



Where is the Evidence? Comparing the Effects of Evidence Strength and Demographic Characteristics on Plea Discounts

Kevin Petersen¹ · Allison D. Redlich¹ · David B. Wilson¹

Accepted: 26 July 2022 / Published online: 13 August 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Objectives It is well established that defendants who plead guilty receive reduced sentences compared to the likely outcome if convicted at trial. Prominent theories of plea bargaining posit that the plea discount is determined by the strength of the evidence against the defendant. Research on this claim has produced mixed findings, however, and others have suggested that discounts may be influenced by extra-legal characteristics such as race, age, and sex. To date, there have been few attempts to directly compare the effects of these factors on plea discount estimates.

Methods This study uses a penalized ridge regression to predict counterfactual trial sentences for a sample of defendants who pled guilty. Plea discounts are estimated using each defendant's predicted trial sentence and observed plea sentence. Discount estimates are then regressed on variables related to case evidence and the demographic characteristics of the defendant.

Results Results suggest that increases in the amount of evidence associated with a case lead to decreases in the size of the plea discount. Both main and interaction effects are observed for race/ethnicity and sex, with Hispanic and male defendants receiving significantly smaller discounts than White or female defendants. Calculation of standardized effect sizes further indicates that demographic characteristics exert larger effects on plea discount estimates than evidentiary variables.

Conclusions Plea discounts appear to be influenced by both evidence and extra-legal factors. Legal participants may indeed consider the strength of the evidence when determining acceptable plea discounts, but this alone appears to be an insufficient explanation.

Keywords Plea discount · Trial penalty · Strength of evidence · Guilty pleas · Sentencing

Introduction

In an era where the criminal justice system is largely defined by the practice of plea bargaining (see Lafler v. Cooper 2012), there has been increasing attention given to the factors that lead defendants to plead guilty. Central to these theories is the notion

✉ Kevin Petersen
kpeter4@gmu.edu

¹ Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA

that plea takers receive substantial discounts in exchange for their guilty pleas. That is, a defendant who pleads guilty often receives a discounted jail or prison sentence compared to what they would have received if convicted at trial (see Alschuler 1981; Brereton and Casper 1981; Kipnis 1976). Empirical research has generally supported this claim (see Abrams 2011 for a noteworthy exception), finding that average plea sentences are significantly shorter than average trial sentences, even after controlling for a host of legal and extra-legal factors thought to be salient components of sentencing decisions (Bushway and Redlich 2012; LaFree 1985; Smith 1986; Ulmer and Bradley 2006; Ulmer et al. 2010; Yan and Bushway 2018). However, this research has also found that plea discounts tend to display significant variation across individuals, with some defendants receiving discounts approaching 100% and others receiving plea sentences that are longer than would be expected if convicted at trial (Bushway and Redlich 2012; Smith 1986; Yan and Bushway 2018). Despite the consistency of the ‘trial penalty’ effect, it is still relatively unclear how the magnitude of the plea discount is determined and what factors correspond to larger or smaller discounts.

A widely held belief is that plea discounts are proportional to the strength of the case, or in other words, the “strength of the evidence” (Brereton 1981, p. 50) against the defendant (see also LaFree 1985; Landes 1971; Smith 1986). Though logical in theory, research on the impact of evidentiary factors on plea discounts remains rare, and the research that does exist is largely inconclusive. While some studies find that inculpatory evidence either produces no effect or significantly *increases* the size of the plea discount (see Bushway and Redlich 2012), other studies find that evidentiary factors account for small but significant *decreases* in plea discounts (Redlich et al. 2016; Smith 1986). Complicating matters is the inherent difficulty involved in measuring and obtaining real-world information on case evidence, which is often missing from administrative data sets (Kutateladze et al. 2015). Accordingly, much remains unknown about the relationship between the strength of the evidence and plea discounts. If evidentiary factors strongly predict the size of the plea discount, then legal participants would be expected to make rational plea decisions based on their predicted trial outcomes (see Bibas 2004; Bushway and Redlich 2012). If not, the implication is that plea discounts may be influenced by other (possibly extra-legal) factors that may create bias against certain groups of defendants (see generally Steffensmeier et al. 1998; Ulmer et al. 2010). Indeed, there is a highly consistent body of research demonstrating that minority defendants are disadvantaged throughout the criminal justice system, including at the plea phase (see Kurlychek and Johnson 2019). Given the centrality of plea discounts to the practice of plea bargaining, understanding the factors that affect these discounts remains an important area of criminological research.

In this study, we seek to provide further insight into the relationship between evidence strength, extra-legal characteristics, and plea discounts. In doing so we build on prior efforts in several ways. First, we analyze novel data sets that provide recent estimates of plea and trial sentence lengths. Second, we use unique combinations of evidentiary variables systematically coded from observations of case summaries (prosecutorial proffers of evidence) presented during plea hearings. Third, we directly compare the effects of legal and extra-legal variables in an attempt to determine the primary drivers of plea discounts. In what follows, we first discuss the prior literature on plea discounts and the factors that may influence them. We then turn to our data, methods, results, and the implications of our study for plea bargaining research moving forward.

Plea Discounts

The sentencing differential observed between defendants convicted at trial and defendants who plead guilty has been referred to by several names, most notably the “plea discount” or “trial penalty” (Kim 2015, p. 1200; Ulmer et al. 2010; Yan and Bushway 2018). While meaningful differences exist between these two terms (see Grunwald 2020; Yan and Bushway 2018), the plea discount is most often defined as the difference between the sentence that a defendant would have received if convicted at trial and the sentence they received as part of their plea, relative to the sentence they would have received if convicted at trial (see Yan and Bushway 2018; see also Redlich et al. 2016).¹ Prior estimates have suggested that approximately 80–90% of defendants receive a plea discount (Bushway and Redlich 2012; Redlich et al., in press), and while it is not possible to directly observe the sentence that a plea bargainer would have received if convicted at trial, researchers have used several different approaches to estimate the average difference between plea and trial sentence lengths and the size of the resulting discount that plea bargainers may receive.

Criminologists have often studied the discrepancy between plea and trial sentences through the estimation of regression models that incorporate adjudication type as an independent variable (for a discussion of this approach see Bushway et al. 2014). These models seek to determine the average difference in sentence lengths comparing cases that end in trial conviction to those disposed of via guilty plea, while controlling for various legal and extra-legal factors. Generally, these efforts find significant but variable differences between plea and trial sentence lengths (Elder 1989; LaFree 1985; Ulmer and Bradley 2006; Ulmer et al. 2010). Using data on serious violent offenses in Pennsylvania, Ulmer and Bradley (2006) found that bench and jury trial convictions increased sentence lengths by 22% and 57% over guilty pleas, controlling for defendant, case, and court-level characteristics. Similarly, Ulmer et al. (2010) found that trial convictions were associated with a 45% increase in sentence lengths for federal offenders, which given a baseline average sentence of 62 months, equated to a plea discount of approximately 31%.²

Others have attempted to estimate the size of plea discounts more directly, either through the use of regression models to predict counterfactual trial sentences for those who pled guilty (Bushway and Redlich 2012; Smith 1986; Yan and Bushway 2018; Yan 2020; see also Piehl and Bushway 2007) or by comparing observed plea sentences to the maximum statutory penalty that a defendant could have received if convicted at trial (Redlich et al., in press; Zottoli et al. 2016). In either case, researchers have consistently found support for the presence of a plea discount. For instance, both Smith (1986) and Bushway and Redlich (2012) estimated the probability of an incarceration sentence if convicted at trial for defendants who pled guilty. In both studies, the observed probability of incarceration for defendants who pled guilty was roughly 75% of the estimated probability of incarceration if convicted at trial, suggesting a discount of approximately 25% in the probability of incarceration for those who pled guilty. Using a similar

¹ Here we focus on sentencing discounts, however, discounting practices can take several forms, including changes in the severity or number of charges that a defendant faces in exchange for a guilty plea (Kelly and Pitman 2018) or bargaining over the facts of the case (Redlich et al. 2017).

² While Ulmer et al. (2010) do not directly report this estimate, they do report a grand mean sentence of 62 months and suggest that the trial penalty would be approximately 28 months. This would correspond to a trial sentence of roughly 90 months and a plea discount of 31%. This discount decreased when Ulmer et al. controlled for various forms of guideline departures, though it remained statistically significant.

approach to predict sentence lengths rather than incarceration probabilities, Yan and Bushway (2018) estimated plea discounts across 40 distinct crime types, finding an average 67% discount between predicted trial sentences and observed plea sentences (see also Yan 2020). Other studies have reported average plea discounts as high as 75–80% when considering the maximum penalty that a defendant could have received if convicted at trial (see Redlich et al. 2018; Zottoli et al. 2016), or even as high as 95% when considering active (rather than suspended) sentences alone (Redlich et al., in press). Thus, the distinction between the maximum potential trial sentence and the expected (or predicted) trial sentence has meaningful implications for the size of plea discount estimates, though the former is likely an over-estimate given that most defendants do not receive the statutory maximum penalty (see Reitz 1993; Tonry 1988).

An important codicil to the findings of these studies concerns the considerable variance in plea discount estimates that has been observed both within and across samples. While Bushway and Redlich (2012) found that the average ratio between the probability of incarceration following a guilty plea and the predicted probability of incarceration at trial was approximately 77%, this ratio ranged from a low of 2% to a high of over 200% across defendants. Yan and Bushway (2018) also found that plea discounts varied significantly across crime types, ranging from an average discount of less than 4% to an average discount of roughly 90%, with several crime types having predicted trial sentences that were less than their average plea sentences (for similar findings at the defendant-level see also Yan 2020). Likewise, Redlich et al. (2018) reported that 22% of defendants in their sample did not receive any plea discount, and that those who did received discounts ranging from 27.27% to 99.72% of their statutory maximum sentence.

What Drives the Plea Discount?

Understanding the variability in plea discount estimates is critical because the plea discount is one of the key mechanisms through which prosecutors are believed to incentivize or even “coerce” (Kipnis 1976, p. 97) defendants into pleading guilty. That is, the assurance that a guilty plea will result in a sentencing reduction when compared to the likely outcome if convicted at trial provides motivation for defendants to plead guilty. Likewise, the assurance that a plea bargain will secure a conviction provides prosecutors with an opportunity to avoid the costs and inherent risk of an acquittal at trial (see generally Alschuler 1981; Brereton and Casper 1981; Kipnis 1976).

A necessary component of this relationship, however, is that there must exist some perceived risk of a conviction or an acquittal if a case is taken to trial. If there is little to no risk of conviction, a defendant may be unwilling to plead guilty entirely, or a large discount may be required to convince them to do so. Similarly, if there is little to no risk of being acquitted at trial, a prosecutor may be either unwilling to offer a plea discount or only willing to offer a small discount relative to the expected trial sentence (see Alschuler 1968; Bibas 2004). From a strictly rational perspective, it then follows that the level of risk associated with taking a case to trial should dictate the level of incentive (i.e., the size of the discount) needed to reach a plea agreement. This level of risk has often been referred to as the “probability of conviction at trial” (Bushway and Redlich 2012, p. 438), and central to this concept is the strength of the evidence that the

prosecution holds against the defendant (see Bibas 2004; Bushway et al. 2014; Landes 1971; Rhodes 1979).

Evidence Strength

Evidence strength, or “the quantity and quality of the evidence presented by the plaintiff/prosecution” (Devine et al. 2001, p. 684), is a frequently cited explanation for variations in the probability of conviction at trial and the size of the plea discount. One of the leading theories of plea bargaining, the “shadow of the trial model” (Bushway and Redlich 2012, p. 437), directly connects these concepts by positing that the strength of the evidence impacts the size of the plea discount through its effect on the perceived probability of conviction at trial (see Bibas 2004). More specifically, as evidence strength increases so too should the probability of conviction at trial, which in turn should produce plea discounts that become proportionally smaller relative to the expected trial sentence (see Bartlett and Zottoli 2021; Bushway et al. 2014; Landes 1971; Mnookin and Kornhauser 1979; Petersen et al. 2020). In support of this proposition, research has indicated that the probability of conviction at trial, as perceived by legal participants, is highly responsive to the strength of the evidence in a case (Bushway et al. 2014; Elder 1989). Thus, evidence strength has often been used as a proxy for probability of conviction in legal scholarship (see Kramer et al. 2007; McAllister and Bregman 1986).

Indeed, research suggests that evidence strength is an important predictor of numerous court outcomes. Cases with stronger evidence are significantly more likely to be prosecuted (Albonetti 1987; Frederick and Stemen 2012; McCoy et al. 2012; Schmidt and Steury 1989; Spohn and Holleran 2001) and result in guilty trial verdicts (Bushway and Redlich 2012; Eisenberg et al. 2005; Devine et al. 2009; Taylor and Hosch 2004). Prior studies also suggest that evidence strength influences the likelihood of a guilty plea/plea offer (see Elder 1989; Emmelman 1998; Henderson 2021; Kramer et al. 2007; Kutateladze et al. 2015; Luna and Redlich 2020), and even the severity of the sentence received as part of a plea offer (Elder 1989; LaFree 1985). However, limited attention has been given to the direct relationship between evidence strength and plea discounts, despite this relationship being a core component of prominent plea-bargaining theories such as the shadow of the trial.

Using data collected by Miller et al. (1980), Smith (1986) examined the impact of various pieces of evidence on plea discounts while accounting for jurisdictional, case, and defendant characteristics. Results indicated that, while the presence of physical evidence and the number of witnesses in a case significantly decreased plea discounts, the presence of eyewitness evidence led to significant increases in the size of plea discounts. Additionally, evidentiary factors appeared to be less important in predicting the size of the plea discount than factors such as jurisdiction, the defendant’s probation/parole status, and whether the defendant had a drug history or pending charges. Using a similar design, Bushway and Redlich (2012) failed to find any significant effects of physical or eyewitness evidence on plea discounts, however, they did find that the number of witnesses in a case and the presence of confession evidence significantly increased plea discounts by 1% and 12%, respectively. These results were in light of the finding that eyewitness, confession, and physical evidence all significantly increased the probability of conviction at trial in their models.³

³ LaFree (1985) and Elder (1989) examined the impact of evidence on plea sentences but not plea discounts. While LaFree (1985) found physical evidence and number of witnesses to significantly increase plea sentences, confession evidence had no effect across all model specifications. Additionally, Elder (1989)

Other studies report findings that are more suggestive of a consistent (albeit limited) impact of evidence strength on plea discounts. Redlich et al. (2016; see also Bushway et al. 2014) experimentally manipulated combinations of confession, witness identification, and DNA evidence among a sample of legal actors given a hypothetical case scenario and asked to make plea recommendations. Findings indicated that all three pieces of evidence produced small but significant decreases in the size of the discounts that participants were willing to offer (often only between 4 and 6%). In addition, Redlich et al. (2018) coded district attorney cases files for the presence/absence of confession evidence, finding that defendants who denied their allegations received significantly larger plea discounts than those who both confessed and partially confessed to the crime, though the effect sizes for these differences were considered small by conventional standards. Relatedly, Kutateladze et al. (2015) found that evidentiary factors such as the presence of video/audio recordings and whether pre-recorded buy money or currency were recovered during an investigation significantly increased the likelihood of a custodial plea sentence offer for drug offenses. While these analyses did not directly examine plea discounts, custodial plea sentence offers are, by definition, smaller discounts than non-custodial plea sentence offers. However, after controlling for court, defendant, and other case characteristics, only the effect of recovered currency remained a significant predictor of custodial plea offers.

Interpreting the results of these studies is further complicated by inconsistent measurements of evidence strength (see generally Devine et al. 2001). Factors such as “theoretical ambiguity” and the “lack of an accepted metric” have made defining this concept difficult (see Devine et al. 2001, p. 686), which is further compounded by the lack of evidentiary variables typically found in administrative data sets (Kutateladze et al. 2015). As a result, prior efforts to measure evidence strength using real-world data have often relied on several different approaches. One approach, common in the plea discount literature, involves the analysis of individual pieces of evidence in isolation. This contrasts with the cumulative approach often used in research on jury verdicts, where individual pieces of evidence are aggregated or Likert-type scales are used to represent evidence strength more holistically (see Eisenberg et al. 2005; Garvey et al. 2004; Taylor and Hosch 2004). For example, Taylor and Hosch (2004) considered the presence of physical evidence, confession, positive eyewitness identification, weapon, number of indictment charges, degree of injury to the victim, and number of witnesses. The end measure of evidence strength was a summary variable of these factors, which was significantly related to both jury verdicts and trial sentence lengths.

While these approaches may seem at odds, it is important to consider both the aggregate amount of evidence in each case along with the individual pieces of evidence in isolation. At the individual level, different forms of evidence are likely to have varying levels of influence, and this insight may be lost when using aggregate measures of evidence alone. Indeed, research has indicated that legal participants value certain kinds of evidence over others. For instance, direct evidence, or evidence that “proves a fact without an inference or presumption” (Heller 2006, p. 248), such as confessions and eyewitness testimony, is consistently overvalued by judges and jurors (Devine et al. 2001; Heller 2006; Kassin and Neumann 1997; Niedermeier et al. 1999; Wells 1992). In contrast, circumstantial evidence,

Footnote 3 (continued)

found no significant effect of physical evidence, identification evidence, or confession evidence on plea sentence length. However, neither study directly examined the relationship between evidence and plea discounts.

or evidence “from which the fact-finder can infer whether the facts in dispute existed or did not exist” (Heller 2006, p. 250), such as forensic/medical evidence, is often undervalued (Kassin 2012; Koehler 2001; Schklar and Diamon 1999). Even within these broad categories, however, evidence in specific forms may carry different levels of importance. Interviews with New York District Attorneys conducted by Kutateladze et al. (2015) suggested that video/audio recordings and eyewitness testimony may be particularly salient types of direct evidence, while others have suggested that the presence of a confession alone may be powerful enough to render other forms of evidence unnecessary (see Kassin 2012; McCormick 1972).

Yet, there is both a quality and quantity component to evidence strength, and individual forms of evidence do not often exist in a vacuum (see Devine et al. 2001, 2009).⁴ As such, examining pieces of evidence in isolation may risk ignoring additive effects that can occur when multiple forms of evidence exist simultaneously. Prosecutors may become more confident in their chances of success at trial if they are not reliant on any single piece of evidence, knowing that they have additional evidence to fall back on. For instance, in analyzing the role of eyewitness evidence on the dispositions of 725 felony cases, Flowe et al. (2011) noted that eyewitness identification was rarely the only form of evidence involved in a case. They further suggested that the police may be reluctant to forward cases to the prosecution until multiple forms of evidence have been collected. Additionally, results from Bushway et al.’s (2014) experimental survey cited earlier suggested that combinations of evidentiary variables exerted stronger effects on probability of conviction estimates and acceptable plea sentences than any single piece of evidence alone. Thus, a deeper understanding of the effects of evidence on plea discounts may require both individual and aggregate-level assessments.

Extra-Legal Factors

Another explanation for the limited and inconsistent relationship between evidence strength and plea discounts seen in prior research is that plea discounts may simply be driven by other factors. The focal concerns theory of judicial decision-making identifies offender “blameworthiness”, “protection of the community”, and “practical constraints and consequences” (Steffensmeier et al. 1998, pp. 766–767) as the central factors considered by criminal justice actors during sentencing decisions. While determining a defendant’s blameworthiness and protecting the community involve the consideration of legal factors such as case characteristics, evidence, and prior criminal record, the focal concerns theory also asserts that these determinations require the prediction of future behavior, which is often based on inadequate or incomplete information. Given the uncertainty involved in these decisions, legal actors may rely on “perceptual shorthand” (Steffensmeier et al. 1998, p. 767; see also Albonetti 1987), or stereotypes associated with extra-legal factors such as race, sex, and age to determine an offender’s level of blameworthiness or risk of future offending.

⁴ It is also important to note that individual pieces of evidence can be of varying strength. For example, not all eyewitness evidence is equally strong; one witness may be a person with poor eyesight who witnessed the crime 100 yards away at night, whereas another witness may previously know the defendant and witnessed the crime up close and personal. To our knowledge, such indicators of the reliability of evidence are not available in archival datasets.

Research has often supported the propositions of focal concerns theory as it applies to sentencing decisions. Steffensmeier et al. (1998) found the interaction of race, sex, and age to be particularly important in explaining sentencing disparity, with young Black males receiving significantly harsher sentences than other demographic sub-groups. Other studies have found that extra-legal factors impact decisions to plead guilty (Albonetti 1990; Elder 1989; Frenzel and Ball 2008; Omori and Peterson 2020) and the length and probability of carceral sentences for plea bargainers (Kutateladze et al. 2014; Johnson and Larroulet 2019; LaFree 1985; Piehl and Bushway 2007). More relevant to the present study, Ulmer and Bradley (2006) noted evidence of a significant relationship between racial composition and trial penalties, Piehl and Bushway (2007) found a correlation between race and plea value, and Kutateladze et al. (2015) reported that defendant age significantly impacted the likelihood of receiving a custodial plea sentence offer. Research has further suggested that defense attorneys may consider the demographic characteristics of their clients when making plea recommendations (Redlich et al. 2016), and that this may lead to significantly longer plea recommendations for minority clients relative to White clients (Edkins 2011).

Thus, it is possible that plea discounts are responsive to extra-legal characteristics through their effect on the way that defendants are perceived by legal actors. To date, however, we are not aware of any studies that directly compare the impact of evidence strength and extra-legal variables on the size of plea discount estimates. Such a comparison has both practical and theoretical importance. If evidence strength exerts stronger effects on the size of the plea discount than extra-legal variables, this would support rational models of plea decision-making. In other words legal participants would appear to be making plea decisions in anticipation of their expected trial outcomes (see Bibas 2004; Bushway and Redlich 2012). If non-legal factors such as a defendant's race, age, or sex exert stronger effects on plea discounts, however, then discounting practices may be subject to the same concerns and subjective biases found in other judicial processes. Alternatively, plea discounts may be responsive to both evidence strength and extra-legal factors, in which case the analysis of interest involves determining which factors produce the strongest effect.

The Current Study

In this study, we test the effects of both evidence strength and extra-legal factors on the size of plea discount estimates. We first use administrative circuit court data from the state of Virginia for all defendants convicted of at least one felony charge via bench or jury trial during 2017–2018 to estimate predicted trial sentence lengths. We then apply the coefficients from this model to a sample of plea bargainers from one mid-sized jurisdiction in Virginia whose plea hearings were systematically observed and coded for the presence of various evidentiary items (i.e., summaries of the evidence that the prosecution would have presented at trial were observed and coded). Taking the predicted trial sentences and the observed plea sentences for this target sample, we estimate plea discounts and examine the effects of multiple forms of evidence (eyewitness, confession, forensic/medical, video/photo/audio) and extra-legal factors (race, sex, age) on the size of these discount estimates. In doing so, we expand on the approach used by Smith (1986; see also Piehl and Bushway 2007; Bushway and Redlich 2012; Yan and Bushway 2018) by using a penalized ridge regression to account for the effects of multicollinearity between terms and the potential for model overfitting.

Based on prior literature and theory, we hypothesized that evidence strength would be significantly and inversely related to plea discounts, such that stronger evidence would lead to smaller plea discounts. More specifically, we hypothesized that direct evidence (e.g., witness and confession evidence) would have a stronger effect on plea discounts than circumstantial evidence (e.g., forensic/medical evidence), and that increases in the number of different forms of evidence would reduce plea discounts by the largest amount. In line with focal concerns theory, we also hypothesized that defendants who are young, male, and/or belong to a minority group would receive significantly smaller discounts than defendants of other demographic backgrounds. However, we anticipated that evidentiary factors would produce the largest effects on plea discounts, given that evidence strength should directly relate to case convictability (Bibas 2004; Spohn 2018).

Methods

Data Sources and Collection

Two primary data sets were used in this study. The first data set included all hearings that took place in Virginia Circuit Courts during the years of 2017–2018.⁵ Circuit courts in Virginia have jurisdiction over all felony prosecutions, any misdemeanor charges that accompany felony prosecutions, and any direct indictments. Processing information for these courts is made publicly available through Virginia’s judicial system website.⁶ These data contain information on hearing dates, charge type, jurisdiction, disposition, sentencing, and certain defendant characteristics. To generate models that would explain variation in trial sentence lengths, we first isolated all cases concluded by bench or jury trials that contained at least one felony conviction (roughly 71% of all trials during this timeframe). This resulted in a valid sample size of 5,671 cases, which represented our population-level trial conviction data.⁷

The second data set was our plea sample, which was obtained from systematic observations of plea hearings that took place between February 2017 and August 2018 in one mid-sized circuit court in Virginia (see Dezember et al. 2022). The data collection site was a suburban county of approximately 450,000 people that processes around 7,500 adult arrests yearly (Federal Bureau of Investigation 2016). A team of six researchers attended and coded plea hearings during this time, providing access to defendant, case, and sentencing information for this sample. Most importantly, the research team observed case summaries of the evidence that the prosecution proffered they would have presented at trial, allowing researchers to code for various pieces of direct and circumstantial evidence. Interrater reliability for these observations was assessed using a sample of 206 hearings observed by multiple coders. The Krippendorff’s alpha value for these hearings was 0.90, which indicates very strong agreement between raters (Hayes and Krippendorff 2007).

As a component of the plea observations, researchers also examined official case records, which provided information on the charges that each defendant was indicted on

⁵ This two-year timeframe was chosen to match the data collection period for our target data set described below.

⁶ These data are available to download in bulk from <https://virginiacourtdata.org/>.

⁷ Two jurisdictions in Virginia do not make their case processing data publicly available. However, our data still represents over 98% of all jurisdictions in the state.

and those that they pled to. This is particularly important as the indicted charges allow us to better account for any potential charge bargaining that may have occurred during case processing (see Piehl and Bushway 2007; Yan and Bushway 2018). More specifically, in exchange for a plea of guilty, defendants often benefit from a reduction in the nature or severity of their charges that would not have otherwise occurred. As such, the charges to which a defendant pleads guilty may represent only a subset of those that they would have faced at trial. By basing predictions on the indicted charges rather than the plea charges, we are able to better capture the full range of charges that the defendant would have likely at faced trial.⁸

Thus, the inclusion of evidentiary variables and information on the charges that each defendant faced at indictment make this data set uniquely suitable for the current study. In total, 611 plea hearings were fully observed and coded. However, here we use only those cases where the defendant was indicted on at least one felony charge, where the judge accepted the defendant's plea, and where an incarceration sentence was imposed, resulting in a valid sample size of 535 cases (for more information on the data collection procedures see Dezimmer et al. 2022).

Dependent Variables

To estimate sentencing discounts for our plea-bargaining sample we first needed trial sentence predictions. Thus, our initial dependent variable was the carceral sentence length received by defendants who were convicted at trial. Prior studies have often used the top charge (e.g., most severe charge) associated with each case as the unit of analysis (see Johnson and Larroule 2019; Yan and Bushway 2018), however, our intended unit of analysis for this study is at the case-level, where a case represents all charges that a defendant faced in a single trial. As such, our sentence length measure is operationalized as the aggregate sentence received across all charges within a case, including both felonies and misdemeanors. We do this based on the possibility that individual charges nested within a case may not be independent, but rather judges may consider the characteristics of the entire case when making sentencing decisions. Additionally, it is not uncommon for felony charges to be accompanied by secondary felonies or misdemeanors that can add substantial variability to the total sentence received. Our aggregated sentence lengths were measured in months and were top coded at 600 (50 years) to avoid excessive skew.⁹ Finally, after estimating these trial sentences, our primary dependent variable became the plea discount, defined here as

⁸ We acknowledge that this is an imperfect solution and that it is also possible for the severity and number of charges to be increased or decreased prior to trial. Given the parameters of our data, this approach provided the best adjustment for charge bargaining and is consistent with prior efforts of this kind (see Yan & Bushway 2018).

⁹ Only 214 (4%) of cases were censored at 600 across the full trial conviction sample. However, it is important to note here that we only included cases where a carceral sentence was imposed. While this approach improves the distribution of the outcome variable, it can create concerns regarding selection bias for any inferences aimed at the broader population of defendants (Bushway and Piehl, 2007; Elder 1989; Winship and Mare 1992). However, likely because we aggregated sentence lengths across all charges within a case, all cases in our trial conviction sample received some length of incarceration. Additionally, only 11% ($n=68$) of eligible plea cases did not receive a carceral sentence. Thus, our estimation of trial sentences should not suffer from selection bias, though we note that our plea discount results are limited to defendants who pled guilty *and* received some term of incarceration.

the proportional difference between the observed plea sentence and the predicted trial sentence (see section "[Plea Discount Estimation](#)" below).

Independent Variables

A primary goal of this study was to generate accurate trial sentence predictions rather than make any direct inferences regarding the effects of individual factors on sentencing outcomes. As such, we used a wide array of independent variables related to case, defendant, and external characteristics to predict trial sentences. Case characteristics included the number of charges that the defendant was convicted on, the types of charges that the defendant was convicted on (e.g., murder, rape, robbery, etc.), the minimum and maximum potential sentences that the defendant could have received across all charges, the statutory class of the charges that the defendant was facing (e.g., class 1 felony, class 2 felony, misdemeanor, etc.), and whether any of the charges that the defendant faced included a mandatory minimum sentence. Given that each case can contain multiple crime types and charge classifications, these characteristics were added as dummy variables (0/1) to allow for multiple categories to exist simultaneously for a single case. Class 1 felonies were extremely rare in our trial conviction data ($n=6$, 0.1%), and so these charges were merged with Class 2 felonies. Additionally, we added squared terms for both the minimum and maximum potential trial sentences that the defendant could have received across all charges to account for any potential curvilinear relationships.

Defendant characteristics included multiple measures of criminal history. While we did not originally have this information, we used the same publicly available Virginia Circuit Court data for the years 2000–2016 to calculate the number of felony convictions, misdemeanor convictions, and prior active incarceration sentences for each defendant. We then merged this data into our trial conviction sample to match criminal history with each defendant.

Considering our interest in the relationship between plea discounts and extra-legal variables, we omit demographic information from our prediction models for both theoretical and practical reasons. First, our goal in this study was to generate the most valid counterfactual trial sentences possible *without* the inclusion of demographic and evidentiary variables. In theory, trial sentence lengths should not be based on the strength of evidence or demographic characteristics (see Bushway & Redlich 2012), meaning that cases and defendants who are legally similar should receive similar sentences. The degree to which the magnitude of the differences between our observed plea outcomes and our predicted trial outcomes relates to evidentiary and demographic variables is the substantive interest of our analyses. Thus, omitting these factors from our prediction model is what allows us to introduce them during our plea discount analysis. The inclusion of evidence and demographic information in the trial sentence predictions would inevitably include these factors in the plea discount measure, potentially creating artificial correlations between our variables of interest.¹⁰ However, as a robustness check, we did estimate trial sentences with the inclusion of demographic variables, and their omission had little effect on our trial sentence predictions (see Results section).

¹⁰ More specifically, these factors would be explicitly modeled in both the dependent and independent variables, which may lead to correlations that are an "artifact of the model" (Bushway & Redlich 2012, p. 449). Because evidence and demographic factors are inherently included in the plea sentence but not in the trial sentence prediction, variation in the difference between these values may be attributed to these variables.

Given potential differences in regional and cultural sentencing practices across the state (see Kramer & Ulmer 2009), we also included the jurisdiction/area where the trial took place. At the county or city level there were several jurisdictions with no trial convictions or very small numbers of trial convictions. To account for this, we combined jurisdictions into their natural regions of the state (North, Central, West, East, and Southwest) and included these regions as categorical independent variables in our prediction model, with the North region (region from which our plea sample was taken) designated as the reference category. This addition is important because, despite our plea sample being drawn from only one region in the state, our trial sentence predictions are adjusted for variation in sentencing practices unique to this region. Lastly, while both of our data sets contained hearings that took place during 2017–2018, approximately 6% ($n=365$) of eligible trial convictions carried over into 2019. As such, we included disposition year as a categorical independent variable given the potential for changing sentencing practices and statutes over time (see Porter 2020).

In explaining our estimated plea discounts, we first employed a variety of different evidentiary measures. These included witness, forensic/medical, video/audio/photo, and confession evidence. Witness evidence involved any situation in which a witness was prepared to testify to observing the alleged offense or some element material to the offense. Forensic/medical evidence included any type of medical or forensic information that would have been introduced, such as fingerprints, DNA, blood analyses, or accompanying expert testimony. Video/photo/audio evidence involved any form of media that tended to directly implicate the defendant, and confession evidence involved direct, incriminating admissions given by the defendant. Individual pieces of evidence were introduced as dummy variables (0/1). We also created a summary measure of evidence by calculating the number of different pieces of evidence contained in a case. Finally, we introduced extra-legal factors such as race/ethnicity, sex, and age (and interactions between these characteristics) to test their comparative effects with those of evidentiary factors.

Plea Discount Estimation

Plea discount estimation followed the general approach first proposed by Smith (1986). Specifically, we generated counterfactual trial sentences using a regression model that considered the case, defendant, and external characteristics described above. Here, however, we used a penalized ridge regression (see Hoerl and Kennard 1970) rather than a standard ordinary least squares regression (OLS). While the coefficients estimated by an OLS regression may be unbiased in the absence of multicollinearity, strong correlations between independent variables can lead to coefficients that are artificially extreme in either direction (Hoerl and Kennard 1970; Yoo et al. 2014). Examination of variance inflation factors (see Marquardt and Snee 1975; Miles 2005) following the estimation of an OLS regression for our data indicated multicollinearity across several independent variables, however, these variables still added significant predictive ability to our model.¹¹ One approach to deal with

¹¹ The inclusion of both a maximum potential sentence across all charges and a minimum potential sentence across all charges led to VIF values that were greater than 10 for these factors. To test the importance of including both variables, we ran an OLS regression including only the maximum potential sentence and its squared term and then added the minimum potential sentence and its squared term to the model. Results of an F -test indicated that the addition of these variables significantly improved model fit ($F(2, 5635) = 365.02, p < 0.001$).

this issue would involve variable deletion (McDonald 2009; Tabachnick and Fidell 2007), however, this approach would reduce the amount of information contained in the model and would lead to less accurate predictions.

Ridge regression attempts to solve these issues by applying a penalty parameter to the regression coefficients, shrinking them toward 0 to mitigate the effects of multicollinearity while still allowing them to remain in the model (see Fox 2016; McDonald 2009).¹² This, in turn, introduces bias into the regression coefficients in exchange for a decrease in variance. That is, this shrinkage has the effect of smoothing the regression coefficients, which may reduce the variance that the model explains on the data from which it was generated but serves to increase the generalizability of the model to other data sets (see Agresti 2019; Melkumova and Shatskikh 2017). A key distinction is that, while the standard OLS regression seeks to minimize the sum of the squared errors between predictions and observations (see Tabachnick and Fidell 2007), the ridge regression attempts to minimize the sum of the squared errors plus the additional penalty parameter, as depicted in Eq. 1 (see Tibshirani 1996):

$$\sum_{i=1}^N \left(y_i - \sum_j \beta_j x_{ij} \right)^2 + \lambda \sum_j \beta_j^2 \quad (1)$$

Here, y_i represents the observed trial sentence for a given defendant, $\sum_j \beta_j x_{ij}$ represents the predicted trial sentence for that defendant, and $\lambda \sum_j \beta_j^2$ represents the penalty parameter applied to the regression coefficients. The predicted trial sentences themselves resemble a standard OLS formula as seen in Eq. 2:

$$\sum_j \beta_j x_{ij} = \beta_0 + \beta_1 \text{Offense}_{ij} + \beta_2 \text{Criminal History}_{ij} + \beta_3 \text{External}_{ij} \quad (2)$$

where β_1 represents a vector of regression coefficients related to the characteristics of the offense, β_2 represents a vector of coefficients related to the defendant's criminal history, and β_3 represents a vector of coefficients related to external characteristics (e.g., jurisdiction and year of the trial). All independent variables here are standardized prior to model estimation and are then converted back to their original scale prior to reporting the results (see Hastie et al. 2016). We chose not to report standardized coefficients as the majority of our independent variables are binary in nature. Additionally, we did not have a substantive interest in comparing the strength of the coefficients across differing scales, particularly given that our coefficients are penalized.

A critical component of the ridge regression is the tuning parameter (λ in Eq. 1). This parameter determines how much shrinkage will be applied to the coefficients, with more shrinkage leading to increased bias and decreased variance (and vice versa; see Melkumova and Shatskikh 2017). To select an appropriate value for the λ parameter, we first randomly divided our trial conviction data into a training data set and a test data set using a

¹² We chose to use a ridge regression over similar methods such as the LASSO regression because, while the ridge regression shrinks parameters toward 0, it will never fully eliminate any coefficients from the model. In contrast, the penalty parameter in a LASSO regression is free to shrink coefficients to 0 and eliminate them from the model (Tibshirani 1996). While LASSO regressions may lead to more parsimonious models, we did not want any of our coefficients eliminated from the model, given that each variable has a theoretical connection to sentencing outcomes.

70%—30% split. Here, the training and test data sets were statistically independent of each other and neither data set contained cases from the target sample (i.e., plea sample). After dividing the trial conviction data in this way, we conducted a tenfold cross-validation on the training data set to identify the λ value that produced the minimum mean squared error (MSE).¹³ However, we used the “one standard error rule” (Ternès et al. 2016, p. 2564) to select the largest λ that still produced a MSE value that was within one standard error of the minimum MSE across all models (see also Hastie et al. 2016). This approach is intended to provide a compromise between the amount of penalization applied to the coefficients and the overall error of the estimated model (see Melkumova et al. 2017).

Using our selected λ value, we calculated our regression coefficients using the training data set. While some have recommended the use of a natural log transformation when modeling sentence lengths (see Freiburger et al. 2019), we found that our untransformed sentence lengths led to substantially higher predictive accuracy, and thus we report results from the prediction model using the untransformed dependent variable.¹⁴ As recommended (Hastie et al. 2016; Tibshirani 1996), we then applied the coefficients from this model to the test data set to evaluate the accuracy of the predictions before ultimately applying the model to our plea sample. When applying these coefficients to the plea sample, we used the charges at indictment to account for any potential charge bargaining that may have occurred between indictment and plea. Finally, after generating predicted trial sentences for our plea sample, we calculated the plea discount using Eq. 3, where \widehat{TS} represents the predicted trial sentence and PS represents the observed plea sentence:

$$\text{Plea Discount} = \frac{\widehat{TS} - PS}{\widehat{TS}} * 100 \quad (3)$$

It is important to note that, while each individual in our sample received a trial sentence prediction, our results and inferences are concerned with marginal effects rather than individual estimates. That is, our predictions are based on the average effect of group membership across our collection of independent variables. This approach is consistent with those of prior efforts to examine the impact of evidence on plea discounts and the estimated probability of conviction at trial (see Bushway & Redlich 2012; Smith 1986; Yan & Bushway 2018).

All analyses were conducted in R statistical software (R Core Team 2020). Data partitioning was conducted using the *rsmple* package (Silge et al. 2021) while cross-validation and ridge regression models were estimated using the *glmnet* package (Friedman et al. 2010). The R code is available from the authors upon request.

¹³ To do this we used the “glmnet.cv” function in the *glmnet* package (Hastie et al. 2016) in R statistical software. This function automatically generates a λ sequence between the minimum and maximum λ values.

¹⁴ For example, an OLS regression estimated using the non-logged dependent variable produced an R^2 of 0.66 while the same model estimated with the logged dependent variable produced an R^2 of 0.51. In this regard, it is important to reiterate that our dependent variable did not contain any 0 s and was top-coded, which reduced skew. Additionally, by including all charges associated with a case in cumulative fashion, our dependent variable may have better resembled an additive process.

Table 1 Descriptive statistics for trial sentence predictors

	Trial Conviction (<i>N</i> =5,671)		Guilty Plea (<i>N</i> =535)	
	Mean/Proportion	SD	Mean/Proportion	SD
<i>Case characteristics</i>				
Sentence length	129.350	146.771	94.778	105.671
Number of charges	2.037	2.554	3.120	2.590
Maximum potential sentence	276.622	214.549	343.761	208.771
Minimum potential sentence	53.438	89.213	59.731	87.271
Murder	0.037	–	0.022	–
Rape	0.021	–	0.017	–
Robbery	0.044	–	0.047	–
Assault	0.149	–	0.092	–
Kidnap and arson	0.030	–	0.017	–
Other sex crime	0.048	–	0.067	–
Other person crime	0.200	–	0.164	–
Property crime	0.308	–	0.428	–
Drug crime	0.277	–	0.359	–
Minor crime	0.169	–	0.286	–
Class 1 or 2 felony	0.052	–	0.034	–
Class 3 felony	0.048	–	0.043	–
Class 4 felony	0.046	–	0.065	–
Class 5 felony	0.274	–	0.426	–
Class 6 felony	0.360	–	0.346	–
Class U felony	0.443	–	0.544	–
Misdemeanor	0.182	–	0.348	–
Mandatory minimum	0.118	–	0.144	–
<i>Criminal history</i>				
Prior felony convictions	2.049	4.890	0.884	2.00
Prior misdemeanor convictions	0.495	1.650	0.166	0.727
Prior incarceration sentence	0.366	–	0.236	–
<i>External characteristics</i>				
Trial in 2017	0.477	–	0.482	–
Trial in 2018	0.459	–	0.518	–
Trial in 2019	0.064	–	–	–
Northern	0.035	–	1.000	–
Central	0.364	–	–	–
Eastern	0.379	–	–	–
Western	0.046	–	–	–
Southwest	0.176	–	–	–

Crime types and classes are not mutually exclusive and do not sum to 1. Standard deviations are not reported for proportions.

Results

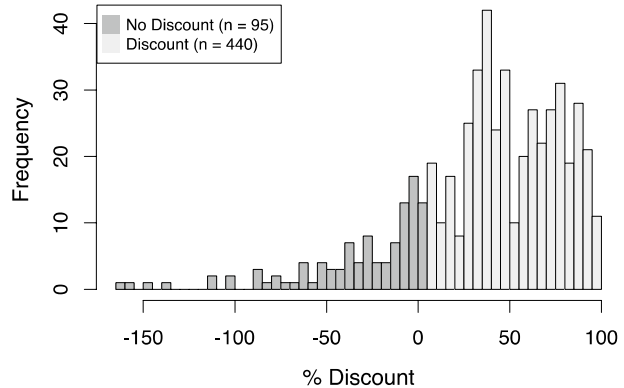
Descriptive statistics for the full trial conviction and plea samples can be seen in Table 1. On average, defendants convicted at trial received an additional 34 months of incarceration compared to defendants who pled guilty, however, there were significant differences in the characteristics of plea and trial cases. Trial conviction cases were more likely to include murder, rape, and assault charges than plea cases (20.7% vs. 13.1%, $\chi^2(1) = 17.15$, $p < 0.001$), while plea cases were more likely to include drug charges and property crime (78.7% vs. 58.5%, $\chi^2(1) = 82.45$, $p < 0.001$). Plea cases were also more likely to include misdemeanors and class 6 felonies than trial cases (69.4% vs. 54.2%, $\chi^2(1) = 45.12$, $p < 0.001$), and defendants convicted at trial had more extensive criminal records, with over 2.5 prior felony and misdemeanor convictions compared to just over 1 prior conviction for defendants who pled guilty ($t(1281.1) = 11.77$, $p < 0.001$). However, defendants who pled guilty faced more charges per case ($t(6204) = 9.36$, $p < 0.001$) and higher maximum potential sentences if convicted at trial ($t(6204) = 6.93$, $p < 0.001$). Demographic information is not displayed in Table 1 as it was not used as a component of our trial sentence predictions (demographic information for the plea sample can be seen in Table 3). However, the trial conviction sample was predominately Black (49.1%) and male (79.2%); age could not be calculated as date of birth was not provided in these data. Lastly, certain jurisdictions and trial years were not applicable to our plea sample. The coefficients from these empty factor levels simply dropped out of our trial sentence predictions.

As previously discussed, we first estimated our trial prediction model on a training data set comprised of a random sample of 70% ($n = 3,967$) of all trial conviction cases. We then tested the accuracy of our predictions on a test data set, which was the remaining 30% ($n = 1,704$) of trial conviction cases. Results from these models can be seen in Table 2. Of note here, the addition of the penalty parameter in the ridge regression intentionally introduces bias, and in doing so, the standard errors of the regression coefficients lose meaning. As such, inferences regarding statistical significance are often not recommended for ridge coefficients (see Fox 2016; Goeman et al. 2018). Furthermore, statistical significance is not meaningful for our purposes given that we have no specific hypotheses or interest in theory-building related to the trial sentence predictions. Thus, we report only the coefficients themselves and the predictive accuracy of this model.

The coefficients in Table 2 generated predictions that explained over 60% of the total variance in the trial sentence lengths for both the training and test data sets, indicating a predictive ability comparable to that of prior studies (see Yan and Bushway 2018; Yan 2020). Additionally, the consistency across both training and test data sets suggests that the model is generalizable across samples. In general, defendants facing a larger number of charges and more severe charges received longer trial sentences. Murder and rape charges increased trial sentence predictions by over 65 and 82 months, respectively, while class 1 and class 2 felonies increased trial sentences by over 42 months. Prior convictions and incarceration sentences also led to increases in sentence lengths, though criminal history appeared to have a smaller effect than case characteristics (e.g., number of charges per case). There was also notable variation across jurisdictions, with cases in the Western region of the state receiving an average 6-month reduction in sentencing, compared to cases in the Northern region. While the model reported in Table 2 does not account for demographic characteristics, we added race and sex as a robustness check. The R^2 values for both the training and test data sets increased by less than 0.2%, and the predicted values

Table 2 Ridge regression coefficients for trial sentence predictions ($n = 3,967$)

	Coefficient
Intercept	33.542
<i>Case characteristics</i>	
Number of charges	6.259
Maximum potential sentence	0.088
Maximum potential sentence squared	0.0001
Minimum potential sentence	0.231
Minimum potential sentence squared	0.0002
Murder	65.111
Rape	82.750
Robbery	21.759
Assault	2.113
Kidnap and Arson	28.780
Other sex crime	40.738
Other person crime	5.944
Property crime	1.873
Drug crime	-3.393
Minor crime	-4.289
Class 1 or 2 felony	42.492
Class 3 felony	19.004
Class 4 felony	18.462
Class 5 felony	7.405
Class 6 felony	4.680
Class U felony	14.835
Misdemeanor	2.510
Mandatory minimum	12.615
<i>Criminal history</i>	
Prior felony convictions	0.423
Prior misdemeanor convictions	0.323
Prior incarceration sentence	0.776
<i>External characteristics</i>	
Central (ref = Northern)	2.297
Eastern (ref = Northern)	1.231
Western (ref = Northern)	-6.204
Southwest (ref = Northern)	2.632
Trial in 2018 (ref = 2017)	0.138
Trial in 2019 (ref = 2017)	-1.669
Lambda (λ)	142.040
R^2 (training data)	0.625
RMSE (training data)	93.069
R^2 (test data)	0.602
RMSE (test data)	94.209

Fig. 1 Plea discount distribution

with and without these demographic factors were highly correlated on both the training and test data sets ($r=0.999$ for both). Thus, it does not appear as though the omission of demographic characteristics significantly impacted the accuracy of our trial sentence predictions.

Taking the coefficients from Table 2, we then estimated the counterfactual trial sentences and resulting plea discounts for our target sample using the charges at indictment (see Fig. 1). For the full sample, the observed plea sentence was 64.2% of the predicted trial sentence, suggesting an average discount of approximately 36%. This mirrors the results at the individual level as well, where the mean discount across defendants was 36.5% (Table 3). However, like prior studies we find a large degree of variance in the size of the plea discount across defendants, with discounts ranging from a 99% reduction in the plea sentence relative to the expected trial sentence to a 165% increase in the plea sentence relative to the expected trial sentence. This suggests that some defendants are receiving considerably longer plea sentences than would be expected at trial. In fact, we find that nearly 18% of defendants received a plea sentence that was higher than their expected trial sentence (for a similar estimate see Bushway and Redlich 2012).

Given the presence of significant variation in these discount estimates (Fig. 1), our primary interest becomes exploring the sources of this variation. Table 3 reports the distribution of the evidentiary and extra-legal variables used as predictors of the plea discount. On average, these cases contained around two distinct forms of evidence, with the vast majority containing some form of witness evidence. Confession evidence was also common, occurring in nearly 50% of all cases, while both video/photo/audio evidence and forensic/medical evidence occurred in less than a quarter of all cases. Defendants in this sample were predominately male and in their early 30s. Black was the modal racial category, followed by White and Hispanic.

Table 4 reports the results from a series of OLS regressions examining the impact of evidence on the size of the estimated plea discount. In Model 1, we regress the discount on each individual form of evidence. The intercept for this model indicates that defendants who had none of the measured forms of evidence against them received nearly a 50% plea discount. However, we find that only the presence of direct video/photo/audio evidence significantly impacts this discount, leading to a 9.3% decrease in the size of the discount when controlling for other evidentiary factors. The presence of witness evidence also reduced our estimated plea discounts by approximately 13%, however, this effect was only present at a less stringent 0.10 significance threshold.

Table 3 Plea discount and predictor variables for plea sample ($N=535$)

	Mean/proportion	SD	Min	Max
<i>Discount</i>				
Observed plea sentence	94.778	105.671	1	600
Predicted trial sentence	147.534	89.670	48.775	609.484
Plea discount	0.365	0.443	-1.648	0.988
<i>Evidence</i>				
<i>N</i> of evidence	1.725	0.726	0	4
Witness	0.930	–	0	1
Confession	0.497	–	0	1
Forensic/medical	0.062	–	0	1
Video/photo/audio	0.236	–	0	1
<i>Extra-legal</i>				
White	0.374	–	0	1
Black	0.415	–	0	1
Hispanic	0.172	–	0	1
Other	0.039	–	0	1
Male	0.796	–	0	1
Age	32.073	11.533	15	80

There were 52 (9.7%) cases where an evidence summary was not presented. Thus, there are only 483 with valid evidentiary variables.

In Model 2 we introduce a continuous measurement of the number of different forms of evidence contained in each case. This variable could not be added to Model 1 as it is simply an additive product of each individual form of evidence, and thus is perfectly collinear with these measures. However, we do introduce this variable alongside video/photo/audio evidence, given that this was the only individually significant form of evidence found in Model 1. As such, Model 2 tests whether the number of pieces of evidence impact plea discounts above and beyond the effect of video/photo/audio evidence alone. Results indicate that each additional form of evidence leads to a statistically significant 7% decrease in the size of the plea discount. The presence of video/photo/audio evidence is no longer significant in this model, suggesting that the cumulative effect of evidence may be a more important determinant of the plea discount than any single piece of evidence in isolation.¹⁵

Model 3 introduces extra-legal demographic variables. The number of pieces of evidence in each case continues to demonstrate a significant relationship with plea discounts. However, even after controlling for this effect, Hispanic defendants and male defendants received significantly smaller plea discounts than other demographic groups. Specifically,

¹⁵ Note that we also tested the number of pieces of evidence as a categorical variable using 0 pieces of evidence as the reference category. Results support the idea that this relationship resembled a linear one. Here, one piece of evidence decreased plea discounts by 28% ($p=0.05$), two pieces of evidence decreased plea discounts by 29% ($p=0.04$), three pieces of evidence decreased plea discounts by 42% ($p<0.01$), and four pieces of evidence decreased plea discounts by 62% ($p=0.02$), controlling for the presence of video/photo/audio evidence. However, operationalizing this measure as a categorical variable did not significantly improve model fit ($F(2, 477)=1.71, p=0.18$).

Table 4 OLS regression results for plea discount

	Model 1 <i>b</i> (SE)	Model 2 <i>b</i> (SE)	Model 3 <i>b</i> (SE)	Model 4 <i>b</i> (SE)	Model 5 <i>b</i> (SE)
<i>Evidence</i>					
Witness	-0.130 ⁺ (0.078)	-	-	-	-0.091 (0.077)
Confession	-0.049 (0.040)	-	-	-	-0.041 (0.039)
Forensic/medical	-0.095 (0.082)	-	-	-	-0.051 (0.081)
Video/photo/audio	-0.093* (0.047)	-0.021 (0.054)	-	-	-0.086 ⁺ (0.046)
<i>N</i> of evidence	-	-0.072* (0.031)	-0.070** (0.027)	-0.062* (0.027)	-
<i>Extra-legal</i>					
Age	-	-	-0.001 (0.002)	-0.0004 (0.002)	-0.0005 (0.002)
Black	-	-	-0.049 (0.044)	-0.100 (0.093)	-0.099 (0.094)
Hispanic	-	-	-0.198*** (0.057)	0.116 (0.137)	0.117 (0.138)
Other	-	-	-0.166 (0.108)	-0.250 (0.253)	-0.265 (0.255)
Male	-	-	-0.122* (0.049)	-0.099 (0.073)	-0.098 (0.074)
<i>Interactions</i>					
Black*Male	-	-	-	0.062 (0.106)	0.060 (0.106)
Hispanic*Male	-	-	-	-0.371* (0.151)	-0.372* (0.152)
Other*Male	-	-	-	0.100 (0.280)	0.121 (0.283)
Intercept	0.497*** (0.079)	0.453*** (0.053)	0.482*** (0.049)	0.465*** (0.063)	0.592*** (0.093)
Observations	483	483	483	483	483
<i>R</i> ²	0.019	0.017	0.059	0.075	0.077
<i>F</i> -Statistic	2.338 ⁺ (<i>df</i> =4, 478)	4.213* (<i>df</i> =2, 480)	4.957*** (<i>df</i> =6, 476)	4.272*** (<i>df</i> =9, 473)	3.246*** (<i>df</i> =12, 470)

Bold values indicates statistical significance at $p < 0.05$ or less

Age and *N* of evidence are mean centered in Models 3–5 but not in Models 1 and 2.

+ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Hispanic defendants' average plea discount is 20% smaller than White defendants, and male defendants' average plea discount is 12% smaller than female defendants.

Given these findings and the recommendations of focal concerns theorists to examine specific demographic subgroupings (see Steffensmeier et al. 1998), we added an interaction term for race/ethnicity and sex to Model 4. The addition of the interaction term necessitates that the main effects for race/ethnicity and sex be ignored in this model, and our results indicated a significant interaction between the two, with Hispanic males receiving an additional 37% decrease in the size of their plea discounts beyond the effect of being male or Hispanic alone. Using the intercept of Model 4 as a reference point, White males of average age and with an average amount of evidence against them received a mean plea discount of roughly 36.6%. In comparison, Black males received a mean discount of 32.8% and Hispanic males received a mean discount of only 11%. The effect of evidence also remained significant in Model 4, with a one-unit increase in the amount of evidence leading to an average 6% decrease in plea discounts, though no other main or interaction effects were found across our other independent variables.

Model 5 replicates Model 4, but controls for each individual piece of evidence rather than the total number of pieces of evidence in each case. Results are substantively similar, indicating a significant interaction between being Hispanic and male, however, no individual piece of evidence achieves statistical significance in this model. Lastly, we tested interaction terms between race/ethnicity and the number of pieces of evidence, as well as between race/ethnicity, sex, and age (not shown). Results for both tests were nonsignificant.¹⁶

The relationship between evidence, race, sex, and plea discount can be further seen in Fig. 2. For both males and females of all racial categories, increases in the number of pieces of evidence lead to decreasing plea discounts. However, plea discounts also appear to be smaller for males than for females across all evidentiary and racial categories. Additionally, there is a clear and significant difference between the average predicted discount for Hispanic males and females. While Hispanic males receive the smallest plea discount across all levels of evidence, Hispanic females appear to receive the largest plea discount across all levels evidence.¹⁷ Hispanic males with 3–4 pieces of evidence against them were predicted to receive plea discounts of approximately 0%. In comparison, no other combinations of race/ethnicity and sex were predicted to receive plea discounts less than approximately 10%, regardless of the number of pieces of evidence against them.

Thus, our findings suggest significant impacts of both evidence and extra-legal factors on plea discounts, however, this does not indicate which factors are producing the largest effect. To compare the magnitude of these effects, we calculated standardized effect sizes for the number of pieces of evidence, the main effects of race/ethnicity and sex, and the interaction between race/ethnicity and sex. Given that both the number of pieces of evidence and the plea discount are continuous measures, correlation coefficients were used as measures of effect size to allow for comparison of evidence with race/ethnicity and sex (see Lipsey and Wilson 2001). Results indicated that the being Hispanic and male produced the

¹⁶ We had a limited sample with which to examine age. For example, 75% of White defendants, 75% of Black defendants, and 81% of Hispanic defendants were between the age of 18–39. Few defendants were in age groups where significantly decreasing punishment might be expected. Thus, our comparisons are predominately between young males in each racial category.

¹⁷ Note that only 12 Hispanic female defendants are represented in this sample, however.

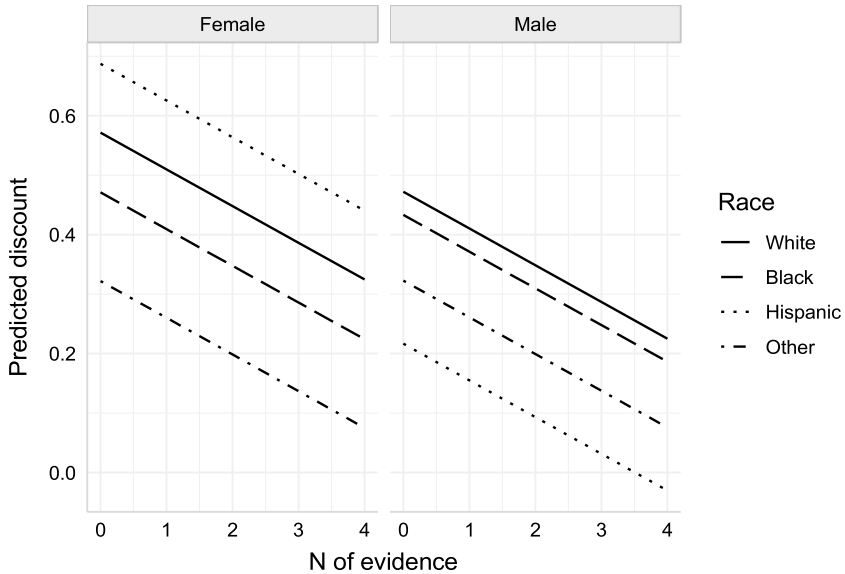


Fig. 2 Predicted discount by evidence, race, and sex

largest effect ($r = -0.271$, 95% CI $[-0.400, -0.143]$), followed by the main effect of being Hispanic alone ($r = -0.228$, 95% CI $[-0.344, -0.113]$), the number of pieces of evidence ($r = -0.130$, 95% CI $[-0.218, -0.042]$), and finally the main effect of being male alone ($r = -0.127$, 95% CI $[-0.215, -0.040]$).¹⁸

Sensitivity Analyses

We conducted several sensitivity analyses on our main findings (see Supplementary Materials). First, given the small sample sizes available for our interaction effects (e.g., 12 Hispanic female defendants), we re-estimated these models using a log transformation for the dependent variable. This required us to flip the distribution such that all values were positive, and to do so we simply divided the plea sentence by the expected trial sentence, which produced a distribution that was perfectly and negatively correlated with our discount measure. When this measure was log transformed, results were substantively similar to those of Model 4, indicating a statistically significant interaction effect between race/ethnicity and sex ($b = 0.66$, $p = 0.02$). The coefficient for this model is positive because the outcome distribution has been flipped, but the results still correspond to a significant decrease in the average plea discount for Hispanic male defendants.

Second, we re-introduced variables related to criminal history and charge severity into our main plea discount models to determine whether these factors may have confounded

¹⁸ The effect size for the number of pieces of evidence here represents a linear effect. If this effect size is calculated comparing defendants with four pieces of evidence to those with none, it becomes the largest effect ($r = -0.60$, 95% CI $[-0.95, -0.26]$). However, there are very few defendants with either zero or four pieces of evidence. A more realistic comparison of defendants with three pieces of evidence to those with one suggests an effect size of $r = -0.15$, 95% CI $[-0.27, -0.02]$, which remains smaller than both the interaction effect for being Hispanic and male, as well as the main effect for Hispanic alone.

our observed effects. Our results remained unchanged, with significant main effects for the number of pieces of evidence, race/ethnicity, sex, and the interaction between race/ethnicity and sex. We urge caution in interpreting these results, however, as the variables related to criminal history and charge severity in these models were inevitably included in both the dependent and independent variables. Nonetheless, these results provide some indication that our findings may be robust to the inclusion of multiple theoretically salient variables.

Finally, because modeling all charges contained within each case produces numerous combinations of charges which may be nonequally constructed across plea and trial cases, we repeated our analyses using only a top charge per case. Here, we defined the top charge as the charge with the longest minimum potential sentence. We chose the minimum potential sentence as there is no clear hierarchy of severity by crime type or charge class in Virginia (e.g., assault charges can range from misdemeanors to class 2 felonies, class U felonies can be both more and less severe than other classes of felonies, etc.). Limiting our models to a single charge resulted in a loss of explanatory power for both training ($R^2=0.49$, $RMSE=73.5$) and test ($R^2=0.49$, $RMSE=70.00$) data sets, and the average plea discount decreased to approximately 14% (down from 36.5%). However, under this specification there remained a significant interaction between race/ethnicity and sex ($b=-0.42$, $p=0.048$). The effect of evidence, while similar in magnitude ($b=-0.046$), was no longer statistically significant. There was also a notable outlier in this distribution, with one defendant receiving a plea sentence nearly five times larger than their predicted trial sentence. When omitting this outlier, the number of pieces of evidence per case was a significant predictor of plea discount estimates ($b=-0.07$, $p=0.03$), as was being Hispanic ($b=-0.19$, $p=0.01$) and male ($b=-0.17$, $p=0.01$), but the interaction between race/ethnicity and sex was only marginally significant ($b=-0.35$, $p=0.08$). Thus, this alternative specification produced substantively similar results to that of our main approach, leading to similar overall conclusions.

Discussion

Rational theories of plea decision-making, such as the shadow of the trial model, posit that evidence strength and the resulting probability of conviction at trial should directly determine the size of the discount necessary to induce a guilty plea (Bibas 2004; Bushway and Redlich 2012). However, research on this effect has been limited and inconsistent (Bushway and Redlich 2012; Smith 1986), calling into question the validity of this theory. If evidence strength is not a primary driver of the plea discount, then it remains possible that, like other sentencing outcomes, plea discounts are influenced by focal concerns related to factors such as race, sex, and age (see Ulmer and Bradley 2010; see also Steffensmeier et al. 1998). The current study explored these issues by directly testing the impact of both evidence strength and extra-legal factors on plea discount estimates. Our results suggest that both factors impact the size of plea discounts, but that demographic characteristics, and particularly the interaction between these characteristics, may produce the largest effects. These findings have important theoretical and policy implications and raise questions about rational theories of plea decision-making.

Compared to prior efforts of this kind (see Bushway and Redlich 2012; Yan and Bushway 2018), our results are more suggestive of a relationship between evidence strength and plea discount magnitude. We find that increases in the amount of evidence associated with a case are associated with significant decreases in the size of the estimated plea discount. However, we also find that this effect appears to be more about

evidence quantity than any specific form of evidence in isolation. Prior research has suggested that direct evidence such as confessions or eyewitness testimony is often overvalued by legal decision-makers (Devine et al. 2001; Heller 2006; Kassin and Neumann 1997; Niedermeier et al. 1999; Wells 1992). Given these findings, we hypothesized that our measures of direct evidence (confession, witness, and video/photo/audio) would significantly reduce plea discount magnitude. This hypothesis was only partially supported, however, as the presence of video/photo/audio evidence was the only individual form of evidence to significantly impact plea discounts. Additionally, when controlling for the number of pieces of evidence associated with each case, the influence of video/photo/audio evidence was no longer significant. Thus, whereas direct evidence, such as confession and eyewitness testimony, has been found to strongly impact trial verdicts (e.g., Devine et al. 2009; Kassin 2012) and to a lesser degree, plea discounts in a hypothetical case (Bushway et al. 2014; Redlich et al. 2016), we did not find such relationships here.

We also hypothesized that the total amount of evidence in each case would exert a larger impact on plea discounts than any individual piece of evidence alone. That is, the totality of the evidence facing a defendant may be a better proxy for evidence strength than any single form of evidence alone. Moreover, multiple forms of evidence may increase legal actors' confidence in a case, providing additional room for error should any individual piece of evidence fall through. Our results supported this hypothesis, as the number of different forms of evidence contained in each case remained the only evidentiary variable to demonstrate significant effects across all model specifications. However, we still expected to see significant variation in the impact of specific forms of evidence, and it is unclear why this was not the case. It is possible, perhaps, that salient individual pieces of evidence (i.e., direct evidence) are more important considerations at trial than they are during plea negotiations. If so, this would stand against the notion that plea decisions are made in the shadow of the trial.

Perhaps our most important finding concerns the discrepancy in the size of the observed effects for evidentiary variables and demographic characteristics. Bushway and Redlich (2012) suggested that "if plea arrangements are not conducted in the "shadow of a trial", we would expect to see factors other than those related to evidence to be influential" (p. 443). Specifically, focal concerns theory suggests that extra-legal factors such as race, sex, and age may impact plea discounts through their effect on the way that defendants are perceived (i.e., young minority males may be considered more dangerous and blameworthy, and thus less deserving of a plea discount; Ulmer et al. 2010; see also Kutateladze et al. 2015; Steffensmeier et al. 1998). It is possible that legal actors in our sample resorted to "perceptual shorthand" (Steffensmeier et al. 1998, p. 767) when presented with incomplete information about a defendant's level of dangerousness or blameworthiness, and that this may have led to improper considerations of race/ethnicity and sex during plea and sentencing decisions. Our findings provide preliminary support for these claims, indicating that Hispanics, males, and particularly the interaction between the two, were associated with significantly smaller plea discounts (i.e., less lenient plea offers) than White and/or female defendants. Moreover, this effect remained significant even after controlling for the amount of evidence in each case.

It is not clear why this effect was so concentrated within Hispanic defendants as opposed to other minorities. While all racial categories were associated with smaller plea discount estimates than White defendants, only the effect of being Hispanic was statistically significant. It is worth noting that these data come from a jurisdiction where about one-quarter of

the population is Hispanic, and it remains possible that there is a unique locational element responsible for this discrepancy. However, as of 2018 the Hispanic population across Virginia jurisdictions ranged from 0.7% to 40.1% (University of Virginia 2018). Thus, while our target jurisdiction has a larger Hispanic population than many others, it is not an anomaly within the larger context of the jurisdictions included in our prediction model.

Recent work has also found strong relationships between disparate court outcomes and ethnic distinctions (Omori and Petersen 2020; Smith et al. 2021). In Miami-Dade County (FL), Omori and Petersen (2020) reported that Black Latinos were disproportionately given pre-trial detention, convicted, and sentenced to incarceration, and that this disparity was greatest when comparing White non-Latinos to Black Latinos. Omori and Petersen argue for the importance of examining both racial and ethnic distinctions in research on court outcomes, and while we are not able to separate our demographic characteristics in this way, it is possible that our results are a combination of both racial and ethnic effects. An alternative explanation is that Hispanic defendants are more likely to lack U.S. citizenship, and that citizenship status exerts a strong impact on sentencing disparities (see Light 2014; Light et al. 2014). Indeed, a 2010 Supreme Court decision, *Padilla v. Kentucky*, mandated that non-US citizens must be made aware of the risk of deportation when pleading guilty. This would be a plausible explanation for our findings, but one that we could not directly test.

The larger question to be asked concerns the implications of our findings for plea bargaining theory and criminal justice inequality. On a cursory level, our study may indicate that there is merit to both rational theories of plea decision-making (i.e., the “shadow of the trial” model) and theories that consider subjectivity and potential bias in these processes (i.e., focal concerns). In either case, our results suggest that plea discounts are not solely (or even primarily) related to the convictability of a case. Instead, our findings paint a more complicated picture, stressing the importance of the way that defendants are perceived by legal actors. If these perceptions are based on characteristics that should have no legal bearing, then plea discounts are not simply the product of legal considerations.

However, the concept of convictability may also involve more than just evidence strength or legal considerations alone. For instance, in an ethnographic study conducted from within a district attorney’s office, Frohmann (1997) found that prosecutors often considered non-legal characteristics such as race and sex when determining the convictability of a case. That is, prosecutors envisioned the way that each case might be perceived by a hypothetical jury when deciding whether to pursue charges, and these projections accounted for extra-legal characteristics of the victim and defendant. Thus, whereas extra-legal factors in theory should not influence convictions (either by trial or plea), they nonetheless do, a fact that most experienced prosecutors are likely aware. In this way, it may be that evidence, race, and sex all become components of convictability assessments and the perceived strength of the case. Prosecutors in our sample may have been more confident that they could secure a trial conviction against defendants who were Hispanic and male, net of any significant differences in evidence strength. If so, these decisions would undoubtedly be based on stereotypes, and could result in systematic biases against certain groups of defendants, but the proposition that plea discounts relate to case convictability might still hold true. It may simply be that the strength of the case, as perceived by legal actors, involves much more than objective evidence alone.

Limitations and Conclusions

Several important limitations of this study deserve mention. First, as with any effort to generate predictions for an outcome that did not occur in reality, our approach inevitably assumes that a valid trial sentence prediction can be derived for defendants who plead guilty (for similar discussions see Bushway and Redlich 2012; Bushway et al. 2014; Yan and Bushway 2018). While we took significant measures to control for an array of defendant, case, and external characteristics, our results did suggest significant differences in case characteristics between defendants who went to trial and those that pled guilty. Additionally, our prediction model did not include several variables that may be important determinants of trial sentence length, such as the type of defense counsel, the relationship between the offender and victim, and whether the defendant was detained prior to trial (see LaFree 1985; Smith 1986). While these limitations may be unavoidable using a counterfactual approach, we note that our model still performed very well across both training and test data sets. Thus, to the degree that this assumption is valid, our model appeared well equipped to generate true-to-life trial sentence predictions.

Our trial sentence predictions were also generated using statewide data, but then applied to only one jurisdiction within the state. This presents concern over the potential that specific factors related to sentencing decisions in our target jurisdiction were improperly accounted for. While we added fixed effects to our model to adjust for sentencing differences between geographic regions of the state, we were unable to add terms specific to each individual jurisdiction. Additionally, this approach assumes that the coefficients for our independent variables are constant across geographic regions (i.e., that there are no interactions between geographic region/jurisdiction and other predictor variables). Nonetheless, the use of statewide data was necessary to provide an adequate number of cases to accurately generate trial sentence predictions and cross-validate those predictions. The use of fixed effects to account for variance attributable to macro-level geographic units is also similar to prior efforts of this kind (Bushway & Redlich 2012; Smith 1986), and our geographic regions largely fall along political and cultural lines that may be particularly salient for sentencing severity (see Ulmer et al. 2008).

Second, our measures of evidence were limited to those that were proffered by prosecutors during plea hearings. While this approach is unique, it is possible that we were lacking important evidentiary variables, or that there was additional evidence that was not proffered. On a related note, we did not have several evidentiary measures that have been common in prior studies, such as the number of witnesses in each case (Bushway and Redlich 2012; Smith 1986), or whether a weapon was used or recovered (LaFree 1985; Kutateladze et al. 2015). Measuring the number of pieces of evidence in additive fashion may also be overly simplistic. No differential weights were given to specific pieces of evidence, and we were unable to account for evidence quality. Nonetheless, coding evidence as proffered during plea hearings is a unique approach to this research that allowed us to create novel evidentiary variables, such as video/audio/photo evidence and forensic or medical evidence.

Another important limitation is the potential for unobserved characteristics to confound the effects of race and sex on plea discount magnitude (see Sensitivity analyses section). Given that elements related to a defendant's criminal history and current case characteristics were used to generate trial sentence predictions, they could not be confidently added as predictors of plea discount estimates. This creates the possibility that Hispanic defendants and/or males were also associated with longer criminal histories, more severe charges, or other characteristics that could have impacted discount estimates. While we explored this

potential in our sensitivity analyses and did not find any evidence to suggest that criminal history or charge severity confounded our findings, these models should be interpreted with caution given the inclusion of the same variables in both the dependent and independent variables. Additionally, the factors included in our sensitivity analyses are only a small selection of those that could potentially impact plea discounts, and there is an ever-present possibility for unmeasured characteristics to confound our observed relationships.

Lastly, the size of our plea-bargaining sample was sufficiently large for aggregate tests, but there were small sample sizes for certain combinations of demographic characteristics. As such, it remains possible that our observed interaction effects were at least partly influenced by this small number of observations. Once again, sensitivity analyses using the natural log of the plea discount estimates returned substantively similar results, suggesting that these findings were robust. However, limited samples of demographic subgroups may still be nonrepresentative of the population of these subgroups more broadly. Relatedly, our plea-bargaining sample was drawn from a single mid-sized jurisdiction in Virginia. It is not clear whether these results would generalize to other jurisdictions in Virginia, or other states in general.

Despite these limitations, this study provides an important contribution to the literature on guilty pleas. Our results suggest that plea discounts are influenced by case evidence to a limited degree, but that the influence of non-legal characteristics is both large and robust. In other words, evidence strength does not appear to be the primary driver of the plea discount. It is not clear what mechanisms lead to these outcomes, but theories such as the “shadow of the trial” and focal concerns provide important insight. It could be that legal participants act on both rational considerations, such as the likelihood of conviction at trial based on the strength of the evidence, and biased considerations, such as the perceived dangerousness of a defendant based on their race/ethnicity and sex. Alternatively, it could be that each of these considerations become components of the perceived convictability of a case (see generally Albonetti 1987; Frohmann 1997). More research is certainly needed, but given how fundamental the plea discount is to current criminal justice practice, it is important to start questioning just who this discount is really for.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10940-022-09555-8>.

Acknowledgements This research was in part supported by the National Science Foundation and the National Institute of Justice. We are particularly grateful to Shawn Bushway, Jodi Quas, Amy Dezember, Skye Woestehoff, Megan Stoltz, and Melissa Manley. We also are indebted to the county court for their cooperation.

Funding This work was partially supported by the National Science Foundation (Grant #1603944) and the National Institute of Justice.

References

- Agresti A (2019) An introduction to categorical data analysis, 3rd edn. John Wiley and Sons, Hoboken, NJ
- Albonetti CA (1987) Prosecutorial discretion: the effects of uncertainty. *Law Soc Rev.* <https://doi.org/10.2307/3053523>
- Albonetti CA (1990) Race and the probability of pleading guilty. *J Quant Criminol* 6(3):315–334
- Alschuler AW (1968) The prosecutor’s role in plea bargaining. *Univ Chicago Law Rev* 36(1):50–112. <https://doi.org/10.2307/1598832>
- Alschuler AW (1981) The changing plea bargaining debate. *Calif Law Rev* 69:652

- Bartlett JM, Zottoli TM (2021) The paradox of conviction probability: mock defendants want better deals as risk of conviction increases. *Law Hum Behav* 45(1):39
- Bibas S (2004) Plea bargaining outside the shadow of trial. *Harv Law Rev* 117(8):2463–2547. <https://doi.org/10.2307/4093404>
- Brereton D, Casper JD (1981) Does it pay to plead guilty? Differential sentencing and the functioning of criminal courts. *Law Soc Rev* 16(1):45–70. <https://doi.org/10.2307/3053549>
- Bushway SD, Redlich AD (2012) Is plea bargaining in the “shadow of the trial” a mirage? *J Quant Criminol* 28(3):437–454. <https://doi.org/10.1007/s10940-011-9147-5>
- Bushway SD, Redlich AD, Norris RJ (2014) An explicit test of plea bargaining in the “shadow of the trial.” *Criminology* 52(4):723–754. <https://doi.org/10.1111/1745-9125.12054>
- Devine DJ, Buddenbaum J, Houps S, Studebaker N, Stolle DP (2009) Strength of evidence, extra evidentiary influence, and the liberation hypothesis: data from the field. *Law Hum Behav* 33(2):136–148. <https://doi.org/10.1007/s10979-008-9144-x>
- Devine DJ, Clayton LD, Dunford BB, Seying R, Pryce J (2001) Jury decision making: 45 years of empirical research on deliberating groups. *Psychol Public Policy Law* 7(3):622
- Dezember A, Luna S, Woestehoff SA, Stoltz M, Manley M, Quas JA, Redlich AD (2022) Plea validity in circuit court judicial colloquies in misdemeanor vs. felony charges. *Psychol Crime Law* 28:268–288
- Edkins VA (2011) Defense attorney plea recommendations and client race: does zealous representation apply equally to all? *Law Hum Behav* 35(5):413–425. <https://doi.org/10.1007/s10979-010-9254-0>
- Eisenberg T, Hannaford-Agor PL, Hans VP, Waters NL, Munsterman GT, Schwab SJ, Wells MT (2005) Judge-jury agreement in criminal cases: a partial replication of Kalven and Zeisel’s *The American Jury*. *J Empir Leg Stud* 2(1):171–207
- Elder HW (1989) Trials and settlements in the criminal courts: an empirical analysis of dispositions and sentencing. *J Leg Stud* 18(1):191–208
- Emmelman DS (1998) Gauging the strength of evidence prior to plea bargaining: the interpretive procedures of court-appointed defense attorneys. *Law Soc Inq* 22:927–955. <https://doi.org/10.1111/j.1747-4469.1997.tb01093.x>
- Federal Bureau of Investigation. (2016). Uniform crime report statistics, 2016. US Department of Justice.
- Flowe HD, Mehta A, Ebbesen EB (2011) The role of eyewitness identification evidence in felony case dispositions. *Psychol Public Policy Law* 17(1):140
- Fox J (2016) *Applied regression analysis and generalized linear models*, 3rd edn. Sage Publications Inc., Thousand Oaks, CA
- Frederick, B., and Stemen, D. (2012). *The anatomy of discretion: An analysis of prosecutorial decision-making*: Tech. Rep. No. Washington, DC: U. S. Department of Justice, National Institute of Justice.
- Freiburger TL, Jordan KL, Hilinski-Rosick CM (2019) A multivariate analysis of incarceration and sentence length decisions for older defendants. *Crim Justice Policy Rev* 30(7):1064–1085
- Frenzel ED, Ball JD (2008) Effects of individual characteristics on plea negotiations under sentencing guidelines. *J Ethnicity in Crim Justice* 5(4):59–82
- Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33(1):1–22
- Frohmann L (1997) Convictability and discordant locales: Reproducing race, class, and gender ideologies in prosecutorial decisionmaking. *Law Soc Rev* 31(3):531
- Garvey SP, Hannaford-Agor P, Hans VP, Mott NL, Munsterman GT, Wells MT (2004) Juror first votes in criminal trials. *J Empir Leg Stud* 1:371–399
- Goeman J, Meijer R, Chaturvedi N (2018) L1 and L2 penalized regression models. *Vignette R Package Penalized*. URL <http://cran.nedmirror.nl/web/packages/penalized/vignettes/penalized.pdf>.
- Grunwald B (2020) Distinguishing plea discounts and trial penalties. *Georgia State Univ Law Rev* 37(2):261–304
- Hastie T, Qian J, Tay K (2016) An introduction to glmnet. <https://cloud.r-project.org/web/packages/glmnet/vignettes/glmnet.pdf>
- Hayes AF, Krippendorff K (2007) Answering the call for a standard reliability measure for coding data. *Commun Methods Meas* 1(1):77–89. <https://doi.org/10.1080/19312450709336664>
- Heller KJ (2006) The cognitive psychology of circumstantial evidence. *Mich Law Rev* 105:241–305
- Henderson KS (2021) Examining the effect of case and trial factors on defense attorneys’ plea decision-making. *Psychology, Crime and Law* 27(4):357–382
- Hoerl AE, Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1):55–67
- Johnson BD, Larroulet P (2019) The “distance traveled”: Investigating the downstream consequences of charge reductions for disparities in incarceration. *Justice Q* 36(7):1229–1257. <https://doi.org/10.1080/07418825.2018.1529250>

- Kassin SM (2012) Why confessions trump innocence. *Am Psychol* 67:431–445. <https://doi.org/10.1037/a002812>
- Kassin SM, Neumann K (1997) On the power of confession evidence: an experimental test of the fundamental difference hypothesis. *Law Hum Behav* 21:469–484
- Kelly WR, Pitman R (2018) *Confronting underground justice: Reinventing plea bargaining for effective criminal justice reform*. Rowan and Littlefield, Lanham, MD.
- Kim A (2015) Underestimating the trial penalty: an empirical analysis of the federal trial penalty and critique of the Abrams study. *Mississippi Law J* 84(5):1195–1256
- Kipnis K (1976) Criminal justice and the negotiated plea. *Ethics* 86(2):93–106
- Koehler JJ (2001) When are people persuaded by DNA match statistics? *Law Hum Behav* 25:493–513. <https://doi.org/10.1023/A:1012892815916>
- Kramer GM, Wolbransky M, Heilbrun K (2007) Plea bargaining recommendations by criminal defense attorneys: Evidence strength, potential sentence, and defendant preference. *Behav Sci Law* 25(4):573–585. <https://doi.org/10.1002/bsl.759>
- Kramer JH, Ulmer JT (2009) *Sentencing guidelines: lessons from Pennsylvania*. Lynne Rienner Publishers, Boulder, CO
- Kurlychek MC, Johnson BD (2019) Cumulative disadvantage in the American criminal justice system. *Ann Rev Criminol* 2:291–319
- Kutateladze BL, Andiloro NR, Johnson BD, Spohn CC (2014) Cumulative disadvantage: examining racial and ethnic disparity in prosecution and sentencing. *Criminology* 52(3):514–551
- Kutateladze BL, Lawson VZ, Andiloro NR (2015) Does evidence really matter? An exploratory analysis of the role of evidence in plea bargaining in felony drug cases. *Law and Human Behavior* 39(5):431–442
- Lafler v. Cooper 132 S. Ct. 1376 (2012). <https://www.supremecourt.gov/opinions/11pdf/10-209.pdf>
- LaFree GD (1985) Adversarial and nonadversarial justice: a comparison of guilty pleas and trials. *Criminology* 23(2):289–312. <https://doi.org/10.1111/j.1745-9125.1985.tb00338.x>
- Landes WM (1971) An economic analysis of the courts. *J Law and Econom* 14(1):61–107. <https://doi.org/10.1086/466704>
- Light MT (2014) The new face of legal inequality: Noncitizens and the long-term trends in sentencing disparities across US district courts, 1992–2009. *Law Soc Rev* 48(2):447–478
- Light MT, Massoglia M, King RD (2014) Citizenship and punishment: the salience of national membership in US criminal courts. *Am Sociol Rev* 79(5):825–847
- Lipsey MW, Wilson DB (2001) *Practical meta-analysis*. Sage Publications Inc, Thousand Oaks, CA
- Luna S, Redlich AD (2020) Unintelligent decision-making? the impact of discovery on defendant plea decisions. *Wrongful Conviction Law Rev* 1:314
- Marquardt DW, Snee RD (1975) Ridge regression in practice. *Am Stat* 29(1):3–20
- McAllister HA, Bregman NJ (1986) Plea bargaining by prosecutors and defense attorneys: a decision theory approach. *J Appl Psychol* 71(4):686
- McCormick CT (1972) *Handbook of the law of evidence*, 2nd edn. West, St. Paul, MN
- McCoy T, Salinas PR, Walker JT, Hignite L (2012) An examination of the influence of strength of evidence variables in the prosecution’s decision to dismiss driving while intoxicated cases. *Am J Crim Justice* 37(4):562–579. <https://doi.org/10.1007/s12103-011-9141-3>
- McDonald GC (2009) Ridge regression. *Wiley Interdiscip Rev Comput Stat* 1(1):93–100. <https://doi.org/10.1002/wics.14>
- Melkumova LE, Shatskikh SY (2017) Comparing ridge and lasso estimators for data analysis. *Procedia Eng* 201:746–755. <https://doi.org/10.1016/j.proeng.2017.09.615>
- Miles J (2005) Tolerance and variance inflation factor. In: Everitt BS, Howell DC (eds) *Encyclopedia of statistics in behavioral science*. John Wiley and Sons, Hoboken, NJ, pp 2055–2056
- Miller HS, McDonald WF, Cramer JA (1980) *Plea bargaining in the United States*. United States Department of Justice, National Institute of Justice, Washington
- Mnookin RH, Kornhauser L (1979) Bargaining in the shadow of the law: The case of divorce. *Yale Law J* 88(5):950–997. <https://doi.org/10.2307/795824>
- Niedermeier KE, Kerr NL, Messe LA (1999) Jurors’ use of naked statistical evidence: exploring bases and implications of the Wells Effect. *J Pers Soc Psychol* 76:533–554
- Omori M, Petersen N (2020) Institutionalizing inequality in the courts: decomposing racial and ethnic disparities in detention, conviction, and sentencing. *Criminology* 58(4):678–713. <https://doi.org/10.1111/1745-9125.12257>
- Padilla v. Kentucky, 130 S. Ct., 1472 (2010). <https://www.supremecourt.gov/opinions/09pdf/08-651.pdf>

- Petersen K, Redlich AD, Norris RJ (2020) Diverging from the shadows: explaining individual deviation from plea bargaining in the “shadow of the trial.” *J Exp Criminol*. <https://doi.org/10.1007/s11292-020-09449-4>
- Piehl AM, Bushway SD (2007) Measuring and explaining charge bargaining. *J Quant Criminol* 23(2):105–125. <https://doi.org/10.1007/s10940-006-9023-x>
- Porter L (2020) Top trends in state criminal justice reform, 2019. The Sentencing Project. <https://www.sentencingproject.org/publications/top-trends-in-state-criminal-justice-reform-2019/>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Redlich AD, Bushway SD, Norris RJ (2016) Plea decision-making by attorneys and judges. *J Exp Criminol* 12(4):537–561. <https://doi.org/10.1007/s11292-016-9264-0>
- Redlich AD, Bibas S, Edkins VA, Madon S (2017) The psychology of defendant plea decision making. *Am Psychol* 72(4):339
- Redlich AD, Yan S, Norris RJ, Bushway SD (2018) The influence of confessions on guilty pleas and plea discounts. *Psychol Public Policy Law* 24(2):147–157. <https://doi.org/10.1037/law0000144>
- Redlich AD, Zottoli T, Dezember A, Schneider R, Catlin M, Shammi S (in press) Emerging issues in the psycho-legal study of guilty pleas. In DeMatteo D, Scherr K (eds) *The oxford handbook of psychology and law*. Oxford University Press, New York, NY
- Reitz KR (1993) Sentencing facts: Travesties of real-offense sentencing. *Stanford Law Review* 45(3):523–574
- Rhodes WM (1979) Plea bargaining: Its effect on sentencing and convictions in the District of Columbia. *J Crim Law Criminol* 70(3):360–375
- Schklar J, Diamond SS (1999) Juror reactions to DNA evidence: errors and expectancies. *Law Hum Behav* 23:159–184. <https://doi.org/10.1023/A:1022368801333>
- Schmidt J, Steury EH (1989) Prosecutorial discretion in filing charges in domestic violence cases. *Criminology* 27(3):487–510
- Silge J, Chow F, Kuhn M (2021) Rsample: General resampling infrastructure. R package version 0.1.0. <https://CRAN.R-project.org/package=rsample>
- Smith DA (1986) The plea bargaining controversy. *J Crim Law Criminol* 77(3):949–968. <https://doi.org/10.2307/1143445>
- Smith CM, Goldrosen N, Ciocanel MV, Santorella R, Topaz CM, Sen S (2021) Racial disparities in criminal sentencing vary considerably across federal judges. Institute for the Quantitative Study of Inclusion, Diversity, and Equity, Inc., 501(c)3. <https://osf.io/preprints/socarxiv/>
- Spohn C (2018) Reflections on the exercise of prosecutorial discretion 50 years after publication of the challenge of crime in a free society. *Criminol Public Policy* 17(2):321–340. <https://doi.org/10.1111/1745-9133.12367>
- Spohn C, Holleran D (2001) Prosecuting sexual assault: a comparison of charging decisions in sexual assault cases involving strangers, acquaintances, and intimate partners. *Justice Q* 18(3):651–688
- Steffensmeier D, Ulmer J, Kramer J (1998) The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, black, and male. *Criminology* 36(4):763–798. <https://doi.org/10.1111/j.1745-9125.1998.tb01265.x>
- Tabachnick BG, Fidell LS (2007) *Using multivariate statistics*, 5th edn. Pearson, Boston, MA
- Taylor TS, Hosch HM (2004) An examination of jury verdicts for evidence of a similarity-leniency effect, an out-group punitiveness effect or a black sheep effect. *Law Hum Behav* 28(5):587–598
- Ternès N, Rotolo F, Michiels S (2016) Empirical extensions of the lasso penalty to reduce the false discovery rate in high-dimensional Cox regression models. *Stat Med* 35(15):2561–2573
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J Roy Stat Soc: Ser B (methodol)* 58(1):267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Tonry M (1988) Sentencing guidelines and the model penal code. *Rutgers Law J* 19(3):823–848
- Ulmer JT, Bader C, Gault M (2008) Do moral communities play a role in criminal sentencing? Evidence from Pennsylvania the *Soc Q* 49(4):737–768
- Ulmer JT, Bradley MS (2006) Variation in trial penalties among serious violent offenses. *Criminology* 44(3):631–670. <https://doi.org/10.1111/j.1745-9125.2006.00059.x>
- Ulmer JT, Eisenstein J, Johnson BD (2010) Trial penalties in federal sentencing: extra-guidelines factors and district variation. *Justice Q* 27(4):560–592. <https://doi.org/10.1080/07418820902998063>
- University of Virginia (2018) Population estimates for age, sex, race & Hispanic, and towns. <https://demographics.coopercenter.org/population-estimates-age-sex-race-hispanic-towns>
- Wells GL (1992) Naked statistical evidence of liability: Is subjective probability enough? *J Pers Soc Psychol* 62:739–752

- Winship C, Mare RD (1992) Models for sample selection bias. *Ann Rev Sociol* 18(1):327–350
- Yan S (2020) What exactly is the bargain? The sensitivity of plea discount estimates. *Justice Q*. <https://doi.org/10.1080/07418825.2019.1707856>
- Yan S, Bushway SD (2018) Plea discounts or trial penalties? Making sense of the trial-plea sentence disparities. *Justice Q* 35(7):1226–1249. <https://doi.org/10.1080/07418825.2018.1552715>
- Yoo W, Mayberry R, Bae S, Singh K, He QP, Lillard JW Jr (2014) A study of effects of multicollinearity in the multivariable analysis. *Int J Appl Sci Technol* 4(5):9
- Zottoli TM, Daftary-Kapur T, Winters GM, Hogan C (2016) Plea discounts, time pressures, and false-guilty pleas in youth and adults who pleaded guilty to felonies in new york city. *Psychol Public Policy Law* 22(3):250–259. <https://doi.org/10.1037/law0000095>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.