jurors more discerning in their reliance on forensic expert testimony—such as judicial instructions (Eastwood & Caldwell, 2015), visual aids (Ribeiro *et al.*, 2023), or educational videos (LaBat *et al*., 2023)—and those methods have also yielded mixed results. Future research should continue to examine how laypeople appraise expert testimony from medical examiners specifically, including how cross-examination might influence those appraisals.

A. Policy Recommendations

Rather than relying on attorneys and jurors to detect, expose, and devalue unreliable manner determinations in the courtroom, it would be more effective to address the problem further upstream—by reforming medicolegal death investigation practices in ways that minimize the risk of unreliable manner determinations in the first place. Dror and Kukucka (2021) have described a procedure called *Linear Sequential Unmasking—Expanded* (LSU-E) for reducing bias in decision-making in both forensic and non-forensic arenas. Under this procedure, decision-makers make thoughtful *a priori* decisions about what information should or should not inform their decision,

prioritize information that is most relevant and objective, and document the information they considered and how their opinion evolved over time. Variants of this procedure have now been endorsed and/or adopted by practitioners in a range of forensic disciplines (e.g., Archer & Wallman, 2016; Dahal *et al*., 2022; Found & Ganas, 2013; Whitehead *et al.*, 2022).

In medicolegal death investigations, LSU-E would advise medical examiners to complete the autopsy and document a tentative manner of death opinion based solely on the observations therein *before* considering any other contextual information that might be relevant to their determination, while also strictly avoiding information that is irrelevant to their decision (see also Simon, 2019, for a proposed typology of information that is always, sometimes, or never relevant to manner determinations). Some pathologists have now endorsed LSU-E in writing, calling it an "important suggestion on how to mitigate the impact of bias" (Graber, 2021; see also Ko & Glusac, 2023). Moreover, Dr. Andrew Baker, tacitly endorsed LSU-E at the murder trial of Derek Chauvin when he testified that he "intentionally chose not to" watch the cell phone video of George Floyd's death prior to performing Mr. Floyd's autopsy because he "did not want to bias [the] exam by going in with any preconceived notions." Although Dr. Baker did not describe it as such, this calculated approach is entirely consistent with what LSU-E prescribes.

Manner determinations should also be as standardized, independent, and transparent as possible. Standardization tends to protect against bias insofar as it lessens room for interpretation and disagreement, which, as noted above, is endemic to manner determinations (Hanzlick *et al*., 2015). In Italy, for example, death certificates include instructions and examples of proper manner determinations in an effort to promote consistency between agencies and examiners (Di Vella & Campobasso, 2015). Medical examiners should also operate independently of law enforcement agencies and prosecutors' offices, so as to minimize the risk of overt pressure to produce a certain determination (Luzi *et al*., 2013) and/or implicit allegiance effects (Murrie *et al*., 2013). Finally, consistent with LSU-E, medical examiners should be transparent about what information informed their manner determination (see also Quigley-McBride *et al*., 2022).

Maryland's Office of the Medical Examiner (OCME) recently agreed to reforms of this nature as part of a settlement pertaining to the death of Anton Black, a Black teenager who died while being restrained by police in 2018. To be exact, the OCME will now (a) have a clear policy for the handling of in-custody deaths that follows NAME standards for homicide determinations (i.e., the "but for" principle), (b) prohibit non-OCME personnel (e.g., law enforcement) from giving input on autopsies, and (c) document the presence of law enforcement during autopsy and any investigative information received therefrom. At the same time, Maryland's Office of the Attorney General is conducting an independent audit of OCME's handling of past deaths involving police restraint for signs of bias and/or inappropriate procedures, with an eye toward improving death investigation practices and creating a model for other states to follow.

In sum, the current studies reveal a problematic disconnect between how experts and non-experts regard manner of death determinations, such that jurors consider them highly scientific and persuasive, whereas practitioners acknowledge their dubiety and question their suitability for courtroom presentation. To bridge this gap, we must better educate factfinders on the limitations of manner determinations and/or reform the processes by which medical examiners make and communicate those determinations. Either will require collaboration between psychologists,

attorneys, and medical examiners who share an ambition to optimize death investigation practices in ways that minimize the risk of miscarriages of justice—and we hope that the current studies will inspire more scholars and practitioners to undertake such efforts.

V References

Anderst, J., Nielsen-Parker, M., Moffatt, M., Frazier, T., & Kennedy, C. Using Simulation to Identify Sources of Medical Diagnostic Error in Child Physical Abuse. (2016) Child Abuse & Neglect, 52, 62-69. Online: doi:10.1016/j.chiabu.2015.12.015

Archer, M. S., & Wallman, J. F. Context Effects in Forensic Entomology and Use of Sequential Unmasking in Casework. (2016) Journal of Forensic Sciences, 61, 1270-1277. Online: doi:10.1111/1556-4029.13139

Crozier, W. E., Kukucka, J., & Garrett, B. L. Juror Appraisals of Forensic Evidence: Effects of Blind Proficiency and Cross-Examination. (2020) Forensic Science International, 315, 110433. Online: doi:10.1016/j.forsciint.2020.110433

Dahal, S., Chaudhary, G. K., & Agrawal, N. K. Operation Makalu air crash: Influence of Cognitive and Human Factors on Decision-Making. (2022) Forensic Sciences Research, 7, 803-807. Online: doi:10.1080/20961790.2022.2095691

Despodova, N. M., Kukucka, J., & Hiley, A. Can Defense Attorneys Detect Forensic Confirmation Bias? Effects on Evidentiary Judgments and Trial Strategies. (2020) Zeitschrift für Psychologie, 228, 216-220. Online: doi:10.1027/2151-2604/a000414

Di Vella, G., & Campobasso, C. P. Death Investigation and Certification in Italy. (2015) Academic Forensic Pathology, 5, 454-461. Online: doi:10.23907/2015.050

Dror, I. E., & Kukucka, J. Linear Sequential Unmasking–Expanded (LSU-E): A General Approach for Improving Decision-Making as Well as Minimizing Bias. (2021) Forensic Science International: Synergy, 3, 100161. Online: doi:10.1016/j.fsisyn.2021.100161

Dror, I. E., Melinek, J., Arden, J. L., Kukucka, J., Hawkins, S., Carter, J., & Atherton, D. Cognitive Bias in Forensic Pathology Decisions. (2021) Journal of Forensic Sciences, 66, 1751-1757. Online: doi:10.1111/1556-4029.14697

Duflou, J. Commentary on Dror *et al*. 'Cognitive Bias in Forensic Pathology Decisions.' (2021) Journal of Forensic Sciences, 66, 2561. Online: doi:10.1111/1556-4029.14836

Eastwood, J., & Caldwell, J. Educating Jurors About Forensic Evidence: Using an Expert Witness and Judicial Instructions to Mitigate the Impact of Invalid Forensic Science Testimony. (2015) Journal of Forensic Sciences, 60, 1523-1528. Online: doi:10.1111/1556-4029.12832

Findley, K. A., & Strang, D. A. Ending Manner-Of-Death Testimony and Other Opinion Determinations of Crime. (2022) Duquesne University Law Review, 60, 302-337.

Found, B., & Ganas, J. The Management of Domain Irrelevant Context Information in Forensic Handwriting Examination Casework. (2013) Science & Justice, 53, 154-158. Online: doi:10.1016/j.scijus.2012.10.004

Garrett, B. L., & Mitchell, G. Error Aversions and Due Process. (2022) Michigan Law Review, 121, 707-751.

Garrett, B. L., Scurich, N., & Crozier, W. E. Mock Jurors' Evaluation of Firearm Examiner Testimony. (2020) Law and Human Behavior, 44, 412-423. Online: doi:10.1037/lhb0000423

Graber, M. L. Commentary on Dror *et al*. 'Cognitive Bias in Forensic Pathology Decisions.' (2021) Journal of Forensic Sciences, 66, 2574. Online: doi:10.1111/15564029.14857

Hanzlick, R. L., Goodin, J., & Haden-Pinneri, K. Mind Your Manners: 20 Years Later. (2015) Academic Forensic Pathology, 5, 380-395. Online: doi:10.23907/2015.042

Innocence Project. (2023). *Misapplication of Forensic Science.* Online: https://innocenceproject.org/misapplication-of-forensic-science/

Jones, C. S., & Kaplan, M. F. The Effects of Racially Stereotypical Crimes on Juror Decision-Making and Information-Processing Strategies. (2003) Basic and Applied Social Psychology, *25*, 1-13. Online: doi:10.1207/S15324834BASP2501_1

Kerstholt, J., Eikelboom, A., Dijkman, T., Stoel, R., Hermsen, R., & van Leuven, B. Does Suggestive Information Cause a Confirmation Bias in Bullet Comparisons? (2010) Forensic Science International, 198, 138-142. Online: doi:10.1016/j.forsciint.2010.02.007

Ko, C. J., & Glusac, E. J. Cognitive Bias in Pathology, As Exemplified in Dermatopathology. (2023) Human Pathology, 140, 267-275. Online: doi:10.1016/j.humpath.2023.03.003

Koehler, J. J. Intuitive Error Rate Estimates for the Forensic Sciences. (2017) Jurimetrics, 57, 153-168.

Kukucka, J., & Dror, I. E. Human Factors in Forensic Science: Psychological Causes of Bias and Error. In D. DeMatteo & K. C. Scherr (Eds.), (2023) The Oxford Handbook of Psychology and Law (pp. 621-642). Oxford University Press.

Kukucka, J., & Findley, K. A. (2023). Cognitive Bias in Medicolegal Judgments. In K. A. Findley, C. Rossant, K. Sasakura, L. Schneps, W. Squier, & K. Wester (Eds.), *Shaken Baby Syndrome: Investigating the Abusive Head Trauma Controversy* (pp. 205-217). Cambridge University Press.

Kukucka, J., Hiley, A., & Kassin, S. M. Forensic Confirmation Bias: Do Jurors Discount Examiners Who Were Exposed to Task-Irrelevant Information? (2020) Journal of Forensic Sciences, 65, 1978-1990. Online: doi:10.1111/1556-4029.14546

Kukucka, J., Kassin, S. M., Zapf, P. A., & Dror, I. E. Cognitive Bias and Blindness: A Global Survey of Forensic Science Examiners. (2017) Journal of Applied Research in Memory and Cognition, 6, 452-459. Online: doi:10.1016/j.jarmac.2017.09.001

LaBat, D. E., Goldfarb, D., Evans, J. R., Compo, N. S., Koolmees, C. J., LaPorte, G., & Lothridge, K. Improving Juror Assessments of Forensic Testimony and Its Effects on Decision-Making and Evidence Evaluation. (2023) Law and Human Behavior, 47, 566–578. Online: doi:10.1037/lhb0000539

Loos, M. L. H., Allema, W. M., Bakx, R., Stoel, R. D., van Rijn, R. R., & Karst, W. A. Paediatric Femur Fractures—The Value of Contextual Information on Judgement in Possible Child Abuse Cases: Are We Bias? (2021) European Journal of Pediatrics, 180, 81-90. Online: doi:10.1007/s00431-020-03704-6

Luzi, S. A., Melinek, J., & Oliver, W. R. Medical Examiners' Independence Is Vital for The Health of The American Legal System. (2013) Academic Forensic Pathology, 3, 84-92. Online: doi:10.23907/2013.012

Martire, K. A., Ballantyne, K. N., Bali, A., Edmond, G., Kemp, R. I., & Found, B. Forensic science evidence: Naïve Estimates of False Positive Error Rates and Reliability. (2019) Forensic Science International, 302, 109877. Online: doi:10.1016/j.forsciint.2019.10987 7

McKimmie, B. M., Masters, J. M., Masser, B. M., Schuller, R. A., & Terry, D. J. Stereotypical and Counterstereotypical Defendants: Who Is He and What Was the Case Against Her? (2013) Psychology, Public Policy, and Law, 19, 343–354. Online: doi:10.1037/a0030505

McQuiston-Surrett, D., & Saks, M. J. The Testimony of Forensic Identification Science: What Expert Witnesses Say And What Factfinders Hear. (2009) Law and Human Behavior, 33, 436-453. Online: doi:10.1007/s10979-008-9169-1

Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. Are Forensic Experts Biased by The Side That Retained Them? (2013) Psychological Science, 24, 1889-1897. Online: doi:10.1177/0956797613481812

National Registry of Exonerations. (2023). *Exonerations in the United States.* Online: https://www.law.umich.edu/special/exoneration/Pages/Exonerations-in-the-United-States-Map.aspx

Nisbett, R. E., & Wilson, T. D. Telling More Than We Can Know: Verbal Reports on Mental Processes. (1977) Psychological Review, 84, 231-259. Online: doi:10.1037/0033-295X.84.3.231

Obenson, K. Commentary on Dror *et al*. 'Cognitive Bias in Forensic Pathology decisions.' (2021) Journal of Forensic Sciences, 66, 2582-2584. Online: doi:10.1111/1556-4029.14855

Oliver, W. R. Manner Determination in Forensic Pathology. (2014) Academic Forensic Pathology, 4, 480-491. Online: doi:10.23907/2014.062

Oliver, W. R. Commentary on Dror *et al*. 'Cognitive Bias in Forensic Pathology Decisions.' (2021) Journal of Forensic Sciences, 66, 2563-2564. Online: doi:10.1111/1556-4029.14841

Ophoven, J. (2022, March 18). *Declaration of Janice Jean Ophoven, M.D. in State of Texas v. Melissa Lucio*. 138th District Court of Cameron County Texas. Online: https://www.documentcloud.org/documents/21657863-dr-janice-ophoven-declaration

Peat, M. A. JFS editor-in-chief preface. (2021) Journal of Forensic Sciences, 66, 2593-2540. Online: doi:10.1111/1556-4029.14844

Peterson, B. L., Arnall, M., Avedschmidt, S., Beers, D., Bell, M., Burton, S. & Wright, R. Commentary on Dror *et al*. 'Cognitive Bias in Forensic Pathology Decisions.' (2021a) Journal of Forensic Sciences, 66, 2541-2544. Online: doi:10.1111/1556-4029.14843

Peterson, B. L., Gill, J., & Oliver, W. Peterson *et al*. Response to Authors' Response. (2021) Journal of Forensic Sciences, 66, 2549-2552. Online: doi:10.1111/1556-4029.14849

Phillips, B., Little, D., McDougall, J., & Langlois, N. E. The Coronial System and Determining Manner of Death in Australia: An Overview. (2015) Academic Forensic Pathology, 5, 436-442. Online: doi:10.23907/2015.047

Pronin, E., & Hazel, L. Humans' Bias Blind Spot and Its Societal Significance. (2023) Current Directions in Psychological Science, 32, 402-409. Online: doi:10.1177/09637214231178745

Quigley-McBride, A., Dror, I. E., Roy, T., Garrett, B. L., & Kukucka, J. A Practical Tool for Information Management in Forensic Decisions: Using Linear Sequential Unmasking-Expanded (LSU-E) in Casework. (2022) Forensic Science International: Synergy, 4, 100216. Online: doi:10.1016/j.fsisyn.2022.100216

Quintana, D. S., & Williams, D. R. Bayesian Alternatives for Common Null-Hypothesis Significance Tests in Psychiatry: A Non-Technical Guide Using JASP. (2018) BMC Psychiatry, 18, 1-8. Online: doi:10.1186/s12888-018-1761-4

Ribeiro, G., Likwornik, H., & Chin, J. M. Visual Decision Aids: Improving Laypeople's Understanding of Forensic Science Evidence. (2023) Journal of Applied Research in Memory and Cognition, 12, 230-240. Online: doi:10.1037/mac0000026

Scanlon, P., Banyai, B., Hart, E., & Cooper, S. L. Juror Certainty About Expert Firearms Identification Evidence and The Impact of Cross-Examination. (2021) Southern California Interdisciplinary Law Journal, 31, 91-105.

Scurich, N., Nguyen, K. D., & John, R. S. Quantifying the Presumption of Innocence. (2016) Law, Probability and Risk, 15, 71-86. Online: doi:10.1093/lpr/mgv016

Scurich, N., & John, R. S. Jurors' presumption of innocence. (2017) The Journal of Legal Studies, 46, 187-206.

Sharma, S. K., & Bajpai, S. Medicolegal Death Investigation in India: An overview. (2015) Academic Forensic Pathology, 5, 443-446. Online: doi:10.23907/2015.048

Simon, D. Minimizing Error and Bias in Death Investigations. (2018) Seton Hall Law Review, 49, 255-305.

Smalarz, L., Madon, S., Yang, Y., Guyll, M., & Buck, S. The Perfect Match: Do Criminal Stereotypes Bias Forensic Evidence Analysis? (2016) Law and Human Behavior, 40, 420-429. Online: doi:10.1037/lhb0000190

Smalarz, L., Madon, S., & Turosak, A. Defendant Stereotypicality Moderates the Effect of Confession Evidence on Judgments of Guilt. (2018) Law and Human Behavior, 42, 355–368. Online: doi:10.1037/lhb0000286

Thompson, W. C., & Scurich, N. How Cross-Examination on Subjectivity And Bias Affects Jurors' Evaluations Of Forensic Science Evidence. (2019) Journal of Forensic Sciences, 64, 1379-1388. Online: doi:10.1111/1556-4029.14031

Weedn, V. W. Bases of Forensic Pathology Expert Testimony With Emphasis on Iowa v Tyler. (2021) Academic Forensic Pathology, 11, 185-195. Online: doi:10.1177/19253621211060961

Whitehead, F. A., Williams, M. R., & Sigman, M. E. Decision Theory and Linear Sequential Unmasking in Forensic Fire Debris Analysis: A Proposed Workflow. (2022) Forensic Chemistry, 29, 100426. Online: doi:10.1016/j.forc.2022.100426

In Pursuit of Innocence: A Study of Race and Ethnicity Differences in Time-to-Exoneration

Virginia Elizabeth Braden
University of Louisville
Louisville, Ky
U.S.A.

The purpose of this study is to examine the impact of race and ethnicity on time-to-exoneration through the lens of focal concerns theory. Focal concerns theory has been used to demonstrate that criminal justice actors are influenced by legal and extralegal factors in decision making and rely on stereotypes to assess blameworthiness, protection of the community, and in navigating practical constraints and consequences. Utilizing data obtained from the National Registry of Exonerations (N =507) survival analysis was performed. The findings indicate that Black exonerees experienced a longer time-to-exoneration than did White exonerees and that Hispanic exonerees experienced the shortest time-to-exoneration of all. The findings offer support for focal concerns theory in the demonstration that racial and ethnic differences are present in time-to-exoneration resulting in disparities which disadvantage minorities. Further support for focal concerns theory is found in that the legal components of a case are shown to be associated with racial and ethnic differences in time-to-exoneration.

I. Introduction
II. Literature Review
    A. Racial and ethnic differences in exonerations
    B. Focal Concerns Theory
III. Purpose of the Study
IV. Methods
    A. Measures
    B. Data analysis plan
V. Results
    A. Policy and Programming
VI. Discussion
VII. Limitations
VIII. Conclusion

I Introduction

In recent decades, the phenomena of exonerations have become increasingly more visible thanks to the ability of DNA testing to prove innocence and professional exonerators, such as the Innocence Project, who are dedicated to representing the wrongfully convicted. An exoneration occurs when an individual who has been convicted of a crime is officially cleared based on new evidence of innocence in any form with no unexplained physical evidence of that individual's guilt remaining.[1] An exoneration may occur while a person is living or be awarded posthumously.

[1] National Registry of Exonerations "Glossary" (2019), online:

A known and established process of law by which an individual can be convicted and found guilty of a crime exists. However, no process exists set by law that lays forth how a convicted individual may be proven innocent. An individual seeking an exoneration faces significant obstacles arising from procedural restrictions, as well as the reluctance of the justice system to admit error.[2] Consequently, the journey to exoneration may take years and even decades to complete and present significant obstacles.[3]

Analyses of the current data on exonerations suggests racial differences are present in exonerations.[4] Additionally, available data also shows there are racial and ethnic differences in time-to-exoneration.[5] However, an examination of the empirical literature regarding time-to-exoneration reveals there is a dearth of scholarly attention to this particular topic. The few studies that do address time-to-exoneration do not focus on race or ethnicity as the central issue but include it as a control variable.[6] Additionally, none of the empirical literature regarding race and ethnicity and time-to-exoneration provide a theoretical premise.

<https://www.law.umich.edu/special/exoneration/Pages/glossary.aspx> [*NRE Glossary*]; Brandon Garrett, "Actual Innocence and Wrongful Convictions" (2017) Academy for Justice, A Report on Scholarship and Criminal Justice Reform (Erik Luna ed., 2017 Forthcoming); Samuel R Gross & Michael Shaffer, "Exonerations in the United States, 1989–2012" (2012) SSRN Journal, online: <http://www.ssrn.com/abstract=2092195>.

[2]Jon B Gould, & Richard A Leo, "The Path to Exoneration" (2015) Albany Law Review, 2016 Forthcoming, American University School of Public Affairs Research Paper No 2016-02 , online: <https://papers.ssrn.com/abstract=2694709> [Gould & Leo]; Jeffrey S Gutman, "An Empirical Reexamination of State Statutory Compensation for the Wrongly Convicted" (2017) 82 Mo L Rev 369, online: <https://papers.ssrn.com/abstract=3422444>.

[3] Gould & Leo, *ibid*.

[4] Marvin Free & Mitch Ruesink, "Flawed justice: A study of wrongly convicted African American women" (2018) 16:4 Journal of Ethnicity in Criminal Justice 333–347, online: <https://www.tandfonline.com/doi/full/10.1080/15377938.2015.1015199>; Samuel R Gross, Maurice Possley & Klara Stephens, Race and Wrongful Convictions in the United States, by Samuel R Gross, Maurice Possley & Klara Stephens (The National Registry of Exonerations, Newkirk Center for Science and Society, 2017), online: <http://www.law.umich.edu/special/exoneration/Documents/Race_and_Wrongful_Convictions.pdf> [*Gross et al 2017*]; National Registry of Exonerations, "Milestone: Exonerated defendants spent 20,000 years in prison" (2018), online: <https://www.law.umich.edu/special/exoneration/Documents/NRE.20000.Years.Report.pdf> [*NRE 2018*].

[5] *NRE 2018 Ibid*.

[6] Gould & Leo, *supra* note 2; Maeve Olney & Scott Bonn, "An Exploratory Study of the Legal and Non-Legal Factors Associated With Exoneration for Wrongful Conviction: The Power of DNA Evidence" (2015) 26:4 Criminal Justice Policy Review 400–420, online: <http://journals.sagepub.com/doi/10.1177/0887403414521461> [*Olney*]; Patrick Rafail & Margaret Mahoney "A Long Road to Freedom: The Exoneration Pipeline in the United States, 1989–2015" (2019) 60:4 TSQ 537-558, online: <https://doi-org/10.1080/00380253.2018.1547175> [*Rafail & Mahoney*].

A theoretical framework is important to provide context for the understanding of results.[7] Theories provide organization for data and deepen understanding of empirical study.[8] Little is known about the decision-making process regarding exonerations. Therefore, a theoretical premise is necessary.

This study examines racial and ethnic differences in time-to-exoneration through the lens of Steffensmeier's (1980) Focal Concerns Theory (FCT). FCT posits that judges and other decision makers are driven by concerns with blameworthiness, protection of the community, and the practical constraints and consequences of their decision. Further, that they rely on heuristics derived from stereotypes to aid them in their decisions. FCT is utilized as the theoretical premise to contextualize the likelihood of racial and ethnic differences in time-to-exoneration. The ability to understand time-to-exoneration using FCT provides insight into how these concerns affect racial and ethnic differences in time-to-exoneration that have not been previously tested. The results of the current study can aid in developing policy and guide in shaping reform measures.

II Literature Review

A. Racial and ethnic differences in exonerations

Research on exonerations demonstrates racial and ethnic differences are present in all major categories of crimes for which data are collected.[9] A review of the most comprehensive data on known exonerations in the US maintained by the National Registry of Exonerations, indicates racial and ethnic differences are also present in time-to-exoneration. Specifically, Black innocent defendants spend 45% more time wrongfully imprisoned before being exonerated than do White

[7] George E Higgins, Gennaro F. Vito, & Elizabeth L. Grossi, "The impact of race on the police decision to search during a traffic stop: A focal concerns theory perspective" (2012) 28:2 J Contemp Crim Justice 166-183, online: <https://doi-org./10.1177/1043986211425725> [*Higgins 2012*]; Robert J Norris *et al*, "Thirty Years of Innocence: Wrongful Convictions and Exonerations in the United States, 1989-2018" (2020) 1:1 WCLR 2-58, online: <https://wclawr.org/index.php/wclr/article/view/11/23>.

[8] George E Higgins & Catherine Davis Marcum, Criminological theory, Aspen College series (New York: Wolters Kluwer, 2016).

[9] Marvin D Free & Mitch Ruesink, Race and justice: wrongful convictions of African American men (Boulder, Colo: Lynne Rienner Publishers, 2012); Marvin Free & Mitch Ruesink, "Flawed justice: A study of wrongly convicted African American women" (2018) 16:4 Journal of Ethnicity in Criminal Justice 333–347, online: <https://www.tandfonline.com/doi/full/10.1080/15377938.2015.1015199>; Gross *et al* 2017, supra note 4; Karen F. Parker, Mari A. DeWees, & Michael L. Radelet, "Race, the death penalty, and wrongful convictions" (2003) 18:1 Crim. Just 49, online: <https://heinonline.org/HOL/P?h=hein.journals/cjust18&i=51>; Arthur Rizer "The race effect on wrongful convictions" (2003) 29:3 Wm Mitchell L Rev 845, online: <http://open.mitchellhamline.edu/wmlr/vol29/iss3/5>; Earl Smith & Angela J Hattery, "Race, Wrongful Conviction & Exoneration" (2011) 15:1 J Afr Am St 74–94, online: <http://link.springer.com/10.1007/s12111-010-9130-5>.

innocent defendants.[10] The data indicates that for all offense types the average time served is 10.7 years for Black exonerees, 7.2 years for Hispanic exonerees, and 7.2 years for White exonerees.[11] Additionally, this data showed that in some categories of crimes, Hispanic innocent defendants served more time before exoneration than did White innocent defendants but less than Black innocent defendants.

Scholarly literature assessing the impact of race and ethnicity on time-to-exoneration is scant. The studies which do exist point to significant disparities. Gross *et al*. reviewed over 1,900 exonerations found that Black innocent defendants spend more time wrongfully imprisoned than White innocent defendants in every single category of crimes for which exoneration data was collected.[12] Similarly, in their examination of the impact of DNA on exonerations, Olney and Bonn also found that Black exonerees encounter the longest time-to-exoneration.[13]

Utilizing survival analysis, *Rafail and Mahoney* focused on the length of time it takes for exonerees to complete the process of exoneration.[14] Their findings exposed significant differences arising from factors of place, evidence type, and race. Further, the temporal gap in achieving exoneration was found to disadvantage Black exonerees.

None of these studies, however, contextualize their results using a theoretical context to the examination of racial and ethnic differences in time-to-exoneration. Leo argued scholarship on innocence is "theoretically impoverished".[15] The application of focal concerns theory provides the context to understand racial and ethnic differences in time-to-exoneration.

B. Focal Concerns Theory

Focal concerns theory originally emerged as a theoretical model within the scholarly exploration of gender differences in sentencing.[16] In later years, Steffensmeier *et al*. expanded the theory to include other social characteristics such as race, ethnicity, and age.[17] The theory's key premise is that judges and other court actors are attuned to three focal concerns in arriving at decisions: blameworthiness, protection of the community, and practical implications of the

[10] NRE 2018, *supra* note 7.

[11] NRE 2018, *supra* note 7.

[12] Gross *et al* 2017, *supra* note 4.

[13] Olney, *supra* note 6.

[14] Rafail & Mahoney, *supra* note 6.

[15] Richard A Leo, "Rethinking the study of miscarriages of justice: Developing a criminology of wrongful conviction" (2005) 21:3 J Contemp Crim Justice 213, online: <https://doi-org.ezproxy.lib.torontomu.ca/10.1177/1043986205277477>.

[16], Darrell J Steffensmeier "Assessing the Impact of the Women's Movement on Sex-Based Differences in the Handling of Adult Criminal Defendants" (1980) 26:3 Crime & Delinquency 344–357, online: <http://journals.sagepub.com/doi/10.1177/001112878002600305> [*Steffensmeier 1980*].

[17]Darrell Steffensmeier, Jeffery Ulmer & John Kramer, "The Interaction of Race, Gender, , and Age in Sentencing: The Punishment Cost of Being Young, Black, and Male" (1998) 36:4 Criminology 763–798, online: <https://onlinelibrary.wiley.com/doi/10.1111/j.1745-9125.1998.tb01265.x> [*Steffensmeier 1998*].

resulting decision".[18] In addition, a guiding principle of FCT is the understanding that judges and other decision makers often do not have enough information or have an overwhelming amount of information, and often a short amount of time in which to consider it when making decisions. In these circumstances, focal concerns theory argues they resort to heuristics to guide their decisions.

Scholarly research demonstrates to facilitate the decision-making process judges and prosecutors do resort to heuristics that incorporate stereotypical beliefs regarding an individual's race or ethnicity when considering blameworthiness, protection of the community, and practical constraints and consequences.[19] Further, this has been shown to result in disparities in the outcomes which often negatively impact minorities.[20]

---

[18] Darrell Steffensmeier, Noah Painter-Davis & Jeffery Ulmer, "Intersectionality of Race, Ethnicity, Gender, and Age on Criminal Punishment" (2017) 60:4 Sociological Perspectives 813, online: <http://journals.sagepub.com/doi/10.1177/0731121416679371>.

[19] G Ford, "The New Jim Crow: Male and Female, South and North, from Cradle to Grave, Perception and Reality: Racial Disparity and Bias in America's Criminal Justice System." (2009) 11 *Rutgers Race & L. Rev.* 324; Celesta Albonetti, "Integration of Theories to Explain Judicial Discretion" (1991) 38:2 Social Problems 247–266; Tina Freiburger, "Race and the Sentencing of Drug Offenders: An Examination of the Focal Concerns Perspective." (2009) 6:2 Southwest Journal of Criminal Justice 163 [*Albonetti*]; Brian D Johnson, "Racial and ethnic disparities in sentencing departures across modes of conviction" (2003) 41:2 Criminol 449-490, online: <https://doi-org/10.1111/j.1745-9125.2003.tb00994.x> [*Johnson*]; Michael J. Leiber & Anita N. Blowers "Race and misdemeanor sentencing" (2003) 14:4 Crim Justice Policy Rev 464-485, online:<https://doi-org.ezproxy.lib.torontomu.ca/10.1177/0887403403254492>;
Darrell Steffensmeier & Stephen Demuth, "Ethnicity and Judges' Sentencing: Hispanic-Black-White Comparisons" (2001) 39:1 Criminology 145–178, online:
<https://onlinelibrary.wiley.com/doi/10.1111/j.1745-9125.2001.tb00919.x> [*Steffensmeier & Demuth 2001*]; Patricia Warren, Ted Chiricos & William Bales, "The Imprisonment Penalty for Young Black and Hispanic Males: A Crime-Specific Analysis" (2012) 49:1 Journal of Research in Crime and Delinquency 56–80, online: <http://journals.sagepub.com/doi/10.1177/0022427810397945> [*Warren*];

[20] George Bridges & Sara Steen, "Racial Disparities in Official Assessments of Juvenile Offenders: Attributional Stereotypes as Mediating Mechanisms" (1998) 63:4 American Sociological Review 554–570 [*Bridges & Steen*]; Stephen Demuth & Darrell Steffensmeier, "Ethnicity Effects on Sentence Outcomes in Large Urban Courts: Comparisons Among White, Black, and Hispanic Defendants *" (2004) 85:4 Social Science Quarterly 994–1011, online:
<https://onlinelibrary.wiley.com/doi/10.1111/j.0038-4941.2004.00255.x> [*Demuth & Steffensmeier 2004*]; Jill K Doerner & Stephen Demuth, "The Independent and Joint Effects of Race/Ethnicity, Gender, and Age on Sentencing Outcomes in U.S. Federal Courts" (2010) 27:1 Justice Quarterly 1–27, online: <https://www.tandfonline.com/doi/full/10.1080/07418820902926197> [Doerner & Demuth]; Susan Sharp, Adrienne Braley & Susan Marcus-Mendoza, "Focal Concerns, Race & Sentencing of Female Drug Offenders" (2000) 28:2 Free Inquiry in Creative Sociology 3–16 [*Sharp*]; Cassia Spohn & Lisa L Sample, "The Dangerous Drug Offender in Federal Court: Intersections of Race, Ethnicity, and Culpability" (2013) 59:1 Crime & Delinquency 3–31, online: <http://journals.sagepub.com/doi/10.1177/0011128708319928>.
Darrell Steffensmeier, Noah Painter-Davis & Jeffery Ulmer, "Intersectionality of Race, Ethnicity, Gender, and Age on Criminal Punishment" (2017) 60:4 Sociological Perspectives 810–833, online: <http://journals.sagepub.com/doi/10.1177/0731121416679371> [*Steffensmeier 2017*]; Sara Steen, Rodney

In recent years, scholars have applied FCT to a variety of settings and actors within the criminal justice system. These include decisions made by prosecutors, police officers, corrections officers, and parole officers, as well as decisions made by judges at all stages of the legal process.[21] The literature demonstrates support for FCT and consequent disparities based on race even when accounting for different modes of conviction and varying types of prosecutorial and judicial discretion.[22]

In recent years, empirical support has been found for combined effects of the interaction of extra-legal factors such as race, ethnicity, gender, age, employment, and education, in decision making.[23] The literature shows that, net of legal factors, the impact of these interactions often disadvantages Blacks and Hispanics in the outcomes realized.[24]

Though FCT has been utilized to provide context in the empirical analysis of decision making and outcomes across a diverse range of legal processes, it has not been utilized in examining the impact of race and ethnicity on time-to-exoneration. The literature review supports the legitimacy of the application of this theory to the decision to exonerate.

L Engen & Randy R Gainey, "Images of Danger and Culpability: Racial Stereotyping, Case Processing, and Criminal Sentencing" (2005) 43:2 Criminology 435–468, online: <https://onlinelibrary.wiley.com/doi/10.1111/j.0011-1348.2005.00013.x>.

[21]Dawn Beichner & Cassia Spohn, "Modeling the Effects of Victim Behavior and Moral Character on Prosecutors' Charging Decisions in Sexual Assault Cases" (2012) 27:1 Violence Vict 3–24, online: <http://connect.springerpub.com/lookup/doi/10.1891/0886-6708.27.1.3>; Higgins 2012, supra note 7 ; Glen A. Ishoy & dean A. Dabney, "Policing and the focal concerns framework: Exploring how its core components apply to the discretionary enforcement decisions of police officers" (2017) 39:7 Deviant Behav 878-895, online: <https://doi-org./10.1080/01639625.2017.1335537>; Anthony Gennaro Vito, Elizabeth L Grossi & George E Higgins, "Analyzing racial profiling from traffic searches: Using focal concerns theory and propensity score matching" (2018) 41:6 PIJPSM 721–733, online: <https://www.emerald.com/insight/content/doi/10.1108/PIJPSM-06-2017-0081/full/html>;
Beth M Huebner, & Timothy S. "An analysis of parole decision making using a sample of sex offenders: A focal concerns perspective" (2006) 44:4 Criminol 961-991, online: <https://doi-org./10.1111/j.1745-9125.2006.00069.x>; Tina Freiburger, Catherine D Marcum & Mari Pierce, "The Impact of Race on the Pretrial Decision" (2010) 35:1–2 Am J Crim Just 76–86, online: <http://link.springer.com/10.1007/s12103-009-9069-z> [*Freiburger 2010*]; Steffensmeier & Demuth 2001, supra note 19.

[22] Johnson, *ibid.*

[23] Steffensmeier 2017, *supra* note 20.

[24] Doerner & Demuth, supra note 20; Ben Feldmeyer *et al*, "Racial, Ethnic, and Immigrant Threat: Is There a New Criminal Threat on State Sentencing?" (2015) 52:1 Journal of Research in Crime and Delinquency 62–92, online: <http://journals.sagepub.com/doi/10.1177/0022427814548488>; *Freiburger 2010*, *supra* note 21; Cassia Spohn & David Holleran, "The Imprisonment Penalty Paid by Young, Unemployed Black and Hispanic Male Offenders" (2000) 38:1 Criminology 281–306, online: <https://onlinelibrary.wiley.com/doi/10.1111/j.1745-9125.2000.tb00891.x>; *Steffensmeier 2017*, *supra* note 20 ; *Warren*, *supra* note 19.

III Purpose of the Study

The purpose of this study is to examine exoneration data for evidence of racial and ethnic differences in time-to-exoneration. Additionally, time-to-exoneration is examined in the context of FCT to guide in greater understanding of any disparities found. The results of this study are useful in understanding the decision-making process of judges in deciding exonerations. The study addresses two hypotheses derived from the focal concerns theoretical perspective:

> Hypothesis 1: Racial and ethnic differences are present in length of time-to-exoneration.
> Hypothesis 2: The legal components of a case are associated with the racial and ethnic differences in the length-of-time-to-exoneration.

IV Methods

The data utilized in this study comes from a comprehensive database maintained by the National Registry of Exonerations (NRE). The data for this study used a subsample (N= 489) of the larger data set of exonerations from the NRE.[25] The data used were of exonerations which occurred from the years 2008 to 2018. This was done to provide a manageable subsample of the population and still provide a decade's worth of exoneration information.

A. Measures

The measures for this study include whether an exoneration took place within a specified number of days, as well as extralegal and legal factors that are associated with studies of FCT. The independent variable for this study is the race or ethnicity of the exoneree. The dependent variable is the length of time-to-exoneration.

A central measure to this study is capturing the number of days that it takes an individual to be exonerated.[26] To calculate this measure, the conviction date was subtracted from the exoneration date. This provided the exact number of days to exoneration.

Another key measure is whether an exoneration took place in a specified amount of time.[27] In other words, this provides a specific time of occurrence for the study. The median was calculated for this measure. For the median of days, the exoneration event was coded as 1 for above the median and 0 for below the median.

---

[25] *Public Spreadsheet [Dataset and Code Manual]* (National Registry of Exonerations, 2019). The original sample (N = 507) contained 18 exonerees whose race or ethnicity was not White, Black, or Hispanic. Those 18 exonerees were excluded from the sample, resulting in the sample size N = 489.

[26] David G Kleinbaum & Mitchel Klein, Survival Analysis: A Self-Learning Text, Statistics for Biology and Health (New York, NY: Springer New York, 2012) [*Kleinbaum*].

[27] Christiana Kartsonaki, "Survival analysis" (2016) 22:7 Diagnostic Histopathology 263–270, online: <https://linkinghub.elsevier.com/retrieve/pii/S1756231716300639>.

The literature is diverse when it comes to measuring the different aspects of this version of FCT. This dissertation follows the sentencing literature and groups the measures that are available in the data into the context of extralegal and legal factors.[28]

The available extralegal factors that exist in the data are age, race, ethnicity, and biological sex. Age is captured in years at the time of crime commission. Race/ethnicity is captured as a nominal level measure. In this study, White refers to non-Hispanic Whites and Blacks refer to non-Hispanic Blacks. Due to potential data constraints (i.e., small n's within categories), three dummy codes for race were used, and they are as follows: 1 = White and 0 = other, 1= Black and 0 = other, 1 = Hispanic and 0 = other. For the purposes of survival analysis, race was coded 0 = White and 1= Black and 0 = White and 1 = Hispanic. Coding race/ethnicity in this way facilitates two comparative analyses of days-to-exoneration.[29] The first analysis compares Whites to Blacks, and the second analysis compares Whites to Hispanics. Additionally, multivariate analyses compared subsamples differentiated by race/ethnicity. Biological sex is captured as the biological sex of the individual at the time of the crime. The measure was recoded so that 0 = female and 1 = male.

Several legal factors were used in this analysis. The legal factors are false or misleading forensic evidence, perjury or false accusation, mistaken eyewitness identification, official misconduct, inadequate legal defense, drug crime, and violent crime. They are as described below. This measure was coded as 0 = no and 1 = yes.

False or misleading forensic evidence: This is a single item indicator that means the individual's conviction was based at least in part on forensic information that consisted of one or more of the following criteria: (1) caused by errors in forensic testing, (2) based on unreliable or unproven methods, (3) expressed with exaggerated and misleading confidence, or (4) fraudulent.[30] This measure was coded as 0 = no and 1 = yes.

Perjury or False Accusation: A person other than the individual committed perjury by making a false statement under oath that incriminated the individual in the crime for which the individual was later exonerated or made a similar unsworn statement that would have been perjury if made under oath (NRE, 2019).[31] This measure was coded as 0 = no and 1 = yes.

Mistaken Eyewitness ID: This refers to at least one eyewitness affirmatively and mistakenly said that he or she saw the individual commit the crime or saw the individual under

---

[28] Albonetti, *supra* note 19; Steffensmeier 1998, *supra* note 17; Steffensmeier & Demuth 2001, *supra* note 19;

[29] The number defendants that were of a race or ethnicity other than Black, Hispanic, or White due was so small (n = 18) that they were excluded from this study.

[30] Gerald LaPorte, "Wrongful Convictions and DNA Exonerations: Understanding the Role of Forensic Science" (2017) 279 NIJ Journal, online: <https://nij.ojp.gov/topics/articles/wrongful-convictions-and-dna-exonerations-understanding-role-forensic-science>; National Registry of Exonerations "Glossary" (2019), online: <https://www.law.umich.edu/special/exoneration/Pages/glossary.aspx>

[31] Jon Gould & Richard Leo, "One Hundred Years Later: Wrongful Convictions after a Century of Research" (2010) 100:3 Journal of Criminal Law and Criminology 825, online: <https://scholarlycommons.law.northwestern.edu/jclc/vol100/iss3/7>.

circumstances that suggest that the individual participated in the crime.[32] This measure was coded as 0 = no and 1 = yes.

Official Misconduct: Police, prosecutors, or other government officials significantly abused their authority or the judicial process in a manner that contributed to the individual's conviction.[33] This measure was coded as 0 = no and 1 = yes.

Inadequate Legal Defense: The individual's lawyer at trial provided obviously and grossly inadequate representation.[34] This measure was coded as 0 = no and 1 = yes.

Drug Crime: Whether the offense was a drug related offense was considered a legal factor.[35] This measure was coded as 0 = no and 1 = yes.

Violent Crime: Whether the offense was violent was considered a legal factor.[36] This measure was coded as 0 = no and 1 = yes.

B. Data Analysis Plan

The data analysis occurred in a series of steps. Step one is a presentation of the overall descriptive statistics. By utilizing univariate statistics, this step offers a brief description of the distribution of the sample.

Step two is a presentation of the survival analysis. This is important in addressing the first hypothesis. Survival analysis is a family of techniques designed to model the time it takes for an event to occur when there is a possibility that the event will not occur for all in a given sample .[37] This type of data is often non-normal. This is because of censoring, a common feature of survival analysis.[38]

---

[32] Jon Gould *et al*, Predicting Erroneous Convictions: A Social Science Approach to Miscarriages of Justice (Washington, DC: National Institute of Justice: 241389, 2012); *NRE Glossary*, *supra* note 1.

[33] Peter A. Joy, "Brady and jailhouse informants: Responding to injustice" (2006) 57:3 Case W Res L Rev 619, online: <https://scholarlycommons.law.case.edu/caselrev/vol57/iss3/14/>; *NRE Glossary*, *supra* note 1.

[34] Sharp, *supra* note 20; NRE Glossary, *supra* note 1.

[35] Demuth & Steffensmeier 2004, *supra* note 20.

[36] Steffensmeier 2017, *supra* note 20; Darrell Steffensmeier & Stephen Demuth, "Ethnicity and Sentencing Outcomes in U.S. Federal Courts: Who is Punished More Harshly?" (2000) 65:5 American Sociological Review 705-729, online: <http://www.jstor.org/stable/2657543?origin=crossref>.

[37] D.R. Cox, "Regression Models and Life-Tables" (1972) 34:2 Journal of the Royal Statistical Society Series B: Statistical Methodology 187–202, online: <https://academic.oup.com/jrsssb/article/34/2/187/7027194>; D R Cox & David Oakes, Analysis of survival data, Monographs on statistics and applied probability (London ; New York: CRC Press, 1984) [*Cox & Oakes*].

[38] Kwang-Moon Leung, Robert M. Elashoff, & Abdelmonem Afifi "Censoring issues in survival analysis" (1997) 18:1 Annu Rev Public Health 83-104, online:

The Kaplan-Meier technique was used to determine the proportion of individuals who were exonerated by the median time-to-exoneration established. In this study time was measured in days. This technique provides a method of estimating the length of time that it will take for someone to be exonerated.[39] The survival and hazard functions are presented graphically.

The Kaplan-Meier also allowed for a direct test of the difference of these functions by groups. In the present study, the groups were differentiated by race and ethnicity. The comparison of the groups took place using the log rank test. Applied to this study, the log rank test allowed for the examination of the survival and hazard functions of exoneration by race/ethnicity.[40] To do this, the log rank test allows the survival and hazard functions to be weighted equally with time. This provided the opportunity for a chi-square test of difference between exoneration by racial/ethnic group.

Step three consists of logistic regression. In addressing the second hypothesis logistic regression is important because of the dichotomous nature of the dependent measure. Specifically, the median time-to-exoneration is dummy coded "0" for exonerated before the median time of 1,000 days and "1" for exonerated after the median time-to-exoneration. The logistic regression model allows for a dichotomous dependent variable while examining the impact of multiple predictor variables.

When interpreting the coefficients in this form of regression, it is important to understand the coefficients represent a link between the covariates and the odds of falling into the group of exonerees who were not exonerated by the median time-to-exoneration. The interpretation of the dummy variables indicated either an increased or decreased likelihood of being exonerated by the median time for that variable. The effect size for this form of logistic regression is the Exp(b). In this form of regression, the Exp(b) is interpreted as an odds ratio.[41] In the present study this form of binomial logistic regression allowed for the proper modeling of the dichotomous dependent measure, days to exoneration, and the legal and extralegal measures representing the focal concerns theory to address the hypotheses of interest.

In this study five models are estimated. The first model consists of all the data and the legal and extralegal measures for a subsample of only Black and White exonerees. The second model is

<https://doi.org/10.1146/annurev.publhealth.18.1.83>.

[39] E. L Kaplan & Paul Meier, "Nonparametric estimation from incomplete observations" (1958) 53:282 JASA 457-481, online: <https://doi.org/10.1080/01621459.1958.10501452> [*Kaplan-Meyer*]; John P Klein & Melvin L Moeschberger, "Refinements of the Semiparametric Proportional Hazards Model" in Survival Analysis (New York, NY: Springer New York, 1997) 269; *Kleinbaum, supra* note 26.

[40] *Kaplan-Meyer Ibid; Cox & Oakes, supra* note 37; Singh, Ritesh & Keshab Mukhopadhyay, "Survival analysis in clinical trials: Basics and must know areas" (2011) 2:4 Perspect Clin Res 145, online: <https://journals.lww.com/10.4103/2229-3485.86872> [*Singh & Mukhopadhyay*].

[41] Cox & Oakes, *supra* note 37; Fox, John, "Cox proportional-hazards regression for survival data. An R and S-PLUS companion to applied regression" (2002); Frank E. Harre, Jr, Kerry L. Lee, & Barbara G. Pollock, "Regression models in clinical studies: determining relationships between predictors and response" (1988) 80:15 JNCI 1198-1202, online: <https://doi.org/10.1093/jnci/80.15.1198>; Singh & Mukhopadhyay Ibid.

comprised of all the data and the legal and extralegal measures for a subsample of only Hispanic and White exonerees. The third model is for the data for White exonerees. The fourth model is for the data for Black exonerees. The fifth model consists of the data for Hispanic exonerees.

To address the hypothesis that there are racial differences in the focal concern measures, the Paternoster *et al.* z-score was applied to the slopes and standard errors of the White, Black, and Hispanic models.[42] Applying the z-score allowed for the understanding of whether the focal concern measures are equal across the racial and ethnic groups.

V Results

The current study is designed to provide an understanding of the impact of race and ethnicity on time-to-exoneration. To provide this understanding of time-to-exoneration, the study makes use of FCT. The results of the study are presented in a series of steps.

The first step is a presentation of the descriptive statistics. The full results of the descriptive statistics measure are shown in Table 1. Black exonerees comprised 46% of the sample, White exonerees 41%, and Hispanic exonerees 13%. Male exonerees were 85% of the sample. This is in line with known statistics regarding gender and incarceration. The mean time-to-exoneration was found to be 1,223.12 days. The median time-to-exoneration was 1,000 days. As expected, it was found that the exoneration event variable needed to be censored, pointing to the necessity for survival analysis.

Table 1. Descriptive Statistics of the Measures

Measure	Mean	Median	Standard Deviation	Min	Max
Biological Sex (Male)	.85	-	-	0	1
Age	32.60	-	11.44	13	83
Race/Ethnicity					
White	.41	-	-	0	1
Black	.46	-	-	0	1
Hispanic	.13	-	-	0	1
Official Misconduct	.35	-	-	0	1
Inadequate Legal Defense	.24	-	-	0	1
Drug Crime	.39	-	-	0	1

[42]Raymond Paternoster, *et al* "Using the correct statistical test for the equality of regression coefficients" (1998) 36:4 Criminology 859-866, online: <https://doi.org/10.1111/j.1745-9125.1998.tb01268.x> [*Paternoster, et al*].

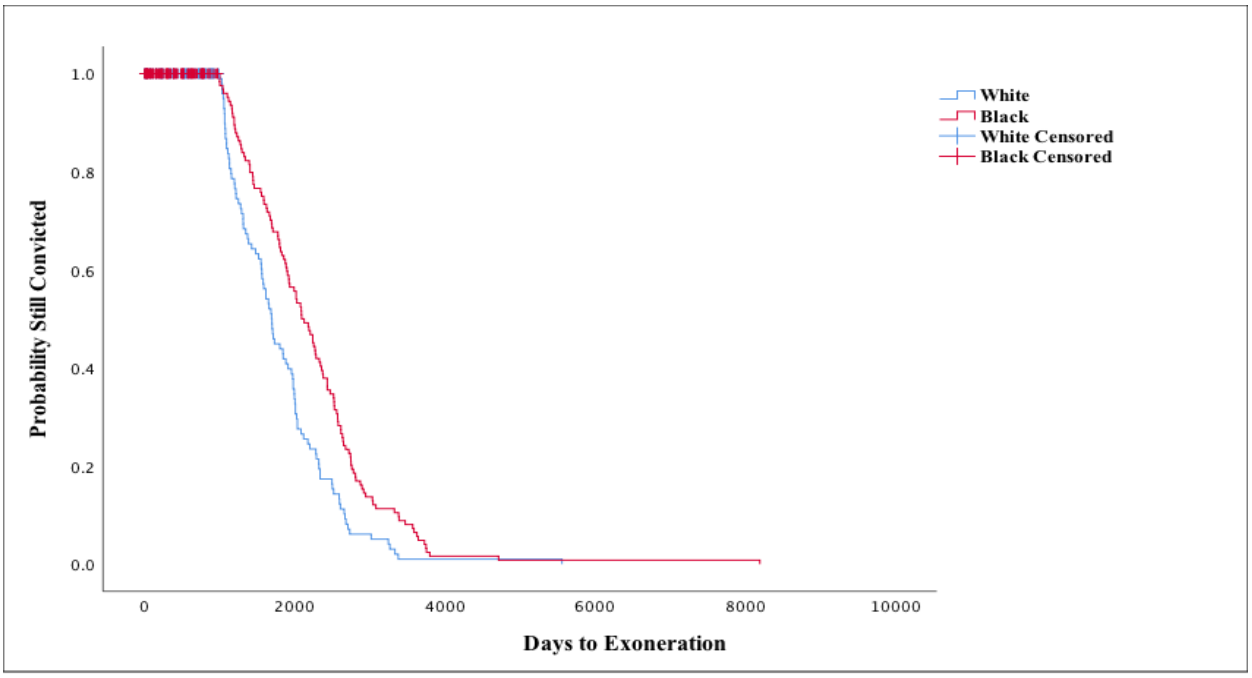
Violent Crime	.29	-	-	0	1
False/Misleading Forensic Evidence	.27	-	-	0	1
Mistaken Eyewitness ID	.10	-	-	0	1
Perjury or False Accusation	.40	-	-	0	1
Dependent Measure		-	-	-	-
Time (days) to Exoneration	1223.12	1000	1013.823	3	8199

N = 489

In step two Kaplan Meier analysis was performed. The results can be seen in Figure 1 and Figure 2 below. The Kaplan-Meier analysis demonstrated that the survival and hazard functions do vary by race and ethnicity. Specifically, the results showed that Black exonerees experienced a longer time-to-exoneration than did White exonerees. The Chi-square statistic indicated the differences were significant. These results provide supporting evidence for the first hypothesis. Additionally, it provides supportive evidence for FCT in the context of time-to-exoneration.

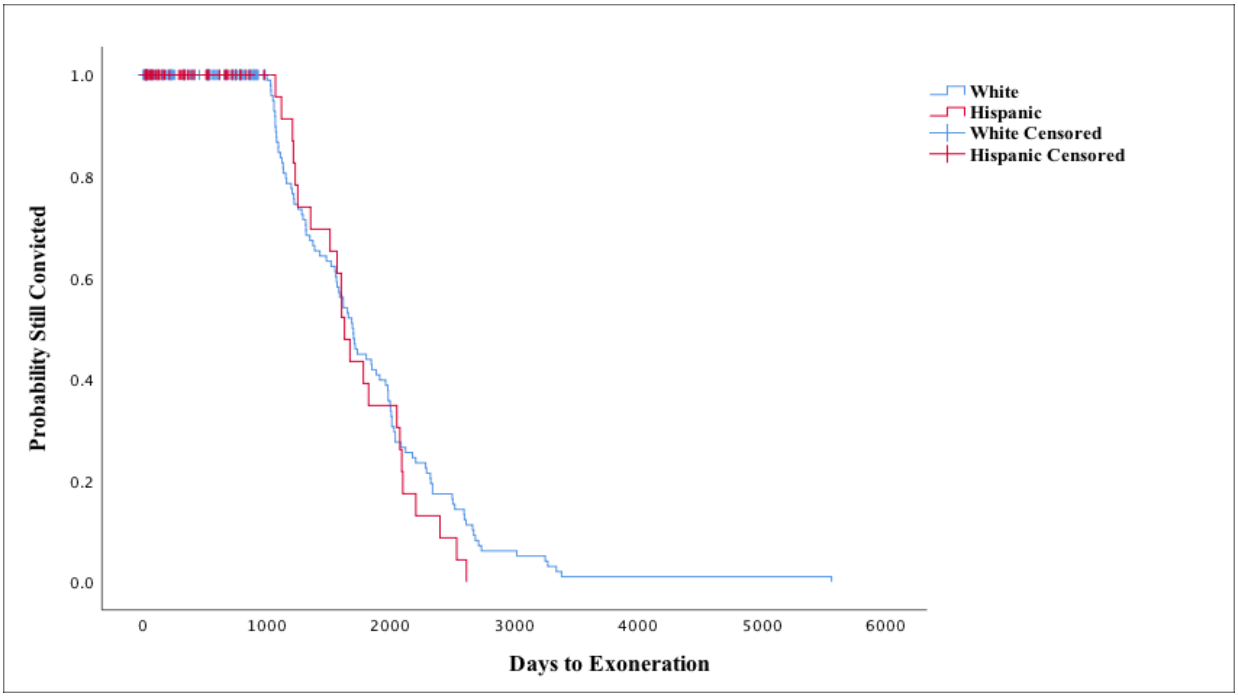
The results of the second Kaplan-Meier analysis indicated that overall, Hispanic exonerees experienced a shorter time-to-exoneration than did White exonerees, though the Chi-square indicated the differences were not significant.

Figure 1. Survival Analysis Results of Blacks versus Whites Median Days to Exoneration



Chi-square = 14.75, 1 *df* *p= .000

Figure 2. Survival Analysis Results of Hispanics versus Whites Median Days to Exoneration



Chi-square = .610, 1 *df* p = .435

In step three, logistic regression analysis was performed to explore how the extralegal and legal measures are associated with the time-to-exoneration. First, logistic regression was performed on a group which included only Black exonerees with White exonerees as the comparison. Second, logistic regression was performed on a group which included only Hispanic exonerees with White exonerees as the comparison.

Table 2 illustrates the results of the Black vs White exoneree group analysis. Results indicate the legal variable of Inadequate Legal Defense (b = .580, Exp(b) = 1.787, p <.05) and being convicted of a Violent Crime (b= .739, Exp(b) = 2.095, p < .01) resulted in an increase of the likelihood of a longer time-to-exoneration. The race of the exoneree was not found to be a significant factor in time-to-exoneration.

The results of this group comparison do not support the first hypothesis but are supportive of the second hypothesis. Additionally, the increase in time-to-exoneration for those convicted of a violent crime supports the premise of FCT that severity of the crime is of importance to decision makers.[43]

[43] Steffensmeier 1998, *supra* note 17.

Table 2. Black and White Group Exonerees Logistic Regression Analysis N= 421

Measure	B	SE	Exp (B)	Tolerance
Biological Sex (Male)	.533	.302	1.704	.933
Age	.008	.010	1.008	.858
Race/Ethnicity (Black)	.405	.228	1.499	.828
Official Misconduct	.477	.261	1.612	.709
Inadequate Legal Defense	.580*	.267	1.787	.830
Drug Crime	-.526	.330	.591	.416
Violent Crime	.739**	.298	2.095	.568
False/Misleading Forensic Evidence	.422	.297	1.524	.620
Mistaken Eyewitness ID	-.449	.421	.639	.659
Perjury or False Accusation	-.168	.276	.845	.584
-2logliklihood = 539.378				
Nagelkerke R^2 = .130				
Cox & Snell R^2 = .097				
Chi-square = 42.994 p =.000				

* p <.05 ** p <.01

Table 3 illustrates the results of the Hispanic vs White exoneree group analysis. The results showed the legal variable False or Misleading Forensic Evidence (b= 1.038, Exp(b) = 2.823, p < .01) and being convicted of a Violent Crime (b= .913, Exp(b) = 2.491, p < .01) resulted in a greater likelihood of being exonerated after the median time. Exonerees who had been convicted of a Drug Crime were shown to be 64.1 % less likely (b= - 1.025, Exp(b) = .359, p < .05) to be exonerated after 1,000 days. The race and ethnicity of the exoneree was not found to impact time-to-exoneration for this group.

The results of this group comparison are not supportive of the first hypothesis. However, these results do show support for the second hypotheses. Additionally, these findings demonstrate support for FCT which proposes legal factors do wield influence on outcomes.

Table 3. Hispanic and White Group Exonerees Logistic Regression Analysis N= 266

Measure	B	SE	Exp (B)	Tolerance
Biological Sex (Male)	.157	.380	1.170	.929
Age	.001	.012	1.001	.824
Race/Ethnicity (Hispanic)	-.390	.338	.677	.885
Official Misconduct	.105	.326	1.110	.724

Measure	B	SE	Exp (B)	Tolerance
Inadequate Legal Defense	.593	.332	1.810	.836
Drug Crime	-1.025*	.431	.359	.510
Violent Crime	.913**	.345	2.491	.730
False/Misleading Forensic Evidence	1.038**	.400	2.823	.798
Mistaken Eyewitness ID	.580	.711	1.786	.885
Perjury or False Accusation	.276	.332	1.318	.657
-2logliklihood = 316.370	-	-	-	-
Nagelkerke R^2 = .230	-	-	-	-
Cox & Snell R^2 = .172	-	-	-	-
Chi-square = 50.216 p =.000	-	-	-	-

* p <.05 ** p <.01

Probing this for racial and ethnic differences among the FCT concepts, next logistic regression analysis was re-estimated for each race and ethnicity. After the estimation of this regression, the Paternoster *et al*. z-score was applied to assess any racial differences that occur in the measures of focal concerns as they relate to time-to-exoneration.[44] The results of the regression analysis and z-scores for the whole sample are depicted in Table 4.

The data shows that 49.5% of White exonerees experienced a time-to-exoneration more than 1,000 days. The results of the regression analysis show that for White exonerees False or Misleading Forensic Evidence resulted in being 2.816 times more likely (b= 1.035, Exp(b) = 2.816, p < .05) to be exonerated after the median time. The legal factor of Drug Crime was found to decrease the odds of having to wait longer than 1,000 days for exoneration by 73%. These results are supportive of FCT in that legal measures were found to impact time-to-exoneration.

The data reveals that 56% of Black exonerees experienced a time-to-exoneration that was greater than the median time. The results indicate for Black exonerees Official Misconduct resulted in being 2.283 times more likely (b= .825, Exp(b) = 2.283, p < .05) to wait longer than 1,000 days to be exonerated. The results also showed Inadequate Legal Defense resulted in being 2.477 times more likely (b=.907, Exp(b) = 2.477, p < .05) to be exonerated beyond the median time. These results support the second hypothesis as well as the FCT measures in that legal measures affected time-to-exoneration for Black exonerees.

The data shows that 34% of Hispanic exonerees were incarcerated more than approximately three years while waiting to be exonerated. The results indicate for every one unit increase in age at the time of crime commission, there was a decrease of 7.9% (b = -.082, Exp(b) = .921, p <.05) in the likelihood of being exonerated after 1,000 days. Inadequate Legal Defense

[44]Paternoster, *et al, supra* note 42.

(b= 4.749, Exp(b) = 115.45, p < .000) resulted in a Hispanic exoneree being 115.45 times more likely to wait more than 1,000 days to be exonerated. The legal variable Violent Crime was shown to result in being 41.801 times more likely (b= 3.733, Exp(b) = 41.801, p < .01) to spend more than 3 years incarcerated before being exonerated. False or Misleading Forensic Evidence resulted in a Hispanic exoneree being 15.956 times more likely (b= 2.770, Exp(b) = 15.956, p < .01) to wait longer than 1,000 days to be exonerated. The findings for the Hispanic exonerees demonstrate support for the second hypotheses that the legal and non-legal measures representing FCT are associated with racial and ethnic differences in time-to-exoneration.

The results of the logistic regression analysis within each racial and ethnic group provide support for the second hypothesis. Further, support for the influence of both the legal and non-legal measures which represent the FCT measures is demonstrated. Specifically, the results indicated that the impact of these variables on time-to-exoneration were not the same for all races and ethnicities.

Next, to establish whether the disparities between racial and ethnic groups was statistically significant, the Paternoster *et al*. z-score was applied.[45] The results of this analysis can be seen in Table 4. The results showed no significant differences in the impact on time-to-exoneration when comparing the disparities in the FCT measures between White exonerees and Black exonerees. However, there were several significant disparities found between Hispanic exonerees and White exonerees. Namely, the differences in the slopes of the extra-legal measure of age, and the legal measures of Inadequate Legal Defense and Violent Crime were found to be statistically significant.

The results indicate that the impact of being younger at the time of conviction was stronger (p <.05) for Hispanics than for White exonerees, resulting in greater time-to-exoneration for younger wrongfully convicted Hispanics than for younger wrongfully convicted White exonerees. Similarly, the detriment to Hispanic exonerees who had Inadequate Legal Defense in their case, was greater (p <.05) than it was for White exonerees with this factor. Lastly, the impact of being convicted of a Violent Crime was greater for Hispanic exonerees (p <.05) than it was for White exonerees, resulting in significantly greater likelihood for Hispanic exonerees of experiencing a time-to-exoneration which exceeded 1,000 days. These results are supportive of the first and second hypothesis. Further, they offer support for FCT premise that legal and non-legal factors affect outcomes.

Table 4. Logistic Regression Analysis of Exonerees by Race/Ethnicity

	White			Black			Hispanic			Black/ White	Hispanic White
Measure	B	SE	Exp (B)	B	SE	Exp (B)	B	SE	Exp (B)		Z-score
Biological Sex (Male)	.271	.415	1.311	.813	.474	2.255	-.302	1.159	.739	.63	.47
Age	.007	.014	1.007	.011	.014	1.011	-.082*	.041	.921	.20	2.1*

[45] *Ibid.*

	White			Black			Hispanic			Black/ White	Hispani White
Measure	B	SE	Exp (B)	B	SE	Exp (B)	B	SE	Exp (B)		Z-score
Official Misconduct	.078	.363	1.081	.825*	.407	2.283	1.025	.984	2.787	.55	1.1
Inad Legal Defense	.286	.359	1.331	.907*	.419	2.477	4.749***	1.604	115.45	1.13	3.1*
Drug Crime	-1.315*	.545	.268	-.055	.457	.947	1.106	1.176	3.022	-1.77	-1.9
Violent Crime	.712	.386	2.037	.841	.499	2.319	3.733**	1.437	41.801	-.20	-2.0*
False/Mis Forensic Ev	1.035*	. 488	2.816	-.033	.404	.968	2.770**	1.098	15.956	1.69	-1.4
Mistaken Eyewit ID	1.166	1.158	3.208	-.838	.557	.433	1.130	1.981	3.095	1.56	.02
Perjury/False Accus	.019	.370	1.020	-.516	.441	.597	1.861	1.059	6.429	.93	-1.6
2logliklihood =	246.085			= 282.711			= 49.145				
Nagelkerke R^2 =	.178			= .135			= .592				
Cox & Snell R^2 =	.134			= .101			= .427				
Chi-square =	28.381 p =.001			= 23.624 p = .005			= 37.876 p =.000				
	N = 198			*N*=223			*N*=68				

*p<.05 **p <.01 ***p <.000

VI Discussion

The results of this study demonstrate support for the first hypothesis that racial and ethnic differences do occur in days-to-exoneration. Specifically, Black exonerees were shown to experience longer times to exoneration than White and Hispanic exonerees. The results of this study also show support for the second hypothesis that the legal components of a case are associated with racial and ethnic differences in time-to-exoneration. Illustrating this, while overall as a group Hispanics were not shown to spend more time before being exonerated than Whites or Blacks, it was demonstrated that certain legal variables resulted in Hispanics being more likely to experience longer times to exoneration than White exonerees with those same factors and the differences were statistically significant. This finding suggests inequity in the way these factors affected the different racial and ethnic groups.

In addition, this study provides support for the FCT premise that both legal and extralegal factors impact outcomes. In this study, it was shown that a violent crime conviction impacted the

length of time-to-exoneration. This is consistent with FCT research .[46] However, it was illustrated that the level of the severity of the crime did not impact all races and ethnicities in the same manner.

The study illustrated Inadequate Legal Defense resulted in a longer time-to-exoneration for Hispanic exonerees than for White exonerees with this same factor. This illustrates support for FCT in that it illustrates how stereotypes surrounding minorities and proclivity for criminal activity may impact access to adequate legal resources and result in detriment to certain racial and ethnic groups. Additionally, this finding may indicate bias towards minorities who are non-English speaking, impacting their ability to secure adequate legal defense.

Support for the FCT was demonstrated in that the non-legal measure of age was found to be significant for Hispanic exonerees when compared to White exonerees. Hispanic exonerees who were younger at the time of conviction spent longer awaiting an exoneration than White exonerees who were younger at the time of conviction. This illustrates the prominent stereotype of young minority's association with drugs, violence, and crime in general which is commonly portrayed through popular media in our society.[47]

The results of this study point to racial and ethnic differences in time-to-exoneration which disadvantages minorities and echoes the findings of other studies which have examined race and time-to-exoneration.[48] Additionally, this study demonstrates that legal components of a case do impact the racial and ethnic groups differently regarding time-to-exoneration. Support is found for the FCT premise that this is as a result of a reliance on stereotypes that suppose criminality to be a persistent attribute of non-Whites. [49]

A. Policy and Programming

It is widely acknowledged that for justice policy and programming to be effective, empirical study must go beyond informing on the state of the issue at hand and provide the critical link between research and practice.[50] Therefore, the following recommendations are offered for consideration.

---

[46] John Kramer & Darrell Steffensmeir, "Race and Imprisonment Decisions" (1993) 34:2 The Sociological Quarterly 357–376, online: <https://www.jstor.org/stable/4120706>; Steffensmeier 1998, supra note 17 ; Jeffery T Ulmer & Brian Johnson, "SENTENCING IN CONTEXT: A MULTILEVEL ANALYSIS" (2004) 42:1 Criminology 137–178, online: <https://onlinelibrary.wiley.com/doi/10.1111/j.1745-9125.2004.tb00516.x>.

[47] Warren, *supra* note 19.

[48] Olney, *supra* note 6; *Gross et al 2017*, *supra* note 4 ; Rafail & Mahoney, *supra* note 6.

[49]Albonetti, *supra* note 19; Steffensmeier 1998, *supra* note 17; Bridges & Steen, *supra* note 20.

[50] Richard R. Bennett, "Comparative criminology and criminal justice research: The state of our knowledge" (2004) 21:1 Justice Quarterly 1–21, online: <https://www.tandfonline.com/doi/full/10.1080/07418820400095721>; John Gregrich, "A note to researchers: Communicating science to policy makers and practitioners" (2003) 25:3 Journal of Substance Abuse Treatment 233–237, online: <https://linkinghub.elsevier.com/retrieve/pii/S074054720300120X>; Daniel P. Mears, "Towards rational and evidence-based crime policy: (2007) 35:6 JCJ 667-682, online: <https://doi.org/10.1016/j.jcrimjus.2007.09.003>; Daniel P Mears & Sarah Bacon "Improving criminal

First, it is recommended that widespread training to address racial and ethnic stereotypes be implemented among the agencies and actors responsible for addressing claims of innocence. Evidence based research from the field of social-cognitive psychology indicates it is possible to reduce unconscious bias and reliance on stereotypes through education and training.[51] Applied to the criminal justice system, this education and training would likely have a positive impact on reducing racial and ethnic disparities.

Second, it is recommended that Congressional legislation be enacted that would require states to regularly assess racial and ethnic disparities in the post-conviction process and report the findings to qualify for available funding. This recommendation is based on a model of reforms within the juvenile justice system that were mandated by the Juvenile Justice Delinquency and Prevention Act (JJDPA). Drawing from this model, all states would be required to address racial and ethnic disparities in the post-conviction processes through identification of the points where racial and ethnic bias are present, the development of action plans, and by performing outcome-based evaluations. Additionally, states would be required to publish the results of the outcome-based evaluations annually, promoting transparency. The states would also be required to establish or designate existing bodies comprised of diverse stakeholders to act in an advisory capacity towards the aims of reducing racial and ethnic bias. In the context of the adult justice system, it is likely that such requirements would be effective for reducing racial and ethnic disparities in the time-to-exoneration of wrongfully convicted minorities.

Third, research has repeatedly exposed the striking invisibility of Hispanic and Latino individuals in the criminal justice data.[52] Capturing data is critical to transparency. The justice system is woefully lacking on consistent data leaving criminal justice actors very much in the dark and forced to rely on their own "gut" instinct in arriving at their decisions.[53] Therefore, data

---

justice through better decision making: Lessons from the medical system" (2009) 37:2 JCJ 142-154, online: <https://doi.org/10.1016/j.jcrimjus.2009.02.001>.

[51] Irene V Blair, "The Malleability of Automatic Stereotypes and Prejudice" (2002) 6:3 Pers Soc Psychol Rev 242–261, online: <http://journals.sagepub.com/doi/10.1207/S15327957PSPR0603_8>; Diana Burgess *et al*, "Reducing Racial Bias Among Health Care Providers: Lessons from Social-Cognitive Psychology" (2007) 22:6 J GEN INTERN MED 882–887, online: <http://link.springer.com/10.1007/s11606-007-0160-1>; Kareem J Johnson & Barbara L. Fredrickson, "We all look the same to me: Positive emotions eliminate the own-race bias in face recognition" (2005) 16:11 Psychol Sci 875-881, online: <https://doi-org/10.1111/j.1467-9280.2005.01631.x>; Devin G. Pope & Joseph-Prince Justin Wolfers "Awareness Reduces Racial Bias" (2013) 64:11 Management Science 4988-4995, online: <https://doi.org/10.1287/mnsc.2017.2901>.

[52]Neelum Arya *et al*, "America's Invisible Children: Latino Youth and the Failure of Justice" (2009) SSRN Journal, online: <http://www.ssrn.com/abstract=1892966>; R A Valencia, "Latinos and the Criminal Justice System: An Overview of the Invisible/Visible Minority" (1994) 27:1 Harv Latino L Rev, online: <https://escholarship.org/uc/item/3664v2kc>.

[53] Pierre H Bergeron & Michael P Donnelly, "How a Spreadsheet Could Change the Criminal-Justice System", (14 December 2020), online: The Atlantic

collection guidelines and procedures that record ethnicity consistently throughout the justice system are of paramount importance to facilitate accurate analysis of criminal justice data and guide reforms that address ethnic disparities.[54]

Lastly, it is suggested that considerable attention be devoted to the problems inherent in communications with minorities who speak little to no English. Research shows language barriers pose significant hurdles for Hispanics and Latinos, which often impacts their ability to comprehend what is transpiring in the legal process at all stages.[55] This study illustrated the impact of certain factors which resulted in a longer time to on the length of time a Hispanic exoneree experienced, including inadequate legal defense. Problems with communication due to language barriers prevent non-English speaking defendants from accessing adequate representation, from responding to authorities in ways that could be beneficial to them, and from accessing information needed to pursue and exoneration.[56] Therefore, it is recommended that policies be put in place that guarantee translation services that have been vetted will be provided to all individuals who require them. Furthermore, it is recommended that the appropriate steps be taken to ensure easy access to legal forms, transcripts, and other media in the language of the individual seeking an exoneration.

VII Limitations

While the current dissertation contributes to the literature on time-to-exoneration, it is not without limitations. This study uses secondary data that does not directly or completely measure the concepts of FCT. Second, the possibility exists the data in the study are not accurate. To date, these data have been considered the most comprehensive and representative data on exonerations.[57] Third, consistent with the biases of exonerations, which are more likely to occur with more severe crimes, violent crimes are overrepresented in the data which presents a statistical limitation. Fourth, the data considered Hispanic as a separate category but did not specify what races were captured within this category. It is therefore possible that race and ethnicity could be crossed. Fifth, the number of Hispanic exonerees in this study was relatively small, which could impact the estimates. Sixth, the data on exonerations represented actual exonerations as captured by the NRE and, as such, was not equally distributed across geographical locations. This presents the possibility that factors associated with place could have an impact on the findings. Lastly, the data does not provide any measures on access to legal services and other support that may have an impact on time-to-exoneration.

<https://www.theatlantic.com/ideas/archive/2020/12/how-a-spreadsheet-could-change-the-criminal-justice-system/617370/>.

[54] Steffensmeier & Demuth 2001, *supra* note 19.

[55] Cecilia Menjívar & Cynthia Bejarano "Latino immigrants' perceptions of crime and police authorities in the United States: A case study from the Phoenix metropolitan area" (2010) 21:7 Ethn Racial Stud 120-148, online: <https://doi-org.ezproxy.lib.torontomu.ca/10.1080/0141987032000147968>.

[56] Cruz Reynoso, "Hispanics in the Criminal Justice System" in *Hispanics in the United States: An Agenda for the Twenty-first Century* (New York: Routledge, 2017); Martin G Urbina, "Language Barriers in the Wisconsin Court System: The Latino/a Experience" (2004) 2:1–2 Journal of Ethnicity in Criminal Justice 91–118, online: <http://www.tandfonline.com/doi/abs/10.1300/J222v02n01_06>.

[57] Rafail & Mahoney, *supra* note 6.

Future research should investigate the impact of other extralegal factors on time-to-exoneration. This may yield important insights into time-to-exoneration. Additionally, it may provide understanding of the interaction of factors.

VIII Conclusion

This study examined the racial and ethnic differences in time-to-exoneration using FCT. These results are limited by their secondary nature, validity, and cross-sectional nature. Despite these limitations, this study is the first study to assess the time-to-exoneration using a theoretical perspective via survival analysis among a national sample of individuals and shows racial and ethnic disparities in exonerations do exist. This study makes a modest contribution to the literature on the racial and ethnic differences in time-to-exoneration. Moreover, there is little theoretical driven research in this area, and this study has shed some light on how FCT can help explain the racial and ethnic differences in time-to-exoneration. The results can be utilized in guiding policy and developing reform measures.

A Computational Decision-Tree Approach to Inform Post-Conviction Intake Decisions

Kalina Kostyszyn, Carl J. Wiedemann, Rosa M. Bermejo, Amie Paige, Kristen W. Kalb-DellaRatta, Susan E. Brennan
Stony Brook University
Stony Brook, New York
U.S.A.

How might data analytic tools support intake decisions? When faced with a request for post-conviction assistance, innocence organizations' intake staff must determine (1) whether the applicant can be shown to be factually innocent, and (2) whether the organization has the resources to help. These difficult categorization decisions are often made with incomplete information (Weintraub, 2022). We explore data from the National Registry of Exonerations (NRE; 4/26/2023, N = 3,284 exonerations) to inform such decisions, using patterns of features associated with successful prior cases. We first reproduce Berube et al. (2023)'s latent class analysis, identifying four underlying categories across cases. We then apply a second technique to increase transparency, decision tree analysis (WEKA, Frank et al., 2013). Decision trees can decompose complex patterns of data into ordered flows of variables, with the potential to guide intermediate steps that could be tailored to the particular organization's limitations, areas of expertise, and resources.

Acknowledgements: Special thanks to Dr. Nancy Franklin for content expertise on this project. This material is based upon work supported by the National Science Foundation under Award No. 2125295 (NRT-HDR: Detecting and Addressing Bias in Data, Humans, and Institutions). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

I. Introduction
  A. The Promise and Pitfalls of Data-Intensive Methods
  B. Wrongful Convictions and the Intake Process
  C. The NRE and the Six Canonical Factors
  D. The Current Analysis
II. Method
  A. Sample
  B. Variables
    a. Covariates
  C. Data Analysis Plan
    a. Latent Class Analysis (LCA)
    b. Decision Tree Analysis
III. Summary of Results
  A. Latent Class Analysis Reproduction
  B. Decision Trees
    a. Six-Factor Model
    b. Extended Model

c. Removing ‘No-Crime’ and ‘DNA’ Cases
d. Removing Innocence Organization and Conviction Integrity Unit Information
e. Adding State-Wise Information
IV. Discussion
A. Classification Disagreements
B. Next Steps
a. Coping with Continuous Variables
b. Re-Assessing Latent Class Labels
c. Potential Uses
C. Policy Implications
D. Conclusion
V. References
VI. Appendix A

I Introduction

A. The Promise and Pitfalls of Data-Intensive Methods

Data and data-intensive methods are increasingly promoted—and indeed, sometimes mandated—as solutions within domains that call for people to make difficult decisions about pressing human problems. The fair and ethical use of such methods requires transparency, especially when the stakes are high. However, AI/machine learning tools can be notoriously opaque to human users. Opacity in algorithms can result in biased decisions that, once made, are difficult to challenge. The damage done can be life-changing and difficult to reverse (e.g. firing good teachers for the wrong reasons; Turque, 2012; O’Neil, 2016). When data-intensive algorithms are “black boxes,” it’s difficult to understand the reasoning behind the outcomes. Therefore, it is necessary to advocate for transparency in two key ways: first, the variables included in the training data for algorithms should be justifiable, and second, it should be clear how these variables are evaluated or weighted in outcomes or predictions. This is particularly important for algorithm-aided decisions in the domain of criminal justice, which can have profound impacts on vulnerable individuals.

Bias can creep into algorithms in different ways. When machine learning models are trained on datasets that are missing relevant information, the models produce outcomes that are unreliable for those cases that are underrepresented in the datasets; this may result in reproducing the biases of the past, or in otherwise unreliable outcomes. For example, “state-of-the-art” facial recognition algorithms learned to detect White faces better than Black faces and male faces better than female faces (Buolamwini & Gebru, 2018), making errors when classifying new cases that were not well represented in the training data. And when intersectional identity is sparsely represented and unevenly distributed in training data, “fairness gerrymandering” may result (Kearns & Roth, 2020), as it did when the faces of Black women were recognized least accurately of all (Buolamwini & Gebru, 2018). Transparency in the kind of data used to develop, train, and test algorithms is necessary to understand and ultimately prevent the potential misclassification of underrepresented individuals down the line.

Bias can also stem from the variables chosen for inclusion in training data. During training, models learn to represent underlying patterns among variables in the data in ways that are hidden even from their developers, and that may perpetuate undesirable stereotypes. This can occur even when key variables such as race or gender are removed from a dataset (and thus are considered to be "protected"). It may seem sufficient that a protected variable isn't included in training a model, yet it can nevertheless still influence the outcome when other (proxy) variables that are correlated with the protected variable are included (see, e.g., O'Neil, 2016 & Angwin *et al*., 2016 for discussion of risks associated with proxies). For example, Amazon discontinued using an algorithm trained to identify successful job candidates after it was discovered that, despite removing gender as a variable, the algorithm still recommended men over women (Dastin, 2018); the resumés in the training data still included information such as extracurricular activities strongly correlated with gender. As another example, some states require that inmates fill out questionnaires that are used to support automated predictions about recidivism. Although asking about race is illegal in some jurisdictions and therefore avoided, questions about family members' unemployment or welfare status, or about the age at which an individual first interacted with the police, can differentiate privileged from underprivileged individuals (and may divide them along race-based lines; Angwin *et al*., 2016). In this way, variables associated with privilege or lack thereof can serve as a proxy for race while ignoring that one's first contact with the police may be a result of biased policing practices (O'Neil, 2016; see also Harcourt, 2015).

Another source of bias can arise when algorithms are deployed as decision aids without taking into account whether the distribution of errors is equitable and fair, or whether there are disparate impacts on individuals. For example, Northpointe's COMPAS (Correctional Offender Management Profiling for Alternative Sanctions; T. Brennan & Dieterich, 2017) algorithm derives individuals' recidivism risk scores from questionnaires given to them when they are incarcerated. COMPAS does not include an overt race variable, and its developers claimed that it was unbiased because its error rate in predicting recidivism for Black parolees was the same as for White parolees (39% for both). In an adversarial audit, the public interest group ProPublica obtained a dataset through a FOIA (Freedom of Information Act) request of more than 7,000 parolees in Broward County, FL over a 2-year period; all had been given the COMPAS algorithm's lengthy input questionnaire when incarcerated (Angwin *et al.*, 2016). The ProPublica team painstakingly unearthed the ground truth about whether these individuals actually recidivated over the next several years and merged this information with COMPAS's predictions about them (as recounted in Christian, 2020). They found that the *types* of errors were dramatically different for Black and White parolees: approximately two-thirds of errors for White parolees were false negatives (where COMPAS had recommended release, but the parolee had recidivated), whereas two thirds of errors for Black parolees were false positives (where COMPAS had recommended denying parole, but the individual was paroled and there was no recidivism) (Angwin *et al*., 2016). This distribution of errors privileges one group while being grossly unfair to another. Yet data-intensive decision-making aids such as COMPAS are mandated in many jurisdictions around the U.S. (and with limited or no oversight; see Christian, 2020).

Biases can be further compounded when a decision-support algorithm is deployed blindly by those who should ultimately be the ones accountable for a decision, but who don't understand the limitations of the algorithm within their context of use. The COMPAS algorithm was designed to assist with judges' decisions about eligibility for parole or treatment programs (Angwin *et al.*,

2016). Yet it has been applied to decisions about bail, pre-trial detention, and sentencing (uses that even the developers deem to be inappropriate; Angwin *et al*., 2016 & Christian, 2020).

Finally, although decision-support algorithms such as COMPAS are often used to assess risks posed by individuals accused or convicted of crimes (for the benefit and protection of society), these data-intensive methods can also be used to determine when and how to provide benefit and support to the accused or convicted individuals themselves. Whenever algorithms are used to recommend life-changing decisions, transparency is essential, not only to prevent unintended harms, but also to undo the harms that may have resulted from multiple sources of bias. Here, we explore the use of data-intensive methods in the domain of *wrongful convictions.*

B. Wrongful Convictions and the Intake Process

Wrongful convictions, by their very nature, are not readily observable. Accordingly, the true rate of wrongful convictions is a *dark figure*, that is to say, a figure that is typically recognized as unknown or even unknowable, but at the very least, extremely difficult to ascertain (Bedau & Radelet, 1987; Gross *et al.,* 2014). One estimate based on a thoroughly-vetted survey of state prisoners (with non-parametric tests used to account for possible false innocence claims) suggests that 6% of incarcerations are based on wrongful convictions (Loeffler *et al*., 2019). Among capital cases, a conservative estimate of the rate of wrongful convictions is about 4% (Gross *et al.*, 2014). It can therefore safely be said that many people are actively serving prison sentences for crimes that they did not commit, or that did not even occur in the first place.

Through exoneration, the official alleviation of legal culpability for a crime that a person was originally found to be guilty of, victims of wrongful conviction may find an avenue to justice. Innocence organizations are a group of legal representatives and advocates for the wrongfully convicted. With more than 900 contributions to exonerations, innocence organizations play a vital role in exonerating the wrongfully convicted. As of 2023, there are 72 member organizations in the Innocence Network spread across the United States. The Innocence Network serves as a community that provides various forms of support for newly exonerated individuals in addition to providing resources for legal organizations that join its mission in exonerating the innocent. Whether an innocence organization accepts an application can depend on the availability of resources like the number of staff and budget. Innocence organizations can receive anywhere from 20 to 2,400 requests for assistance a year, and moving a case to exoneration is an extremely time-consuming process that intake staff estimate to take around seven years (Weintraub, 2022). The investigative processing of a case alone can take more than a year to complete (Krieger, 2011). Therefore, innocence organizations and staff must strategically allocate resources to cases they determine are most likely to be successful.

The inner workings of individual innocence organizations impact the types of cases they can investigate and litigate. A qualitative study of 19 innocence organizations by Weintraub (2022) found that intake procedures vary among organizations. Such variations include 1) length of application, 2) whether the application is reviewed by either intake staff or directors, attorneys, or law students, 3) whether an organization conducts a screening interview with the applicant, and 4) intake criteria. Common intake criteria of most innocence organizations include factual innocence and geographic restrictions within a certain state or region, but organizations vary on acceptance

or consideration of cases involving child sexual abuse and sustained abuse, whether an applicant was involved in the criminal action, cases with DNA evidence, arson, shaken baby syndrome, guilty pleas, new evidence of innocence at intake, indigent status, and sentence length (Weintraub, 2022).

To support intake staff as they categorize and evaluate post-conviction requests, data-intensive decision support tools should empower them to effectively interpret and communicate about the results of multi-step data-driven analyses.

C. The NRE and the Six Canonical Factors

To understand the myriad of factors that contribute to wrongful convictions, data from successful exoneration cases can be illuminating. To this end, the Innocence Project actively maintains, updates, and consults a national dataset containing information on DNA-based exonerations (Innocence Project, *Cases*, 2023). Through examination of this dataset, the Innocence Project has identified "contributing causes" of wrongful convictions exposed via DNA evidence as: eyewitness misidentification, misapplication of forensic science, false confession or incriminating statement, incentivized informants' statements, misconduct by government actors, and inadequate defense counsel (West & Meterko, 2016). These factors are particularly relevant to the Innocence Project's internal investigations and goals that focused originally on DNA evidence, but are not generalizable to the larger set of wrongful conviction cases that include non-DNA cases as well (Acker & Redlich, 2019).

The much broader National Registry of Exonerations (NRE) database aims to include all exonerations; for this reason, we focus here on the NRE. Founded in 2012, this database is maintained by a dedicated group of scholars, lawyers, and journalists who have cataloged data on successful exonerations both for DNA- and non-DNA-based cases that have occurred since 1989. As of April 26th, 2023 when we did our analyses, the database contained information on 3,284 cases in total, making it the most comprehensive and most-frequently cited (Gross, 2008) source of raw information on known wrongful convictions to date.

Each case in the NRE database includes at least one of six "canonical" factors that have been identified as common contributors to wrongful convictions: False Confession (FC), Mistaken Witness Identification (MWID), False/Misleading Forensic Evidence (F/MFE), Perjury/False Accusation (P/FA), Official Misconduct (OM), and Inadequate Legal Defense (ILD) (Acker & Redlich, 2019). Given the greater diversity of cases in this dataset, the six so-called canonical factors are presumably more appropriate for analyses seeking to shed light on wrongful convictions in general, compared to the causal factors related to DNA-based exonerations (ibid).

Due to their dichotomous nature, the six canonical factors can be used as indicator variables for the technique known as latent class analysis (LCA). LCAs are informative for datasets such as the NRE, as they identify latent (i.e., not directly observable) subgroups within populations (McCutcheon, 2002). This method can be considered analogous to factor analysis albeit for categorical data: Both analyses demonstrate the interrelatedness of indicator variables whose associations are explained by unobserved factors, rather than direct causal relationships (McCutcheon, 2002). Many cases in the NRE include more than one of the six canonical factors,

as they frequently co-occur. A benefit of applying LCA to a dataset such as the NRE is that classes extracted from the analysis would account for co-occurrences of the relevant subsumed canonical factors.

Our present project is inspired by the results of an LCA analysis of the NRE database, reported by Berube *et al.* (2023). In their paper, Berube and colleagues sought to identify patterns that broadly underlie wrongful convictions. Through applying LCA to the NRE, they found that a four-class model best fit the data and named the four extracted classes as follows: *Intentional Errors*, *Witness Mistakes*, *Investigative Corruption*, and *Failures to Investigate*. They then performed correlations with other NRE variables, such as exoneree demographics, measures of case severity, and process/evidence-related variables to examine how trends within the six canonical factors, as represented by the latent classes, related to other case factors.

D. The Current Analysis

Although LCA offers a useful method for extrapolating underlying patterns within a dataset such as the NRE, there are a number of important limitations that should be considered alongside its implementation (Weller *et al*., 2020). According to current best practices for LCA, as described by Weller *et al*. (2020), proper class assignment and percentages of representation within a particular class are not always guaranteed because LCAs rely on probability estimates to assign members of a dataset to a particular latent class (Muthén & Muthén, 2000). Weller *et al*. (2020) also warn of the heightened potential for "naming fallacies" to occur when researchers attempt to create labels for the extrapolated classes. Such labels may fail to appropriately capture the complexities of the determining factors in class memberships. Therefore, we first aimed to reproduce Berube *et al*.'s (2023) analysis, both to demonstrate the stability of their findings with a larger data set, and to simultaneously allow for a possible re-assessment of the labels originally conferred upon the latent classes. Second, to account for the inherent limitations of LCAs, we aimed to use the extracted classes as targets for a predictive analysis that would be more transparent and interpretable.

Models of complex data often use regression-based analyses to make predictions. However, these models can be nonoptimal for guiding human decision-making because of difficulties in interpreting and applying data to ambiguous, novel, and idiosyncratic cases. We introduce what we propose may be a more transparent framework using *decision trees* (Flach, 2012; Duda *et al*., 2001). Decision trees identify and lay out the impacts of variables one by one in a graphical representation similar to a flow chart, in a form that can be scrutinized by a human decision-maker. It may be possible to use decision trees to identify combinations of features relevant at different stages of evaluating a post-conviction case, to help with prioritizing new cases, and to direct attention to the most promising path to pursue next. Once the algorithm segments the dataset based on a particular feature, subsequent branches (or steps) can be interpreted more easily than the outcomes of classic regression analyses. Furthermore, integrating the grouping variables/classes identified by an LCA can improve model fit for decision trees (Gañan-Cardenas *et al*., 2022), making the pairing of these two approaches promising. A decision tree approach may uncover previously undetected trends in the data and increase the interpretability of results derived via LCA.

Ultimately, the framework we propose in this paper uses successful exonerations to evaluate and identify potential pathways that may be used during the intake process for new applicant cases. This might conserve work hours, identify specialized resources needed for a particular applicant, and transparently support efforts to expedite and communicate about decisions within an innocence organization. Of course, this framework will need to be tested within the context of use. To the extent that this framework may reveal previously unknown biases introduced prior to conviction, it may also allow for more effective communication with law enforcement, legislators, and other policy-making entities in an effort to reduce future wrongful convictions.

II Method

A. Sample

The analyses presented here were based on data from the National Registry of Exonerations, downloaded on April 26th, 2023. There was a total of 3,284 exonerees in the database at the time of download. Notably, Berube *et al*.'s (2023) latent class analysis was conducted on data from the same source, but at the time of their download the database included a total of 2,880 exonerees.

B. Variables

In accordance with our goal of assessing the reproducibility of Berube *et al*.'s (2023) analysis, we based our analyses on the same variables and the same coding scheme to the greatest extent possible. We therefore relied on variables included in the NRE dataset, such as the six canonical factors, exoneree demographic information, case severity measures, and process/evidence-related variables.

a. Covariates

Exoneree Demographic Information. Several NRE database variables concern demographic information about the exoneree and geographic information (jurisdiction) about the case; while we initially remove state information, we add it back later while experimenting with manipulations of the tree structure. We excluded exoneree names from analyses, as well as counties.[1] In addition, in the absence of intuition about how these variables are distributed or interact with other variables, we removed any continuous variables from our initial study—ages and dates, for example. We do, however, differentiate juveniles at time of conviction from adults, without further differentiating within those classifications. We retain information about race and sex (in fact, 'female exoneree' is a separate variable listed with the process information).

[1] We discovered that six exonerees were in the database twice, from multiple exonerations pertaining to the same case. For these cases, we used data from the chronologically later exoneration, which generally has more specific characteristics (or "tags" in NRE database parlance). For two of these exonerees, the LCA analysis assigned each entry a different class because of significant differences in tags, despite it being the same individual.

Case Severity Measures. Another set of variables deals with case severity. The 'worst crime display' variable contains values for the single most severe crime associated with each case; there are additional binary variables specifying whether attributes (such as homicide, sexual assault, etc.) were part of a case, distributing this information in a way that makes comparison simpler. Berube and colleagues (2023) also use the sentence length as a measure of case severity; just as we removed ages and dates, we remove this information. The sentence length may interact, in ways we currently are unable to discern, with conviction date, length of incarceration, and details associated with post-conviction actions.

Process/Evidence Related Variables. The bulk of the variables in the data set are divided by the NRE as either 'tags', which include information about the crime or the exoneration, or 'official misconduct tags', which contain more specific information about misconduct that led to wrongful conviction; these are coded as binary true/false values. A small number of them reiterate information in other variables - female exonerees and juvenile defendants have been mentioned, but there is in addition a variable that marks whether a case was held at the federal level and that is repeated in the state information.

Handling of Exceptional Covariates. As we examine the trees, a number of these covariates will be manipulated due to somewhat exceptional status. First, because the NRE dataset represents a snapshot in time (with the outcomes of cases that may have taken decades to adjudicate), we consider a class of variables that are determined only at the end of the exoneration case. We use the 'no crime' and 'DNA used in exoneration' variables as examples of information that did not contribute to the original conviction but was a basis for overturning it. Second, we code based on whether an innocence organization and/or conviction integrity unit participated in the exoneration process, though we recognize that these variables may not be useful for all intended analyses. Manipulating these variables, however, is important to demonstrate the flexibility of these decision trees and how they are able to make similar generalizations even when provided with different input data.

C. Data Analysis Plan

a. Latent Class Analysis (LCA)

We began by reproducing Berube *et al*.'s (2023) latent class analysis (LCA), which identified four underlying classes in the NRE dataset, such that each case could be categorized based on its highest probability of belonging to one of the four classes. Following Berube (ibid), we used the *Six Canonical Factors* that contribute to wrongful convictions (Acker & Redlich, 2019) as latent class indicators, coded dichotomously. These are: Mistaken Witness Identification (MWID), False Confession (FC), Perjury or False Accusation (P/FA), False or Misleading Forensic Evidence (F/MFE), Official Misconduct (OM), and Inadequate Legal Defense (ILD). Goodness of fit was assessed using multiple criteria, including the Bayesian information criterion (BIC; Schwarz, 1978) and Akaike's information criterion (AIC; Akaike, 1987).

b. Decision Tree Analysis

Using decision trees (WEKA, Frank *et al*., 2016), we decomposed and reanalyzed the four classes modeled in the LCA. First, we used decision trees to predict classification from the LCA approach, using only the Six Canonical Factors, to assess the validity in combining these approaches (*six-factor model*) and determine what ordered combinations of features could predict LCA-based classification. Second, we explored other trends within the four latent classes by examining covariates other than the canonical factors (*extended model*). The decision trees determined other associated features that predict the classification of a case. Case tags - such as *withheld exculpatory evidence* or *juvenile defendant* - were recoded as binary features where possible, then ordered by the decision tree to see how accurately combinations of these features could predict LCA-based classification.

III Summary of Results

A. Latent Class Analysis Reproduction

All statistical analyses pertaining to the LCA reproduction were conducted in R (R Core Team, 2020), with associated figures produced via the ggplot2 package (Wickham, 2016). In accordance with recommended best practices for LCAs (Weller *et al*., 2020), we successively fit a series of models, starting with a one-class model. A four-class model provided the best overall fit according to statistical criteria, thus supporting Berube *et al*. (2023) while expanding their analysis to a larger dataset. The optimal BIC value was associated with a four-class model (21052.42), as compared to a three-class model (21180.39) and a five-class model (21054.58). Similar to results reported by Berube *et al*. (2023), fit improvement, as indicated by AIC, from the four-class model to the five-class model (a difference of 40.52) was much smaller than fit improvement from the three-class model to the four-class model (a difference of 170.64). So, we maintain, in agreement with Berube *et al*. (2023), that a four-class model seems to best fit the data, despite a more favorable AIC value being associated with the five-class model. We also note that one of the classes in the five-class model included a membership of only 7% of cases. Such a low representation of the dataset could lead to issues both with generalizability and interpretability. Again, deferring to the four-class model appears to be the optimal solution.

Table 1. Results of Model Fit Comparisons

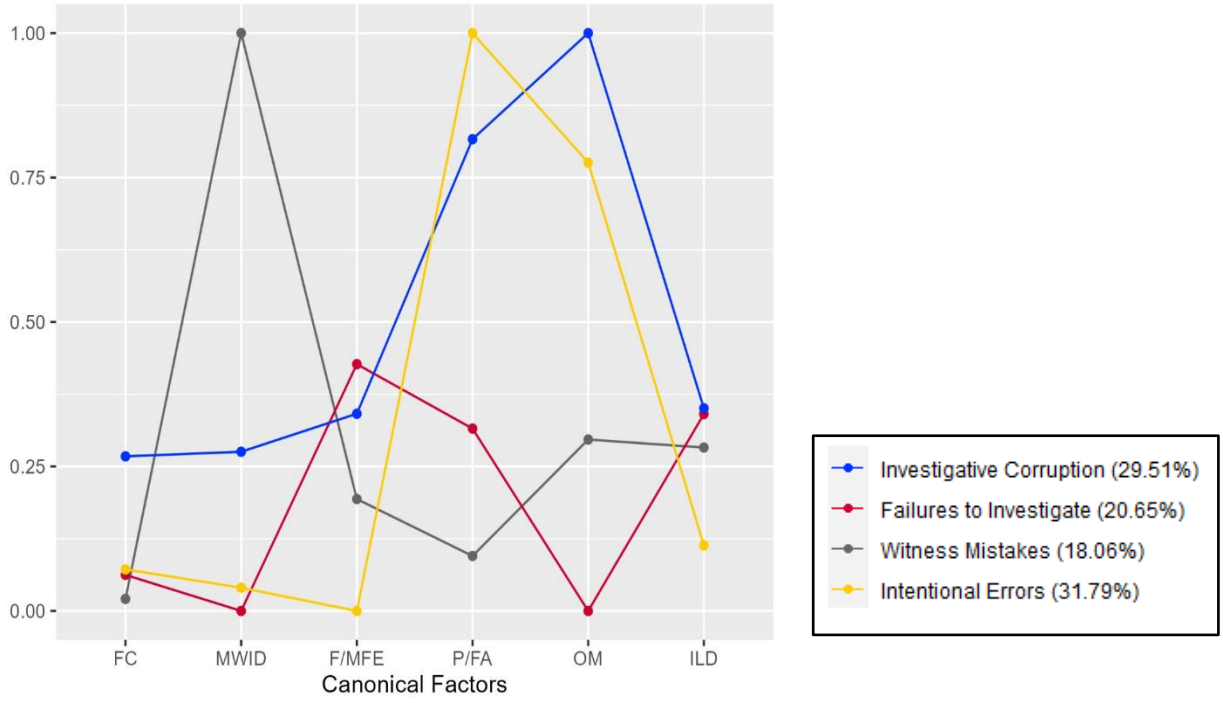
Model	AIC	BIC	BLRT	*P*-value	χ^2	*P*-value	Mixing Proportions
1-Class	22452.86	22489.44	1724.51	-	1849.51	-	-
2-Class	21528.89	21294.40	472.80	<.001	484.89	<.001	.37/.63
3-Class	21058.45	21180.39	302.10	<.001	315.37	<.001	.19/.64/.17
4-Class	20887.81	21052.42	117.46	<.001	115.32	<.001	.18/.21/.30/.32
5-Class	20847.29	21054.58	62.94	<.001	63.00	<.001	.16/.30/.16/.07/.31

Note. AIC = Akaike's information criterion; BIC = Bayesian information criterion; BLRT = bootstrap likelihood ratio test; χ^2 = chi-square. *P*-values were reported for BLRT and χ^2. Mixing proportions were based on the most likely latent class membership.

Distributions of predicted class memberships and the profiles of their respective representations of the Six Canonical Factors were highly similar to those reported by Berube *et al*. (2023). For example, 100% of the cases estimated to be members of the class originally labeled "Witness Mistakes" were associated with MWID. The canonical factor OM (Official Misconduct) was also highly indicative of cases assigned to the "Witness Mistakes" class. The "Investigative Corruption" class was most strongly characterized by OM (100% of assigned cases), P/FA (Perjury/False Accusation), and ILD (Inadequate Legal Defense), while the "Intentional Errors" class was most strongly characterized by P/FA (100% of assigned cases) and OM. The "Failures to Investigate" class was most strongly characterized by F/MFE (False/Misleading Forensic Evidence), but did not have as high of an association with this canonical factor as was observed for the factors that most strongly characterized the other latent classes. These patterns are largely in alignment with Berube *et al*.'s (2023) results. We therefore tentatively retain the labels reflecting Berube *et al*.'s (2023) interpretations, but a critical evaluation of these labels and their interpretability follow in the next paragraph and in subsequent sections of this paper. Percentages of predicted class membership were as follows: Intentional Errors, 31.79%; Witness Mistakes, 18.06%; Investigative Corruption, 29.51%; and Failures to Investigate, 20.65% (see *Figure 1*).

Figure 1. Latent Class Analysis of NRE Data as of April 26th, 2023 (Reproduction of Berube et al., 2023)

Latent Classes (as interpreted by Berube, 2023)



Note. FC = False Confession; MWID = Mistaken Witness Identification; F/FME = False/Misleading Forensic Evidence; P/FA = Perjury/False Accusation; OM = Official Misconduct; ILD = Inadequate Legal Defense.

Because a goal of the present work is to use these extracted classes as targets of prediction via decision trees, a critical evaluation of the extent to which classes are mutually exclusive is warranted. To that end, it is worth noting that both the Intentional Errors class and the Investigative Corruption class were characterized by high degrees of P/FA and OM. Visually, their patterns of representation were quite similar. The similarities in these patterns suggest that our subsequent decision tree analysis may be more likely to misidentify cases assigned membership to the Intentional Errors class as Investigative Corruption, or vice versa. Examining the LCA's posterior probabilities can help us predict the directionality of errors that a decision tree might make in predicting membership within these two classes. In LCA, posterior probabilities represent the probability of a given case to have otherwise been assigned to one of the alternative classes in the model. We thus ran a Welch's two sample *t*-test, comparing the mean posterior probability of members in the Intentional Errors class to have been categorized as Investigative Corruption (M = 0.18, SD = 0.11), to the mean posterior probability of members in the Investigative Corruption class to have been categorized as Intentional Errors (M = 0.15, SD = 0.18). Results of the *t*-test indicated that there was a statistically significant difference in these mean posterior probabilities, $t(1566.5) = 4.66$, $p < .001$. In other words, cases that were assigned membership to Intentional Errors had a higher mean posterior probability of being assigned to the Investigative Corruption class than the mean posterior probability of cases classified as Investigative Corruption to have been assigned to the Intentional Errors class. It should therefore be expected that, when using decision trees, classification disagreements will manifest such that cases originally assigned as Intentional Errors will be more often classified as Investigative Corruption, as opposed to the converse.

B. Decision Trees

We use WEKA's J48 package (WEKA, Frank *et al*., 2016) to build our decision trees. J48 is an implementation of the C4.5 algorithm (Quinlan, 1993), which is a Classification and Regression Tree (or CART model; Breiman *et al*., 1984) that, given the input database, will make partitions within the data based on how well a partition is able to generalize for classification. Variables used for these partitions have a high *information gain* at that point in the algorithm; the higher information gain a variable has, the more evenly its values subdivide the space, which minimizes the number of additional variables needed to classify a data point. A variable having low information gain does *not* mean that a given value is not representative of a class, but rather that the other values are not sufficiently discriminatory for other classes. At the point these low-gain variables are found in the decision tree, competing branches have been eliminated, making categorization based on the variable's value more likely.

To take advantage of the decision trees' learning ability, we ran several models, training two instances of each with either a 75% train-25% split or with 10-fold cross validation. In the first training regimen, 75% of the dataset was used to train the tree a single time, and evaluation was done over the held-out 25% of the dataset. In the second, we use 10-fold cross-validation (Stone, 1976), where we first partition the dataset into 10 equal sets, train a model over 9 of those ten, and rotate which model we test on the remaining 10th set. All other standard settings are untouched; in particular, we did not 'prune' the tree, or remove low-occurrence branches, as we wanted to examine the breadth of generalizations.

For evaluation, we primarily use precision, recall, and f-score measures. Precision is the percentage of selected items that belong to the target group versus selected items that were not targets; recall is the percentage of target items selected versus target items the model did not select. The f-score is the harmonic mean of these two. Additionally, we will list confusion matrices, which will display how many items in each group were correctly classified and, if not, which other category they were classified into. These values will elucidate the error rates and patterns of classification disagreements (here, disagreements in categorization between the LCA analysis and the decision tree), which we will analyze below.

a. Six-Factor Model

We begin with the six-factor model, which trains itself on the canonical factors associated with each case and predicts which latent class is attributed to each. This demonstrates how well the decision trees are able to interpret the underlying data given to the latent class models. Hearteningly, these models perform near-perfectly, easily using the canonical factors to categorize cases.

b. Extended Model

Next, we created an 'extended' model, in which we train on the set of covariates rather than the six canonical factors, while still predicting the latent class for each case. High performance here will demonstrate that the decision tree is finding underlying patterns in the covariates that align with the latent classes.

Table 2. Evaluative accuracy scores for the 6-factor and baseline extended models

Model	Precision	Recall	F-Score
6-factor, cross validated	1.000	1.000	1.000
6-factor, 75-25	0.999	0.999	0.999
Extended, cross validated	0.722	0.721	0.720
Extended, 75-25	0.737	0.737	0.737

Given high performance of the extended model, we can now manipulate the tree by excluding covariates that offer little predictive power regarding new cases under consideration, or including covariates which were initially set aside.

Table 3. Confusion matrices for baseline extended models (75/25 split)

	Classed FtI	Classed IE	Classed IC	Classed WM
True FtI	131	16	0	12
True IE	19	173	56	11
True IC	0	53	176	12
True WM	10	11	15	124

c. Removing ‘No-Crime’ and ‘DNA’ Cases

In order to more closely model the incomplete information that may be available to intake staff, in this section we remove variables that refer to the outcome of the exoneration. In the NRE, the variable 'DNA' refers specifically to new DNA evidence introduced in post-conviction that directly led to exoneration. In 'no-crime cases,' the exoneree was initially convicted of a crime that did not happen. This could be a crime that was entirely fabricated, or an incident that was mistaken for a crime, such as an accident or a suicide. Because these variables may be unknown at intake, we present a model here that makes predictions without them. These perform comparably to the baseline extended model, which used these variables for partition, suggesting that it is able to use other information in the data set to make similar generalizations.

Table 4. Confusion matrices for no DNA/no-crime extended models (75/25 split)

	Classed FtI	Classed IE	Classed IC	Classed WM
True FtI	114	25	0	20
True IE	7	172	57	23
True IC	0	35	200	6
True WM	18	12	26	104

d. Removing Innocence Organization and Conviction Integrity Unit Information

In this section, we remove variables that refer to the involvement of an innocence organization (IO) or conviction integrity unit (CIU), as they are characteristics of those who are vetting the case post-conviction rather than variables in place at the time of the crime, investigation, or prosecution. However, considering that our analyses may be used in the future to inform whether to accept a case, such information may be of value. It is worth noting that, while these two variables were included in the baseline model, that baseline was outperformed by the 75-25 split of this manipulation. This suggests that the high information gain of these variables may be preventing the model from revealing other, more informative partitions downstream. Because of these differences, it may be useful to observe a model with and without these variables to see where their predictions differ. This is not to say that the claims made by one model are inadequate, but rather that when the data is restructured by removing carefully selected variables, the model will compensate by taking advantage of new generalizations it previously could ignore.

Table 5. Confusion matrices for no IO/CIU extended models (75/25 split)

	Classed FtI	Classed IE	Classed IC	Classed WM
True FtI	136	11	0	12
True IE	20	172	54	13
True IC	0	45	185	11
True WM	11	5	23	121

e. Adding State-Wise Information

In this section, rather than removing variables, we add the state where the case occurred as a variable. Because this includes over 50 discrete values - every state or U.S. territory, plus federal and military cases, which are listed separately - this was removed from the initial tree for interpretability, but reintroduced to assess the predictive power of these variables. These models are comparable to the baseline, though state information, where it appears, partitions trees close to the leaves.

Table 6. Confusion matrices for extended models with state information (75/25 split)

	Classed FtI	Classed IE	Classed IC	Classed WM
True FtI	127	17	0	15
True IE	18	173	58	10
True IC	0	46	185	10
True WM	17	8	22	113

Table 7. Evaluative accuracy scores for modified extended models

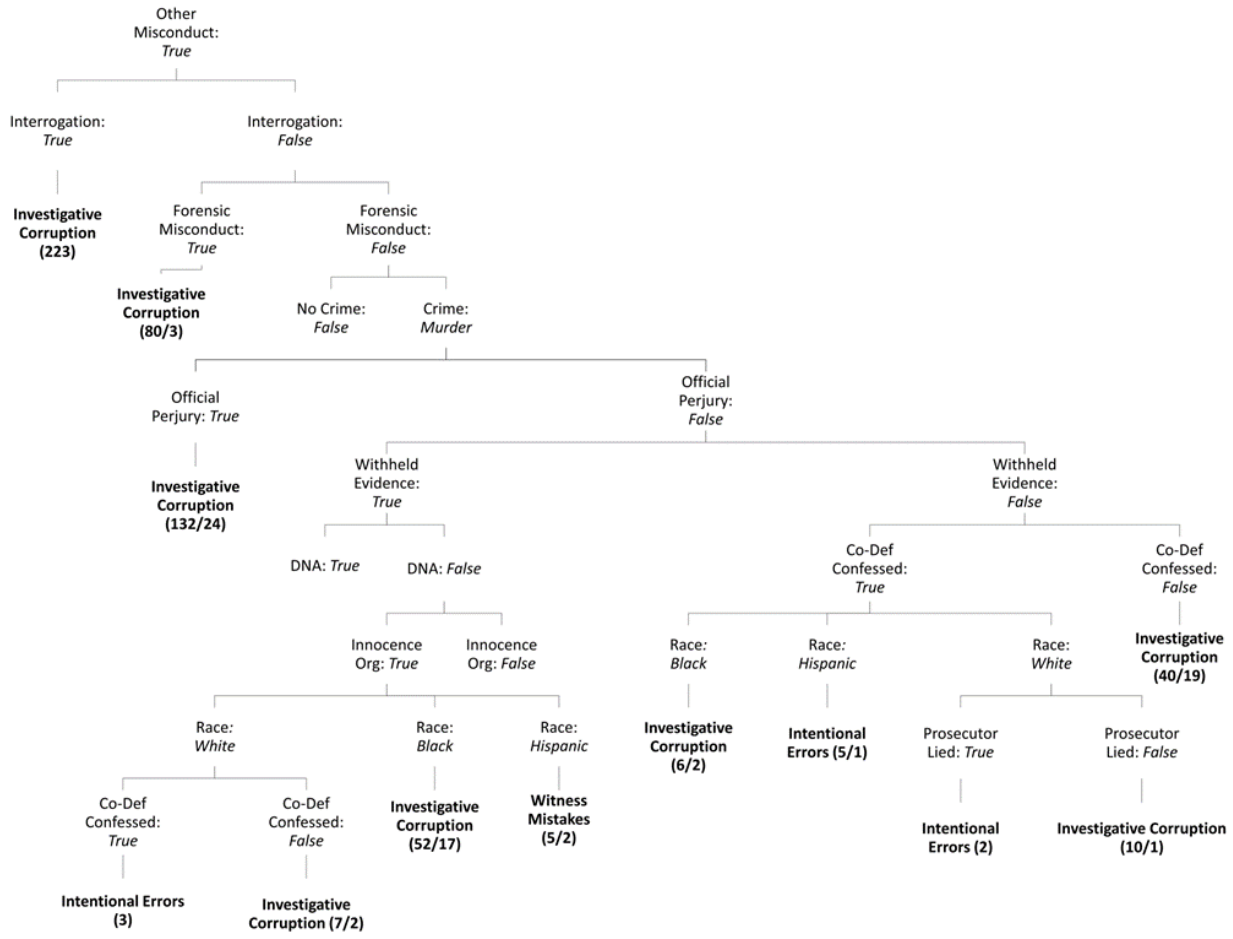
Model	*Precision*	*Recall*	*F-Score*
No DNA, no-crime, cross validated	0.714	0.711	0.711
No DNA, no-crime, 75-25	0.723	0.720	0.719
No IO, CIU, cross validated	0.718	0.717	0.716
No IO, CIU, 75-25	0.750	0.750	0.749
Added state, cross validated	0.722	0.719	0.719
Added state, 75-25	0.731	0.730	0.730

IV Discussion

Once these models are generated, a user can easily traverse them in the order displayed for the variables associated with a case, to examine similar cases and differences in classification. In *Figure 2*, we present an image of certain branches in our extended baseline model, with some sub-branches removed for ease of reading (for the entire tree, see Appendix A). Some of the case classifications are straightforward; if a case in the database has ‘true’ values for the variables ‘Other Misconduct’ (shortened from the NRE variable ‘Misconduct That Is Not Withholding Evidence’) and ‘Interrogation’ (or ‘Misconduct in the Interrogation of the Exoneree’), it is immediately sorted into the Investigative Corruption class, with no differences in classification between the LCA and the decision tree.

There is more nuance deeper within the trees, where more variables must be known for classification. As an example, consider the presence of *race*, referring to the race of the exoneree at the end of many of the branches in *Figure 2*. The generalizations for racial groups depends on the value of 'Withheld Evidence' (referring to exculpatory evidence withheld in the original trial); when this variable is true, the variable 'Co-Defendant Confessed' determines classification, but when that same variable is false, 'DNA' (whether DNA evidence was a deciding factor in exoneration) and 'Innocence Org' (whether innocence organizations were involved in exoneration) are prioritized. Even at that point, *race* is considered before *co-defendants*.

Figure 2. Example Branch of Extended Model



Note. A zoomed-in view of a branch of the extended model (predicted latent classes are in bold). The first number in parentheses reflects cases that follow this branch; after the slash is the number of classification disagreements with the LCA classes, or cases whose assigned class doesn't match the decision tree's generalization.

A. Classification Disagreements

At the ends of these branches, the numbers in parentheses refer to the number of cases whose values follow that branch; when there is a second number after a slash, that number

quantifies instances where the decision tree disagrees with the LCA for how to classify individual cases. Take, for example, the 'Withheld Evidence: True' branch that terminates with 'Investigative Corruption' (52/17). This means that, given cases with a Black exoneree, where an Innocence organization took part in exoneration and DNA was a substantial factor in exoneration, there are 52 individual cases whose facts match this branch. However, only 35 of these cases are agreed to be members of the Investigative Corruption class. Of the 17 items classified differently between the LCA and decision tree, the LCA categorized 15 cases as members of the Intentional Errors class and 2 as Witness Mistakes. The fact that this branch does not subdivide further means that the decision tree algorithm has determined there is no further generalization that can distinguish between the classes. There may be, for example, some variables whose values are distributed evenly within the next branching, but because it does not add further diagnosticity, that branching is not included in the model's output.

B. Next Steps

a. Coping with Continuous Variables

In future models, we plan to consider how best to approach the inclusion of continuous variables. This may be informed by Gañan-Cardenas *et al*. (2022), given the similarity of their methods to our own. In their work, they identify coefficients used to measure dissimilarity in continuous variables. Our goal would be to reintegrate age and chronological dates using these measures to find commonalities between exonerees whose timelines are comparable.

b. Re-Assessing Latent Class Labels

A secondary goal of this project is to critically evaluate the latent class labels conferred by Berube *et al*. (2023) for their informativeness to human users (such as innocence organization intake staff members) in understanding the factorial patterns by which exonerations can be characterized. Toward this end, we might first consider whether the canonical factors listed in the NRE database provide the most useful basis for our analyses. An empirical investigation of the differences between wrongful conviction cases and cases in which the person charged with a crime escaped conviction was carried out by Gould *et al*. (2014). They argued that factors commonly associated with wrongful convictions such as police misconduct, false confessions, eyewitness misidentification, and reliance on jailhouse informants, should perhaps not be considered as "causal" factors, but rather as contributors. Their results pointed instead to age and criminal history of the person charged with a crime, punitiveness of the state, Brady violations, forensic error, weak defense, weak prosecution case, family defense witness, non-intentional misidentification, and lying by a non-eyewitness, as better candidates for "causal" factors (Gould *et al*., 2014; see also Acker & Redlich, 2019, pp. 20-21). However, an important distinction can be made in that the causal factors pointed to by Gould *et al*. (2014) are specifically relevant to processes that result in a wrongful conviction. Although informative for identifying wrongful convictions, the presence of any one of these factors may not be specifically predictive of whether a case will result in exoneration. Because the NRE contains only cases that resulted in successful exonerations, the six canonical factors in the NRE inherently lend themselves to more confident inferences about exonerations as opposed to wrongful convictions in general.

As discussed in the *Introduction*, there are notable differences between the set of six "contributing causes" identified by the Innocence Project as of Aug 1, 2018 and the NRE's six "canonical" contributing factors from 2,253 cases as of that date (Acker & Redlich, 2019, pp. 15-16), with quite different percentages across the two distributions. Recall that the Innocence Project cases all involve DNA evidence (often from sexual assault), whereas the NRE set is much larger and more representative of wrongful convictions in general (ibid, p. 17). A benefit of our decision tree approach, constrained on the latent classes extracted from the six canonical factors in the NRE database, is that it can help determine which combination(s) of factors present in a wrongful conviction are likely to result in an exoneration, while simultaneously prioritizing transparency.

The benefit offered by predicting latent class membership from covariates within the NRE database via decision trees hinges on the extent to which the latent classes themselves are easily distinguished, and thereby interpretable. For the sake of distinguishability, the so-called "Witness Mistakes" and "Failures to Investigate" classes appear to have suitably unique patterns. However, as noted in our results, the so-called "Investigative Corruption" and "Intentional Errors" classes display markedly similar patterns of underlying canonical factors. Accordingly, our post-hoc analysis of the LCA's posterior probabilities suggested that cases originally assigned to the Intentional Errors class in our LCA would be more often classified as Investigative Corruption, as opposed to the converse, in the subsequent decision trees. Indeed, this pattern was borne out by each of the models produced. It appears, therefore, that Intentional Errors cases may be harder to distinguish than Investigative Corruption cases.

This inference is consistent with Berube *et al*.'s (2023) assessment of differences between the Investigative Corruption and Intentional Errors classes. Based on correlational analyses of the covariates, they identified fewer discriminating factors for Intentional Errors than for Investigative Corruption. More specifically, Berube *et al*. (2023) suggested that federal and no-crime cases should be particularly indicative of Intentional Errors. While both of these variables are present in the output trees, there are differences in their patterns of distribution. In our models, no-crime cases appear in the decision tree before federal cases, suggesting that the no-crime label has a higher discriminatory power. In our dataset, 640 of the 1039 cases categorized as Intentional Error are labeled as no-crime, for a rate of about 62%. However, while there are more individual cases with this label in the Intentional Error category, they make up a higher percentage of the Failure to Investigate category - 474 out of 678, for 70% (the percentages for Investigative Corruption and Witness Mistakes, respectively, are 20% and 0.2%). Conversely, the Intentional Errors category does have the highest raw count and percentage of federal cases, with 61 out of 1039 (about 6%). While this means that the label will more likely indicate an Intentional Error case, these are very small proportions of the dataset. This is reflected in the decision tree with no-crime cases being represented at an early point where this information can easily eliminate a case from a category, while federal cases are much later in the hierarchy, providing discriminatory power only when additional information has already been considered.

Berube *et al*. (2023) also pointed out that the second-highest rate of F/MFE was observed in the Investigative Corruption class, so F/MFE might be a helpful distinguishing factor. Our decision tree results supported this; the models in which forensic misconduct was included to predict Investigative Corruption showed high information gain. In Berube *et al*.'s (2023) LCA, about 42% of juvenile defendants were assigned to the Investigative Corruption class, and in our

reproduction, the same phenomena were true for about 44% of juvenile defendants. Indeed, Intentional Errors are less easy to distinguish from Investigative Corruption than vice versa; the confusion matrices in Tables 3 through 5 highlight this, showing that the raw number of cases identified by the LCA as Intentional Errors but classified by the decision tree as Investigative Corruption is always higher than vice versa (although the degree of difference varies).

This underscores a benefit of the concurrent use of decision trees and LCA methods for understanding patterns in the NRE, particularly when the ultimate goal is to inform real-world intake decisions. Decision trees allow for a more nuanced window into how covariates may predict latent class membership, as opposed to inferences made through correlations alone.

c. Potential Uses

Our eventual aim is to understand how data-intensive methods could support post-conviction intake decisions. A qualitative study of 22 innocence organizations in 2011 found that organizations on average reviewed more than a thousand requests for every one successful exoneration (Krieger, 2011). Innocence organizations are often the last resort for wrongfully convicted applicants, and the organizations carefully consider each application with this in mind. When asked about their work practices, innocence organizations estimated that initial reviews of applications took around 21% of their time, and investigations, 50% (ibid). It can be difficult to decide whether to conduct an investigation without advance knowledge of the eventual outcome. Organizations burdened by the need to conduct excessive numbers of investigations actually achieved a lower rate of successful exonerations (ibid). Accordingly, Krieger (2011) recommended that future studies should focus on identifying patterns and trends of characteristics among cases that required serious investigation, in order to assist innocence organizations with reviewing new cases:

> A future study should attempt to analyze all the cases seriously investigated (within a particular project or from many projects) to determine if particular characteristics or trends can be found that will help projects improve their selection of cases for serious investigation or review (p. 378, footnote # 240).

Some innocence organizations are already using patterns of data from previous exonerations to help them identify cases with high likelihoods of success (Weintraub, 2022). However, in their raw form, these patterns of data are not easy to observe, nor easy to infer, through rote case-by-case examination. The use of a decision tree algorithm in this context extracts patterns that exist within the available data and presents such patterns in a fashion that is easily readable, interpretable, nuanced, and transparent. Therefore, innocence organizations may refer to decision trees to inform and perhaps deepen their understanding of these patterns, such that they might be better equipped to identify cases with high likelihoods of success.

Innocence organizations may also use this framework as a training tool for law students undergoing internships/practicums, newly admitted lawyers working in post-conviction litigation, or newly hired intake staff. For example, law students are an invaluable resource for innocence organizations in providing support to their applicants and their cases (Ricciardelli *et al*., 2012). These students spend the majority of their time screening applicant cases, which provides an

increased educational benefit (Stiglitz *et al*., 2002). By having a better understanding of the pathways that lead to wrongful convictions, students may be better equipped to assess and apply their knowledge to these cases. Yet, certain critical case elements may be missing, or overlooked, in the initial legal proceedings of a criminal investigation (Findley & Scott, 2006). Accordingly, law students working at innocence projects through internships/practicums are often tasked with finding and collecting this information (Ricciardelli *et al*., 2012). A potential benefit of the decision tree framework is that it might make this process more efficient; it focuses on the factors most relevant to a particular applicant's case. An arduous information search could therefore be bolstered by efficient and communicable data-intensive methods.

C. Policy Implications

The analytical approach we outline here facilitates adding and removing variables in data-intensive models, which in theory could allow both users and policymakers to better understand the models on which they rely for life-altering decisions. It could make it easier for policymakers to audit and monitor algorithms for biases, especially for those that negatively impact vulnerable individuals (see Kalluri, 2020). It could shape the policies underlying decision-making by innocence organizations to be more efficient, as well as empower them to audit their own practices for bias if they wish (e.g., intake staff members are well aware that deciding to take on a client who has submitted a complete questionnaire is much easier than one for whom key information is missing, ambiguous, or incoherent). It could allow policymakers to communicate about AI and data-intensive models with lawmakers (as well as with the general public) by demonstrating that removing a sensitive variable (such as race) from consideration by a model does not ameliorate the bias that can be created by proxies. It could allow stakeholders who may be injured or disadvantaged by the outcome of a particular algorithm to discover, document, and contest that decision. This sort of transparency is not present in black box approaches such as deep learning and complex regressions (which cannot be explained even by their developers).

Bias is not simply a characteristic that exists "in" the algorithms and their training data; the emergence of biases (as well as their unintended consequences) depends on the context of use. Recall the Amazon algorithm that recommended qualified men but not qualified women for hiring (Dastin, 2018). If the context of use had simply been to find men to hire (reflecting the data patterns of the past), that algorithm would have been deemed successful. The point is that decision-support algorithms should be monitored and evaluated regularly, as the impacts can change over time. In the domain of exonerations, such changes may include changes in the law, new developments in forensic techniques and caveats about reliability (e.g., Fabricant, 2022), evolving precedents about the use of predictive technologies, and public literacy about such technologies.

Moreover, policy should provide regulators with the tools and "teeth" to establish transparency baselines and standards in the use of AI/machine learning, even (or especially) from tech corporations and other powerful institutions who commonly claim that their data-intensive methods and algorithms are proprietary trade secrets. Any black-box methods should be linked to laws requiring accountability from those in power, as well as clear and available policies for stakeholders who wish to contest the decisions recommended by an algorithm. In another high-impact domain for the use of AI healthcare systems, frameworks for ethical use have identified algorithm monitoring and de-implementation as a final phase in mitigating bias in an algorithm's

life cycle (Chin *et al*., 2023). While agencies such as Health and Human Services (HHS) have begun to take steps towards regulatory monitoring of AI (HealthIT, 2023), similar transparency and regulatory power should be established to monitor and de-implement potentially harmful tools used in criminal justice decision making.

For the post-conviction context of use, the methods proposed here might be useful not only to innocence organizations, but also to CIUs (conviction integrity units) and others who review potentially wrongful convictions. This will require developing easy-to-use tools that users can understand when they explore large datasets—an area for further research.

D. Conclusion

Although data-intensive methods make promises of efficiency and accuracy, resulting decisions may be biased when an algorithm's inner workings are neither transparent nor interpretable. Our approach uses LCA coupled with decision tree analysis on successful exoneration data. This reverse-engineering approach to intake data relies on patterns already present in the data to clarify trends within the latent class categorization and find further similarities between successful cases. These commonalities may be useful to determine what information would be needed for future post-conviction cases, while also directing resources for policy reform or educating staff in the use of data-driven frameworks.

V References

Acker, J. R. & Redlich, A. D. (2019). *Wrongful Conviction: Law, Science, and Policy* (2nd ed.). Durham, North Carolina: Carolina Academic Press.

Akaike, H. (1987). Factor Analysis and AIC. *Psychometrika, 52*(3), 317–332. Online: https://doi.org/10.1007/BF02294359

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. Machine bias. (2016, May 23) ProPublica. Online: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bedau, H. A., & Radelet, M. L. Miscarriages of Justice in Potentially Capital Cases. (1987) Stanford Law Review, 40(1), 21–179. Online: https://doi.org/10.2307/1228828

Berube, R., Wilford, M. M., Redlich A. D., & Wang, Y. Identifying Patterns Across the Six Canonical Factors Underlying Wrongful Convictions. (2023) Wrongful Conviction Law Review, 3(3), 166-195. Online: https://doi.org/10.29173/wclawr82

Breiman, L., Friedman, J., Olshen, R., & Stone, C. Classification and regression trees. (1984) Wadsworth Int. Group, 37(15), 237-251.

Brennan, T., & Dieterich, W. Correctional Offender Management Profiles for Alternative Sanctions (COMPAS). (2017) Handbook of Recidivism Risk/Needs Assessment Tools, 49–75. Online: https://doi.org/10.1002/9781119184256.ch3

Buolamwini, J. & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparity in Commercial Gender Classification. In *Conference on Fairness, Accountability, and Transparency* (pp. 77-91). PMLR. Online: https://proceedings.mlr.press/v81/buolamwini

18a/buolamwini18a.pdf

Chin, M. H., Afsar-Manesh, N., Bierman, A. S., Chang, C., Colón-Rodríguez, C. J., Dullabh, P., Duran, D. G., Fair, M., Hernandez-Boussard, T., Hightower, M., Jain, A., Jordan, W. B., Konya, S., Moore, R. H., Moore, T. T., Rodriguez, R., Shaheen, G., Snyder, L. P., Srinivasan, M., Umscheid, C. A., … Ohno-Machado, L. Guiding Principles to Address the Impact of Algorithm Bias on Racial and Ethnic Disparities in Health and Health Care. (2023) JAMA Network Open, 6(12). Online: https://doi.org/10.1001/jamanetworkopen.2023.45050.

Christian, B. (2020). *The Alignment Problem: Machine Learning and Human Values*. WW Norton & Company.

Dastin, J. (2018, October 10). Insight - Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women. *Reuters*. Online: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2nd ed.). John Wiley & Sons.

Fabricant, M. C. (2022). *Junk Science and the American Criminal Justice System.* Akashic Books.

Findley, K., Scott, M. S. The Multiple Dimensions of Tunnel Vision in Criminal Cases. (2006) Wisconsin Law Review, 2, 291–397. Online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=911240#

Flach, Peter. 2012. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data* (1st ed.). Cambridge University Press.

Frank, E., Hall, M. A., & Witten, I. H. (2016). The WEKA Workbench. Online Appendix for *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Morgan Kaufman.

Gañan-Cardenas, E., Pemberthy, J. I., Rivera, J. C., & Mendoza-Arango, M. C. Operating Room Time Prediction: An Application of Latent Class Analysis and Machine Learning. (2022) Ingeniería y Universidad, 26, 1-23. Online: https://doi.org/10.11144/Javeriana.iued26.ortp

Gould, J. B., Carrano, J., Leo, R. A., & Hail-Jares, K. Predicting Erroneous Convictions. (2014) Iowa Law Review, 99(2), 471-522.

Gross, S. R. Convicting the innocent. (2008) Annual Review of Law and Social Science, 4(1), 173-192. Online: https://doi.org/10.1146/annurev.lawsocsci.4.110707.172300

Gross, S. R., O'Brien, B., Hu, C., & Kennedy, E. H. Rate of False Conviction of Criminal Defendants Who are Sentenced to Death. (2014) Proceedings of the National Academy of Sciences of the United States of America, 111(20), 7230–7235. Online: https://doi.org/10.1073/pnas.1306417111

Harcourt, B. E. Risk as a Proxy for Race: The Dangers of Risk Assessment. (2015) Federal Sentencing Reporter, 26(4), 237-243.

HealthIT. (2023, December 20). *Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing (HTI-1) final rule*. Online: https://www.healthit.gov/topic/laws-regulation-and-policy/health-data-technology-and-interoperability-certification-program?et_rid=941383285&et_cid=5033060

Innocence Project. (2023). *Cases*. Online: https://innocenceproject.org/all-cases/#exonerated-by-dna

Kalluri, P. (2020). Don't Ask if Artificial Intelligence is Good or Fair, Ask How It Shifts Power. *Nature, 583*, 169. Online: https://doi.org/10.1038/d41586-020-02003-2

Kearns, M. & Roth, A. (2020). *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford University Press.

Krieger, S. A. Why Our Justice System Convicts Innocent People, And the Challenges Faced by Innocence Projects Trying to Exonerate Them. (2011) New Criminal Law Review, 14(3), 333–402. Online: https://doi.org/10.1525/nclr.2011.14.3.333

Loeffler, C. E., Hyatt, J. & Ridgeway, G. Measuring Self-Reported Wrongful Convictions Among Prisoners. (2019) J Quant Criminol, 35, 259–286. Online: https://doi.org/10.1007/s10940-018-9381-1

McCutcheon, A. (2002). Basic Concepts and Procedures in Single-and Multiple-Group Latent Class Analysis. In J. Hagenaars & A. McCutcheon (Eds.), *Applied Latent Class Analysis* (pp. 56-86). Cambridge: Cambridge University Press. Online: https://doi.org/10.1017/CBO9780511499531.003

Muthén B. O., Muthén L. K. Integrating Person-Centered and Variable-Centered Analyses: Growth Mixture Modeling with Latent Trajectory Classes. (2000) Alcoholism: Clinical & Experimental Research, 24(6), 882-891. Online: https://doi.org/10.1111/j.1530-0277.2000.tb02070.x

National Registry of Exonerations. (n.d.). Online: https://www.law.umich.edu/special/exoneration/Pages/detaillist.aspx

O'Neil, C. (2016). *Weapons of Math Destruction*. New York: Broadway Books.

Quinlan, R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA.

R Core Team (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Online: http://www.R-project.org/

Ricciardelli, R., Bell, J. G., & Clow, K. A. "Now I See It for What It Really Is": The Impact of Participation in an Innocence Project Practicum on Criminology Students. (2012) Albany Law Review, 75(3), 1439–1466.

Schwarz, Gideon. Estimating the Dimension of a Model. (1978) The Annals of Statistics, 6(2), 461–464. Online: https://doi.org/10.1214/aos/1176344136

Stiglitz, J., Brooks, J., & Shulman, T. The Hurricane Meets the Paper Chase: Innocence Projects New Emerging Role in Clinical Legal Education. (2002) California Western Law Review, 38:2(5), 413–431. Online: https://scholarlycommons.law.cwsl.edu/cwlr/vol38/iss2/5

Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. (1976) Journal of the Royal Statistical Society (Series B - Methodological), 36, 111-133.

Turque, B. (2012, March 6). 'Creative... motivating' and fired. *The Washington Post*. Online: https://www.washingtonpost.com/local/education/creative--motivating-and-fired/2012/02/04/gIQAwzZpvR_story.html

Weintraub, Jennifer N. (2022). *The Dark Figure of Wrongful Convictions: How Intake Decisions Impact Exonerations*. [Unpublished doctoral dissertation]. The University at Albany, State University of New York, NY.

Weller, B. E., Bowen, N. K., & Faubert, S. J. Latent Class Analysis: A Guide to Best Practice. (2020) Journal of Black Psychology, 46(4), 287-311. Online: https://doi.org/10.1177/0095798420930932

Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. Springer New York.

West, E., & Meterko, V. Innocence Project: DNA Exonerations, 1989-2014; Review of Data and Findings From the First 25 Years. (2016) Albany Law Review, 79(3), 717-795, 2015/2016. Online: https://ssrn.com/abstract=2986970

VI Appendix A.

Expanded Decision Tree, Figure 2.

```
OTHER_MISCONDUCT = True
|   INTERROGATION = True: Investigative_Corruption (223.0)
|   INTERROGATION = False
|   |   FORENSIC_MISCONDUCT = True: Investigative_Corruption (80.0/3.0)
|   |   FORENSIC_MISCONDUCT = False
|   |   |   NO_CRIME = False
|   |   |   |   CRIME = Murder
|   |   |   |   |   OFFICIAL_PERJURY = True: Investigative_Corruption (132.0/24.0)
|   |   |   |   |   OFFICIAL_PERJURY = False
|   |   |   |   |   |   WITHELD_EV = True
|   |   |   |   |   |   |   DNA = True
|   |   |   |   |   |   |   |   PROSECUTOR_MISCONDUCT = True: Investigative_Corruption (28.0/2.0)
|   |   |   |   |   |   |   |   PROSECUTOR_MISCONDUCT = False
|   |   |   |   |   |   |   |   |   WINTESS_TEMP = True: Intentional_Errors (14.0/6.0)
|   |   |   |   |   |   |   |   |   WINTESS_TEMP = False: Investigative_Corruption (2.0)
|   |   |   |   |   |   |   DNA = False
|   |   |   |   |   |   |   |   INNOCENCE_ORG = True
|   |   |   |   |   |   |   |   |   RACE = White
|   |   |   |   |   |   |   |   |   |   CO_DEF_CONFESSED = True: Intentional_Errors (3.0)
|   |   |   |   |   |   |   |   |   |   CO_DEF_CONFESSED = False: Investigative_Corruption (7.0/2.0)
|   |   |   |   |   |   |   |   |   RACE = Black: Investigative_Corruption (52.0/17.0)
|   |   |   |   |   |   |   |   |   RACE = Hispanic: Witness_Mistakes (5.0/2.0)
|   |   |   |   |   |   |   |   INNOCENCE_ORG = False
|   |   |   |   |   |   |   |   |   PERMITTING_PERJURY = True
|   |   |   |   |   |   |   |   |   |   INTEGRITY_UNIT = True: Investigative_Corruption (12.0/4.0)
|   |   |   |   |   |   |   |   |   |   INTEGRITY_UNIT = False
|   |   |   |   |   |   |   |   |   |   |   JAIL_INFORMANT = True
|   |   |   |   |   |   |   |   |   |   |   |   CO_DEF_CONFESSED = True: Intentional_Errors (4.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   CO_DEF_CONFESSED = False: Investigative_Corruption (14.0/5.0)
|   |   |   |   |   |   |   |   |   |   |   JAIL_INFORMANT = False: Intentional_Errors (51.0/14.0)
|   |   |   |   |   |   |   |   |   PERMITTING_PERJURY = False
|   |   |   |   |   |   |   |   |   |   CHILD_VICTIM = True: Investigative_Corruption (12.0/3.0)
|   |   |   |   |   |   |   |   |   |   CHILD_VICTIM = False
|   |   |   |   |   |   |   |   |   |   |   PROSECUTOR_MISCONDUCT = True
|   |   |   |   |   |   |   |   |   |   |   |   JUVENILE_DEF = True: Intentional_Errors (7.0/3.0)
|   |   |   |   |   |   |   |   |   |   |   |   JUVENILE_DEF = False: Investigative_Corruption (43.0/18.0)
|   |   |   |   |   |   |   |   |   |   |   PROSECUTOR_MISCONDUCT = False: Intentional_Errors (37.0/11.0)
|   |   |   |   |   |   WITHELD_EV = False
|   |   |   |   |   |   |   CO_DEF_CONFESSED = True
|   |   |   |   |   |   |   |   RACE = White
|   |   |   |   |   |   |   |   |   PROSECUTOR_LIED = True: Intentional_Errors (2.0)
|   |   |   |   |   |   |   |   |   PROSECUTOR_LIED = False: Investigative_Corruption (10.0/1.0)
|   |   |   |   |   |   |   |   RACE = Black: Investigative_Corruption (6.0/2.0)
|   |   |   |   |   |   |   |   RACE = Hispanic: Intentional_Errors (5.0/1.0)
|   |   |   |   |   |   |   CO_DEF_CONFESSED = False: Investigative_Corruption (40.0/19.0)
```

Innocence Discovery Lab - Harnessing Large Language Models to Surface Data Buried in Wrongful Conviction Case Documents

Ayyub Ibrahim
Innocence Project New Orleans
U.S.A.

Huy Dao
Innocence Project New Orleans
U.S.A.

Tarak Shah
Human Rights Data Analysis Group
San Francisco, CA
U.S.A.

The recent advent of commercial artificial intelligence (AI), especially in natural language processing (NLP), introduces transformative possibilities for wrongful conviction research. NLP, a pivotal branch of AI that forms the basis for Large Language Models (LLMs), enables computers to interpret human language with a nuanced understanding. This technological advancement is particularly valuable for analyzing the complex language found in case documents associated with wrongful convictions. This paper explores the effectiveness of LLMs in analyzing and extracting data from case documents collected by the Innocence Project New Orleans and the National Registry of Exonerations. The diverse and comprehensive nature of these datasets makes them ideal for assessing the capabilities of LLMs. The findings of this study advance our understanding of how LLMs can be utilized to make wrongful conviction case documents easily accessible by automating the extraction of relevant data.

I. Introduction
II. Defining a Wrongful Conviction
III. Framework for Exoneration Document Analysis
  A. Metadata Compilation
  B. Page Classification
  C. Unstructured Data Extraction with Large Language Models
  D. Deduplication
  E. Cross-referencing
  F. Structured Data Extraction with Regular Expressions
  G. Structured Data Extraction with a Large Language Model
  H. Hypothetical Document Embeddings
  I. Example of Model's Workflow
  J. Performance Evaluation
  K. Preprocessing Parameters
  A. Model-Specific Parameters
IV. Fine-Tuning the Large Language Model
V. Entity Resolution and Entity Matching
VI. Future Research

I Introduction

Thousands of wrongfully convicted people have been exonerated across the globe in the last four decades. These wrongful convictions have not only uncovered and undone mistakes in investigations and prosecutions, they have exposed profound and systemic injustice in criminal legal systems themselves. The patterns and practices that inform - indeed create - wrongful convictions have largely remained hidden. They often start with how communities are policed, from the moment a crime is reported or observed. As exonerations continue to happen, a growing body of documents and data remains untapped. They have the potential to reveal not only the identities of law enforcement involved in wrongful convictions, but also their roles in each case, their patterns of misconduct and migration, and how they are connected not only to each other but to wrongful conviction cases not yet found or investigated. The Innocence Discovery Lab, born of Innocence Project New Orleans (IPNO) and the Louisiana Law Enforcement Accountability Database (LLEAD), seeks to leverage the advent of large language models to transform unstructured documents from case documents into a structured, accessible format.

Historically, the expansive volume of documentation associated with exoneration cases has presented significant analytical challenges. The sheer quantity and complexity of data, ranging from legal transcripts and police reports to witness statements and forensic analyses, have long made it difficult to extract meaningful insights and make important connections. These difficulties are compounded by the diverse nature of the documents, which often include varying formats and levels of detail. Large language models, like ChatGPT, now make processing these documents at scale feasible due to their ability to process and interpret vast amounts of text. These capabilities include analyzing the language used in legal documents to detect biases, examining patterns in policing practices, and cross-referencing details across cases to identify systemic issues. Furthermore, the ability of these models to learn and adapt over time means that they become more efficient and accurate as they process more data, continuously enhancing their analytical power.

Key to our research is the integration of wrongful conviction data with police databases, including IPNO's internal database developed by its case management team, and its public counterpart, the Louisiana Law Enforcement Accountability Database (LLEAD). IPNO’s internal database, which houses case data from IPNO’s clients and applicants, offers detailed insights into individual wrongful conviction cases. LLEAD, with data on over 100,000 officers in Louisiana, provides a broader view of police misconduct and migration patterns. The combination of these databases enables a more comprehensive analysis that connects the specifics of individual cases to systemic issues in law enforcement. By correlating data from these two sources, our research aims not only to reveal specific links between wrongful convictions and law enforcement misconduct but also to shed light on broader trends and practices contributing to these injustices.

II Defining a Wrongful Conviction

Within the scope of our research, a 'wrongful conviction' is defined as the conviction of an individual who is factually innocent of the crime charged. This can result from a trial verdict or a plea. An exoneration is the official overturning of a wrongful conviction. These typically occur

through pardons or acquittals at retrial, often initiated by the emergence of new evidence that proves innocence and was not available or presented during the trial phase of the case.

III Framework for Exoneration Document Analysis

In Orleans Parish, Louisiana, where Innocence Project New Orleans is based, 78%[1] of wrongful convictions have been linked to law enforcement's failure to share exculpatory evidence with the defense, a rate more than double the national average.

Our research, recognizing the explicit relationship between law enforcement misconduct and wrongful convictions, aims to establish best practices for transforming unstructured wrongful conviction case data into structured, accessible formats. This transformation is crucial for lawyers, advocates, and community members who are committed to leveraging insights from past wrongful convictions to prevent future occurrences. Our methodology is built on a multi-stage process:

A. Metadata Compilation

The foundation of our research involves compiling metadata into an index, essential for effectively managing our extensive and growing corpus of exoneration documents. In organizing the metadata, we have focused on collecting details crucial for document identification and management. This includes capturing the file path and name, file type, file size, number of pages, and creating a unique identifier for each document by truncating the SHA1 content hash. A case ID is also assigned to each document, derived from the directory names used during the scanning process.

B. Page Classification

In the course of an exoneration case, a wide variety of documents are accumulated, reflecting materials that may extend over many decades. These documents are not only extensive in volume but also varied in nature, each offering a unique lens into the intricacies of the case. A challenge we frequently encounter is the presence of inaccurately or inconsistently named files, which prevents the immediate identification of their contents. Moreover, it is not uncommon to find lengthy documents, e.g. exceeding a thousand pages, containing a collection of different document types.

After consultation with the IPNO's case management team, we decided to focus on a specific subset of documents considered most relevant to wrongful conviction research. These documents included police reports, court transcripts, and court testimonies. To accurately classify the array of documents produced over the course of an exoneration case, we have developed an automated page classification model to overcome the limitations of traditional manual review methods. This model utilizes a pre-trained convolutional neural network, optimized through training on thumbnail images of key document types. The thumbnails' smaller sizes ensure

[1] Samuel R Gross *et al*, "Government Misconduct and Convicting the Innocent: The Role of Prosecutors, Police and Other Law Enforcement" (1 Sept 2020) online (pdf): Online: <law.umich.edu/special/exoneration/Documents/Government_Misconduct_and_Convicting_the_Innocent.pdf>.

efficient processing and minimal computational load. We used the FastAI library to adapt the ResNet34 architecture, initially trained on the ImageNet[2] database, for the identification of these document types from their thumbnails. This approach significantly streamlines the document classification process, overcoming the challenges posed by the volume and diversity of the documents.

C. Unstructured Data Extraction with Large Language Models

In the current stage of our research, which is the primary focus of this paper, we are constructing a database from unstructured text found within over 300,000 pages of exoneration documents collected from IPNO and the National Registry of Exonerations[3]. Our objective in creating this database is to index the identities of all police officers, prosecutors, and laboratory personnel featured in these documents, along with extracting detailed information about their actions and the events they were involved in. To facilitate this extraction, we are utilizing large language models for their advanced capabilities in parsing and interpreting complex text structures. This approach is particularly crucial given the diverse range of document types typically encountered in wrongful conviction cases. After the information is extracted, it will be converted into formats optimized for analysis. This structuring is expected to significantly enhance the accessibility and utility of the data, enabling more rigorous research into wrongful convictions.

D. Deduplication

After the structured data extraction, a significant challenge we face is the extensive issue of data redundancy, exacerbated by the nature of our dataset. Officers involved in wrongful convictions are often mentioned in various documents that span different periods of time. This results in instances where the same individual may appear with different names or in different contexts, leading to multiple entries in our initial index.

To address this issue, our approach involves filtering this index down to a table of unique identities. This process is designed to accurately identify and consolidate instances where the same officer is referenced in disparate capacities or times within the dataset. By implementing this filtration, we aim to effectively remove duplicate entries, thus preserving the accuracy and integrity of our dataset. This step is essential to ensure that the representation of each officer's involvement in wrongful convictions is consistent and precise in our analysis, thereby facilitating a clearer and more comprehensive understanding of each officer's role within a case.

E. Cross-referencing

Our research will culminate in a comparison of data from our exoneration case documents, soon to be indexed into a wrongful conviction database, with the Louisiana Law Enforcement Accountability Database (LLEAD) and the Innocence Project New Orleans' (IPNO) internal database. LLEAD, with data on approximately 100,000 police officers in Louisiana, including

---

[2] Kaiming He *et al*, "Deep Residual Learning for Image Recognition" (2015) ArXiv 1512.03385, online (pdf): <arxiv.org/pdf/1512.03385.pdf>

[3] University of Michigan Law School, "Spread-Sheet Request Form" (last accessed 5 Dec 2023), online: <law.umich.edu/special/exoneration/Pages/Spread-Sheet-Request-Form.aspx>.

records of over 40,000 misconduct allegations, will be a key resource. We will focus on identifying officers involved in both misconduct and wrongful convictions by analyzing data such as names, ranks, department affiliations, hire dates, and departure dates.

After initially cross-referencing our wrongful conviction data with the Louisiana Law Enforcement Accountability Database (LLEAD), we will proceed to a further stage of analysis that involves cross-referencing the officers we've identified as having histories of misconduct and associations with wrongful convictions with the Innocence Project New Orleans' (IPNO) internal database. This internal database is particularly significant, as it contains names of officers flagged during case review and by potential clients.

Through this additional layer of cross-referencing, our objective is to deepen our understanding of each officer's potential involvement in wrongful convictions, past and potentially future. This comprehensive approach, which includes cross-referencing with both the LLEAD and IPNO's internal database, will guide IPNO's case management team in scrutinizing cases involving officers with direct ties to wrongful convictions as well as those indirectly connected, such as their partners, supervisors, departments/divisions they've moved to, or their trainees. A particular focus will be placed on those associated officers with a substantial history of misconduct or identified involvement in wrongful conviction cases.

F. Structured Data Extraction with Regular Expressions (Regex)

Our research into wrongful convictions begins with the extraction of structured, meaningful data from a vast store of documents containing unstructured text. While our current research focuses on employing large language models (LLMs) for this task, it's essential to acknowledge the various methodologies that have historically shaped Information Extraction (IE) in Natural Language Processing (NLP), a field that has experienced significant growth over the past few decades[4] in response to the growth in complexity and volume of data.

The earliest iterations of IE were predominantly rule-based and dictionary-based systems. These initial methods involved the application of manually created rules and the use of curated dictionaries for phrase matching. While capable of high accuracy in small and well-defined datasets, these systems lacked flexibility and scalability. The requirement for extensive manual input to develop rules and dictionaries rendered them less practical for large or dynamically changing datasets. Moreover, the language specificity of these methods restricted their effectiveness across different linguistic contexts.

With the general increase in data availability and complexity, the field of IE gravitated towards statistical machine learning methods. This shift responded to the limitations of rule-based systems, as these newer methods utilized algorithms that could learn directly from data, moving beyond the confines of manually programmed rules. Techniques like Hidden Markov Models (HMM), Maximum Entropy Models (MEM), Support Vector Machines (SVM), and Conditional Random Fields (CRF) were increasingly employed to extract statistical features from manually

[4] Yang Yang *et al*, "A Survey of Information Extraction Based on Deep Learning" (2022) 12:19 Applied Sciences 9691, online: <mdpi.com/2076-3417/12/19/9691>.

labeled corpora. This advancement in scalability came with its own challenges, most notably the need for manual annotation and complex feature engineering.

In the evolving landscape of IE, deep learning methodologies represent the latest advancement. Models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have revolutionized IE with their proficiency in feature extraction and ability to learn from large datasets. Automating the feature engineering process, these models are adept at identifying complex patterns in data, significantly enhancing fields like NLP, speech recognition, and visual object recognition. However, despite their groundbreaking potential, deep learning models, including advanced large language models (LLMs), are not without limitations. They require substantial datasets for training and significant computational resources, presenting barriers in certain contexts. This contrast between the capabilities of deep learning and the practical constraints of various IE methods, from rule-based to deep learning approaches, highlights the considerations in selecting an appropriate technique for specific research challenges.

Acknowledging the range of information extraction techniques, we opted for a rules-based approach for our baseline model, specifically regex, due to its immediate deployability and efficiency in pattern recognition.

The primary task of extracting officer information, however, quickly highlighted the limitations of regex when handling the complexities of natural language found in wrongful conviction case documents. Consider the challenges posed by a complex sentence from a court transcript: "John Ruiz was mentioned as being involved in the joint investigation with Detective Martin Sholtz regarding the Seafood City burglary and the murder of Randy Gray." In this scenario, regex's capabilities are notably limited. While regex can effectively identify 'Detective Martin Sholtz' as a key entity, owing to the clear pattern of a recognized title followed by a name, it may fail to recognize 'John Ruiz' in a similar capacity. This limitation arises because regex operates on predefined patterns and lacks the ability to understand semantic nuances. In our example, regex is programmed to detect titles like 'Detective' followed by names, which it does successfully for 'Detective Martin Scholtz'. However, without the explicit title 'Detective' preceding 'John Ruiz', regex overlooks this name, despite its relevance in the narrative. This shortfall illustrates a critical aspect of regex's nature: its inability to infer context or understand relational connections between entities in text. Consequently, important information like the involvement of John Ruiz in the investigation, equally crucial to the case narrative, might be missed, underscoring the need for more sophisticated methods capable of semantic comprehension in legal text analysis.

To demonstrate the limitations of regex, we created a baseline regex model designed to extract officer names. This model’s intent was not to comprehensively extract all details related to officers but to evaluate the effectiveness of regex in extracting a specific type of structured data. The model, captured in the pattern below, was tested on police reports and court transcripts.

Listing 1.0: Regular Expression Pattern

```
pattern = .compile(r”(detective|sergeant|lieutenant|captain|corporal|deputy|
investigator|criminalist|technician|det\.|sgt\.|lt\.|cpt\.|cpl\.|dty\.|tech\.|dr\.)\s+([A-Z][A-Za-
z]*(\s[A-Z][A-Za-z]*)?)", re.IGNORECASE)
```

Evaluating the performance of our regex model involved analyzing metrics such as precision, recall, F1 score, and F-beta score. Precision, which measures the accuracy of the model in correctly identifying officer names, was found to be 84.5% in police reports, suggesting a high level of reliability in the model's identifications.

However, recall, which assesses the model's ability to detect all relevant instances of officer names, was only 51.8% in police reports. This indicates that the model missed almost half of the actual officer names present in the documents. The F1 score, a metric combining precision and recall, stood at 0.614, reflecting a moderate balance between these two aspects.

We also considered the F-beta score, a variation of the F1 score that gives more weight to recall. Given the critical nature of not missing any true positives in our research, recall was weighted more heavily. In the context of police reports, the F-beta score was 0.549, highlighting the model's limitations in recall.

The model's performance in court transcripts followed a similar trend but with more noticeable shortfalls. It achieved a precision of 86.56%, signifying relatively accurate identifications. However, its recall dropped to 42.81%, indicating that the model failed to detect more than half of the officer names in these documents. Consequently, the F1 score was recorded at 0.5461, and the F-beta score, prioritizing recall, further decreased to 0.4663.

Figure 1. Baseline Regular Expression Model Evaluation

Metric	Police Reports	Court Transcripts
Precision	0.845	0.8656
Recall	0.518	0.4281
F1 Score	0.614	0.5461
F-Beta	0.549	0.4663

The overall performance points to the inherent limitations of regex in handling the nuanced and context-rich language of legal documents. The low F-Beta score, particularly in court transcripts, emphasizes the need for more sophisticated data extraction techniques that can capture the full range of necessary information.

The challenges we encountered with regex highlighted the need for more advanced data extraction methods, specifically those capable of comprehending and interpreting the semantic context within wrongful conviction case documents. This realization has steered our research towards the integration of large language models with regex. Large language models, with their advanced capabilities in understanding natural language, offer a promising solution for the complexities we face in extracting structured data from legal texts. By combining the pattern-

matching strengths of regex with the deep learning and contextual understanding[5] of large language models, we aim to significantly enhance our recall capabilities.

G. Structured Data Extraction with a Large Language Model

The process of structured data extraction using large language models (LLMs) presented a unique set of challenges, particularly in managing the substantial length of the documents we are analyzing. Single documents often extend to hundreds of pages, posing a significant contrast to the prompt length constraints of LLMs.

To effectively utilize LLMs in our data extraction process, we developed a strategy to isolate specific text segments within each document. These segments were then used to create more focused and efficient prompts for the language model.

We approached this task by identifying the relevant text segments, and secondly, by extracting structured information about officers from these identified segments. For managing this two-step process, we employed Langchain[6], a natural language processing library, and OpenAI's GPT.

For the first step, identifying the relevant chunks of text within the larger document, we used the approach outlined in Precise Zero-Shot Dense Retrieval without Relevance Labels[7]. This approach splits our information retrieval task into multiple steps:

1. A query requesting names and roles of mentioned officers was entered into a large language model, which then generated a "hypothetical" document in response.
2. This hypothetical document was embedded.
3. The document text was divided into overlapping chunks, with each chunk receiving an embedding using the same system as the hypothetical document.
4. Facebook's AI Similarity Search (FAISS)[8], a nearest-neighbor search implementation, was then used to identify relevant text content by comparing chunk embeddings with those of the hypothetical document.

H. Hypothetical Document Embeddings

Hypothetical Document Embeddings (HyDE) transform raw text into a structured, searchable format. This process begins with a large language model generating a hypothetical

---

[5] Somin Wadhwa, Silvio Amir & Byron C Wallace, "Revisiting Relation Extraction in the era of Large Language Models" (2023) 1 Proc 61st Annual Meeting of the Association for Computational Linguistics 15566, online (pdf): <aclanthology.org/2023.acl-long.868.pdf>.

[6] LangChain, "Get Started" (last accessed Dec 5, 2023) online: <python.langchain.com/docs/get_started/introduction>

[7] Gao *et al*, "Precise Zero-Shot Dense Retrieval without Relevance Labels" (2022) ArXiv:2212.10496, online: <arxiv.org/abs/2212.10496>.

[8] Hervé Jegou, Matthijs Douze & Jeff Johnson, "Facebook AI Similarity Search (FAISS)" (29 Mar 2017) Engineering at Meta, online: <engineering.fb.com/2017/03/29/data-infrastructure/faiss-a-library-for-efficient-similarity-search/>.

document in response to a query. The strength of this document lies in its pattern-rich content, essential for locating documents with similar content in a database, despite potential inaccuracies.

The next step involves converting the hypothetical document into an embedding vector. This conversion translates the text into vector representations in a multi-dimensional space. These embeddings capture more than simple word counts or keyword matches; they encapsulate the text's contextual nuances and underlying intent. Thus, searches leveraging these embeddings focus on contextual similarity and semantic connections between documents, surpassing traditional keyword-based search methods in depth and relevance.

Listing 2.0: Hypothetical Document Embeddings Query

```
PROMPT_TEMPLATE_HYDE = PromptTemplate(input_variables=["question"],
template="""
You're an AI assistant specializing in criminal justice research.
Your main focus is on identifying the names and providing detailed context of mention for
each law enforcement personnel.

This includes police officers, detectives, deputies, lieutenants, sergeants, captains,
technicians, coroners, investigators, patrolmen, and criminalists, as described in court
transcripts and police reports.

Question: {question}

Responses:
""",
)
```

Listing 2.1: Hypothetical Document Embeddings Implementation

```
def generate_hypothetical_embeddings():
llm = OpenAI()
prompt = PROMPT_TEMPLATE_HYDE
llm_chain = LLMChain(llm=llm, prompt=prompt)
base_embeddings = OpenAIEmbeddings()
embeddings = HypotheticalDocumentEmbedder(
llm_chain=llm_chain, base_embeddings=base_embeddings
)
return embeddings
```

To create the vector database, we utilize our ‘process_single_document’ function. This function initiates by loading the text of a document and segmenting it. For segmentation, we use LangChain’s RecursiveCharacterTextSplitter, which divides the document into word chunks. The chunk size and overlap are chosen to ensure that each segment is comprehensive enough to maintain context while being sufficiently small for efficient processing. Post-segmentation, these chunks are transformed into high-dimensional vectors using the hypothetical document's embedding scheme. The concluding step involves the ‘FAISS.from_documents function’, which compiles these vectors into an indexed database. This database enables efficient and context-sensitive searches, allowing for the quick identification of documents that share content similarities with the hypothetical document.

Listing 3.0: **Storing the Document in a Vector Database**

```
def process_single_document(file_path, embeddings):
logger.info(f"Processing document: {file_path}"
loader = JSONLoader(file_path)
text = loader.load()
logger.info(f"Text loaded from document: {file_path}")
text_splitter = RecursiveCharacterTextSplitter(chunk_size=500,
chunk_overlap=250)
docs = text_splitter.split_documents(text)
db = FAISS.from_documents(docs, embeddings)
return db
```

Following the creation of our vector database, the document becomes fit for structured information extraction. This task is carried out by the get_response_from_query function, which is designed to transform pre-processed, unstructured data into structured outputs.

Initial Query Processing: The extraction phase begins when a user sends a query to the vector database. Once the query is received, the database conducts a search within its embedding space, identifying and retrieving text chunks that best match the query's contextual and semantic criteria. This retrieval process is carried out using the ‘db.similarity_search_with_score’ method, which selects the top 'k' relevant chunks based on their high similarity to the query.

Sorting of Retrieved Chunks: After their retrieval, the chunks are sorted according to relevance using the ‘sort_retrieved_documents’ function. This step ensures that the most relevant chunks are appropriately organized within the model’s context window. This approach is supported by findings from 'Lost in the Middle: How Language Models Use Long Contexts,[9] which emphasize that language models typically yield better performance when pertinent information is positioned at the beginning or end of their input contexts. After sorting, the chunks are concatenated into a single string, eliminating the overhead of processing multiple individual strings and reducing unnecessary tokens.

Model Initialization and Response Generation: The processing begins with the instantiation of an OpenAI model and the LLMChain class. This setup allows the chain to process the combined document content along with the original query. Following this, the LLMChain executes its run method, using the inputs of prompt, query, and document content to generate a structured and detailed response. The model then extracts information relevant to the query and structures the output according to the specifications in the prompt template.

Listing 4.0: Template for Model

```
PROMPT_TEMPLATE_MODEL = PromptTemplate(input_variables=["question",
"docs"],template=""
```

[9] Nelson F Liu *et al*, “Lost in the Middle: How Language Models Use Long Contexts” (2023) ArXiv:2304.03173, online: <arxiv.org/abs/2307.03172>.

As an AI assistant, my role is to meticulously analyze criminal justice documents and extract information about law enforcement personnel.

Query: {question}
Documents: {docs}

The response will contain:

1) The name of a police officer.
Please prefix the name with "Officer Name: ".
For example, "Officer Name: John Smith".

2) If available, provide an in-depth description of the context of their mention.
If the context induces ambiguity regarding the individuals role in law enforcement, note this.
Please prefix this information with "Officer Context: ".

3) Review the context to discern the role of the officer. For example, Lead Detective.
Please prefix this information with "Officer Role: "
For example, "Officer Role: Lead Detective"

The full response should follow the format below, with no prefixes such as 1., 2., 3., a., b., c.:

Officer Name: John Smith
Officer Context: Mentioned as officer at the scene of the incident.
Officer Role: Patrol Officer

Officer Name:
Officer Context:
Officer Role:

Additional guidelines:
Only derive responses from factual information found within the police reports.
""",)

Listing 4.1: Function for Generating Responses

```
def get_response_from_query(db, query):
# Set up the parameters
prompt = PROMPT_TEMPLATE_MODEL
roles = ROLE_TEMPLATE
temperature = 1
k = 20

# Perform the similarity search
doc_list = db.similarity_search_with_score(query, k=k)

# Sort documents by relevance scores as suggested in the literature
```

```
docs = sorted(doc_list, key=lambda x: x[1], reverse=True)

third = len(docs) // 3
highest_third = docs[:third]
middle_third = docs[third:2*third]
lowest_third = docs[2*third:]
highest_third = sorted(highest_third, key=lambda x: x[1],reverse=True)
middle_third = sorted(middle_third, key=lambda x: x[1], reverse=True)
lowest_third = sorted(lowest_third, key=lambda x: x[1], reverse=True)
sorted_docs = highest_third + lowest_third + middle_third
# Join documents into one string for processing
docs_page_content = " ".join([d[0].page_content for d in sortedocs])

# Create an instance of the OpenAI model
llm = ChatOpenAI(model_name="gpt-4")

# Create an instance of the LLMChain
chain = LLMChain(llm=llm, prompt=prompt)

# Run the LLMChain and print the response
response = chain.run(question=query, docs=docs_page_content,
temperature=temperature)
print(response)

return response
```

I. Example of Model's Workflow

Query Input:

"In the transcript, identify individuals by their names along with their specific law enforcement titles, such as officer, sergeant, lieutenant, captain, commander, sheriff, deputy, detective, inspector, technician, analyst, and coroner. Alongside each name and title, note the context of their mention. This includes the roles they played in key events, decisions they made, actions they took, their interactions with others, responsibilities in the case, and any significant outcomes or incidents they were involved in."

Document Retrieved from the Vector Database:
(1 of 20 documents identified by the Faiss similarity search as relevant)
"Mark MeCrarey, New Orleans police sergeant. A 16 .01 Sergeant DeHarmen Cade, where are you assigned now? - A Second Police District. 13. And in October, September of 1979 and in Q 19 September and October of 1980, where were you assigned? :1 Homicide division. A. And how long have you been on the police department right now? Thirteen and a half years. A Dwight Fernandez, when did you or did you ever take over the investigation of ... murder? A", metadata={'source': '../../data/convictions/transcripts/iterative\\(C) Det. Martin Scholtz Testimony - Trial One.docx, 'page_number': 150}'"

Figure 2. Model's Response

Officer Name	Officer Context	Officer Role
Mark MeCrarey	Mentioned as the patrol officer of CAR 802 at the scene in the Garden district.	Patrol Officer
DeHerman Cade	DeHerman Cade is referenced multiple times throughout the transcript, indicating his active involvement in various aspects of the case. His roles include being assigned to the homicide division, where he appears to play a lead role in the investigation. The transcript details instances where he is actively engaging in investigative duties, analyzing evidence, and interacting with other law enforcement personnel.	Investigating Detective
Dwight Fernandez	Mentioned as being an Officer (from Upper Quarter Crime Prevention on detail) in various contexts, including when he was at the homicide office with Cade.	Officer

J. Performance Evaluation

To assess the model's effectiveness in extracting officer names from wrongful conviction case documents, we developed a detailed evaluation framework using the GPT-4 model. The framework was structured as a series of tests where the same query was executed six times to evaluate the consistency and reliability of the model. These tests were influenced by several key parameters, including preprocessing parameters like chunk size, which determines the volume of consecutive text units processed, and chunk overlap, indicating the number of shared words between consecutive text chunks. Additionally, the evaluation considered model-specific parameters such as the impact of Hypothetical Document Embeddings (HYDE) on model effectiveness, the 'k' value specifying the number of text chunks per query, and the temperature parameter, which controls the variability of outputs.

K. Preprocessing Parameters

Chunk Size: Determines the volume of consecutive text units processed in each instance.
Chunk Overlap: Indicates the number of words shared between consecutive text chunks. For example, a 250-word overlap means the subsequent chunk initiates 250 words before the end of the preceding one.

L. Model-Specific Parameters

Hypothetical Document Embeddings (HYDE): Their influence on the model's overall effectiveness was assessed.
'k' Value: Specifies the number of text chunks inputted into the model for each query.

Temperature Parameter: Controls the degree of variability in the model's outputFor the evaluation of our model, we chose the F-beta score as the primary metric because of its capacity

to balance and differentially weigh precision and recall. In our model's context, we prioritized recall, assigning it twice the importance of precision. This decision is in line with our goal to achieve thorough identification of relevant data, accepting the possibility of occasionally including some irrelevant information. Such an approach is particularly valuable in scenarios where missing key information is more critical than avoiding irrelevant data.

Our model achieved optimal performance with the following parameters: a chunk size of 500 words, a chunk overlap of 250 words, the integration of HYDE embeddings, and a 'k' value of 20. Utilizing these parameters, the model attained an F-beta score of 0.864909 for police reports and 0.813397 for court transcripts. Notably, larger chunk sizes (1000 and 2000 words) and greater overlaps (500 and 1000 words) reduced the F-beta score, despite providing more contextual information. HYDE embeddings consistently enhanced performance, proving vital to the model's effectiveness. Additionally, a temperature setting of 1 generally improved F-beta scores. However, it is crucial to manage this parameter carefully for extracting accurate officer context mentions in subsequent phases. Higher temperature settings, while increasing variability, risk generating inaccurate or fabricated content, an issue often termed 'hallucination' in language models.

Figure 3: Model Performance Evaluation

chunk_size	chunk_overlap	temperature	k	hyde	filetype	FN	FP	TP	n_files	precision	recall	F1	F-beta
500	250	1	20	1	transcript	3	27	34	4	0.557	0.919	0.694	0.813
500	250	0	20	1	report	6	56	60	5	0.517	0.909	0.659	0.789
2000	1000	1	5	0	report	12	32	71	5	0.689	0.855	0.763	0.816
2000	1000	1	5	1	transcript	3	11	17	3	0.607	0.85	0.708	0.787
500	250	1	20	1	report	20	2	105	5	0.981	0.84	0.905	0.865
1000	500	0	10	1	report	13	70	61	5	0.466	0.824	0.595	0.714
1000	500	0	10	1	transcript	15	31	57	6	0.648	0.792	0.712	0.758
2000	1000	0	5	1	report	13	37	49	5	0.57	0.79	0.662	0.734
2000	1000	0	5	0	report	15	13	54	5	0.806	0.783	0.794	0.787
500	250	0	20	1	transcript	19	29	53	6	0.646	0.736	0.688	0.716
2000	1000	0	5	1	transcript	22	18	60	7	0.769	0.732	0.75	0.739
2000	1000	1	5	1	report	37	19	86	5	0.819	0.699	0.754	0.72
1000	500	1	10	1	report	34	10	78	5	0.886	0.696	0.78	0.728
1000	500	1	10	1	transcript	16	32	19	4	0.373	0.543	0.442	0.497
2000	1000	0	5	0	transcript	44	36	50	9	0.581	0.532	0.556	0.541

Following the initial evaluations based on varying parameters, we continued to measure the model's performance through multiple iterative runs by employing two distinct query strategies. Given the inherent stochastic nature of large language models, which often yield varied outputs when processing the same document multiple times, these strategies were designed to assess the model's performance under static parameters. These additional tests were conducted using three models: GPT-4, GPT-3.5-Turbo-16K, and GPT-3.5-Turbo-4K.

The first strategy involved using six different queries, each specifically crafted to extract distinct types of information. This approach enabled us to evaluate the model's versatility in handling a variety of information extraction tasks. The second strategy used a single, comprehensive query, designed to extract all relevant information, over six iterations. This approach enabled us to measure the model's performance across successive iterations.

Figure 4. Average Cumulative F-Beta Score by Iteration for Police Reports GPT-4 (Standard Model)

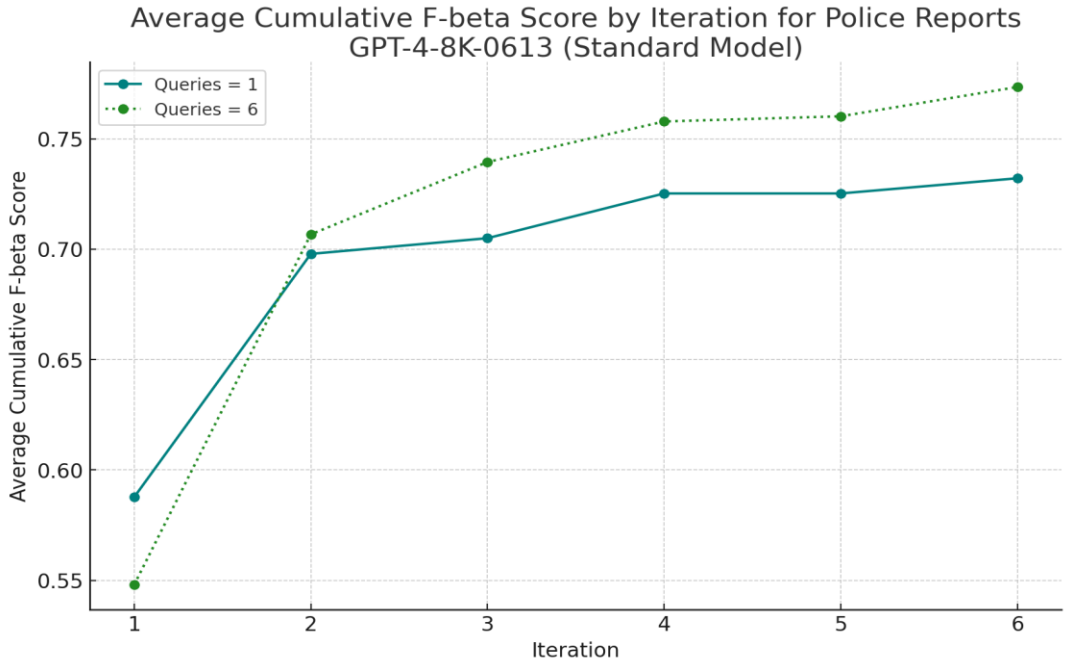


Figure 5. Average Cumulative F-Beta Score by Iteration for Transcripts Reports GPT-4 (Standard Model)

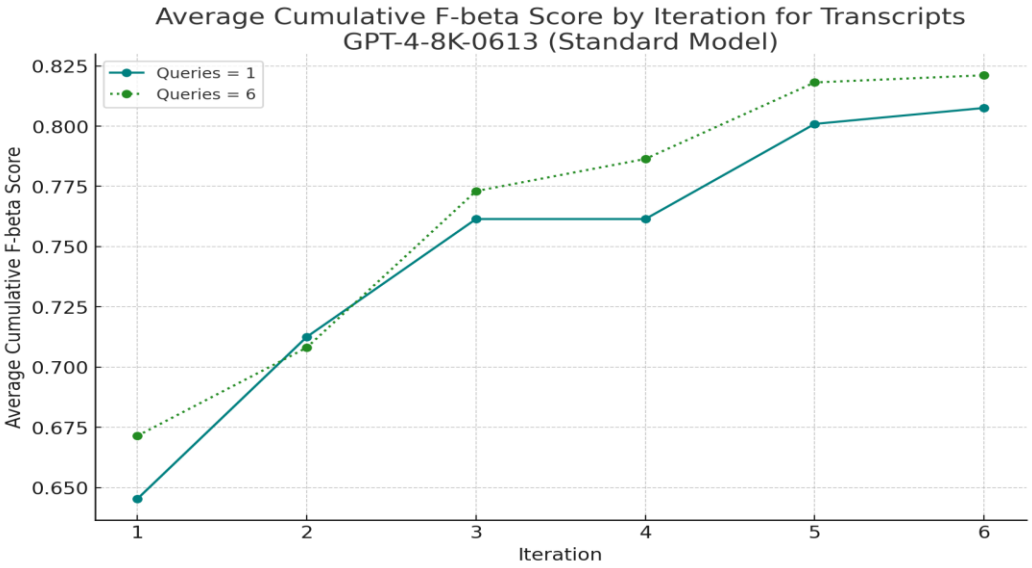


Figure 6. Average Cumulative F-Beta Score by Iteration for Police Reports GPT-3.5-Turbo-16K-0613 (Standard Model)

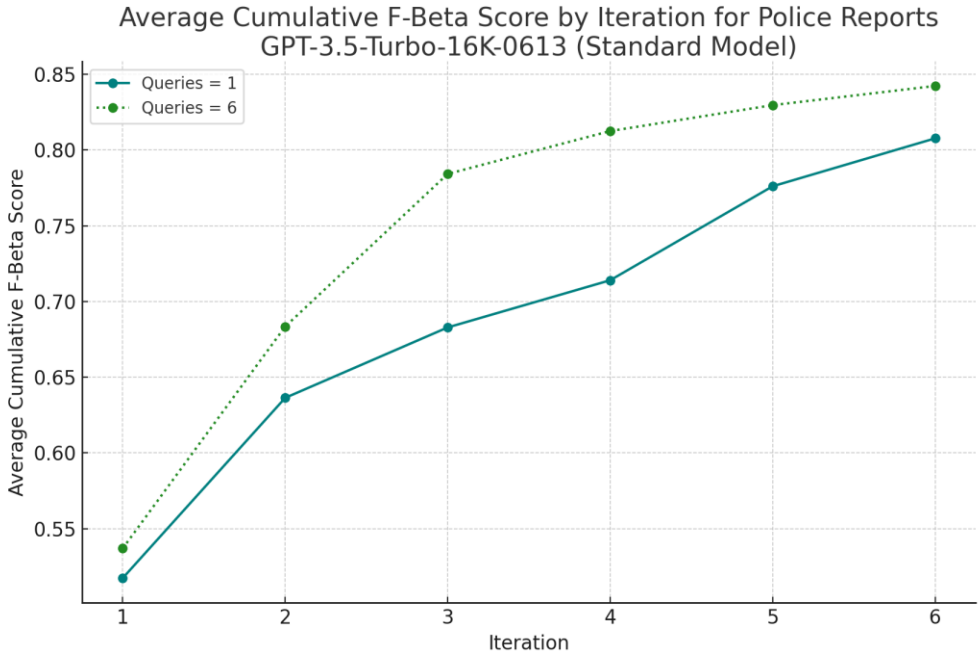


Figure 7. Average Cumulative F-Beta Score by Iteration for Transcripts GPT-3.5-Turbo-16K-0613 (Standard Model)

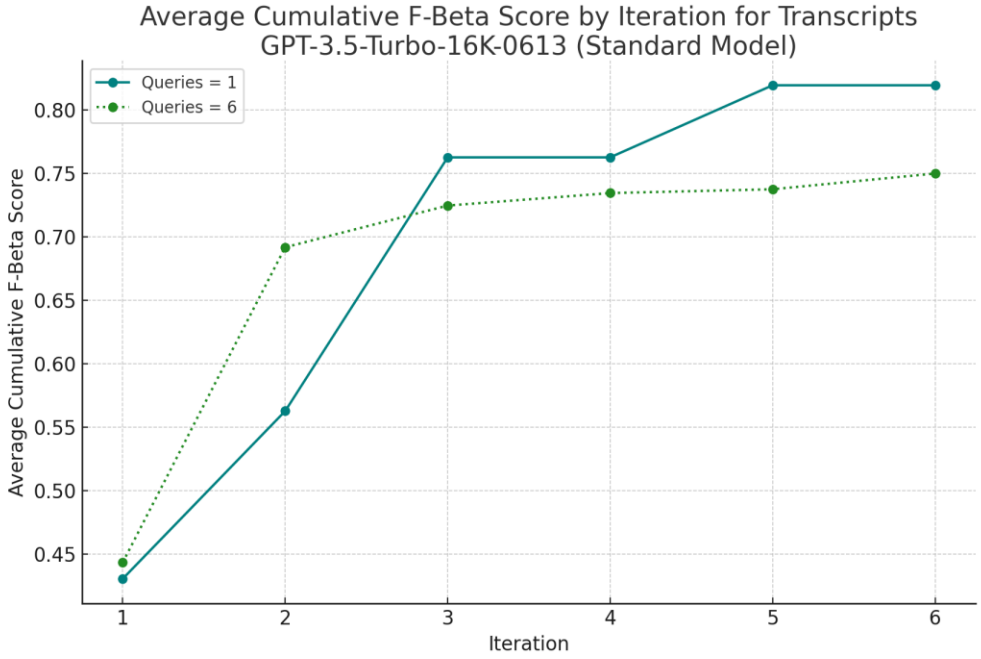


Figure 8. Average Cumulative F-Beta Score by Iteration for Police Reports GPT-3.5-Turbo-4K-0613 (Standard Model)

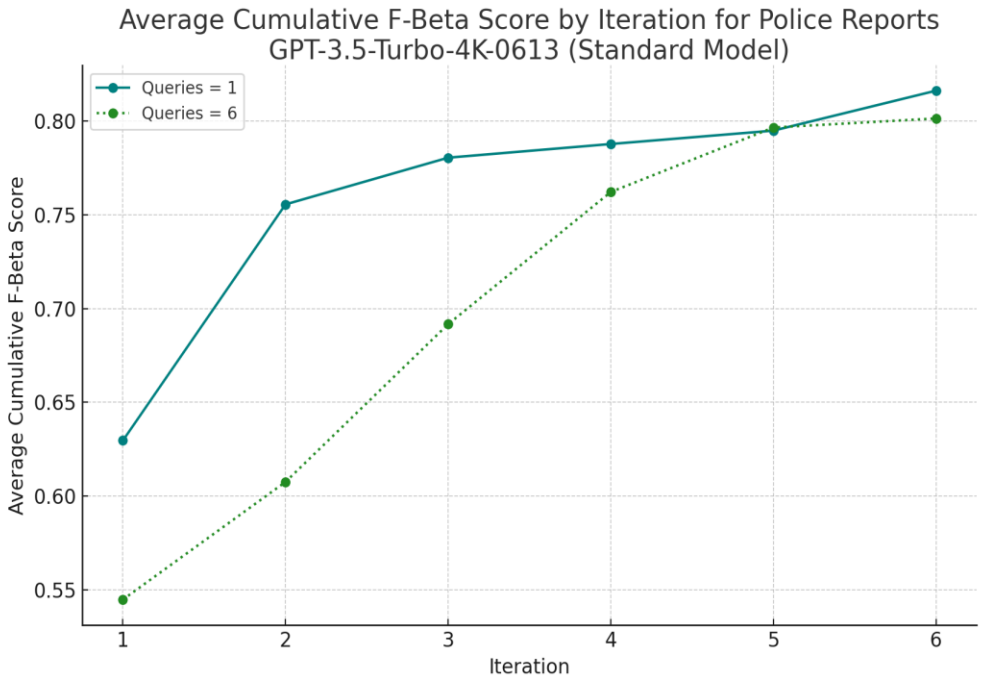
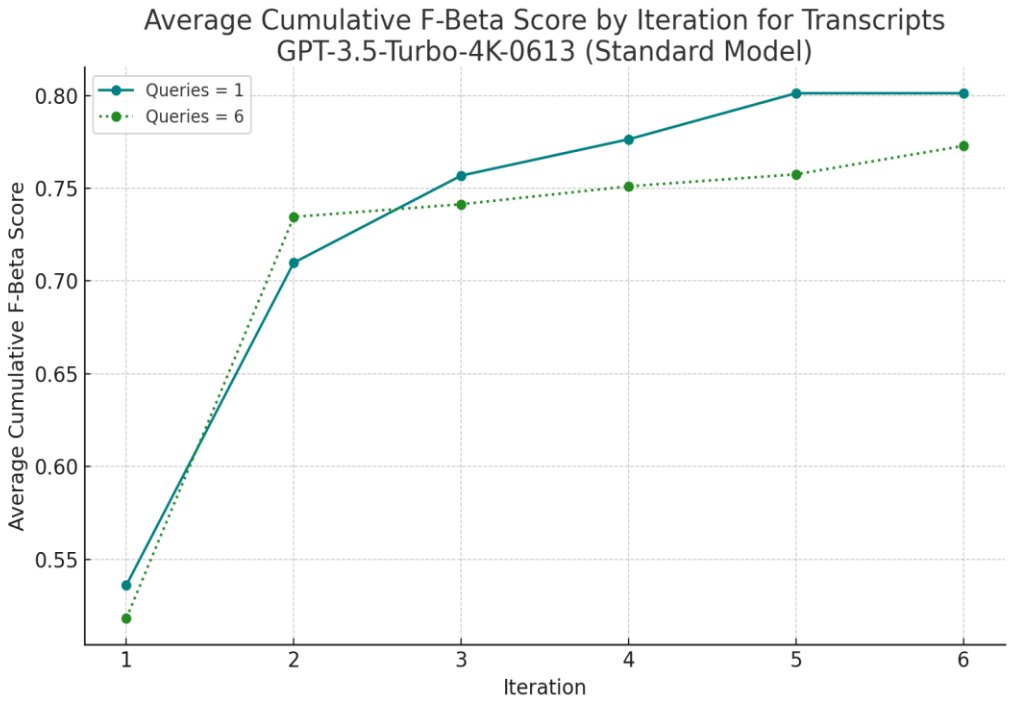


Figure 9. Average Cumulative F-Beta Score by Iteration for Transcripts GPT-3.5-Turbo-4K-0613 (Standard Model)



In evaluating the model's iterative performance, we observed that the point of diminishing returns generally occurs after the fourth iteration. While subsequent iterations do yield improvements, the incremental benefits should be weighed against computational and economic costs. Our comparative analysis using six unique queries versus one comprehensive query revealed that neither approach consistently outperformed the other. Instead, their effectiveness varied depending on the document type and the specific test.

The cumulative F-beta scores from the final iteration indicate that each model's proficiency varied based on the testing conditions. The GPT-4-8K model excelled in handling transcripts, while the GPT-3.5-Turbo-16K model demonstrated a slight advantage in processing reports. On the other hand, the GPT-3.5-Turbo-4K model showcased a balanced performance across different document types, with notable improvement over both of the more advanced models when employing a single query in reports. These findings underscore the importance of a strategic approach to model selection, as specific models may yield better results with particular types of documents or query strategies.

IV Fine-Tuning the Large Language Model

At the time of our tests, the cost differences between GPT-4-0613 and GPT-3.5-0613 were significant, with GPT-4's cost per input and output token being 1900% and 2900% higher, respectively. Given these substantial economic factors, our next iteration of model evaluation considered fine-tuning the GPT-3.5-Turbo-4K-0613, which was then the only model available for this purpose.

Fine-tuning involves tailoring a pre-trained model to better suit specific tasks or datasets[10]. By adjusting the model's parameters through additional training on a targeted dataset, fine-tuning aims to enhance the model's performance for particular applications[11]. In our case, the objective of fine-tuning GPT-3.5-Turbo-4K-0613 was to approximate the advanced performance of GPT-4 and GPT-3.5-Turbo-16k-0613, leveraging GPT-3.5-Turbo-4K-0613's existing capabilities while optimizing it for the specific nuances and complexities found in exoneration documents.

To train the GPT-3.5-Turbo-4K-0613 for handling exoneration documents, we generated training data using GPT-4-0613. GPT-4-0613, with its advanced capabilities, produced outputs that replicate the issues found in actual exoneration documents, such as inconsistent OCR quality, typos, and complex syntactic structures. These outputs were then utilized to train GPT-3.5-Turbo-4K to be more adept at analyzing exoneration documents.

Listing 5.0: **JSON Training Data Example for Fine-Tuning GPT-3.5-Turbo-4K-0613**

{
"messages":
[{ "role": "system", "content":

- "As an AI assistant, my role is to meticulously analyze criminal justice documents and extract information about law enforcement personnel. The response will contain: 1) The name of a law enforcement personnel. The individual's name must be prefixed with one of the following titles to be in law enforcement: Detective, Sergeant, Lieutenant, Captain, Deputy, Officer, Patrol Officer, Criminalist, Technician, Coroner, or Dr. Please prefix the name with 'Officer Name: I will derive this data from the following paragraph: On September 13, DET. X. Y. Allen responded to claims of counterfeit money circulating in the Westside Market. Primary informants were Mrs. Jacobs, a vendor, and Mr. Silva, a customer. FORWARD COPY TO: DETECTIVE DIVISION. INFORMANT DOCS."},

{ "role": "user", "content":

- 'Identify each individual in the transcript, by name, who are directly referred to as officers, sergeants, lieutenants, captains, detectives, homicide officers, and crime lab personnel." },

{ "role": "assistant", "content":"Officer Name: DET. X. Y. Allen

- Officer Context: On September 13, DET. X. Y. Allen responded to claims of counterfeit money circulating in the Westside Market.
- Officer Role: Detective" }]

}

To assess the impact of fine-tuning on the GPT-3.5-Turbo-4K-0613 model's performance, we established a set of benchmarks using three baseline models: the Standard GPT-4-8K-0613, the Standard GPT-3.5-Turbo-16K-0613, and the Standard GPT-3.5-Turbo-4K-0613. These models provided a comparative framework from which we could measure the incremental performance gains achieved through fine-tuning efforts on datasets of 25 to 300 labels.

[10] Evani Radiya-Dixit & Xin Wang, "How fine can fine-tuning be? Learning efficient language models" (2024) ArXiv: 2004.14129, online: <arxiv.org/abs/2004.14129>.

[11] Alexander Dunn *et al*, "Structured information extraction from complex scientific text with fine-tuned large language models" (2022) ArXiv:2212.05238, online: <arxiv.org/abs/2212.05238>.

Initially, the results of fine-tuning were mixed; the GPT-3.5-Turbo-4K-0613 models trained on smaller datasets often performed worse than the baseline models, including the standard GPT-3.5-Turbo-4K model. However, once fine-tuned with 300 labels, the performance of the GPT-3.5-Turbo-4K-0613 model showed significant advancement. For six queries in police reports, it attained a score of 0.859, surpassing the previous high baseline score of 0.842. In transcripts with one query, the fine-tuned model achieved a score of 0.849, outperforming all baseline models, which had a high of 0.819. Across all other test conditions, the fine-tuned model performed comparably to the baseline models, clearly demonstrating that fine-tuning can significantly enhance model performance beyond the established standards for analyzing exoneration documents.

As we continue to explore the potential of large language models, understanding their performance variability—especially under different conditions and datasets—is important. Our current comparative analysis selects the highest scores from five iterations for each standard and fine-tuned model variant. Moving forward, we plan to conduct a more rigorous investigation involving 100 individual runs per model. This approach will provide a robust dataset that captures the full scope of each model's capabilities. We will record and analyze the performance data for each run, focusing on police reports and transcripts, with one and six queries, to accurately assess the models' stability and reliability. Statistical measures such as mean performance scores, confidence intervals, and standard deviations will be employed to quantify the models' consistency. The data obtained will be instrumental in enhancing model performance and fine-tuning application strategies, ensuring that the models deliver consistent and reliable results.

Figure 10. Comparative Performance Analysis of GPT-3.5-Turbo Variants on Reports and Transcripts

Model Variant	Data Type	Performance with 1 Unique Query	Performance with 6 Unique Queries
GPT-4-8k-0613 (Standard)	Reports	0.732	0.773
GPT-4-8k-0613 (Standard)	Transcripts	0.807	0.821
GPT-3.5-Turbo-16k-0613 (Standard)	Reports	0.807	0.842
GPT-3.5-Turbo-16k-0613 (Standard)	Transcripts	0.819	0.749
GPT-3.5-Turbo-4k-0613 (Standard)	Reports	0.816	0.801
GPT-3.5-Turbo-4k-0613 (Standard)	Transcripts	0.801	0.772
GPT-3.5-Turbo-4k-0613 (Fine-Tuned with 25 labels)	Reports	0.771	0.827

GPT-3.5-Turbo-4k-0613 (Fine-Tuned with 25 labels)	Transcripts	0.737	0.815
GPT-3.5-Turbo-4k-0613 (Fine-Tuned with 50 labels)	Reports	0.821	0.8
GPT-3.5-Turbo-4k-0613 (Fine-Tuned with 50 labels)	Transcripts	0.793	0.759
GPT-3.5-Turbo-4k-0613 (Fine-Tuned with 100 labels)	Reports	0.794	0.803
GPT-3.5-Turbo-4k-0613 (Fine-Tuned with 100 labels)	Transcripts	0.727	0.802
GPT-3.5-Turbo-4k-0613 (Fine-Tuned with 200 labels)	Reports	0.816	0.793
GPT-3.5-Turbo-4k-0613 (Fine-Tuned with 200 labels)	Transcripts	0.815	0.797
GPT-3.5-Turbo-4k-0613 (Fine-Tuned with 300 labels)	Reports	0.819	0.859
GPT-3.5-Turbo-4k-0613 (Fine-Tuned with 300 labels)	Transcripts	0.849	0.771

V Entity Resolution and Entity Matching

After extracting structured information about individuals involved in exonerations from the case documents, we will have a database of *raw mentions*. A mention consists of a name, the natural language context of the mention, a role, and various pieces of case-level metadata, including case number, date, and jurisdiction.

The same entity can be mentioned repeatedly, within a single document, across many documents related to a single case, and even across cases. Due to natural variations in how names are reported, including the varying contexts in which they appear and even OCR noise, mentions that refer to the same entity – "co-referent" mentions – will look different. For example: "Detective Tom Jones" in one mention may be referred to as "Det. Jones" or even just "Jones" in other mentions. Our goal is to match every mention to a canonical reference record in LLEAD. However, some mentions in isolation do not provide sufficient information to match the mention directly to LLEAD. A mention such as "Detective Jones" could correspond to multiple candidate records. By clustering co-referent mentions, we can aggregate information about distinct entities to give ourselves the best chance at a successful match. For example, knowing that the mention of "Detective Jones" refers to the same person as a previous mention of "Detective Tom Jones of the New Orleans PD" provides us with the context necessary to match *both* mentions to the correct reference record.

Entity resolution, or deduplication, and entity matching both rely on an appropriate pairwise similarity measure – for two given mentions, or for one mention and a reference record in the database, we want to be able to measure how "similar" they are, where we are defining "similarity" in terms of how likely they are to co-refer to the same real-life entity. Using this metric, we then want to cluster mentions that are similar to each other so that we can treat mentions that get clustered together as co-referent. By combining information within clusters, we can build an aggregate mention record that is as complete as possible. Finally, we can match the aggregate mention to our canonical database, once again using a pairwise similarity metric.

The entity resolution process is described in detail in Peter Christen's *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection (Data-Centric Systems and Applications)*. Traditionally, the inputs to this process are structured records. A novel challenge in our work is that a significant portion of each mention, the "context", consists of unstructured text. String similarity, using metrics such as Jaccard or edit distance, may be sufficient to capture similarity in names, but in order to compare two mention contexts, we need to capture semantic similarity. By *embedding* – or, mapping the text to vectors of real numbers – the mentions using natural language models, we can translate the mentions into a format that allows us to measure semantic similarity using vector databases that enable efficient nearest-neighbor searching. Embedding mentions for the purposes of entity normalization are described in Learning Text Similarity with Siamese Recurrent Networks[12] and NSEEN: Neural Semantic Embedding for Entity Normalization[13]. As those examples demonstrate, we seek an embedding that captures both the syntactic similarity of names across mentions as well as the semantic similarity of contexts. Given the appropriate embeddings, we can compare mentions using cosine similarity. As those readings also describe how to learn appropriate embeddings without much training data, we propose once again using embeddings from large language models to embed the contexts, further reducing our need for hand-labeled data.

At this point in the deduplication process, scale becomes a concern. Even with a good pairwise similarity metric between mentions, given the number of distinct mentions that we are extracting, we cannot realistically calculate similarity for every combination of two extracted mentions. However, such a pairwise *similarity matrix* is usually a precondition for clustering. A common solution is blocking[14]. This involves identifying "candidate" pairs that could be co-referent, while filtering out sufficiently dissimilar pairs without ever considering them. One approach to blocking that captures our notion of syntactic similarity in names is to only consider pairs that match on some substring. Representing our mentions as embeddings allows us to unlock another family of powerful candidate selection techniques, known as *locality sensitive hashing* (LSH). LSH describes a family of techniques that are based on mapping input mentions to buckets in such a way that similar mentions get mapped to the same bucket. These methods are described

[12] Paul Neculoiu, Maarten Versteegh & M Rotaru, "Learning Text Similarity with Siamese Recurrent Networks" (2016) Proc 1st Workshop on Representation Learning for NLP 148, online (pdf): <aclanthology.org/W16-1617.pdf>.

[13] Shobeir Fakhraei, Joel Mathew & Jose Luis Ambite, "NSEEN: Neural Semantic Embedding for Entity Normalization" (2018) ArXiv:1811.07514, online: <arxiv.org/abs/1811.07514>.

[14] Patrick Ball, "How do we find duplicates among multiple, giant datasets?" (last accessed 5 Dec 2023) online (blog): <hrdag.org/tech-notes/adaptive-blocking-writeup-1.html>.

in *Mining of Massive Datasets*[15]. In some instances, we may even be able to skip a second clustering step altogether, treating the LSH buckets as our final clusters.

Assuming we still need a separate clustering step, we have a range of unsupervised and semi-supervised clustering algorithms we can make use of. However, as described in "Theoretical Limits of Record Linkage and Microclustering"[16] and "Flexible Models for Microclustering with Application to Entity Resolution,"[17] the data we want to cluster is different from the data assumed by most clustering models in ways that affect the quality of solutions that those models can provide. Specifically most generative models for clustering implicitly assume that the number of data points in each cluster grows linearly with the total number of data points… However, for some applications, this assumption is inappropriate. For example, when performing entity resolution, the size of each cluster should be unrelated to the size of the data set, and each cluster should contain a negligible fraction of the total number of data points."[18]

Ball introduces hierarchical agglomerative clustering as a robust solution to the specific problem of clustering for entity resolution.[19] NSEEN, on the other hand, relies on the "sketch" technique described in Mining of Massive Datasets for a type of locality sensitive hashing that approximates clusters based on cosine distance in the embedding space.[20] In addition to these techniques, we will try to build on our success using HyDE and large language models to improve the matching and clustering steps.

Once we have co-referent clusters, we will attempt to use the same pairwise similarity metric, as well as a new one specific to the LLEAD data, in order to match resolved entities to canonical records in the reference database.

VI Future Research

Future research will concentrate on fine-tuning and optimizing the latest developments from OpenAI, particularly the GPT-3.5-Turbo-16K-1106 and GPT-4-Turbo models. Our experiments will focus on running similar tests, such as testing the effects of chunk size, chunk overlap, and 'k' value, in addition to fine-tuning, as the new GPT-3.5-Turbo-16K-1106 model is available for fine-tuning.

---

[15] Jure Leskovec, Anand Rajaraman & Jeff Ullman, "Mining of Massive Datasets" (last accessed 5 Dec 2023), online: <mmds.org>.

[16] James E Johndrow, Kristian Lum & David B Dunson, "Theoretical Limits of Record Linkage and Microclustering" (2017) ArXiv:1703.04955, online: <arxiv.org/abs/1703.04955>.

[17] Giacomo Zanella *et al*, "Flexible Models for Microclustering with Application to Entity Resolution" (2016) ArXiv:1610:09780, online: <arxiv.org/abs/1610.09780> at 1.

[18] Giacomo Zanella, Brenda Betancourt, Hanna Wallach, Jeffrey Miller, Abbas Zaidi, Rebecca C. Steorts, "Flexible Models for Microclustering with Application to Entity Resolution," [online] available: <https://www.microsoft.com/en-us/research/uploads/prod/2016/11/Flexible-Models-for-Microclustering-with-Application-to-Entity-Resolution.pdf> (accessed 12/5/2023).

[19] Patrick Ball, "Clustering and solving the right problem" (25 Jul 2016) online (blog): <hrdag.org/tech-notes/clustering-and-solving-the-right-problem.html>

[20] Shobeir Fakhraei, Joel Mathew & Jose Luis Ambite, "NSEEN: Neural Semantic Embedding for Entity Normalization" (2018) ArXiv:1811.07514, online: <arxiv.org/abs/1811.07514>.

While our current fine-tuned GPT-3.5-Turbo-4K-0613 model stands as the most efficient iteration so far, it's important to acknowledge that the new GPT-4-Turbo outperforms this model. In its baseline setup with a chunk size of 500, an overlap of 250, and a 'k' value of 20, the GPT-4-Turbo has achieved F-beta scores of up to 0.85. Further adjustments of its parameters to a chunk size of 20000, an overlap of 6000, and a 'k' value of 50 have led to F-Beta scores reaching as high as 0.95. This impressive performance highlights the potential for substantial improvements in structured data extraction tasks, offering a promising direction for future research and model optimization.

When pleas precede evidence: Using Bayesian analyses to establish the importance of a *reasonable* standard for evidence prior to plea offers

Miko M. Wilford
Department of Psychology, Iowa State University
Ames, Iowa
U.S.A.

Joseph E. Gonzales
Department of Psychology, University of Massachusetts Lowell
Lowell, Massachusetts
U.S.A.

Annmarie Khairalla
Forensic Psychology, Ontario Tech University
Oshawa, Ontario
Canada

In most U.S. jurisdictions, prosecutors are not required to clearly establish a reasonable basis for guilt prior to offering defendants plea deals. We apply Bayesian analyses, which are uniquely suited to illuminate the impact of prior probability of guilt on the informativeness of a particular outcome (i.e., a guilty plea), to demonstrate the risks of plea offers that precede evidence. Our primary prediction was that lower prior probabilities of guilt would coincide with a significantly higher risk for false guilty pleas. We incorporated data from Wilford, Sutherland[1] *into a Bayesian analysis allowing us to model the expected diagnosticity of plea acceptance across the full range of prior probability of guilt. Our analysis indicated that, as predicted, when plea offers are accepted at lower prior probabilities of guilt, the probability that a plea is actually false is significantly higher than when prior probabilities of guilt are higher. In other words, there is a trade-off between prior probability of guilt and information gain. For instance, in our analysis, when prior probability of guilt was 50%, posterior probability of guilt (after a plea) was 77.8%; when prior probability of guilt was 80%, posterior probability of guilt was 93.3%. Our results clearly indicate the importance of ensuring that there is a reasonable basis for guilt before a plea deal is extended. In the absence of shared discovery, no such reasonable basis can be established. Further, these results illustrate the additional insights gained from applying a Bayesian approach to plea-decision contexts*.

[1] Miko M Wilford et al., "Guilt status influences plea outcomes beyond the shadow-of-the-trial in an interactive simulation of legal procedures." (2021) 45:4 Law and Human Behavior 271–286, online: <https://doi.apa.org/doi/10.1037/lhb0000450> [*Wilford Sutherland*].

I. Introduction
A. The Purpose of Plea Bargaining
B. In the Shadow-of-the-Trial
C. Quick Pleas
II. Plea Decision Making as a Bayesian Problem
A. Comparing Two Plea Decision Points
B. Everyone is Blind Without Evidence
III. Policy Recommendations
A. Conclusion

I Introduction

In 1963, the United States Supreme Court made a landmark ruling in favor of John L. Brady.[2] Specifically, the Court found that prosecutors' failure to turn over potentially exculpatory evidence to the defense violated his 14th Amendment right to due process of law. The *Brady* doctrine was later applied to cases in which a prosecutor denied having any knowledge of the exculpatory evidence (e.g., it was not turned over by law enforcement[3]). Critically, this established doctrine set a clear deadline for evidence disclosure: trial. Unfortunately, as the Supreme Court acknowledged almost fifty years later, our criminal justice system "…is for the most part a system of *pleas*, not a system of *trials*…".[4] Consequently, a natural question arises: Are prosecutors required to disclose any potentially exculpatory evidence prior to the adjudication of a case (i.e., by trial or by guilty plea)? And, in the absence of such a requirement, how can a reasonable standard for evidence be established prior to a guilty plea?

In the current paper, we discuss the dangers of plea offers that precede evidence (or evidentiary discovery). To illustrate these risks, we apply Bayesian analyses, which are uniquely suited to illuminate the impact of the prior probability of guilt (or base rates of guilt) on the informativeness of a particular outcome (i.e., a guilty plea). We conclude by discussing the implications of the results from our Bayesian analyses and offering relevant recommendations for reform.

A. The Purpose of Plea Bargaining

In 1970, the United States Supreme Court made another landmark *Brady* ruling—this time ruling against the petitioner, Robert M. Brady.[5] In a unanimous opinion, the Court ruled that the threat of death did not render a guilty plea involuntary. The Court defended plea negotiations by stating that they allow "… scarce judicial and prosecutorial resources [to be] conserved for those cases in which there is a substantial issue of the defendant's guilt or in which there is substantial doubt that the State can sustain its burden of proof." Here, the Court clearly suggests that prosecutors should be taking cases they are *less certain* of to trial and that plea bargaining should

[2] *Brady v Maryland*, 1963 373 U.S. 83 [*Brady v Maryland*].

[3] *Kyles v Whitley*, 1995 514 U.S. 419 [*Kyles*].

[4] *Lafler v Cooper*, 2012 566 U.S. 156 [*Lafler*].

[5] *Brady v United States*, 1970 397 U.S. 742 [*Brady*].

be reserved for cases in which there is strong evidence of guilt. In other words, the original purpose of plea bargaining was to accelerate the pace of cases for which the accused person's guilt was essentially certain. In so doing, the State's resources could be reserved to try those cases for which reasonable doubt might exist.

Accordingly, to deem a plea conviction valid, the Court must theoretically establish a *sufficient* factual basis for the plea. In practice, courts can rely solely on the police report, or even the accused person's own guilty plea, to meet this standard.[6] Further, the U.S. Supreme Court later ruled (unanimously) that the prosecution was permitted to require that accused persons' "…waive their right to impeachment information relating to any informants or other witnesses…" as a condition of accepting a plea offer.[7] The Court recognized the right (to potentially exculpatory *Brady* material) as cloaked in one's right to a fair trial—when that right is waived (by pleading guilty), all accompanying rights are also waived.

> It could require the Government to devote substantially more resources to trial preparation prior to plea bargaining, thereby depriving the plea-bargaining process of its main resource-saving advantages. Or it could lead the Government instead to abandon its heavy reliance upon plea bargaining in a vast number – 90% or more – of federal criminal cases. We cannot say that the Constitution's due process requirement demands so radical a change in the criminal justice process in order to achieve so comparatively small a constitutional benefit – Justice Breyer.[8]

Thus, while failing to sufficiently define the factual-basis-for-guilt requirement of pleas, the Court has also made it clear that the *Brady* doctrine does not apply (at least not fully) to plea negotiations. This conclusion seems to conflict with the Court's earlier justification of plea bargaining as a means of more efficiently processing cases for which guilt is not in question. If guilty pleas are to be reserved for essentially unequivocal cases, why deprive the defense of potentially exculpatory evidence? *Ruiz* seems to represent a significant change in the Court's original views of plea-bargaining such that the increased efficiency it confers outweighs potential threats to due process.

Unsurprisingly, criminal attorneys now readily acknowledge the occurrence of guilty pleas in weak cases.[9] In fact, many have noted that the "system of pleas" is designed to encourage

---

[6] Stephanos Bibas, "Plea bargaining's role in wrongful convictions" in Examining wrongful convictions: Stepping back, moving forward (Durham, NC: Carolina Academic Press , 2014) 157 - 167; Amy Dezember et al., "Plea validity in circuit court: judicial colloquies in misdemeanor vs. felony charges" (2022) 28:3 Psychology, Crime & Law 268–288, online: <https://www.tandfonline.com/doi/full/10.1080/1068316X.2021.1905813>; Allison D Redlich et al, "Guilty plea hearings in juvenile and criminal court." (2022) 46:5 Law and Human Behavior 337–352, online: <https://doi.apa.org/doi/10.1037/lhb0000495> [*Redlich 2022*].

[7] *United States v Ruiz*, 2002 536 U.S. 622 [*Ruiz*].

[8] *Ibid* at para 23.

[9] Albert W Alschuler, "A Nearly Perfect System for Convicting the Innocent" (2016) 79:3 Albany Law Review, online: <https://www.albanylawreview.org/article/69791-a-nearly-perfect-system-for-convicting-the-innocent>.

prosecutors to take stronger cases to trial (to preserve high conviction rates) while pleading weaker cases away[10]—completely antithetical to the Supreme Court's original justification for guilty pleas (in *Brady*).[11] When prosecutors' confidence in securing convictions would generally be lowest (or at least the most unclear), they can still negotiate convictions. The faster the guilty plea, the faster the case resolution (maximizing judicial efficiency).[12] Judges also rarely question guilty pleas.[13] Presumably, legal actors defend these quick pleas (that precede evidence) via the assumption that guilty pleas are themselves sufficient evidence for conviction. Yet, a guilty plea (as the prevalence of demonstrably false guilty pleas illustrates) is a far cry from an assurance of true guilt.[14]

B. In the Shadow-of-the-Trial

Many legal scholars have further defended the practice of plea-bargaining with the supposition that it operates in the shadow of the trial.[15] Specifically, accused persons can choose to accept a plea offer by evaluating the sentence it confers against the potential sentence after trial.[16] The shadow-of-the-trial (or shadow) model is essentially utility theory, applied to the context of plea decision-making. When offered a plea, accused persons compare the known utility associated with pleading guilty (i.e., the plea sentence) with the expected utility associated with going to trial (i.e., the estimated risk of conviction and the expected trial sentence). In this way, plea outcomes (and discounts) are theoretically still influenced by the trial process.[17] Research testing the predictive validity of the shadow model has been mixed, with some finding support for it at an aggregate but not individual level.[18] Others have highlighted more systematic weaknesses

---

[10] Jed S Rakoff, "Why Innocent People Plead Guilty" The New York Review of Books (20 November 2014), online: <https://www.nybooks.com/articles/2014/11/20/why-innocent-people-plead-guilty/> [*Rakoff*].

[11] *Brady*, *supra* note 5.

[12] Miko M Wilford & Annmarie Khairalla, "Innocence and Plea Bargaining" in A System of Pleas (Oxford University Press, 2019) 132 - 152 [*Wilford & Khairalla*].

[13] Rakoff, *supra* note 10.

[14] Miko M Wilford & Brian H Bornstein, "The disappearing trial: how social scientists can help save the jury from extinction" (2023) 29:1 Psychology, Crime & Law 1–24, online: <https://www.tandfonline.com/doi/full/10.1080/1068316X.2021.1984482> [*Wilford & Bornstein*].

[15]William M Landes, "An Economic Analysis of the Courts" (1971) 14:1 The Journal of Law and Economics 61–107, online: <https://www.journals.uchicago.edu/doi/10.1086/466704> [*Landes*].

[16]Robert H Mnookin & Lewis Kornhauser, "Bargaining in the Shadow of the Law: The Case of Divorce" (1979) 88:5 The Yale Law Journal 950-997, online: <https://www.jstor.org/stable/795824?origin=crossref>.

[17] Shi Yan, "Estimating the Size of Plea Discounts" in Cassia Spohn & Pauline K Brennan, eds, Handbook on Sentencing Policies and Practices in the 21st Century, 1st ed (Abingdon, Oxon ; New York, NY : Routledge, 2019. | Series: American Society of Criminology Division on Corrections and Sentencing handbook series ; Volume 4: Routledge, 2019) 188-207; Shi Yan, "What Exactly Is the Bargain? The Sensitivity of Plea Discount Estimates" (2022) 39:1 Justice Quarterly 152–173, online: <https://www.tandfonline.com/doi/full/10.1080/07418825.2019.1707856>.

[18] See for examples: Shawn D Bushway & Allison D Redlich, "Is Plea Bargaining in the 'Shadow of the Trial' a Mirage?" (2012) 28:3 J Quant Criminol 437–454, online:

of the model[19], or the significant impact of other decision-making biases or strategies unaccounted for in the shadow model (e.g., anchoring;[20] discounting;[21] framing;[22] fuzzy-trace[23]).

Of course, even if we assume that most accused persons are rational decision-makers (tenuous given growing critiques of utility theory),[24] rational decision-making relies on comprehensive information.[25] Without access to evidentiary discovery, how can we expect accused persons to accurately estimate their probability of conviction? Put another way, how can plea bargaining occur in the shadow of the trial when trial evidence is kept in the dark? Further,

---

<http://link.springer.com/10.1007/s10940-011-9147-5> ; Shawn D Bushway, Allison D Redlich & Robert J Norris, "An Explicit Test of Plea Bargaining in the "Shadow of the Trial"" (2014) 52:4 Criminology 723–754, online: <https://onlinelibrary.wiley.com/doi/10.1111/1745-9125.12054>.

[19] See Jennifer M Bartlett & Tina M Zottoli, "The paradox of conviction probability: Mock defendants want better deals as risk of conviction increases." (2021) 45:1 Law and Human Behavior 39–54, online: <https://doi.apa.org/doi/10.1037/lhb0000432> ; see Kevin Petersen, Allison D Redlich & Robert J Norris, "Diverging from the shadows: explaining individual deviation from plea bargaining in the 'shadow of the trial'" (2022) 18:2 J Exp Criminol 321–342, online: <https://link.springer.com/10.1007/s11292-020-09449-4>; See *Wilford Sutherland, supra* note 1.

[20] Stephanie A Cardenas, "Charged up and anchored down: A test of two pathways to judgmental and decisional anchoring biases in plea negotiations." (2023) 29:4 Psychology, Public Policy, and Law 435–456, online: <https://doi.apa.org/doi/10.1037/law0000390>.

[21] Lauren Clatch & Eugene Borgida, "Plea Bargaining: A Test of Dual Discounting Preferences for Non-Monetary Losses" (2021) 47:7 Pers Soc Psychol Bull 1039–1056, online: <http://journals.sagepub.com/doi/10.1177/0146167220952228>.

[22] Laura M Garnier-Dykstra & Theodore Wilson, "Behavioral Economics and Framing Effects in Guilty Pleas: A Defendant Decision Making Experiment" (2021) 38:2 Justice Quarterly 224–248, online: <https://www.tandfonline.com/doi/full/10.1080/07418825.2019.1614208>; Rebecca K Helm & Valerie F Reyna, "Logical but incompetent plea decisions: A new approach to plea bargaining grounded in cognitive theory." (2017) 23:3 Psychology, Public Policy, and Law 367–380, online: <https://doi.apa.org/doi/10.1037/law0000125> [*Helm & Reyna*].

[23] Rebecca K Helm et al., "Too young to plead? Risk, rationality, and plea bargaining's innocence problem in adolescents." (2018) 24:2 Psychology, Public Policy, and Law 180–191, online: <https://doi.apa.org/doi/10.1037/law0000156> [*Helm et al.*]; Tina M Zottoli et al., "Developing a model of guilty plea decision-making: Fuzzy-trace theory, gist, and categorical boundaries." (2023) 47:3 Law and Human Behavior 403–421, online: <https://doi.apa.org/doi/10.1037/lhb0000532>.

[24] Daniel Kahneman, Thinking Fast and Slow (Macmillan, 2011);Richard H Thaler, "Behavioral Economics: Past, Present, and Future" (2016) 106:7 American Economic Review 1577–1600, online: <https://pubs.aeaweb.org/doi/10.1257/aer.106.7.1577>.

[25] Niv Ahituv, Magid Igbaria & A Viem Sella, "The Effects of Time Pressure and Completeness of Information on Decision Making" (1998) 15:2 Journal of Management Information Systems 153–172, online: <https://www.tandfonline.com/doi/full/10.1080/07421222.1998.11518212>; Rudolf Grünig & Richard Kühn, Solving Complex Decision Problems (Berlin, Heidelberg: Springer Berlin Heidelberg, 2017) at pg. 25-34.

the shadow model provides a clear prescription for prosecutors with weak cases: offer larger plea discounts.[26]

C. Quick Pleas

Depriving accused persons of evidentiary discovery during the plea process eliminates their ability to evaluate plea offers rationally, whether they are innocent or guilty. Even when states require openness,[27] individuals can still be offered a plea prior to seeing their case evidence. Arizona, for instance, is an open-file state. Prosecutors are required to share discovery as soon as charges are formally filed (i.e., at a preliminary or probable cause hearing).[28] But Maricopa County (including Phoenix) now has Early Disposition Courts (EDCs), which are designed to "fast-track" cases by consolidating or skipping steps in the legal process. The original purpose of these courts was to allegedly provide persons accused of low-level, non-violent offenses incentives to plead quickly and receive treatment (e.g., drug rehabilitation) earlier. Yet, data from the Maricopa County Attorney's Office indicates that in a 4-year period (from January 2017 to January 2021), only 6.7% of all EDC cases diverted convicted persons to treatment programs.[29] Many criminal cases are routed into this system as soon as an accused person has been arrested.[30] Typically, the only "evidence" defense attorneys are provided by the EDC is a police report, their client's criminal record (which can be inaccurate when, for instance, the client is confused with someone else of the same name), and a plea offer from the prosecutor's office.[31] Thus, there is no evidence to support that cases being routed to EDCs involve stronger evidence or a higher probability of conviction. In fact, even if prosecutors are in possession of additional evidence when initial pleas are offered, they often refuse to turn it over until the preliminary hearing.

Attorneys and their clients then have until the preliminary hearing to accept, reject, or renegotiate the offer. Once the case has advanced to a preliminary hearing or an indictment, Maricopa County has clearly established that the EDC plea offer expires and will not be matched; any subsequent offer (which is not guaranteed) will be significantly worse. In fact, written plea offers for EDC cases are often accompanied by the following text:

> The offer is withdrawn if the witness preliminary hearing is set or waived. The offer may be changed or revoked at any time before the court accepts the plea. The offer may be changed or revoked at any time before the court accepts the plea. *Note: County attorney

[26] Landes, *supra* note 15.

[27] Jenia Turner & Allison Redlich, "Two Models of Pre-Plea Discovery in Criminal Cases: An Empirical Comparison" (2016) 73 Washington and lee Law Review 285–408 [*Turner*].

[28] See *Ibid*.

[29] *ACLU v Maricopa County Attorney's Office*, 2021 United states district court District of Arizona, online: <https://tmsnrt.rs/36hLSKX>.

[30] Dave Biscobing, "ACLU Lawsuit: MCAO 'coerces' guilty pleas by fast-tracking cases", (7 July 2021), online: ABC15 Arizona in Phoenix (KNXV) <https://www.abc15.com/news/local-news/investigations/protest-arrests/aclu-lawsuit-mcao-coerces-guilty-pleas-by-fast-tracking-cases> [*Biscobing*].

[31] Carissa Byrne Hessick, *Punishment Without Trial: Why Plea Bargaining Is a Bad Deal* (Abrams Press, 2021) [*Hessick*].

policy dictates that if the defendant rejects this offer, any subsequent offer tendered will be *substantially* harsher.

Thus, in jurisdictions like Maricopa County, accused persons are essentially being asked to take a plea offer while completely blind as to the strength of the prosecution's case. The preliminary hearing is when prosecutors are required to share discovery: to demonstrate that they have sufficient evidence to show that a crime occurred, and that the accused person is guilty of that crime.[32] Encouraging individuals to plead guilty before even this preliminary bar has been met is unquestionably increasing the risk of false guilty pleas. It is allowing prosecutors to move forward with convictions even when there is no way they could (yet) meet a significant burden of proof. Accordingly, the ACLU is pursuing a class action lawsuit against Maricopa County alleging that EDCs are punishing and threatening accused persons for exercising their constitutional rights.

Quick pleas have allowed prosecutors and judges to use guilty pleas, in lieu of evidence, to establish guilt. Presumably, these legal actors believe that accused person's decision to plead guilty is fully diagnostic of actual guilt, regardless of when pleas occur. In an opinion, the former Supreme Court Justice Antonin Scalia estimated a wrongful conviction rate of 0.027%, clearly signaling his faith in the veracity of guilty pleas.[33] Former Judge Paul G. Cassell later used a "components parts approach" to calculate a wrongful conviction range of 0.016 to 0.062%;[34] notably, this analysis was informed by untested assumptions like, "… the risk of a wrongful conviction is, unexpectedly, greater for rape-homicides than for less serious crimes".[35] These presumptions ignore the pressure prosecutors can exert on accused persons to plead guilty. Prosecutors possess a substantial toolbox and significant discretion during the plea process. Accused persons can be incarcerated pretrial, face dramatic sentencing discrepancies,[36] even including qualitatively different punishments (e.g., probation versus incarceration).[37]

Yet, the perceived diagnosticity of guilty pleas makes plea convictions even harder to overturn (in many ways) than trial convictions.[38] While accused persons do not automatically waive their right to appeal by pleading guilty, the Supreme Court has noted that avenues for appeal *can* be waived as a condition of pleading guilty.[39] Further, without a trial, records to

[32] Biscobing, *supra* note 30.

[33] *Kansas v Marsh*, 2006 548 U.S. 163 198.

[34] Paul G Cassell, "Overstating America's Wrongful Conviction Rate? Reassessing the Conventional Wisdom About the Prevalence of Wrongful Convictions" (2018) 60:4 ArizLRev 815–864, online: <https://arizonalawreview.org/overstating-americas-wrongful-conviction-rate-reassessing-the-conventional-wisdom-about-the-prevalence-of-wrongful-convictions/> [*Cassell*].

[35] *Ibid* 829.

[36] Wilford & Bornstein, *supra* note 14.

[37] Miko M Wilford et al., "Innocence in the shadow of COVID-19: Plea decision making during a pandemic." (2021) 27:4 Journal of Experimental Psychology: Applied 739–750, online: <https://doi.apa.org/doi/10.1037/xap0000367>; See *Helm et al.*, *supra* note 23.

[38] Wilford & Khairalla, *supra* note 12.

[39] *Class v United States*, 2018 583 U.S.

support an appeal (e.g., prosecutorial misconduct) are extremely limited.[40] In the forthcoming analyses and discussion, we further challenge the assumption that a guilty plea is diagnostic of guilt. By illustrating the relationships among prior probability of guilt, diagnosticity, and wrongful convictions (using a Bayesian approach), we hope to give all legal actors involved in the plea process an opportunity to reevaluate the value of quick pleas against their potential costs.

Notably, we intentionally conflate a lower probability of *proving* guilt with a lower probability of *being* guilty. Our system of justice is predicated on a presumption of innocence—accused persons are presumed innocent until proven guilty. Thus, if there is less evidence of guilt, there is an inherently higher probability the State will fail to prove guilt (resulting in dismissals and acquittals, regardless of whether the individual is innocent or guilty). In the absence of evidence, the presumption is that the accused person is innocent, not that they are guilty (i.e., people are innocent unless the State can prove otherwise). The withholding of potentially exculpatory evidence (i.e., discovery) is a due process concern for which factual guilt is essentially irrelevant. The concern is whether the State is being held to a reasonable burden of proof prior to convicting accused persons of crimes. As such, our analysis focuses on the prior probability of the State's ability to *prove* factual guilt beyond a reasonable doubt at trial.

II Plea Decision-Making as a Bayesian Problem

Given these assumptions and parameters, a Bayesian approach is uniquely well-suited to demonstrate both the diagnostic utility of a guilty plea and the increased risk of false guilty pleas when pleas are entered quickly (and blindly). *Wells* used the same approach to model the diagnosticity of eyewitness identifications in multiple contexts noting, "… that the conditional probabilities of interest to the legal system naturally map into Bayesian formulations".[41] In their treatise, *Wells* argued that the base rate of guilt (i.e., the prior probability that a suspected person is actually guilty) is a system variable;[42] a system variable represents something that is under the control of the legal system (e.g., lineup instructions).[43] The justice system can control the prior probability (or base rate) of guilt for any legal procedure by pre-determining some acceptable criterion for that procedure (e.g., an *evidence-based suspicion* prior to putting

[40] Thea Johnson, *Plea bargain task force report*, by Thea Johnson (American Bar Association, 2023), online:<https://www.americanbar.org/content/dam/aba/publications/criminaljustice/plea-bargain-tf-report.pdf> [*Johnson*].

[41] Gary L Wells, Yueran Yang & Laura Smalarz, "Eyewitness identification: Bayesian information gain, base-rate effect equivalency curves, and reasonable suspicion." (2015) 39:2 Law and Human Behavior 99–122, online: <https://doi.apa.org/doi/10.1037/lhb0000125> [*Wells*].

[42] *Ibid*.

[43] Gary L Wells, "Applied eyewitness-testimony research: System variables and estimator variables." (1978) 36:12 Journal of Personality and Social Psychology 1546–1557, online: <https://doi.apa.org/doi/10.1037/0022-3514.36.12.1546>; Miko M Wilford & Gary L Wells, "Eyewitness system variables." in Brian L Cutler, ed, Reform of eyewitness identification procedures (Washington: American Psychological Association, 2013) 23.

suspected persons in lineups or interrogating them;[44] a *reasonable basis* for guilt prior to offering an individual a plea).

While the majority of plea decision-making experiments in which guilt status is manipulated employ a 50-50 ratio (i.e., there is a 50% prior probability that any given participant-defendant is guilty), the real-world base rate is unknown. However, Bayesian analyses allow us to calculate the posterior probability of guilt, given a guilty plea, across the full spectrum of base rates simultaneously.[45] In other words, we can observe the impact a guilty plea has on the posterior probability of guilt for each possible base rate. Importantly, in the absence of evidence, attorneys (both defense and prosecution) cannot accurately assess the probability that an accused person is *actually* guilty. Consequently, it is important to examine the impact prior probability could have on posterior probability to underscore the importance of a *reasonable* standard for evidence.

For the purposes of the current analyses, we relied on data from *Wilford, Sutherlan*d to inform plea acceptance rates for innocent and guilty participants.[46] We chose this study because of the design utilized (i.e., 2 [guilt status: innocent or guilty] x 3 [plea sentence: 6 months, 12 months, or 18 months] x 3 [conviction probability: 20%, 50%, or 80%] repeated-measures design), as well as the large and diverse sample recruited (i.e., 525 Prolific Academic participants and 596 student participants).[47] *Wilford, Sutherland*'s pattern of findings were also

---

[44] Gary L Wells et al., "Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence." (2020) 44:1 Law and Human Behavior 3–36, online: <https://doi.apa.org/doi/10.1037/lhb0000359> [*Wells et al*]; Sarah A Moody et al., "Evidence-based suspicion and the prior probability of guilt in police interrogations." (2023) 47:2 Law and Human Behavior 307–319, online: <https://doi.apa.org/doi/10.1037/lhb0000513> [*Moody*].

[45] Andrew M Smith, R C L Lindsay & Gary L Wells, "A Bayesian analysis on the (dis)utility of iterative-showup procedures: The moderating impact of prior probabilities." (2016) 40:5 Law and Human Behavior 503–516, online: <https://doi.apa.org/doi/10.1037/lhb0000196>.

[46] Wilford, Sutherland, *supra* note 1.

[47] Wilford, Sutherland tested the predictive power of an expanded shadow-of-the-trial model that incorporated guilt status. Participants were undergraduate students (half completed the study in-person and half completed the study online) and community participants recruited through Prolific Academic. Participants had to be U.S. residents who were 18 years of age or older. Community members had an average age of 30.9 years and were 51.8% male, 45.1% female, and 1.9% transgender or gender nonconforming. They were 65.5% White, 12.4% Asian, 7.6% Black, 5.9% Hispanic or Latinx, 5.9% bi- or multiracial, and 0.8% American Indian or Alaska Native. Student participants had an average age of 19.9 years and were 51.2% male, 45.8% female, and 1.5% transgender or gender nonconforming. They were 60.6% White, 14.6% Asian, 9.6% Black, 8.2% Hispanic or Latinx, 4.2% bi- or multiracial, and 0.2% American Indian or Alaska Native. The study employed two counterbalanced crime scenarios: a hit-and-run and a theft. Participants saw both scenarios (via an interactive computer simulation) and were randomly assigned one of the eighteen experimental conditions for each scenario. Participants started the study by giving consent and completing a demographics questionnaire. They then saw a simulated legal scenario of either the hit-and-run or theft: they were accused of the crime, summoned to court, where the prosecutor laid out the charges and the evidence, then remanded to a holding cell. A flashback then revealed to the

largely consistent with the extant plea literature (e.g., guilty participant-defendants were much more likely to accept the plea offer than innocent participant-defendants).[48] Using this data allowed us to ground our estimation of plea behavior in empirical research for which guilt status was known.

There are two noteworthy limitations of this approach. First, the exact plea acceptance rates from *Wilford, Sutherland* (and plea research broadly)[49] emanate from study-specific parameters (e.g., evidence was constant across conditions in this study) that may not generalize consistently across all aspects of criminal trials.[50] This limitation would be better addressed if (like *Wells*) we used a meta-analysis to inform our behavioral estimates.[51] However, to-date there is no meta-analysis of the plea research literature. Second, existing empirical research has not sufficiently captured the potential dynamism of plea decision-making: plea decisions are not assessed before and after changes to the case parameters (e.g., participant-defendants are typically not asked to accept or reject a plea offer before and after evidentiary discovery is shared). Regarding both limitations, changes in associated plea rates would necessarily produce differences in our empirical analysis. However, these experimental data combined with Bayesian analysis provide an important, initial demonstration of the diagnostic value of plea offers and the risk of false conviction via plea as a function of guilt status and prior probability of guilt. Thus, we strongly encourage future researchers to continue adopting a Bayesian approach with experimental data to further test the generalizability of these trends in plea contexts, as well as other legal contexts (e.g., interrogations).[52]

We used General Linear Mixed Modeling (GLMM) to estimate the log odds of plea acceptance (vs. rejection; e.g., *Wilford, Sutherland*).[53] The model evaluated the overall probability of plea acceptance as a function of guilt (vs. innocence; *G*) status (see Equation 1) while controlling for order effects (*O*), crime type (*C*), subpopulation (*S*), and within-subject effects (β_{0_i}).[54]

$$log_e\left(\frac{P(PA_{it})}{1-P(PA_{it})}\right) = \beta_{0_i} + \beta_1 * O + \beta_2 * C + \beta_3 * S + \beta_4 * G \qquad (1)$$

participant whether they were innocent or guilty. After, they met with their defense attorney who told them their conviction probability and the terms of the plea deal: plead guilty for 6/12/18 months in jail or risk a maximum of 24 months if convicted at trial. After making their plea decision, participants answered manipulation check questions, as well as subjective questions (e.g., how guilty they thought they were, their perceived probability of conviction). All data are available at https://osf.io/k9amw/files/

[48] Wilford, Sutherland, *supra* note 1.

[49] *Ibid*.

[50] *Ibid*.

[51] Wells, *supra* note 41.

[52] See Moody, *supra* note 44.

[53] Wilford, Sutherland, *supra* note 1.

[54] We did not run models in which the overall probability of plea acceptance was examined as a function of guilt, plea discount, and conviction probability. Instead, for the purpose of simplicity, the effect of the plea discount and conviction probability manipulations were collapsed across conditions.

These models were used to estimate the expected, condition-specific log odds of plea acceptance, which were then converted to condition-specific plea acceptance probabilities (see Equation 2) for experimentally manipulated guilt and innocence status (47.9% and 13.7% plea acceptance rates, respectively).

$$P(PA_{it}) = \frac{e^{log_e\left(\frac{P(PA_{it})}{1-P(PA_{it})}\right)}}{(1+e^{log_e\left(\frac{P(PA_{it})}{1-P(PA_{it})}\right)})} \quad (2)$$

We were then able to use experimentally observed conditional response rates to calculate the posterior probability of participant guilt (*G*). Specifically, we calculated the probability of participant guilt given their acceptance of a plea offer (*PA*), *P*(*G*|*PA*) (see Equation 3) using the experimentally derived, condition-specific probabilities of plea acceptance when participant-defendants were guilty, *P*(*PA*|*G*), or innocent (*NG*), *P*(*PA*|*NG*). As previously discussed, we do not know what the base rates of guilt, $P(G)$, and innocence,

$$P(PA) = \frac{P(PA|G)*P(G)}{P(PA|G)*P(G)+P(PA|NG)*P(NG)} \quad (3)$$

$P(NG)$, are in the population, but using this formula we estimated the posterior probability of guilt across the entire range of possible guilt base rates (0-100%). Once we estimated the posterior probability of guilt, we then calculated information gain about guilt probability as a function of plea acceptance by subtracting corresponding baseline probabilities of guilt from our posterior probabilities of guilt (see Equation 4).

$$Information\ Gain = P(G|PA) - P(G) \quad (4)$$

A. Comparing Two Plea Decision Points

Consequently, we were able to compare diagnosticity of plea acceptance (i.e., the posterior probability of guilt) across varying base rates (i.e., prior probabilities) of guilt. When considering plea diagnosticity, it is important to note that the base rate for (provable) guilt varies by both jurisdiction and timepoint in the legal process (e.g., from arrest to adjudication or dismissal). In these analyses we assume that case duration and evidence strength are related. While we acknowledge that the relationship between case duration and evidence strength is not entirely monotonic, we argue that generally, the longer a case survives the process, the more likely the accused person is to be *proven* guilty (and consequently, *plead* guilty). When a case is opened, only the prosecution has had the opportunity to acquire evidence; the defense typically begins building its case only after charges are filed. Thus, as time passes, the chances that the defense can raise motions to limit or suppress evidence, or even dismiss charges entirely, increase as they conduct their own investigations and/or eventually receive evidentiary discovery. As a result, the odds that weak or tenuous cases will drop out naturally increase. Further, because convictions can never outnumber charges, and because cases referred will never outnumber cases filed, we can assume that the cases being dropped as the system progresses are those less likely to conclude with a conviction. In other words, the number of cases that make it to pre-trial motions (for instance) will necessarily be smaller than

the number of cases that make it to a preliminary hearing. Thus, cases that persist through the legal procedure will most likely have a relatively higher prior probability of *demonstrable* guilt (again, we do not distinguish demonstrable prior probability of guilt from actual guilt); these are the cases that prosecutors have not dropped, and judges have not dismissed.

Consider a jurisdiction like Maricopa County, the base rate of guilt for accused persons being offered an initial plea could be relatively low due to their Early Disposition Courts. But, individuals who reject those initial pleas can still be offered subsequent plea deals, and the base rate of guilt for those persons could be significantly different from the base rate for those who accepted initial pleas; weaker cases are more likely to be dismissed or dropped with additional time for investigation. In other words, the base rate of guilt for first-round pleas versus second-round pleas, in the same jurisdiction, could be significantly lower due to changes in the pool of cases prosecutors choose to continue pursuing (versus dismissing). But, there will also be jurisdictional differences. In San Francisco County, for instance, case rejection rates (i.e., cases in which prosecutors choose not to file charges after an initial assessment or screening process) are relatively high (i.e., between 40-60% from 2017 to 2021; Prosecutorial Performance Indicators, 2022). Thus, one would expect that the base rate of guilt among those offered an initial plea in San Francisco County would be relatively higher than those offered an initial plea in Maricopa County (given the seemingly higher criterion, or higher confidence in conviction, for prosecutors to file charges).

A recent analysis of five years of cases from 15 United States prosecutor's offices found that approximately 28% of cases are rejected after initial screening, and another ~28% of those cases are eventually dismissed (Prosecutorial Performance Indicators, 2022).[55] Thus, about 51.8% of cases referred to prosecutor's offices (typically by law enforcement) are fully prosecuted (resolved at trial or by plea). In Maricopa County specifically, between 2019 and 2021, an average of only 36.7% of referred cases were fully prosecuted (excluding those still pending; Maricopa County Attorney's Office, AZ). That said, it is unclear whether cases routed to Early Disposition Courts go through the same initial screening as cases routed to other court systems. Thus, we believe a conservative estimate regarding the base rate among those offered early initial pleas, in jurisdictions like Maricopa County, is around 50%. Note again that in this context, 50% does not necessarily represent the base rate of *actual* guilt, but rather the base rate for those who could be *proven* guilty at trial (which we use as a proxy for *actual* guilt). Thus, we can use 50% to calculate the posterior probability of guilt given plea offer acceptance (see Equation 5) and the information gained from a plea decision (see Equation 6).

$$P(PA) = \frac{.479 * .50}{.479 * .50 + .137 * .50} = \frac{.2395}{.3080} = .778 \quad (5)$$

$$Information\ Gain = .778 - .50 = .278 \quad (6)$$

[55] Please note that these Prosecutorial Performance Indicators are only available from jurisdictions in which the District Attorney's office voluntarily opts into reporting the relevant measures. Thus, there could be self-selection biases that impact the results and trends observed.

When an accused person accepts a plea offer with a 50% prior probability (or base rate) of guilt, there is a corresponding increase of 27.8% of their likelihood of being guilty (to 77.8% total probability).

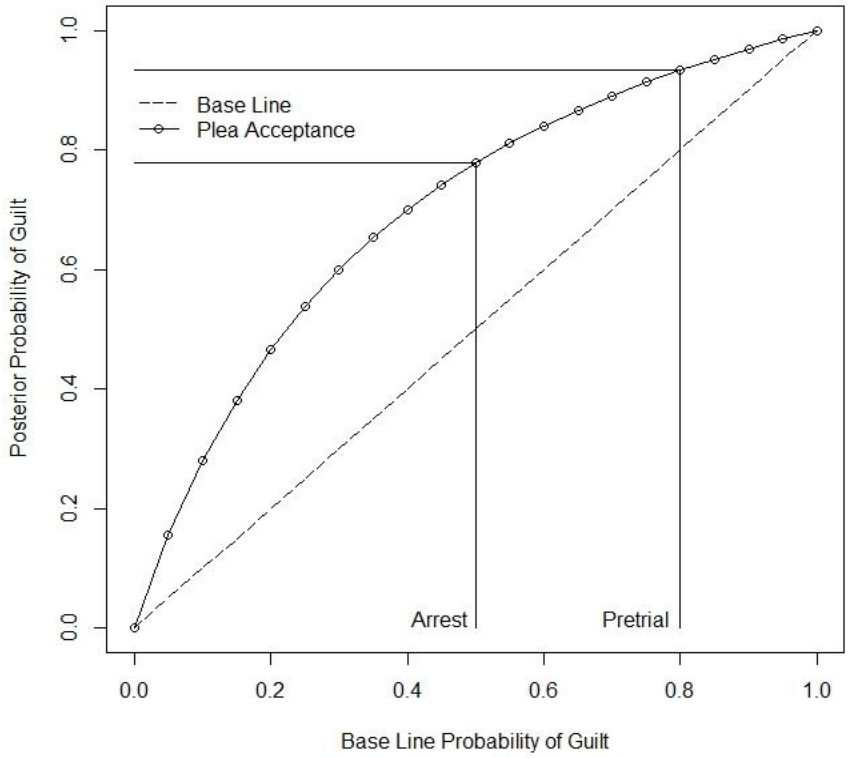
In contrast, the base rate for plea offers in more conservative jurisdictions (e.g., San Francisco County) or those extended later in the process (e.g., shortly before trial), might be closer to 80% (in FY 2019, ~20% of bench and jury trials resulted in non-convictions; *Federal Justice Statistics*, 2019).

$$P(PA) = \frac{.479 * .80}{.479 * .80 + .137 * .20} = \frac{.3832}{.4106} = .933 \quad (7)$$

$$Information\ Gain = .933 - .80 = .133 \quad (8)$$

In this example, the overall probability of guilt, if accepting a plea offer, is high (93.3%; see Equation 7), but information gain drops to 13.3% (see Equation 8). To further illustrate the impact that these base rates can have on the posterior probability of guilt, Figure 1 displays the prior-by-posterior guilt probability. The diagonal, dashed line indicates the posterior probability of guilt if no information was gained from plea acceptance, when the posterior probability of guilt would be equivalent to baseline probability of guilt. The solid curved line with circles indicates the posterior probability of guilt for any given baseline probability of guilt given an accepted plea offer. Finally, the intersecting vertical and horizontal solid lines indicate the posterior probability of guilt when baseline probability of guilt is 50% (e.g., at arrest) and 80% (e.g., shortly before trial); resulting in a 77.8% and 93.3% posterior probability of guilt, respectively.

Figure 1. Posterior probability of guilt as a function of plea acceptance.



Note. The dashed line indicates the posterior probability of guilty if no information was gained from plea acceptance. The solid line with circles indicates the posterior probability of guilt for any given baseline probability of guilt given a person accepted a plea offer. Vertical and horizontal solid lines indicate the posterior probability of guilt when baseline probability of guilt is 50% (at arrest) and 80% (shortly before trial); 77.8% and 93.3% posterior probability of guilt, respectively.

B. Everyone is Blind Without Evidence

The practice of coercing accused persons to accept pleas immediately (e.g., after arrest) is worrying for several reasons. The first being that an arrest can then serve as sufficient evidence to threaten an accused person with an immediate criminal conviction. Notably, the standard of proof for arrest (and most initial phases of prosecution) is "probable" cause, and the determination of probable cause is typically one-sided. Although a judge is required to agree that probable cause exists for an arrest, they have no resources by which to investigate the State's claims at these early phases. Thus, they are likely to defer to the opinion of law enforcement.

A system that can entice individuals to plead guilty as soon as they are accused of a crime, when the only burden the State has met is probable cause, looks like a system presuming *guilt*, not *innocence*. In such jurisdictions, it appears that the presumption of innocence is yet another Constitutional right cloaked in one's right to a trial; a presumption that our more efficient *system of pleas* cannot tolerate. Once accused, the State need only convince an individual to accept a plea offer and its burden of proof has been met.

We can clearly observe the impact of policies such as these: as the base rate of guilt increases (from 50%), information gain decreases (see Figure 1). In other words, there is a trade-off between prior probability of guilt and information gain when treating guilty pleas as diagnostic of guilt. When we rely more on plea outcomes (rather than evidence) to conclude an individual is guilty, we increase the information gained from a guilty plea at the increased risk of false guilty pleas. But, guilty pleas are not evidence, they are convictions.[56] Thus, ideally (as originally envisioned in *Brady*),[57] the system would already be confident in one's guilt prior to offering a guilty plea.

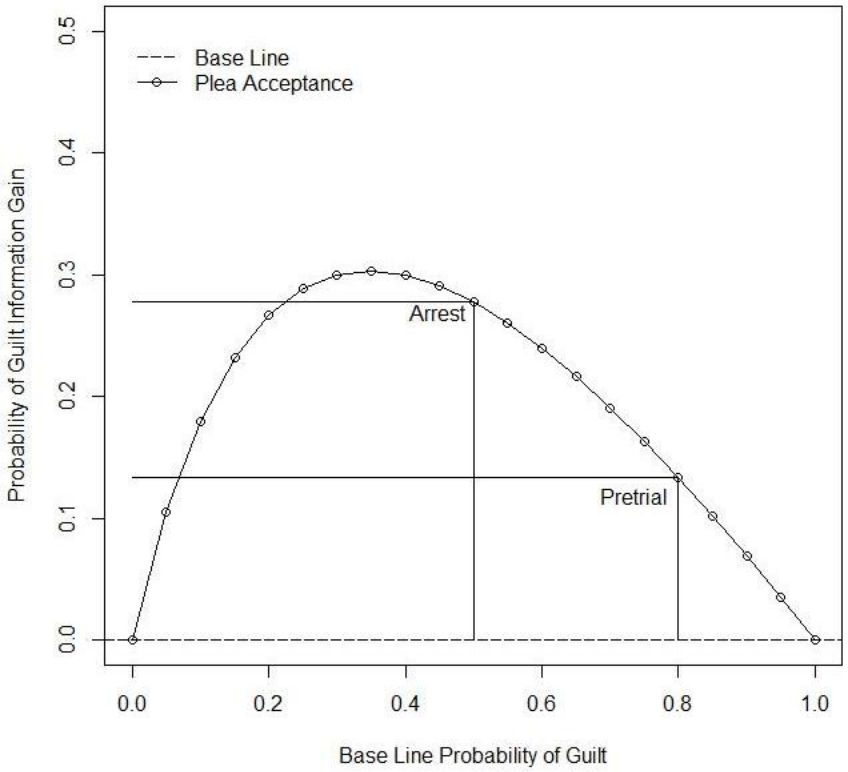
Consequently, it is important to examine the information gained from a guilty plea at various potential points in the legal process across different jurisdictions—not to maximize information gain, but to question whether the plea outcome is replacing the role of evidence in adjudications. To further examine the impact of prior guilt probability on the information gained from the outcome of a plea offer, we constructed an information-gain curve (see Figure 2). The dashed horizontal line indicates information gained (none) if plea acceptance does not provide additional information concerning the probability of guilt, while the solid line with circles indicates the relative information gained about guilt status as a function of plea acceptance and baseline probability of guilt. Similar to Figure 1, the intersecting solid vertical and horizontal lines indicate

[56] *Boykin v Alabama,* 1969 395 U.S. 238; and Miko M Wilford & Gary L Wells, "Bluffed by the dealer: Distinguishing false pleas from false confessions." (2018) 24:2 Psychology, Public Policy, and Law 158–170, online: <https://doi.apa.org/doi/10.1037/law0000165>.

[57] *Brady*, *supra* note 5.

the information gained when baseline probability of guilt is 50% (e.g., at arrest) and 80% (e.g., shortly before trial).

Figure 2. Probability of guilt information gained as a function of plea acceptance

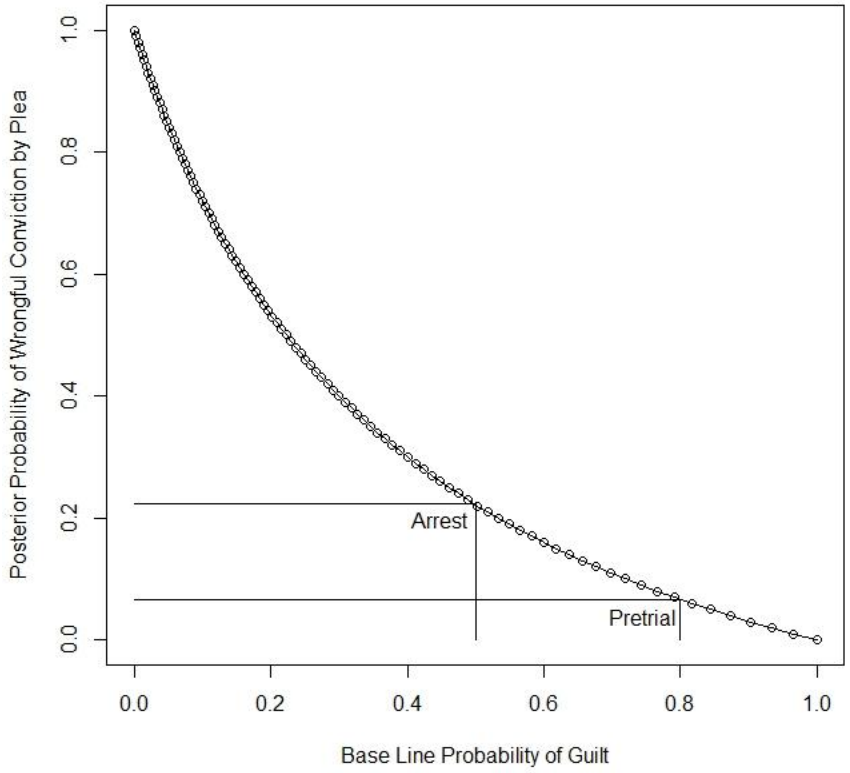


Note. The dashed line indicates the information gained about guilt probability if no information was gained from plea acceptance. The solid line with circles indicates the information gained about guilt probability for any given baseline probability of guilt given plea offer acceptance. Vertical and horizontal solid lines indicate the information gained about guilt probability when baseline probability of guilt is 50% (at arrest) and 80% (shortly before trial); 27.8% and 13.3% increased probability of guilt, respectively.

We observed that information gain from plea acceptance is greatest (30.3%) when the prior probability of guilt is 35%, and that information gain is approximately 20% or higher for prior probability of guilt ranging from approximately 12 to 68%. Overall, Figure 2 indicates that plea acceptance is informative of guilt status for a wide range of prior probability of guilt. However, returning to Figure 1, we inverted the posterior probability of guilt trajectory to become a posterior probability of wrongful conviction (see Figure 3). Figure 3 shows that across that same range of prior probability of guilt (12-68%), the percent of plea acceptance by the innocent ranges from approximately 68 to 12%. Taken together, Figures 1-3 illustrate that when the State extends plea offers too early in the process (i.e., when the probability of case dismissal is greatest and the likelihood of acquittal would be highest), innocent people will necessarily be caught; sometimes at a rate greater than guilty persons (i.e., when prior probability of guilt is 22% or less). When guilty pleas are the primary (or only) piece of evidence against the accused persons, their veracity should be questioned. While these results rely on one study of plea decision-making and should

be replicated via additional plea experiments manipulating guilt status, they nonetheless highlight grave concerns about our current *system of pleas*.

Figure 3. Probability of wrongful conviction as a function of plea acceptance



Note. The solid line with circles indicates the posterior probability of wrongful conviction by plea for any given baseline probability of guilt given a person accepted a plea offer. Vertical and horizontal solid lines indicate the posterior probability of wrongful conviction by plea when the baseline probability of guilt is 50% (at arrest) and 80% (shortly before trial); 22.2% and 6.7% probability of wrongful conviction, respectively.

Some may argue that the diagnostic value of plea decisions presented in this paper is justification for their use, particularly in cases where prosecutorial confidence is low. As has been regularly emphasized, the legal system is overburdened—we should, therefore, preserve limited resources as much as possible. Further, the more cases that must be tried, the longer accused persons will wait before their case is tried (thus, threatening their Constitutional right to a speedy trial). So, why not let them identify themselves as guilty as early in the process as possible, particularly given evidence that in those circumstances, plea acceptance provides greater confidence in *actual* guilt status?

Unfortunately, as is indicated by the extant body of experimental plea research,[58] and this Bayesian analysis, the diagnostic value of pleas is inexorably intertwined with an unacceptable wrongful conviction rate, except at the highest prior probabilities of accused persons' guilt. While previous studies have shown that innocent people accept plea offers, none have shown how relatively weak their correspondence to actual guilt can be (in relation to prior probability of guilt). These results clearly undermine the assumption many legal actors must be making with regard to the diagnostic value of guilty pleas.[59] When significant incentives are offered to accused persons for a guilty plea, they will plead guilty (whether actually guilty or not). Using guilty pleas to inform the system of accused persons' guilt (rather than evidence) will increase the information gained from guilty pleas; but, this increase in information gain is necessarily linked to an increased risk for false guilty pleas.

More importantly, there is an inherent problem with using plea acceptance as a diagnostic test of guilt; acceptance of a plea offer cannot be used as evidence given that it serves as conviction (not as evidence). Given the problematic nature of posthoc justifications for conviction and sentencing and the increasingly high rate of wrongful convictions, solicitation of early plea offers (e.g., prior to discovery) is legally unjustifiable. Further, as the [Supreme] Court has clearly acknowledged, weak cases should not conclude with a conviction (even if that conviction carries a minimum sentence), they should conclude with a dismissal.

III Policy Recommendations

In light of the analyses presented in this manuscript and our review of today's *system of pleas*,[60] we conclude by offering a few recommendations that would unquestionably reduce the rate of false guilty pleas. First, plea offers should not be extended to accused persons until the State can meet a *reasonable* standard for guilt. Although we will not attempt to articulate what this standard should be specifically, we will refer back to Figure 3, which illustrates the posterior probability of wrongful conviction across base rates of guilt. In a jurisdiction in which accused persons are offered pleas immediately after arrest (when prior probability of guilt is lower; e.g., 50%), approximately 22.2% of those accepting pleas are innocent; in contrast, a jurisdiction that pushes pleas closer to trial (when probability of guilt is presumed to be greater; e.g., 80%) will result in approximately 6.7% of those accepting pleas being innocent. Although these exact percentages are based on estimates of baseline guilt probability, they clearly demonstrate that the higher the prior probability that an accused person is guilty, the more diagnostic of guilt the guilty plea will be (and consequently, the fewer innocent people who will be swept up in the process). This recommendation is very similar to one articulated in the American Psychology-Law Society's most recent scientific review paper outlining ideal "Policy and Procedure Recommendations for

[58] Helm & Reyna, *supra* note 22; Allison D Redlich & Reveka V Shteynberg, "To plead or not to plead: A comparison of juvenile and adult true and false plea decisions." (2016) 40:6 Law and Human Behavior 611–625, online: <https://doi.apa.org/doi/10.1037/lhb0000205>; and Wilford Sutherland, *supra* note 1.

[59] *Cassel*, *supra* note 34.

[60] Vanessa A Edkins & Allison D Redlich, eds, *A system of pleas: social science's contributions to the real legal system*, American Psychology-Law Society series (New York, NY: Oxford University Press, 2019), online: <https://doi.org/10.1093/oso/9780190689247.001.0001>.

the Collection and Preservation of Eyewitness Identification Evidence".[61] Specifically, the paper includes a recommendation that law enforcement have an "evidence-based suspicion" prior to putting a suspect in a lineup. If some probability of guilt is important to protect a suspected person from possible wrongful *identification*, we believe a higher probability of guilt is critical to protect a suspected person from possible wrongful *conviction*.

Second, as soon as prosecutors are confident enough in their case to offer a plea deal, they should also be sufficiently confident to share their case file. In other words, all jurisdictions should be open-file as soon as a plea offer is on the table. Allowing closed-file policies makes it too easy for prosecutors to avoid disclosing potentially exculpatory information to accused persons when attempting to get a guilty plea.[62] Consequently, cases survive longer than the evidence supports (artificially inflating case duration). We cannot assume a reasonable standard of guilt has been met unless case evidence is made transparent. Similarly, we would call for additional efforts to collect real-world data to open the "black box" of the plea process (such as the Plea Tracker Project housed at the Wilson Center for Science and Justice at Duke Law School). For instance, we know relatively little regarding how many offers accused persons typically receive and what evidence they are provided prior to each offer.[63]

And finally, if these recommendations burden the system (as has been previously alleged), then: 1) put more money into courtrooms, and/or 2) stop criminalizing so many behaviors that the courts are flooded with so-called criminals. Each year federal and state governments increase spending on law enforcement.[64] However, most of that money goes toward hiring more police officers, building new jails, and funding prosecutors' offices.[65] While more money is going toward arresting and incarcerating people, judicial staffing has only increased 11%, and public defense staffing has only increased by 4%. Further, several courts are being under-utilized,[66] with some courts recording guilty plea rates of 100%.[67] As plea deals are increasing, judges are going long periods of time without trying a case in a courtroom; clerks are leaving clerkships without any trial experience. Instead of spending large portions of the day in a courtroom, judges are saying they

---

[61] Wells *et al*, *supra* note 44.

[62] Samantha Luna & Allison Redlich, "Unintelligent Decision-Making? : The Impact of Discovery on Defendant Plea Decisions" (2020) 1:3 The Wrongful Conviction Law Review 314–335, online: <https://wclawr.org/index.php/wclr/article/view/24>.

[63] Allison D Redlich, Miko M Wilford & Shawn Bushway, "Understanding guilty pleas through the lens of social science." (2017) 23:4 Psychology, Public Policy, and Law 458–471, online: <https://doi.apa.org/doi/10.1037/law0000142>.

[64] "Office of Public Affairs | Justice Department Announces $139 Million for Law Enforcement Hiring to Advance Community Policing | United States Department of Justice", (18 November 2021), online: <https://www.justice.gov/opa/pr/justice-department-announces-139-million-law-enforcement-hiring-advance-community-policing>.

[65] Hessick, *supra* note 31.

[66] *The Trial Penalty: The Sixth Amendment Right to Trial on the Verge of Extinction and How to Save It* (National Association of Criminal Defense Lawyers, 2018), online: <https://www.nacdl.org/Document/TrialPenaltySixthAmendmentRighttoTrialNearExtin> [*NACDL*].

[67] *Redlich 2022*, *supra* note 6.

now spend all their time in their chambers.[68] Thus, courtrooms are not currently overburdened, at least not universally.[69]

Our recommendations dovetail with those published recently in the American Bar Association's *Plea Bargain Task Force Report*.[70] Notably, this task force included representatives from both adversarial sides of the system. The *Report* includes fourteen principles designed to guide future plea policies—two of which are particularly relevant to the current work. Principle Four emphasizes the importance of prosecutorial charging decisions and specifically recommends that, "The prosecutorial mindset should not focus on what the prosecutor can charge, but rather what the prosecutor should charge in light of the **evidence** and interests of justice… Prosecutors should dismiss weak cases rather than seek to resolve them through plea bargaining".[71] Principle Eight focuses on issues relating to discovery recommending that, "Defendants should receive all available discovery, including exculpatory materials, prior to entry of a guilty plea, and should have sufficient time to review such discovery before being required to accept or reject a plea offer".[72] Overall, this Task Force's recommendations also overlapped in many ways with those from the National Association of Criminal Defense Lawyers[73] and Fair and Just Prosecution. Clearly, more and more legal actors are becoming wary of quick pleas and their role in adjudicating cases.

---

[68] Benjamin Weiser, "Trial by Jury, a Hallowed American Right, Is Vanishing", *The New York Times* (8 August 2016), online: <https://www.nytimes.com/2016/08/08/nyregion/jury-trials-vanish-and-justice-is-served-behind-closed-doors.html>.

[69] Regarding increasing decriminalization, a study conducted by the Department of Justice found that states that decriminalized marijuana had substantially fewer marijuana-related arrests and court cases (Farley & Orchowsky, 2019). For example, there was a 90% reduction in Massachusetts and an 86% reduction in California (Neil & Martin, 2015). There have been concerns that legalizing marijuana would result in increased collateral financial costs, such as fatalities as a result of DUIs (Ahrens, 2020). These concerns have not been realized in states that have legalized marijuana; alcohol remains the primary concern for DUIs. In 2015, marijuana arrests accounted for 1-6% of all arrests in Oregon, Colorado, and Washington–three states that have since legalized marijuana. Those 1-6% of arrests required 3 to 4.5 million dollars of each states' budget for policing, correctional, and judicial funding (Miron, 2018). At the federal level, legalizing marijuana is seen as another step toward remedying the negative effects of the War on Drugs, which disproportionately affected Black people. A Black person is more than 3x more likely to be arrested for marijuana possession compared to a White person despite comparable usage rates (American Civil Liberties Association, 2020). In fact, this was one of the reasons President Biden cited for pardoning all individuals charged with simple marijuana possession under federal law in October 2022 (Kanno-Youngs, 2022). Perhaps America should take a page from the Netherlands' approach to criminal behaviors: Dutch culture includes a conception of tolerance that promotes more leniency and selectivity in the prosecution of crimes (Buruma, 2007).

[70] Johnson, *supra* note 40.

[71] *Ibid*, at 18.

[72] *Ibid*, at 24.

[73] NACDL, *supra* note 66.

A. Conclusion

Our current *system of pleas* allows prosecutors to acquire convictions with very little proof of guilt. For decades, U.S. courts have protected plea-bargaining practices by arguing that accused persons are free to make the choice that best serves their own interest. Many legal actors have further justified the practice by presuming that guilty pleas are highly indicative of guilt.

Several amicus briefs were filed to the Supreme Court concerning *Mansfield v. Williamson County, Texas*.[74] The issue in the case pertained directly to discovery policies during pretrial plea negotiations. Specifically, the petitioner (Troy Mansfield) alleged that Williamson County's closed-file policy violated his Fourteenth Amendment right to due process. Prosecutors in his case arguably possessed evidence of *factual innocence* (i.e., even stronger than exculpatory-level evidence) at the time of plea negotiations and did not disclose it while pressuring Mansfield to accept a plea offer. Consequently, Mansfield accepted a sentence of 120 days (rather than risk a potential life sentence). Unfortunately, the Supreme Court refused to hear oral arguments in the case (denying the petition for certiorari); thus, the question as to what evidence (if any) must be disclosed prior to a plea agreement remains unanswered by the highest American court. In other words, the Court has refused to address how accused persons can be expected to make decisions that serve their interests in the absence of evidence. The current analysis has further demonstrated the dangers of a system that replaces substantive evidence with guilty pleas; it also directly undermines assumptions regarding the diagnosticity of a guilty plea. Without a reasonable demonstration of guilt, the system should not be permitted to use its wealth of resources to pressure (plea) convictions directly from accused persons.

[74] *Mansfield v. Williamson County*, 2022 22 U.S. 186.