

EXAMINING INTERJUDGE PUNISHMENT DISPARITIES AND JUDICIAL SENTENCING  
PATTERNS WITHIN COURT COMMUNITIES

By

Michael B. Cassidy

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Criminal Justice – Doctor of Philosophy

2017

## **ABSTRACT**

### **EXAMINING INTERJUDGE PUNISHMENT DISPARITIES AND JUDICIAL SENTENCING PATTERNS WITHIN COURT COMMUNITIES**

By

Michael B. Cassidy

The current study examines individual judges' punishment decisions and sentencing patterns within court communities. Focal concerns perspective states that judges consider offender blameworthiness, community threat, and practical constraints when sentencing offenders, and assessment of the focal concerns is likely to vary across judges. In part, differences are due to judges' subjective decision-making, but additional theories suggest sentencing outcomes are also influenced by the court community in which punishment decisions occur. Extant research generally relies on multilevel models to assess interjudge variation, but more recent work indicates multilevel analysis obscures variation at the judge-level, and provides limited information about how individual judges within court communities consider offender and case characteristics.

The current work offers a more comprehensive examination of interjudge disparity and judicial sentencing patterns within court communities. The research uses seven years (2004-2010) of data collected from the Pennsylvania Commission on Sentencing and a sample of large, medium, and small courts. Findings from multilevel models and individual judge regression models show individual judge analyses provide a better understanding of variation across judges, and whether differences associated with key predictors of punishment are meaningful. The current research also finds little consistency in the ways judges in the same court communities consider extralegal factors in sentencing decisions. This work highlights the need to further

develop theories to explain why offender and case characteristics influence punishment decisions for some judges, but not others, and the role court communities play in sentencing decisions.

## ACKNOWLEDGEMENTS

I would like to thank Dr. Carole Gibbs, my graduate advisor and committee chairperson, who has provided invaluable guidance and support since the first day I entered the doctoral program. You offered me a number of opportunities to learn and grow as a scholar, and I would not have been able to complete this dissertation without you. I would also like to thank my committee members Dr. Steven Chermak and Dr. J. Kevin Ford for their insightful feedback throughout this process, and Dr. Chris Melde for his statistical and methodological advice over the years.

To the friends I made at Michigan State, Jason Rydberg, Rebecca Stone, Kimberly Bender, John Hakola, Alexis Norris, Julie Yingling, Derrick Franke, Gio Circo, Ellen Jesmok, Brianna Bermudez, Alexandria Anstett, and Alyssa Badgley, thank you for making graduate school a truly enjoyable experience. I also want to thank Brian Durant for his friendship and support.

Finally, I owe the greatest thanks to my family. My parents, Wayne and Roseanne, grandmother Anna, sister Laura, brother-in-law Steve, and nieces Charlotte and Audrey have provided constant love, inspiration, and encouragement.



## TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: REVIEW OF THE LITERATURE	6
Early Theories of Sentencing	6
Recent Literature	7
Court Communities and Focal Concerns	10
Sentencing Variation within Court Communities	13
Summary	15
Purpose of the Research	16
Research Questions and Hypotheses	22
CHAPTER 3: DATA AND METHODOLOGY	25
Data	25
Data Reduction and Missing Data	25
Sample Selection	26
Dependent and Independent Variables	28
Control Variables	29
Analytic Strategy	30
CHAPTER 4: RESULTS	34
Descriptive Statistics	34
Multilevel Analysis	36
Individual Judge Analysis	41
Incarceration Models	41
Sentence Length Models	50
Analysis of Judges within Court Communities	64
CHAPTER 5: DISCUSSION	85
The Current Inquiry	85
Theoretical and Methodological Implications	86
Multilevel Analysis of Judge Variation	86
Individual Analysis of Judge Variation	87
Sentencing within Court Communities	92
Implications for Policy	94
Limitations	96
Directions for Future Research	97
REFERENCES	101

## LIST OF TABLES

Table 1. Cases and Number of Judges by Court	28
Table 2. Descriptive Statistics for Overall Sample	35
Table 3. Unconditional Models of Incarceration and Sentence Length	37
Table 4. Random Coefficients Models of Incarceration and Sentence Length	38
Table 5. Number of Judges with Significant Effects for Sentencing Outcomes	64
Table 6. Percent of Large Court Judges with Significant Effects	68
Table 7. Percent of Medium Court Judges with Significant Effects	75
Table 8. Percent of Small Court Judges with Significant Effects	83

## LIST OF FIGURES

Figure 1. HGLM and Logit Effects for Offense Severity	42
Figure 2. HGLM and Logit Effects for Prior Record	44
Figure 3. HGLM and Logit Effects for Age	45
Figure 4. HGLM and Logit Effects for Female Offenders	47
Figure 5. HGLM and Logit Effects for Black Offenders	48
Figure 6. HGLM and Logit Effects for Trial Convictions	49
Figure 7. LMM and OLS Effects for Offense Severity	51
Figure 8. LMM and OLS Effects for Prior Record	52
Figure 9. LMM and OLS Effects for Age	54
Figure 10. LMM and OLS Effects for Female Offenders	55
Figure 11. LMM and OLS Effects for Black Offenders	57
Figure 12. LMM and OLS Effects for Trial Convictions	58
Figure 13. Logit and OLS Effects for Age	60
Figure 14. Logit and OLS Effects for Female Offenders	61
Figure 15. Logit and OLS Effects for Black Offenders	62
Figure 16. Logit and OLS Effects for Trial Convictions	63
Figure 17. Individual Judge Effects in Large Court 2	66
Figure 18. Individual Judge Effects in Medium Court 3	70
Figure 19. Individual Judge Effects in Medium Court 6	72
Figure 20. Individual Judge Effects in Small Court 5	76
Figure 21. Individual Judge Effects in Small Court 9	78



## **CHAPTER 1: INTRODUCTION**

Prior to the 1970s, sentencing systems at both the state and federal level offered judges a wide range of punishment options that could be tailored to individual offenders' specific needs (MacKenzie, 2001; Reitz, 1998). Providing judges with nearly unfettered discretion, however, resulted in unwarranted race and gender disparities in criminal sanctions (Tonry, 1996). Calls for reforms that would promote uniformity and fairness led several states and the federal government to enact sentencing guidelines (Reitz, 1998). Guidelines are designed to limit judicial discretion by providing a sentencing range based on the legally relevant factors of offense severity and prior criminal history, thus ensuring that similarly situated offenders receive similar punishment outcomes (Miethe & Moore, 1985; Tonry, 1996).

Over the last few decades, scholars have developed a substantial body of literature examining the key predictors of sentencing under guidelines systems. Research consistently shows that legal factors are the primary determinants of sentence severity, but disparities based on extralegal factors (e.g., age, race/ethnicity, and gender) remain (for reviews, see Pratt, 1998; Spohn, 2000; Ulmer, 2012; Zatz, 1987, 2000). In recent years, the dominant theoretical framework used to explain sentencing decisions is the focal concerns perspective (Hartley, Maddan, & Spohn, 2007; Kramer & Ulmer, 2009). Focal concerns states that judges' sentencing decisions are based on assessments of offender blameworthiness, protection of the community, and practical constraints. Judges rely primarily on legal factors (e.g., offense severity, prior record) when assessing the focal concerns, but also engage in subjective decision-making based on attributions associated with offender age, race/ethnicity, and gender when determining the appropriate punishment (Kramer & Ulmer, 2009).

Focal concerns theorists also recognize the factors judges consider when assessing the focal concerns, as well as the weight afforded to these factors, is likely to vary across judges (Kramer & Ulmer, 2009). In part, this is due to differences in judges' subjective decision-making, but additional theories suggest that judges' decisions are also influenced by the court community in which punishment decisions occur. Court communities consist of court actors who share a common workplace and develop unique case processing and sentencing norms (Eisenstein, Flemming, & Nardulli, 1988; Kramer & Ulmer, 2009). Extant work also notes, however, that court size plays an important role in shaping court communities. For example, court actor autonomy is highest in large courts, moderate in medium courts, and lowest in small courts (Eisenstein, Flemming, & Nardulli, 1988). As a result, the effects of key predictors of sentencing may differ across courts (Kautt, 2002; Kramer & Ulmer, 2009), though this variation may be conditioned by court size.

Some studies have used the focal concerns perspective to explore variation in judges' subjective sentencing decisions, and findings from multilevel analyses show legal and extralegal effects vary significantly across judges (Anderson & Spohn, 2010; Johnson, 2006; Wooldredge, 2010). However, Wooldredge (2010) compared results from multilevel models and individual judge regression models and found that multilevel analyses masked extralegal effects found in the individual judge models. As such, Wooldredge (2010) highlighted the need for additional judge-level analyses to gain a better understanding of interjudge disparity. Additional research suggests that differences in judges' subjective punishment decisions can be explained, at least in part, by the court community context in which sentencing occurs (Ulmer, 1997). Early studies showed that the influence of court communities may vary based on court size, but this work is limited to examination of a small number of courts, and primarily focused on court processes

(e.g. docket management, case assignment) as opposed to sentencing decisions (Eisenstein, Flemming, & Nardulli, 1988; Eisenstein & Jacob, 1977). More recent work employing multilevel analyses of offenders nested within courts offers support for the court community perspective, finding that legal and extralegal effects vary significantly *across* courts (Britt, 2000; Kautt, 2002; Kramer & Ulmer, 2009; Ulmer & Johnson, 2004). However, these studies provide little information about how individual judges *within* court communities consider offender and case characteristics when determining the appropriate sentence. Thus, while prior work has begun to explore interjudge punishment disparity and the influence of court communities on sentencing decisions, a more comprehensive examination is needed.

The current work fills this gap in the prior literature to further advance knowledge of interjudge disparity and judicial sentencing patterns within court communities. Using seven years (2004-2010) of data collected from the Pennsylvania Commission on Sentencing and a sample of large, medium, and small courts, the current work employs multilevel modeling to replicate the prior work that shows variation in offender and case characteristics across judges. Next, the present study uses individual judge regression models to examine judges' contributions to the legal and extralegal effects found in the multilevel analysis. Finally, analysis of individual judges grouped by court will be used to assess whether and how judges in the same court communities consider legal and extralegal factors. This research extends previous work (Wooldredge, 2010) using multilevel analysis and individual judge models to assess differences in findings produced by these methodological approaches, and is the first to examine individual judges' sentencing patterns within a relatively large sample of court communities.

This work is important from a theoretical and practical standpoint. Focal concerns theory suggests that judges will vary in the specific factors considered in sentencing decisions.

However, pooling of judge estimates for legal and extralegal factors in multilevel random effects models is contrary to the idea that judges vary in their subjective assessments of focal concerns. Further, contemporary sentencing theories are not only concerned with whether judges vary in legal and extralegal effects, but also whether these factors matter in punishment decisions. Yet, multilevel random effects models may not be appropriate when predictors are expected to produce effects for some groups, but not others (Gelman & Hill, 2016). Thus, the current work incorporates individual judge models to more directly examine interjudge variation, and assess whether differences between judges are meaningful. This kind of analysis is a necessary first step in gaining a better understanding of the extent to which judges vary, which can then be followed by further developing theories to explain this variation (see Ulmer, 2012; Wooldredge, 2010).

In addition, much of the extant research on focal concerns either ignores whether court communities affect judges' subjective decisions, or concludes that variation across courts is indicative of court context influencing judges' punishment decisions (Kramer & Ulmer, 2009). However, findings from the latter provide limited information about what is driving this variation. For example, differences across court communities may be the result of similar sentencing patterns among judges within courts, similarities for judges in some courts but not others, and/or differences between judges across courts. As such, the current work explores the significance of court communities on judges' subjective sentencing decisions, and whether varying levels of court actor autonomy present in large, medium, and small courts conditions this relationship.

By testing the focal concerns and court community perspectives using the analytic strategies outlined above, the current work offers implications for theory. Differences between findings from the multilevel analysis and the individual judge regression models may indicate



that analysis at the judge level, as opposed to the jurisdiction or state, is necessary to gain a better understanding of how judges consider legal and extralegal factors when assessing the focal concerns (Wooldredge, 2010). In addition, consistency among judges within the same courts may suggest the court community influences punishment decisions, whereas interjudge disparity in large courts versus small courts, for example, may indicate court contextual influences are conditioned by court size. Conversely, substantial variation within all courts could suggest that court communities influence some processes (e.g., docket management, case assignment), but have less of an impact in the punishment phase. Differences within courts would also indicate a need for further theoretical development to explain why some judges rely on certain legal and extralegal factors more than others when determining the appropriate sentence.

The current work also has potential implications for sentencing law and policy. Sentencing guidelines were developed to reduce unwarranted disparity and increase uniformity in punishment (Kramer & Scirica, 1986). Examining individual judges' sentencing decisions will provide some indication of whether guidelines' have achieved these goals. To the extent disparity continues to exist, implications may include additional training for judges on guidelines implementation, more rigorous research to assist policy makers with identifying sources of disparity, and stricter appellate review standards.

## CHAPTER 2: REVIEW OF THE LITERATURE

### *Early Theories of Sentencing*

Early theories on the influence of legal and extralegal factors on sentencing decisions were based on macro-level relationships between law and society. Conflict theory states that the social and political structures of society are the result of conflict between the ruling class and those with little or no power (Chambliss & Seidman, 1982). Criminal law reflects the empowered class's attempt to continue its political and social dominance, and the courts "tend to produce solutions in the interest of the wealthy" (Chambliss & Seidman, 1982: 237). Thus, conflict theory predicts that extralegal factors such as race/ethnicity, gender, and socioeconomic status play a significant role in sentencing outcomes (Lizotte, 1978). In contrast, consensus theory posits that laws represent broadly shared societal norms and values (Dixon, 1995). These laws govern the sentencing process and are applied uniformly to cases, irrespective of offender class or status (Chiricos & Waldo, 1975). As a result, consensus theory suggests that legal factors, such as offense severity and criminal history, are the primary determinants of sentencing outcomes (Dixon, 1995).

While conflict and consensus theories were useful for framing much of the research in the 1970s and 1980s on the effects of offender and case characteristics on sentencing outcomes, neither perspective garnered significant empirical support (Hagan, 1989). Studies testing conflict theory often failed to properly control for legally relevant variables (e.g., offense severity and prior criminal record) (Hagan, 1974; Kleck, 1981). When these variables were included in the analyses, some researchers found that race and ethnicity effects either did not exist or were inconsequential (Kleck, 1981; Kleck, 1985; Wilbanks, 1987). Yet, reviews of subsequent studies using newer data and more sophisticated methodology raised questions about whether legal

factors fully moderated extralegal effects (Spohn, 2000; Zatz, 1987, 2000). Zatz (1987: 70) argued that even when controlling for offense severity and criminal history, data from determinate sentencing systems, including guidelines, “show subtle if no longer overt bias against minority defendants.” These findings suggest that conflict and consensus theories, which link sentencing outcomes to *either* extralegal *or* legal factors based on macro-level relationships between law and society are too limited to explain the complex nature of judicial decision-making. Consequently, sentencing scholars developed theoretical perspectives that consider the influence of both legal and extralegal factors on sentencing outcomes.

### *Recent Literature*

Albonetti’s (1991) attribution theory of judicial decision-making states that fully rational decision-making is only possible when the decision-maker can accurately identify all of the potential benefits, costs, and alternatives associated with the decision. Since decision-makers rarely have access to all of this information, they are forced to engage in a process characterized by “bounded rationality,” where they search for a solution that will limit the uncertainty of obtaining the desired outcome (Albonetti, 1991: 249). Albonetti (1991) theorized that judges operate under bounded rationality because information about offenders is often incomplete and contradictory. As a result, judges develop decision-making shortcuts or “patterned responses” to address uncertainty (Albonetti, 1991: 17). These patterned responses are the result of judicial attributions of offenders’ recidivism risk and rehabilitation potential. In contrast to conflict and consensus theories, Albonetti (1991) suggested that judicial attributions are influenced by both legally relevant variables and offender characteristics. For example, having a prior record is likely to increase sentence severity because it triggers an attribution of a “stable and enduring offender disposition to commit future criminal activity” (Albonetti, 1991: 257). Similarly,

judges may impose harsher punishments for certain offenders based on attributions linking race and gender stereotypes with recidivism risk and rehabilitation potential (Albonetti, 1991, 1997, 2002).

Other developments in theories of sentencing take into account the shift by several states and the federal government from an indeterminate system of punishment to a determinate structure, which often includes sentencing guidelines. Drawing on the work of Max Weber, Savelsberg (1992) argued that sentencing guidelines attempt to balance two competing interests: formal rationality and substantively rational decision-making. Formal rationality refers to the laws, policies, and sentencing ranges outlined in the guidelines, which limit judicial discretion and promote uniformity in punishment outcomes (Savelsberg, 1992). However, sentencing has traditionally been a substantively rational individualized process, where punishment decisions are guided by judicial consideration of individual offenders' characteristics, needs, or circumstances (Savelsberg, 1992; Ulmer & Kramer, 1996). Under sentencing guidelines, formal rationality provides judges with the guidelines range, but uncertainty remains over selecting the appropriate sentence within that range. Judges engage in substantively rational decision-making based on assessments of individual offenders when selecting the actual sentence, which may result in disparate punishment for similarly situated offenders (i.e., those with similar prior records convicted of similar crimes) (Kramer & Ulmer, 2009).

Sentencing scholars have drawn from Albonetti's (1991) and Savelsberg's (1992) work to develop the focal concerns perspective. Similar to Albonetti (1991), focal concerns contends that sentencing outcomes are the result of a multifaceted and complex decision-making process, where judges make attributions based on assessments of offender blameworthiness, protection of the community, and practical constraints (Kramer & Ulmer, 2009; Steffensmeier, Ulmer, &

Kramer, 1998). Blameworthiness stems from a retributive philosophy of punishment, and is associated with offender culpability and the amount of harm caused. Common factors that influence blameworthiness are offense severity, the offender's role in the crime, and prior victimization of the offender. Protection of the community draws from incapacitation and deterrence philosophies of punishment. Consequently, court actors make attributions about offenders' future behavior based on the crime of conviction (e.g., violent versus property), prior criminal history, and stereotypes that suggest certain offenders pose a greater threat to the community. The third focal concern addresses practical constraints, which consist of ensuring regular case flow, relationships among courtroom actors, and assessment of criminal justice system resources, such as local jail capacity (Kramer & Ulmer, 2009; Ulmer & Johnson, 2004).

In line with Savelsberg's (1992) conceptualization of formal and substantive rationality, proponents of focal concerns recognize that formal sentencing criteria (e.g., sentencing guidelines) are the primary determinants of sentencing outcomes, but judges also engage in substantively rational decision-making (Kramer & Ulmer, 2009). Thus, assessments of the focal concerns are mostly influenced by offense severity and prior criminal history, but attributions linked to offender demographics and social status also play a role in punishment decisions. Several studies have applied the focal concerns framework when researching sentencing decision-making, and findings are consistent with the tenets of this perspective (Doerner & Demuth, 2010; Kramer & Ulmer, 2009; Steffensmeier & Demuth, 2001, 2006; Steffensmeier, Ulmer, & Kramer, 1998). Offense severity and criminal history are the primary factors judges consider when assessing blameworthiness and protection of the community, but extralegal factors also influence sentencing outcomes. In particular, these studies consistently show that female offenders are less likely to be incarcerated and receive shorter sentences than male

offenders, and offenders convicted after trial are punished more harshly than those that enter a guilty plea. Results concerning offender age and race/ethnicity are somewhat less consistent, but when significant effects are found they generally show younger offenders and black/Hispanic offenders are sentenced more severely than older offenders and white offenders, respectively. However, proponents of this perspective also note that judges are likely to vary in the ways they assess the focal concerns, as well as the factors they use in their assessments. Though judicial “reliance on the three concerns is said to be universal, ... the meaning, emphasis, and interpretation of them is local” (Ulmer & Johnson, 2004: 142; see also Kautt, 2002; Kramer & Ulmer, 2009). More specifically, scholars have argued that whether and how judges consider legal and extralegal factors in sentencing outcomes is influenced in part by the court community in which punishment decisions occur (Anderson & Spohn, 2010; Dixon, 1995; Johnson 2006; Kramer & Ulmer, 2009; Ulmer & Johnson, 2004).

### *Court Communities and Focal Concerns*

Theorizing regarding variation in assessment of the focal concerns is based on work that views courts as communities. Court communities consist of court actors who share a common workplace and develop working relationships (Eisenstein, Flemming, & Nardulli, 1988; Eisenstein & Jacob, 1977). The structure of status and power among participants, as well as the characteristics and values of group members, shape these relationships. Most literature on court communities explores courtroom actors’ working relationships and case processing norms (e.g., case assignment, charge bargaining) (Eisenstein, Flemming, & Nardulli, 1988; Eisenstein & Jacob, 1977), but some studies suggest that court communities also influence sentencing (Ulmer, 1997; Ulmer & Johnson, 2004; Ulmer & Kramer, 1996). Court communities develop distinctive

case processing and sentencing norms, which suggests that both sentence severity and the effects of key predictors of sentencing may vary across courts (Kautt, 2002; Kramer & Ulmer, 2009).

Overall, this literature suggests that court communities are unique, and the ways in which they shape processes and outcomes are not necessarily generalizable across multiple communities. However, Eisenstein and colleagues (1988) note that court size plays an important role in shaping court communities. In general, court actor autonomy is lowest in small courts, moderate in medium courts, and highest in large courts. Small courts (one to two judges) are composed of few judges, prosecutors, and defense attorneys and lack the resources needed for trials. Consequently, court actors work closely with one another and most cases are settled through guilty pleas. Conversely, additional resources available in large courts (at least fifteen judges) allow for more trials, and the greater number of court personnel creates an environment where mutual dependence between court actors is more fractured compared to small courts (Eisenstein, Flemming, & Nardulli, 1988). For example, Jacob's (1997) work on the largest court in Illinois (Cook County) showed a tight connection between court personnel regarding courtroom assignment, case assignment, and docket management. Concerning sentencing, however, Jacob (1997: 28) noted "[i]ndividual judges emphasize their need to exercise discretion in order to do justice. The court setting permits them to give free reign to their individual traits and invites them to render their own reading of the law; their rulings have slight if any impact on other courtrooms." Case processing and sentencing practices associated with medium-sized courts (four to fourteen judges) tend to fall somewhere between small and large courts (Eisenstein, Flemming, & Nardulli, 1988).

Qualitative research supports the notion that differences in case processing and sentencing practices are associated with court size. Interviews conducted by Ulmer (1997) and

Ulmer and Kramer (1996) with judges, prosecutors, defense counsel, probation officers, and court administrative personnel in three courts (one small, one medium, and one large) in Pennsylvania revealed substantial differences in workgroup structure and culture, sentencing goals, and guideline adherence among these courts. For example, assistant district attorneys had very little discretion to negotiate plea agreements in the small court, but a great deal of discretion in the medium-sized court (Ulmer, 1997). In the large court, discretion was related to experience (i.e., as assistant district attorneys gained experience discretion increased). Concerning sentencing goals, small court judges relied on rehabilitation, just deserts, and deterrence punishment philosophies, while large court judges focused on rehabilitation and just deserts. In the medium-sized court, deterrence, incapacitation, and just deserts influenced judges' sentencing decisions (Ulmer, 1997).

However, quantitative analysis of sentencing outcomes in these courts showed mixed results regarding extralegal effects. All three courts sentenced female offenders more leniently, and offenders convicted after trial (as opposed to pleading guilty) more harshly (Ulmer, 1997). Only the large and medium-sized courts sentenced black offenders more harshly than whites; no significant differences were found in the small court (Ulmer, 1997). This research provides some evidence of differences in legal and extralegal effects across court communities, but it is limited to only three courts and interviews with a small sample of court actors. More recent research testing the focal concerns and court community perspectives using multilevel modeling and larger samples offers additional support for variation in sentencing patterns across courts.

Several studies have used multilevel models of pooled cases with offenders nested within courts to examine variation in predictors of punishment severity across courts. In contrast to prior work that focused on a small number of courts differing in size (Ulmer, 1997; Ulmer &



Kramer, 1996), this body of research examines cases across all courts at the state or federal level. Findings from this work offer support for integrating the focal concerns and court community perspectives. For example, using Pennsylvania guidelines data, Britt (2000), Ulmer and Johnson (2004), and Kramer and Ulmer (2009) employed multilevel models with random effects that allow the slopes of legal and extralegal predictors to vary randomly across courts. Significant variance components indicated that effects associated with offense severity, prior record, age, race, gender, and mode of conviction differed significantly across courts (Britt, 2000; Kramer & Ulmer, 2009; Ulmer & Johnson, 2004). Kautt's (2002) analysis of federal guidelines data provided similar results for U.S. District Courts. Based on these findings, scholars have concluded "decisionmakers in different courts differentially weight the importance of these various individual-level case characteristics at sentencing" (Kramer & Ulmer, 2009: 129; see also Kautt, 2002; Ulmer & Johnson, 2004).

Overall, the results from multilevel analyses suggest that judicial assessments of the focal concerns, and the legal and extralegal factors used in these assessments, are conditioned by the court community in which sentencing occurs. However, while these studies are useful for gaining a better understanding of sentencing disparity *across* courts, they provide little information about whether and how judges consider legal and extralegal factors *within* these court communities.

### *Sentencing Variation within Court Communities*

Integration of the focal concerns and court community perspectives suggests that effects associated with key predictors of sentencing vary across courts, but judges within these courts may also assess legal and extralegal factors in different ways. According to Johnson and colleagues, while "court actors use legal factors such as offense seriousness and prior record as

initial punishment benchmarks [they] then make situational attributions about defendants' character and risk based on more subtle, subjective, decision-making schema" (Johnson, Ulmer, & Kramer, 2008: 745). Prior work on interjudge disparity, which "occurs when judges in the same jurisdiction sentence similarly situated offenders differently" (Kim, Spohn, & Hedberg, 2015: 5), offers support for the notion that judges engage in subjective decision-making. Johnson (2006) analyzed offenders sentenced in Pennsylvania courts using a three-level mixed model (e.g., offenders nested within judges, nested with courts) with random effects and found that effects associated with offense severity, criminal history, age, gender, and race/ethnicity varied significantly across both judges and courts. Similarly, Anderson and Spohn (2010) analyzed federal guidelines data using a model with offenders nested within judges and found significant variation between judges in three U.S. district courts for effects associated with gender, employment status, and pretrial status.

Wooldredge (2010) also used multilevel analysis for a sample of felony conviction cases in a single Ohio court. Similar to Johnson (2006) and Anderson and Spohn (2010), results from multilevel random effects models showed that with the exception of offender race, legal and extralegal effects differed significantly across judges (Wooldredge, 2010). To further advance understanding of these differences, Wooldredge (2010) employed individual judge logistic regression models to examine how judges contribute to extralegal disparities found in the overall model. The individual models revealed that significant findings from the multilevel model provided limited information about the ways in which judges consider extralegal factors in sentencing decisions. For example, six of the 18 judges showed significant effects for race in the individual models (four were positive, two were negative), despite race being non-significant in the overall model. Further, though multilevel analysis showed that males were 1.5 times more

likely to be incarcerated than females, only five of the 18 judges showed significant effects for gender, and odds ratios for these judges ranged from 2.75 to 16.78. Wooldredge (2010) concluded that pooling cases results in masked effects, which are likely to arise when judges yield null or weak effects, or effects that are significant but in opposite directions, for a given variable. Collectively, this body of literature offers support for the idea that differences among judges are the result of subjective decision-making, and provides some evidence of legal and extralegal effects varying not only across court communities, but also within court communities.

### *Summary*

Extant research consistently shows that judicial assessments of the focal concerns are primarily driven by offense severity and criminal history, but extralegal effects associated with offender age, race/ethnicity, and gender remain. More recently, scholars have argued that effects associated with key predictors of sentencing vary across court communities. This is because court communities develop distinctive case processing and sentencing norms that influence judicial consideration of legal and extralegal factors in assessing offender blameworthiness and community threat. Research offers support for integrating the focal concerns and court community perspectives, finding that effects associated with offense and offender characteristics vary significantly across courts.

Yet, other work shows that legal and extralegal effects also vary significantly *across judges* within courts. According to Ulmer (2012: 17), “[b]etween-actor variation is certainly congruent with and implied by the focal concerns ....” Though variation across judges would seem to indicate that the court community is less influential than theory suggests, “this kind of variation between judges and prosecutors, for example, does not contradict the notion that court community contexts shape actors’ decisions, and that between-actor variation occurs relative to

court community norms” (Ulmer, 2012: 17). Thus, variation between judges may suggest that court communities differ in the presence of, and adherence to, shared case processing and sentencing norms. With some research suggesting court actor autonomy differs across large, medium, and small courts (Eisenstein, Flemming, & Nardulli, 1988), this variation may be linked to the size of the court.

Though prior work has assessed variation across judges and courts using multilevel modeling, this methodological approach does not allow for a detailed examination of interjudge disparity and sentencing patterns of judges in the same court community. Only one study (Wooldredge, 2010) analyzed individual judges’ sentencing decisions in this way, but this research was limited to a single court. To date, no research has applied the focal concerns and court community perspectives to the punishment decisions of individual judges in more than one court to examine whether and how judges in the same court communities consider legal and extralegal factors. The current study addresses this gap in the literature.

### *Purpose of the Research*

The current research seeks to build on the prior literature to further advance knowledge of interjudge disparity and judicial sentencing patterns within court communities. Prior work suggests judges rely primarily on offense severity and criminal history when assessing offender blameworthiness and community threat, but they also engage in subjective decision-making that reflects their sentencing philosophy and attributions associated with offender extralegal characteristics (Johnson et al., 2008). To the extent that subjective decision-making varies across judges, the legal and extralegal factors associated with sentencing outcomes should vary as well.

Yet, the court community perspective suggests that interjudge disparity may be conditioned by the context in which sentencing occurs. Court communities develop distinctive

case processing strategies and sentencing norms that not only affect sentencing outcomes, but also the predictors that influence these outcomes (Kautt, 2002; Kramer & Ulmer, 2009). Despite these assertions, extant work does not offer a comprehensive examination of judges' sentencing patterns within court communities.

Though prior work has examined court communities and variation in sentencing predictors across courts and judges, these studies are limited in several ways. Qualitative studies provide in-depth information about court communities, but are limited to a small number of courts (typically fewer than three). Further, this work has generally focused on court structure, workgroup relationships, and case processing strategies, with less attention paid to predictors associated with sentencing outcomes (Eisenstein, Flemming, & Nardulli, 1988; Jacobs, 1997). Ulmer (1997) and Ulmer and Kramer's (1996) work is an exception, but their analysis of legal and extralegal effects on punishment decisions used pooled case models for each of the three courts. Thus, findings concerning sentencing disparity reflected all judges' decisions within each court, as opposed to each individual judges' decisions within these courts.

Multilevel analysis of offenders nested within courts allows for large-scale comparisons across multiple jurisdictions, and findings from these studies show that effects associated with offender and case characteristics vary significantly across courts (Britt, 2000; Kautt, 2002; Kramer & Ulmer, 2009; Ulmer & Johnson, 2004). Researchers conclude that these findings offer support for the idea that court communities influence sentencing decisions, but these studies provide no information about judges' sentencing patterns within these courts. Additional research using multilevel models provides some information about judges' sentencing patterns within courts, and this work suggests effects associated with legal and extralegal factors vary between judges (Anderson & Spohn, 2010; Johnson, 2006). Though this research raises some

questions about the extent to which court communities influence sentencing decisions, Wooldredge (2010) demonstrated the limitations of using multilevel modeling to assess interjudge disparity. Comparing findings from a two-level model (offenders nested within judges) with individual judge regression models showed that the multilevel analysis obscured individual judges' contributions to sentencing disparities. Thus, Wooldredge's (2010) work suggests that individual judge models may be more appropriate than multilevel analysis for assessing judges' sentencing patterns within courts. However, Wooldredge's (2010) research was limited to one court with 18 judges, and focused on differences in extralegal effects on judges' decisions to incarcerate offenders. In addition, some of the variation found in his analysis may be due to Ohio's relatively lax sentencing guidelines, which explicitly allow judges to consider different sentencing goals (Wooldredge, 2010).

The current research seeks to address these limitations through a more comprehensive examination of interjudge disparity and judicial sentencing decisions within court communities. First, multilevel analysis with offenders nested within judges nested within courts for a sample of large, medium, and small Pennsylvania courts is used to replicate prior work that shows effects associated with legal and extralegal factors vary significantly across judges. Next, findings from this model will be compared with results from individual judge regression models to examine judges' contributions to legal and extralegal effects found in the multilevel analysis. Finally, individual judges will be grouped by court to assess whether and how judges in the same court communities consider offender and case characteristics.

Employing both multilevel models and individual regression models is important for theoretical and methodological reasons. Focal concerns perspective states that differences in judges' subjective decision-making influences the ways in which judges consider offender and

case characteristics when assessing offender blameworthiness and community threat. In addition, sentencing theories are not only concerned with whether judges vary in effects associated with legal and extralegal factors, but also whether these factors significantly influence sentencing decisions. However, extant research generally relies on fixed effects estimates from models with cases pooled at the jurisdiction or state level (e.g., Steffensmeier & Demuth, 2006; Tillyer, Hartley, & Ward, 2015; cf. Wooldredge, 2010), which limits exploration of judges' variation in their subjective assessments of the focal concerns. According to Wooldredge, when relying on overall estimates from pooled case models, "there is a risk of conveying the impression that the problem is pervasive across all judges" (Wooldredge, 2010: 540). Studies that examine individual judges' estimates from the random effects portion of multilevel analyses (e.g., Anderson & Spohn, 2010; Johnson, 2006) provide more information about each judges' value for a given variable, but limitations remain.

The random group effects in multilevel models are obtained by combining information about the specific group effect and the overall model coefficient (Gelman & Hill, 2016; Hox, 2010; Snijders & Bosker, 2012). Less reliable group estimates are "shrunk" closer to the overall mean for the dataset, resulting in biased, but also more precise, estimates (Gelman & Hill, 2016; Hox, 2010). However, scholars conducting school achievement research have urged caution in relying solely on multilevel modeling when the purpose of the research is to evaluate the performance of individual teachers or schools (Fitz-Gibbon, 1991; Tate, 2004; Teddlie & Reynolds, 2000). Since multilevel models pool the group estimates around the average, schools with very good results may be pulled down towards the mean, while schools that are underperforming will be pulled up (de Leeuw & Kreft, 1995; Fitz-Gibbon, 1996). Examination of multilevel models alone to evaluate judges' sentencing decisions raises similar concerns,

particularly in light of contemporary sentencing theories. The pooling of judge estimates for offender and case characteristics is contrary to the idea that judges vary in their subjective assessments of factors associated with the focal concerns of sentencing. Thus, estimates obtained from individual judge regression models may provide information that allows for a more comprehensive evaluation of each judges' sentencing patterns.

As noted above, sentencing theories are also concerned with whether interjudge variation is meaningful, yet multilevel models are not well suited for providing this kind of information. Gelman and Hill (2016) note that identifying statistically significant results for random effects is not the primary purpose of multilevel analysis; rather, multilevel models are designed to obtain the most precise estimate for each group, while taking into account uncertainty. This analytic strategy may not be appropriate when predictors are hypothesized to produce effects for some groups, but not others (Gelman & Hill, 2016). The latter is particularly salient since sentencing theories suggest judges rarely have enough information to make fully informed punishment decisions (Albonetti, 1991). Consequently, the decision-making process “allows for the subtle influences of experiences, prejudices, and stereotypes, as well as idiosyncratic interpretations by different judges” (Johnson, 2006: 267), which is likely to result in factors such as race and gender affecting sentencing decisions for some judges, but not others.

Further, though multilevel analysis can be used to identify differences in legal and extralegal effects across court communities, it provides little information about how individual judges within court communities consider these factors. The court communities perspective suggests that judges' subjective decision-making may be conditioned by local case processing and sentencing norms, and the presence of, and adherence to these norms may be associated with differences in court actor autonomy based on court size. As such, exploring individual judges'



sentencing patterns within large, medium, and small court communities is needed to assess the significance of court communities in sentencing decisions.

This research has a number of methodological and theoretical implications. Concerning interjudge disparity, though multilevel analysis of sentencing data has identified variation in the factors that influence punishment decisions across judges, the pooled estimates may be more appropriate for drawing general conclusions about differences in judicial decision-making. The individual judge analyses, on the other hand, allow for a more detailed examination of this variation, and provide insight about whether these factors matter in sentencing outcomes. Incorporating both analytic strategies is needed to assess whether individual analyses offer a better test of the focal concerns perspective, which suggests judicial sentencing decisions vary, but has not been fully explored using extant methodologies. The methodological approach used in the current work is also a necessary first step in identifying the extent of the variation and whether it is meaningful, which can then be followed by further developing theories to explain why differences among judges exist.

Concerning the court community perspective, greater consistency in effects associated with the key predictors of sentencing among judges in the same courts versus individual judge model findings would suggest the court community influences punishment decisions. However, greater interjudge disparity in large courts versus small courts, for example, may indicate that the influence of court communities on sentencing is conditioned by court size. Substantial judicial variation within all courts could suggest that court communities influence some court processes (e.g., docket management, case assignment), but have less of an impact in the sentencing phase.

More generally, this research will contribute significantly to an area of sentencing research and theory that has received little empirical attention. According to Wooldredge (2010;

564), “we need more comprehensive quantitative descriptions of how judges differ in their sentencing decisions. This pursuit is necessary for assessing and informing theories of sentencing disparities based on extralegal factors.” Similarly, in assessing the state of the research on between-judge variation, Ulmer (2012: 26-27) notes, “[q]uite simply, the field needs more of such research. It is likely that a substantial portion of the interesting variation in sentence severity ... and the effects of legally relevant, organizational, and extralegal factors on sentencing exists at the level of individual judges.” Much of the extant research on between-judge variation relies on aggregate analysis of cases across jurisdictions or states, which is useful for assessing overall patterns and developing general theoretical perspectives about interjudge variation. Yet, analysis of individual judges and their sentencing patterns within court communities is necessary to further develop these perspectives.

Finally, this research will add to a growing body of literature that explores whether sentencing guidelines systems have achieved their intended goals. Reformers recognized that the sentencing of criminal offenders is a fundamental mechanism of formal social control in society, and disparity in punishment for offenders with comparable criminal records convicted of the same crime raises questions about the legitimacy of legal institutions (Reitz, 1998; Tonry, 1996). Examining individual judges’ sentencing patterns will further our understanding of whether sentencing guidelines promote uniformity and consistency in criminal sanctions.

### *Research Questions and Hypotheses*

The following research questions and hypotheses guide the current inquiry:

- 1) How do legal and extralegal factors affect the decision to incarcerate offenders and the length of the sentence imposed, and do these effects vary significantly across judges?

- 2) How do individual judges contribute to the legal and extralegal effects found in the multilevel analysis?
- 3) To what extent do judges in the same courthouses exhibit similar sentencing patterns, in terms of significant legal and extralegal effects?

The first research question will be explored using a three-level multilevel model with offenders nested within judges nested within courts. The following hypotheses are derived from the prior research (e.g., Johnson, 2006; Kramer & Ulmer, 2009):

*H1:* Offense severity and criminal history will be positively associated with sentence severity.

*H2:* Younger offenders will be sentenced more harshly than older offenders.

*H3:* Female offenders will be sentenced more leniently than male offenders.

*H4:* Black offenders will be punished more harshly than white offenders.

*H5:* Offenders who plead guilty will receive less severe sentences than those convicted after trial.

*H6:* Legal and extralegal effects will vary significantly across judges.<sup>1</sup>

The second research question will be answered using individual judge regression models. The data used in the analysis were obtained from a guidelines state that uses offense severity and prior record to determine the sentencing range. Therefore, these variables should be positively associated with sentence severity for most judges. Further, based on theories that suggest judges engage in subjective decision-making concerning offender characteristics and prior work showing differences in extralegal effects across individual judges (Wooldredge, 2010),

---

<sup>1</sup> Though the analysis uses three-level models, this work focuses on variation at the judge level. The court level is included to control for differences across courts.

significant results for extralegal effects should be less consistent than findings associated with legally relevant variables.<sup>2</sup>

The third research question will be answered by examining individual judges' sentencing patterns within their court communities. Though quantitative research using multilevel analysis shows significant variation across courts (e.g., Britt, 2000; Ulmer & Johnson, 2004), these studies provide little information about judicial decision-making within courts. Additional research suggests interjudge disparity exists within courts (e.g., Anderson & Spohn, 2010; Johnson, 2006), but these studies do not provide a comprehensive assessment of effects associated with individual judges' punishment decisions. Qualitative research, however, offers some evidence of the existence of court communities and their impact on case processing and sentencing (Eisenstein, Flemming, & Nardulli, 1988; Ulmer, 1997; Ulmer & Kramer, 1996), but differences in court actor autonomy in small, medium, and large courts may condition this relationship. Thus, lower autonomy among small court judges may result in judges exhibiting similar sentencing patterns. Variation in sentencing patterns may be greater in medium sized courts, and the largest variation in judicial sentencing patterns is likely in large courts as court actor autonomy is highest.<sup>3</sup>

---

<sup>2</sup> Since sentencing theory and research does not yet provide specific hypotheses about why individual judges vary in effects associated with legal and extralegal factors, more detailed predictions are beyond the scope of this work.

<sup>3</sup> Similar to research question two, prior literature does not offer specific expectations about how court communities influence judicial consideration of legal and extralegal factors.

### CHAPTER 3: DATA AND METHODOLOGY

#### *Data*

To examine the research questions and hypotheses outlined above, seven years of data (2004-2010) were obtained from the Pennsylvania Commission on Sentencing (PCS). Pennsylvania's Courts of Common Pleas (the state's county-level trial courts) are required by law to submit all felony and misdemeanor convictions under the Pennsylvania Sentencing Guidelines to the PCS on a yearly basis (PCS, n.d.). The PCS compiles the data into annual datasets and makes them available to the public for a fee. The data provide detailed information about offense type and severity, offender criminal history, and offender characteristics such as age, race, and gender. The data also include information about mode of conviction, the sentence imposed, the court in which the offender was sentenced, and the name of the judge who imposed the sentence (PCS, n.d.).

#### *Data Reduction and Missing Data*

In line with prior research using the PCS data, the data were restricted to include only the most serious offense per judicial transaction (Britt, 2000, 2009; Johnson, 2003, 2006, 2014; Kramer & Ulmer, 2009; Ulmer & Johnson, 2004). Limiting the data to the most serious offense per judicial transaction is also consistent with the way in which the PCS analyzes the data for its annual reports (PCS, n.d.). Given the focus of this research is on individual judges, the data were further limited to black and white offenders because very few judges outside of the larger jurisdictions sentence Hispanic offenders frequently enough for inclusion in the analysis.

Consistent with prior work using these data, missing values for offense gravity score, prior record score, guideline edition, and offender age, race, and gender were removed (Britt,

2000, 2009; Johnson, 2003, 2006, 2014; Kramer & Ulmer, 2009; Ulmer & Johnson, 2004).<sup>4</sup>

Though only a modest amount of data are missing for these variables, 18 percent of cases were missing information on mode of conviction (i.e., whether the offender entered a guilty plea or was convicted after trial). Because of the information loss associated with removing these cases, extant work has employed a dummy variable adjustment (Johnson, 2003, 2006, 2014; Kramer & Ulmer, 2009). However, research suggests that this approach can produce biased estimates of regression coefficients (Allison, 2001). To address missing values for mode of conviction, the current work utilized the ‘Amelia’ package in R (Honaker, King, & Blackwell, 2011) to implement an iterative expectation-maximization algorithm with bootstrapping to substitute plausible values for this variable. This produced a single dataset with maximum likelihood imputed values for the mode of conviction variable. These criteria resulted in a sample of 532,440 cases, sentenced by 571 judges in 60 courts.<sup>5</sup>

### *Sample Selection*

The current work seeks to examine individual judges’ contributions to legal and extralegal disparity, and assess sentencing patterns among judges in the same courthouses. The sample used for the current research was initially selected based on the number of cases judges sentenced between 2004 and 2010. The ‘pwr’ package in R (Champely et al., 2015) was used to determine that a minimum of 230 cases per judge were needed to detect relatively small effect sizes ( $f^2 = 0.09$ ) in multiple regression models.<sup>6</sup> Of the 571 judges, 287 judges handled 230 or

---

<sup>4</sup> Missing values were minimal for offense gravity, prior record score, guidelines edition, and offender age and gender (for each variable, fewer than 0.002% of all cases were missing). Three percent of cases had missing values for offender race/ethnicity.

<sup>5</sup> Though Pennsylvania has 67 counties, the state has only 60 county-level trial courts. The difference is the result of seven of these courts handling cases from two counties.

<sup>6</sup> R’s ‘pwr’ package uses Cohen’s (1988) suggestions for effect size. For multiple regression, Cohen (1988) used  $f^2$  values of 0.02, 0.15, and 0.35 to represent small, medium, and large effects, respectively.

more cases, though the vast majority of these judges handled many more cases (mean = 1,755 cases, median = 1,396).<sup>7</sup> Notably, these 287 judges imposed sentences in 503,602 of the 532,440 cases (95 percent) across the 60 courts. Removing 284 judges (and retaining 95 percent of the cases) is consistent with prior work that suggests a substantial number of judges serve as senior, retired, or traveling judges who handle a very small number of cases each year (Johnson, 2006, 2014; Levin, 1977).

The second step in selecting the sample included disproportionate stratified sampling to examine sentencing patterns among judges in the same courthouses. Based on extant research that suggests case processing and sentencing practices may be influenced by court size (Eisenstein, Flemming, & Nardulli, 1988; Kramer & Ulmer, 2009; Ulmer, 1997), a sample of small, medium, and large courts were selected for the analysis. Prior research using the PCS data identifies small courts as having seven or fewer authorized judgeships, medium courts having between eight and 15, and large courts having 16 or more (Johnson, 2006; Kramer & Ulmer, 2009; Ulmer, 1997; Ulmer & Johnson, 2004). This criterion results in splitting the 60 Pennsylvania courts into 44 small, 12 medium, and four large courts. Given the smaller number of large and medium-sized courts, a disproportionate stratified sample that includes all four of the large courts, six medium courts (50 percent), and 11 small courts (25 percent) was selected for the analysis. To determine which medium and small courts would be included, R's `sample` function (R Core Team, 2016) was used to randomly generate numbers between the minimum and maximum number of small and medium courts. The final sample includes 312,555 cases sentenced by 161 judges in 21 Pennsylvania county courts. Table 1 shows the total number of cases handled in each court, and the number of cases from each court included in the analysis

---

<sup>7</sup> These 287 judges handled at least 230 cases for both the incarceration decision and the sentence length decision.

after removing judges with fewer than 230 cases.<sup>8</sup> Table 1 also includes the percent of cases handled in these courts after removing judges, and the number of judges in each court.<sup>9</sup>

Table 1. Cases and Number of Judges by Court  
Cases in the Sample (N=312,555)

<u>Court</u>	<u>Total Cases</u>	<u>Cases in Analysis</u>	<u>Percent</u>	<u># of Judges</u>
Large Court 1	64,914	56,247	86.65%	19
Large Court 2	43,024	30,278	70.37%	29
Large Court 3	36,868	34,813	94.43%	11
Large Court 4	36,196	33,581	92.78%	16
Medium Court 1	10,570	9,830	93.00%	7
Medium Court 2	32,746	29,503	90.10%	10
Medium Court 3	9,113	8,158	89.52%	2
Medium Court 4	16,599	10,940	65.91%	6
Medium Court 5	13,591	11,078	81.51%	9
Medium Court 6	31,218	29,108	93.24%	10
Small Court 1	10,756	10,073	93.65%	5
Small Court 2	7,393	5,837	78.95%	4
Small Court 3	3,182	2,497	78.47%	2
Small Court 4	3,294	3,188	96.78%	2
Small Court 5	4,911	3,577	72.84%	2
Small Court 6	5,956	5,694	95.60%	6
Small Court 7	6,934	6,521	94.04%	5
Small Court 8	7,230	6,930	95.85%	5
Small Court 9	3,519	3,471	98.64%	2
Small Court 10	8,151	6,329	77.65%	4
Small Court 11	5,040	4,902	97.26%	4

### *Dependent and Independent Variables*

Extant sentencing research suggests that punishment decisions occur in two stages (Wheeler, Weisburd, & Bode, 1982). The first decision is whether to incarcerate an offender, and the second involves determining the length of confinement for those who receive a custodial

<sup>8</sup> Total cases are limited to the most serious offense per judicial transaction and cases resulting in conviction.

<sup>9</sup> In some courts, the number of judges included in the analysis is less than what is outlined in the criteria concerning court size. For example, though large courts have greater than 16 judges, Large Court 3 has only 11 judges in the analysis. For this court and Medium Courts 1, 3, and 4, the smaller number of judges is due to removing judges with fewer than 230 cases.



sentence. Consistent with prior work, two dependent variables are used to model these decisions separately (e.g., Britt, 2000; Dixon, 1995; Hauser & Peck, 2017; Johnson, 2006, 2014; Ulmer & Johnson, 2004; Wolfe, Pyrooz, & Spohn, 2011). For the incarceration decision, a binary variable indicates whether the offender was incarcerated (1 = incarceration, 0 = no incarceration). The length of the sentence imposed is a continuous variable coded to represent the minimum number of months the offender is sentenced to serve in jail or state prison. Given the positive skew of the sentence length data, this variable was recoded to equal the natural logarithm of the minimum number of months of incarceration (Britt, 2009; Bushway & Piehl, 2001; Johnson, 2006).

Independent variables include *offense severity*, which is based on the offense gravity score developed by the PCS and ranges from 1 (least serious) to 14 (most serious). *Prior record* is a measure of the PCS' prior record score, which is an eight-category scale of prior convictions with points given for prior misdemeanors and felonies based on offense severity. Due to the small number of cases in the two highest categories (repeat felony and repeat violent offenders), these were combined into a single category in all analyses. Offender demographic variables include a continuous variable for the offender's *age* when the sentence was imposed, and binary variables for *female* and *black* (reference = white). In addition, mode of conviction is captured with a binary *trial* variable that combines negotiated and non-negotiated pleas into a plea category (reference), and bench and jury trials into a trial category.

### *Control Variables*

Consistent with prior research using the PCS data (e.g., Johnson, 2014; Kramer & Ulmer, 2009), several control variables are included in the analysis. *Mandatory minimum* represents whether a mandatory sentencing provision was applied (reference = no), and changes to the sentencing guidelines are captured through a *guidelines edition* dummy variable (reference = 6<sup>th</sup>

Edition). In line with extant work using the PCS data, *offense type* is a dummy variable that is coded to reflect whether the offender was convicted of a violent, drug, or property offense, with other offenses that do not fall into these crime types (i.e., bad checks, forgery, DUI) serving as the reference (Johnson, 2006, 2014; Ulmer & Johnson, 2004). Further, two *presumptive sentence* variables are included to capture presumptive guideline sentence recommendations (Engen & Gainey, 2000). For the incarceration decision models, a binary variable is used to indicate whether the guidelines prescribe incarceration (reference = no). In the sentence length models, this variable represents the minimum number of months of incarceration recommended in the guidelines (Ulmer, 2000; Ulmer & Johnson, 2004). Finally, *year* is a dummy variable that controls for annual changes within the courts from 2004 to 2010 (reference = 2004).

### *Analytic Strategy*

Due to the nested structure of sentencing data (i.e., offenders nested within judges nested within courts), multilevel analysis of predictors associated with punishment outcomes has become increasingly popular (Britt, 2000; Kautt, 2002; Johnson, 2006, 2014; Kramer & Ulmer, 2009; Ulmer & Johnson, 2004). With nested data, there is an assumption that contextual factors influence individuals, and individuals from the same context likely share common influences (Steenbergen & Jones, 2002). As a result, traditional regression models are not well suited for analyzing these data. Because of the influence of contextual factors, the observations are not truly independent; rather, they are clustered and duplicate one another to some extent, which violates the regression assumption that errors are independent. When this assumption is violated, incorrect standard errors and Type I errors are likely (Steenbergen & Jones, 2002). By incorporating additional disturbance terms and their associated assumptions, multilevel models produce appropriate error terms that control for potential dependency due to nesting effects

(Snijders & Bosker, 2012). However, while multilevel models offer an improvement over classical regression analyses from a statistical standpoint (Fitz-Gibbon, 1996; Gelman & Hill, 2016; Teddlie & Reynolds, 2000), practical concerns remain, particularly when assessing random group effects.

The random group effects in multilevel models are unobserved variables, as opposed to statistical parameters (Snijders & Bosker, 2012), and are obtained using empirical Bayes estimation. Empirical Bayes estimates are weighted averages of the specific group effect and the overall model coefficient (Gelman & Hill, 2016; Hox, 2010; Snijders & Bosker, 2012). As noted earlier, multilevel estimates are pooled (Gelman & Hill, 2016) or “shrunk” towards the mean for the entire data set (Hox, 2010: 29). Shrinkage is determined by the reliability of the estimate, and reliability is based on the group sample size and the difference between the group estimate and the overall model estimate (Hox, 2010). Groups with small sample sizes and estimates far from the overall estimate shrink more to the overall average, while groups with large sample sizes and estimates near the overall estimate are close to the overall mean (Hox, 2010). For intermediate groups, the multilevel estimates fall between these two (Gelman & Hill, 2016). As a result, more variation is expected when looking at unbiased<sup>10</sup> results from separate classical regression analyses, compared to the more precise, but biased, multilevel values (Hox, 2010).

The current work uses multilevel modeling to replicate the findings of recent work that shows legal and extralegal effects vary significantly across judges, and also employs individual judge logistic and ordinary least squares (OLS) regression models to assess how judges contribute to the results from the multilevel analysis. While extant sentencing research has used either multilevel or single-level regression models for pooled cases (cf. Wooldredge, 2010),

---

<sup>10</sup> Classical regression estimates (e.g., OLS) are unbiased, but error exists due to sampling and measurement (Willms, 1992).

Gelman and Hill (2016) suggest an iterative approach to statistical modeling. With multilevel data structures, this can include separate models to obtain unadjusted values, and multilevel analysis to examine pooling of random group effects (Gelman & Hill, 2016). Findings from these models can also be used to identify groups with high or low estimates, and to get a general sense of any patterns in the data (Kreft & Yoon, 1994; Snijders & Bosker, 2012).

The analytic approach begins with multilevel modeling. As the incarceration decision dependent variable is a binary outcome representing incarceration/no incarceration, hierarchical generalized linear models (HGLM) were selected. For the sentence length outcome, which represents the minimum months of incarceration (logged), linear mixed models (LMM) were employed. Both analyses were conducted using R's 'lme4' package (Bates et al., 2016). The first step in the analysis includes unconditional models (one-way ANOVA with random effects) to assess whether variation in the decision to incarcerate offenders and the length of the sentence imposed exists at the judge- and court-levels. In addition, calculating the Akaike information criterion (AIC) for the unconditional models provides a baseline that can be used to assess model fit in subsequent analyses. The second step in the multilevel analysis includes random coefficients ANCOVA models with independent and control variables to examine the fixed and judge-level random effects of offender and case characteristics on the incarceration decision and the length of the sentence imposed.

To assess judges' contributions to findings from the multilevel analysis, separate logistic and OLS regression models are employed for each of the 161 judges using R's 'stat' package (R Core Team, 2016). Since this portion of the analysis focuses on individual judges' caseloads and sentencing patterns, multilevel models are not needed to account for between-judge variation. Finally, individual judges are grouped by court and results from the individual judge regression

models are used to assess similarities in statistically significant legal and extralegal effects among judges in the same courts.

## CHAPTER 4: RESULTS

The following section contains the findings for the current inquiry. First, descriptive statistics for the sample are discussed. As outlined in Chapter 2, the majority of studies examining interjudge disparity use multilevel models. Thus, results from unconditional and random coefficients models are presented to assess effects associated with key predictors of the decision to incarcerate offenders and the length of sentence imposed, and whether these effects vary significantly across judges. However, since prior research shows this approach offers limited information (Wooldredge, 2010), findings from individual judge regression models are provided to gain a better understanding of whether and how judges use legal and extralegal factors in sentencing decisions. Extant work also suggests that judicial decision-making may be influenced in part by the court community in which sentencing occurs (e.g., Ulmer, 1997). Consequently, the final section in this chapter provides results concerning judges' sentencing patterns within the same courts.

### *Descriptive Statistics*

Table 2 provides descriptive statistics for the sample of offenders used in the analysis. Approximately half of the offenders (49 percent) in the sample were sentenced to either jail or prison, while the others (51 percent) received a non-custodial sentence. For those offenders who were incarcerated, the mean sentence length was just under 10 months. The mean offense gravity score is 3.57 (scale ranging from one to 14), and the mean prior record score is 1.45 (scale ranging from one to six). The average offender is 33 years old at sentencing, and males make up a larger portion of the sample than females (80 percent and 20 percent, respectively). Concerning

Table 2. Descriptive Statistics for Overall Sample (N = 312,555)

<u>Dependent Variable</u>	<u>M (SD) / N (%)</u>
Incarcerated	154,001 (49%)
Not Incarcerated	158,554 (51%)
Sentence Length	9.94 (19.13)
<u>Independent Variables</u>	
Offense Severity	3.57 (2.42)
Prior Record	1.45 (1.90)
Age at Sentence	33.17 (11.07)
Male (reference)	250,766 (80%)
Female	61,789 (20%)
White (reference)	205,590 (66%)
Black	106,965 (34%)
Plea (reference)	295,760 (95%)
Trial	16,795 (5%)
Presumptive Incarceration	69,424 (22%)
Presumptive Sentence	11.30 (23.08)
Mandatory Minimum Applied	73,122 (23%)
Offense Type	
Violent	41,816 (13%)
Drug	74,718 (24%)
Property	62,810 (20%)
Other (reference)	133,211 (43%)
Guidelines Edition	
5 <sup>th</sup> Edition	94,840 (30%)
6 <sup>th</sup> Edition (reference)	165,313 (53%)
6 <sup>th</sup> Edition, Revised	52,402 (17%)
Year	
2004 (reference)	36,926 (12%)
2005	42,483 (14%)
2006	44,817 (14%)
2007	47,158 (15%)
2008	49,155 (16%)
2009	49,752 (16%)
2010	42,264 (14%)

offender race, 66 percent of offenders are white, while the remaining 34 percent are black.

Further, the vast majority (95 percent) of the offenders in the sample entered guilty pleas; only five percent were convicted after trial. Twenty-two percent of offenders committed crimes that prescribed a custodial sentence, and the presumptive minimum sentence for incarcerated

offenders is approximately 11 months. Twenty-three percent were convicted of crimes that included mandatory minimum penalties. Most offenders (43 percent) were convicted of crimes other than drug (24 percent), property (20 percent), and violent (13 percent) offenses. Roughly 50 percent of offenders were sentenced under the sixth edition of the guidelines, which were in effect from June of 2005 until December of 2008, while 30 percent and 17 percent were sentenced under the earlier fifth edition and later sixth edition revised guidelines, respectively. Finally, the percent of offenders sentenced each year is relatively consistent, ranging from 12 to 16.

### *Multilevel Analysis*

The analysis addressing the first research question used multilevel modeling to examine the effects of legal and extralegal factors on the decision to incarcerate offenders and the length of the sentence imposed, and whether these effects varied significantly across judges (while controlling for court). Table 3 displays the results from the three-level unconditional models. The significant variance component for the incarceration model in Table 3 suggests that a portion of the variation in sentence severity is attributable to differences between judges and courts.<sup>11</sup> Results from the sentence length model show that five percent of variation in the sentence length is attributable to differences between judges, while seven percent is accounted for at the court level. The incarceration and sentence length unconditional models also provide baseline AICs of 415,129 and 629,062, respectively.

---

<sup>11</sup> The binary outcome for the incarceration model does not include an individual-level variance component. However, Snijders and Bosker (1999) note that if the level 1 model is viewed as a latent variable, the random effects at level 1 can be assumed to have a standard logistic distribution with a mean of 0 and a variance of  $\pi^2/3$ . If this assumption is valid, three percent and four percent of the variance in the likelihood of incarceration is attributable to differences between judges and courts, respectively.



Table 3. Unconditional Models of Incarceration and Sentence Length

<i>Incarceration</i>			<i>Sentence Length</i>		
Fixed effects	<u>b</u>	<u>SE</u>	Fixed effects	<u>b</u>	<u>SE</u>
Intercept	0.05	0.09	Intercept	0.91	0.12***
Random effects	<u>Variance</u>	<u>SD</u>	Random effects	<u>Variance</u>	<u>SD</u>
Level 1			Level 1	3.46	1.86***
Level 2	0.11	0.33***	Level 2	0.19	.44***
Level 3	0.15	0.39***	Level 3	0.27	.52***
			Intraclass Correlation		
			Level 2		
			0.05		
			Level 3		
			0.07		
AIC		415,129	AIC		629,062

\*\*\*p < .001; \*\*p < .01; \*p < .05

Table 4 provides results for the incarceration and sentence length models with fixed and random effects associated with offender and case characteristics on sentence severity.<sup>12</sup> Findings from the fixed effects portion of the incarceration model are consistent with prior research using the PCS data (e.g., Johnson, 2006, 2014; Kramer & Ulmer, 2009). The AIC for the fixed effects model is lower than the AIC in the unconditional model (324,018 versus 415,129), which indicates including the offender-level variables produces a better model fit. Results for the legally relevant variables show a one-unit increase in offense severity and prior record increases the odds of incarceration by 1.41 and 1.43, respectively. Further, offenders are more likely to be incarcerated when the guidelines prescribe a jail or prison sentence, and being convicted of a crime that requires application of a mandatory minimum sentence greatly increases the odds of

<sup>12</sup> Standard errors in the incarceration model were calculated using the delta method. The delta method is used when reporting transformed regression parameters (Casella & Berger, 1991). Since the current work reports odds ratios transformed from the incarceration model coefficients, the delta method is appropriate.

Table 4. Random Coefficients Models of Incarceration and Sentence Length

<i>Fixed Effects</i>	<i><u>Incarceration</u></i>			<i><u>Sentence Length</u></i>	
<u>Independent Variables</u>	<u>Est.</u>	<u>SE</u>	<u>Odds</u>	<u>b</u>	<u>SE</u>
Intercept	-1.85	0.13	--***	-1.24	0.05***
Offense Severity	0.34	0.00	1.41***	0.44	0.00***
Prior Record	0.36	0.00	1.43***	0.21	0.00***
Age	-0.01	0.00	0.99***	0.00	0.00***
Female	-0.37	0.01	0.69***	-0.11	0.01***
Black <sup>a</sup>	0.31	0.01	1.36***	-0.02	0.01**
Trial <sup>b</sup>	0.31	0.03	1.36***	0.22	0.01***
Presumptive Sentence	0.31	0.02	1.37***	0.04	0.00***
Mandatory Minimum	2.41	0.17	11.19***	-0.85	0.01***
Offense Type <sup>c</sup>					
Violent	0.49	0.03	1.63***	0.01	0.01
Drug	-0.15	0.01	0.86***	0.20	0.01***
Property	0.25	0.02	1.28***	0.30	0.01***
Guidelines Edition <sup>d</sup>					
5 <sup>th</sup> Edition	0.21	0.02	1.23***	-0.01	0.01
6 <sup>th</sup> Edition, Revised	-0.01	0.02	0.99	0.04	0.01**
Year <sup>e</sup>					
2005	-0.09	0.02	0.91***	0.16	0.01***
2006	-0.09	0.02	0.92***	0.11	0.01***
2007	-0.10	0.02	0.91***	0.07	0.02***
2008	-0.19	0.02	0.83***	0.09	0.02***
2009	-0.22	0.02	0.80***	0.06	0.02***
2010	-0.21	0.02	0.81***	0.01	0.02
AIC	324,018			464,775	
N	312,555			154,001	

\*\*\*p < .001; \*\*p < .01; \*p < .05; <sup>a</sup>Reference category is white; <sup>b</sup>Reference category is plea; <sup>c</sup>Reference category for all crime types is other crimes; <sup>d</sup>Reference category for all guidelines editions is the 6th Edition; <sup>e</sup>Reference category for all years is year 2004.

Table 4 (cont'd).

<i>Random Effects</i>	<i><u>Incarceration</u></i>			<i><u>Sentence Length</u></i>		
	<u>Variance</u>	<u>SD</u>	<u>X<sup>2</sup></u>	<u>Variance</u>	<u>SD</u>	<u>X<sup>2</sup></u>
Offense						
Severity	0.03	0.16	3392.60***	0.01	0.09	3384.60***
Prior Record	0.02	0.13	1719.10***	0.00	0.05	787.73***
Age	0.00	0.01	814.90***	0.00	0.00	128.56***
Female	0.02	0.13	47.32***	0.01	0.12	91.65***
Black	0.18	0.43	1210.90***	0.02	0.13	215.40***
Trial	0.22	0.47	211.96***	0.03	0.19	118.62***
AIC	320,629 - 323,975			461,394 - 464,687		

incarceration. Concerning extralegal effects, as age increases the likelihood of incarceration decreases, and female offenders are less likely to be incarcerated than male offenders. Blacks and offenders sentenced after a trial verdict are 1.36 times more likely to be incarcerated compared to whites and those who enter a guilty plea.

Findings from the sentence length model are generally consistent with prior studies using the PCS data (e.g., Britt, 2009; Johnson, 2006, 2014; Kramer & Ulmer, 2009), with one exception. In line with prior work, a one-unit increase in offense severity is associated with a 44 percent increase in sentence length, and a one-unit increase in prior record score results in sentences that are 21 percent longer. The extralegal effects in the sentence length model show age has almost no effect, female offenders are sentenced more leniently than males, and conviction after a trial increases the sentence length by 22 percent. However, in contrast to extant research on Pennsylvania sentencing (Johnson, 2006; 2014), black offenders receive slightly shorter sentences (two percent) than whites. One other notable difference between the models is that offenders sentenced for mandatory minimum crimes receive are treated much more leniently in the sentence length model. This finding is likely due to the large number of DUI offenders in

the data receiving relatively short mandatory sentences (Britt, 2009). In addition, the AIC for the fixed effects model is lower than the AIC in the unconditional model (464,775 versus 629,062), which indicates including the offender-level variables produces a better model fit.

Table 4 also shows the random effects associated with select predictors of sentence severity.<sup>13</sup> For both the incarceration decision and the length of the sentence imposed, all relevant predictors varied significantly across judges. More specifically, effects associated with offense severity, prior record, age, gender, race, and mode of conviction on sentence severity vary depending on the judge handling the case. These findings are consistent with studies analyzing the PCS data using random effects models (Britt, 2000; Johnson, 2006; Kramer & Ulmer, 2009; Ulmer & Johnson, 2004). Including random effects also shows a modest improvement in model fit for both sentencing outcomes. The AICs for the random effects incarceration models ranged from 320,629 to 323,975, compared to the fixed effects model AIC of 324,018. Similarly, the AIC for the fixed effects sentence length model is 464,775, whereas the AICs for the random effects models range from 461,394 to 464,687.<sup>14</sup>

With the exception of one study (Wooldredge, 2010), assessing findings from multilevel analyses is the extent to which prior work has examined the variation in effects associated with legal and extralegal factors across judges. The following section provides results from the multilevel model random group effects and 161 individual judge logistic and OLS regression

---

<sup>13</sup> Given the current study's focus on assessing variation in legal and extralegal effects across judges, only six variables were considered for inclusion as random effects in the models: offense severity, prior record, age, gender, race, and mode of conviction.

<sup>14</sup> Models with six simultaneous random effects would not converge. Consequently, random effects analyses were conducted by running a model with all predictors as fixed effects, and a single predictor as a random effect. This random effects model was then compared to the fixed effects model with no random effects using a likelihood ratio test to assess statistical significance (Snijders & Boskers, 2012). The process was repeated for all of the random effects, and the range of AICs under the random effects in Table 4 reflect the model fit for each of the six random effects models.

models to explore judges' contributions to legal and extralegal effects found in the multilevel analysis.

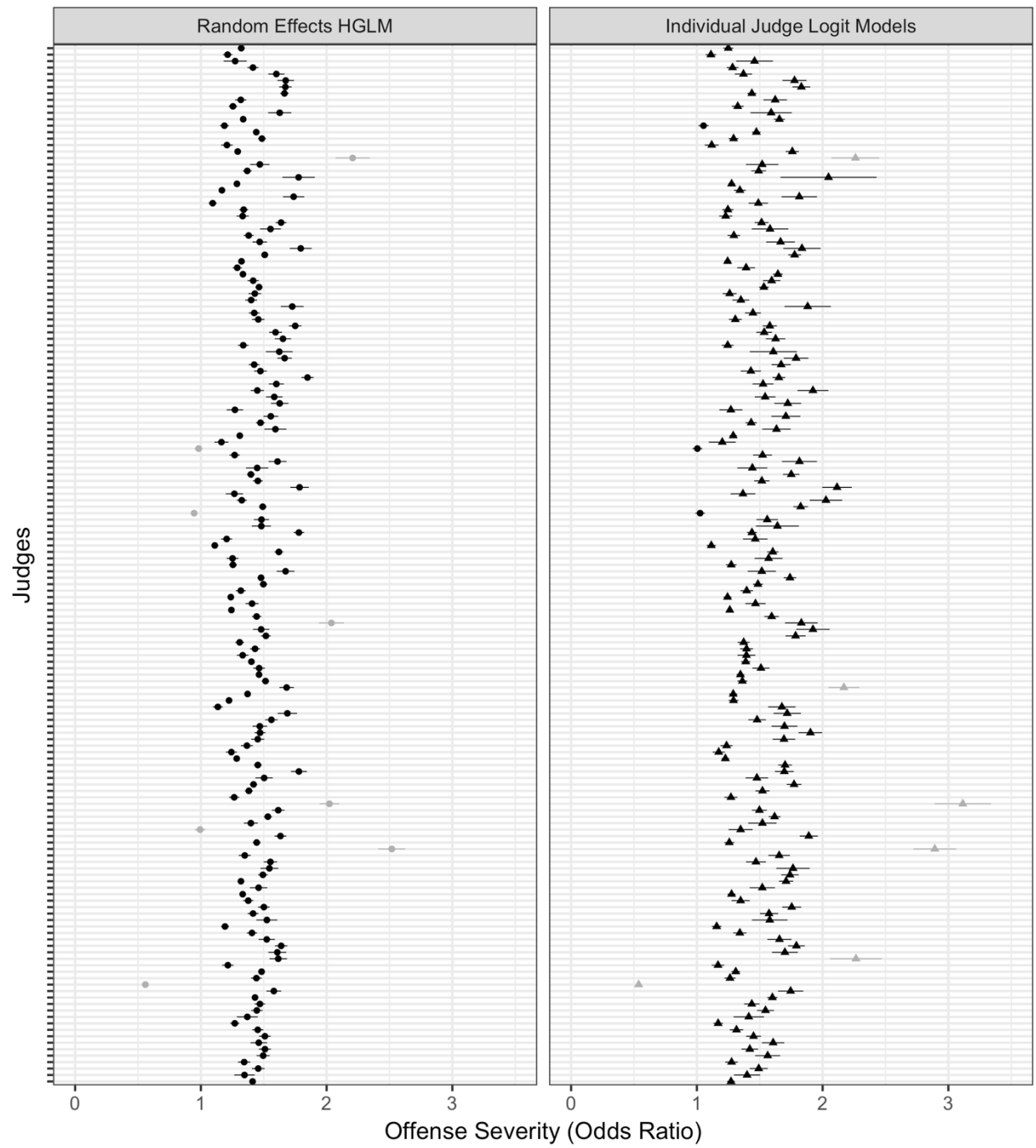
### *Individual Judge Analysis*

Figures 1 through 12 offer a visual representation of the differences in judges' estimates from the multilevel analysis and findings from the individual judge logistic and OLS regression models for offense severity, prior record, age, gender, race and mode of conviction. The Y-axis represents the judges (ranging from 1 to 161 from top to bottom), and the X-axis reflects the odds ratio for the incarceration models, and percents for the logged sentence length models. The panels on the left include each judges' estimates and standard errors obtained from the random effects models, and the panels on the right display odds ratios or percents and standard errors for the individual judge regression models. In both panels, estimates in black lie between two standard deviations below and above the mean, while those in gray are outside this range. The individual judge regression panels on the right also include information about whether a given variable is statistically significant; triangles represent coefficients with a p-value of  $<.05$ , and circles are above this threshold.

### *Incarceration Models*

Figure 1 provides the odds ratio for a one-unit increase in offense severity for the incarceration models. The odds ratios range from 0.56 to 2.51 (standard errors from 0.02 to 0.14) in the HGLM, and 0.54 to 3.11 (standard errors between 0.02 and 0.38) in the judge logit models. Findings also indicate that 95 percent of the values in the HGLM are predicted to lie between 1.02 and 1.94, and within 0.93 and 2.13 for the logit models. As expected, more variation exists in the logit models than the HGLM, where the estimates in the latter show some

Figure 1. HGLM and Logit Effects for Offense Severity



pooling to the overall mean of 1.41. Though this occurs to varying degrees across the judges, it is most noticeable for the more extreme cases towards the bottom half of the judge distributions. To the extent the p-value obtained in the logit models can be used to assess whether individual judges consider offense severity when deciding whether to incarcerate offenders, 158 of the 161 judges are associated with a statically significant increase in the odds of incarceration as offense severity increases. Concerning prior record effects (Figure 2), the HGLM odds ratios range from 1.07 to 2.30 (standard errors 0.02 to 0.04) and 1.15 to 2.25 (standard errors 0.03 to .40) in the logit models. Results also show that 95 percent of the values are predicted to lie between 1.10 and 1.85 and 1.02 and 2.02 in the HGLM and logit models, respectively. Similar to findings for offense severity, the judge logit estimates shrink closer to the overall average of 1.43 in the HGLM. In addition, the individual logit models indicate that 159 of the 161 judges exhibit statistically significant effects associated with harsher punishment for offenders with prior records.

The remaining figures examine differences between the HGLM and logit incarceration models for extralegal factors. Figure 3 shows the findings for offender age are very similar, with the odds ratios ranging just below 0.96 through 1.02 for both the HGLM and logit models. The range of standard errors are also nearly identical (HGLM, 0.00 to 0.01; logit, 0.00 to 0.02). Concerning predicted values across all judges, 95% of the estimates lie between 0.96 and 1.01 for the HGLM, and 0.96 and 1.02 for the logit models. Notably, there are a few judges in the top portion of the HGLM plot that are moving in the opposite direction—rather than shrinking towards the overall mean of 0.99, they moved farther away from it. In addition, the individual judge logit models indicate that statistically significant findings associated with higher odds of

Figure 2. HGLM and Logit Effects for Prior Record

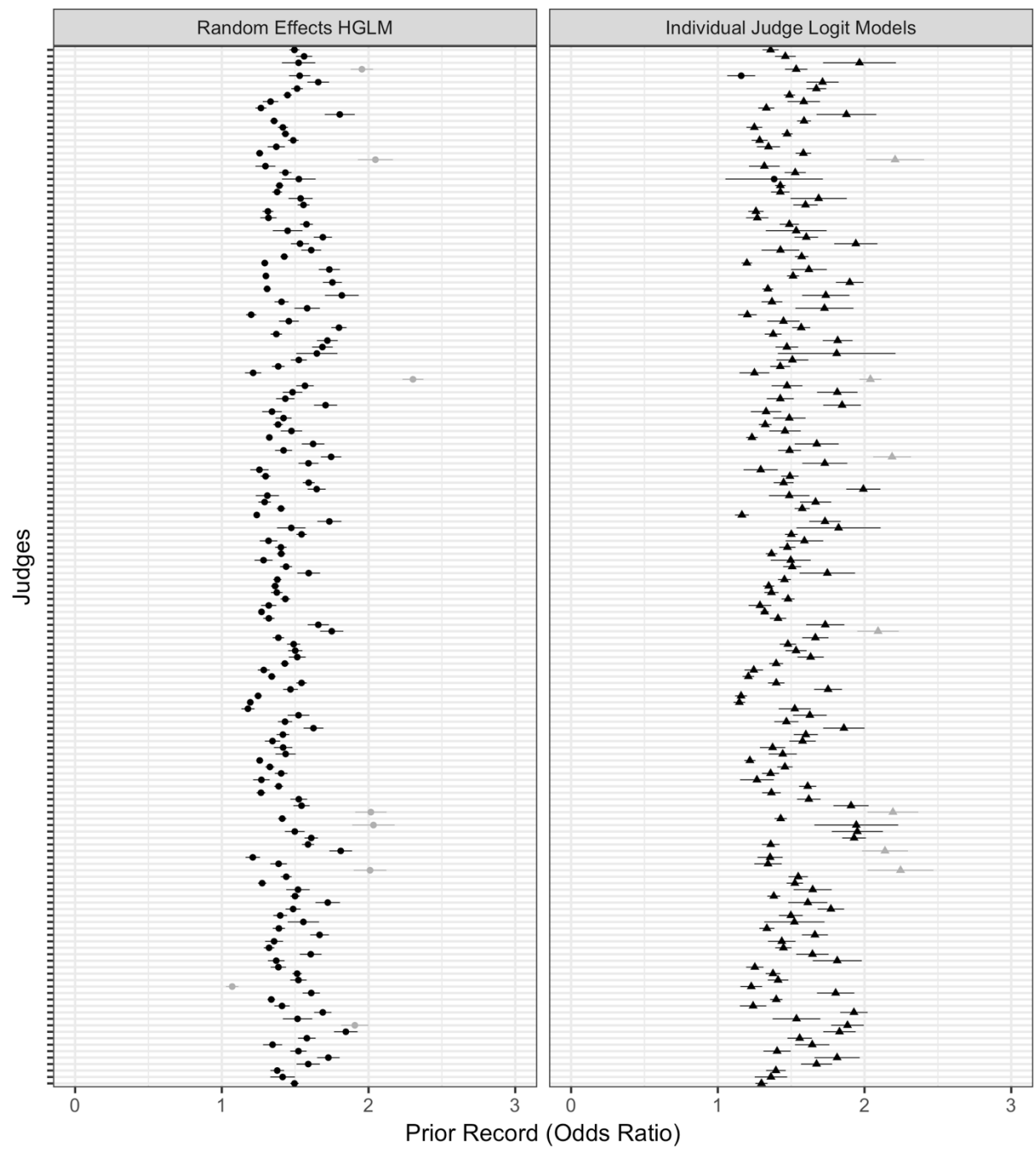
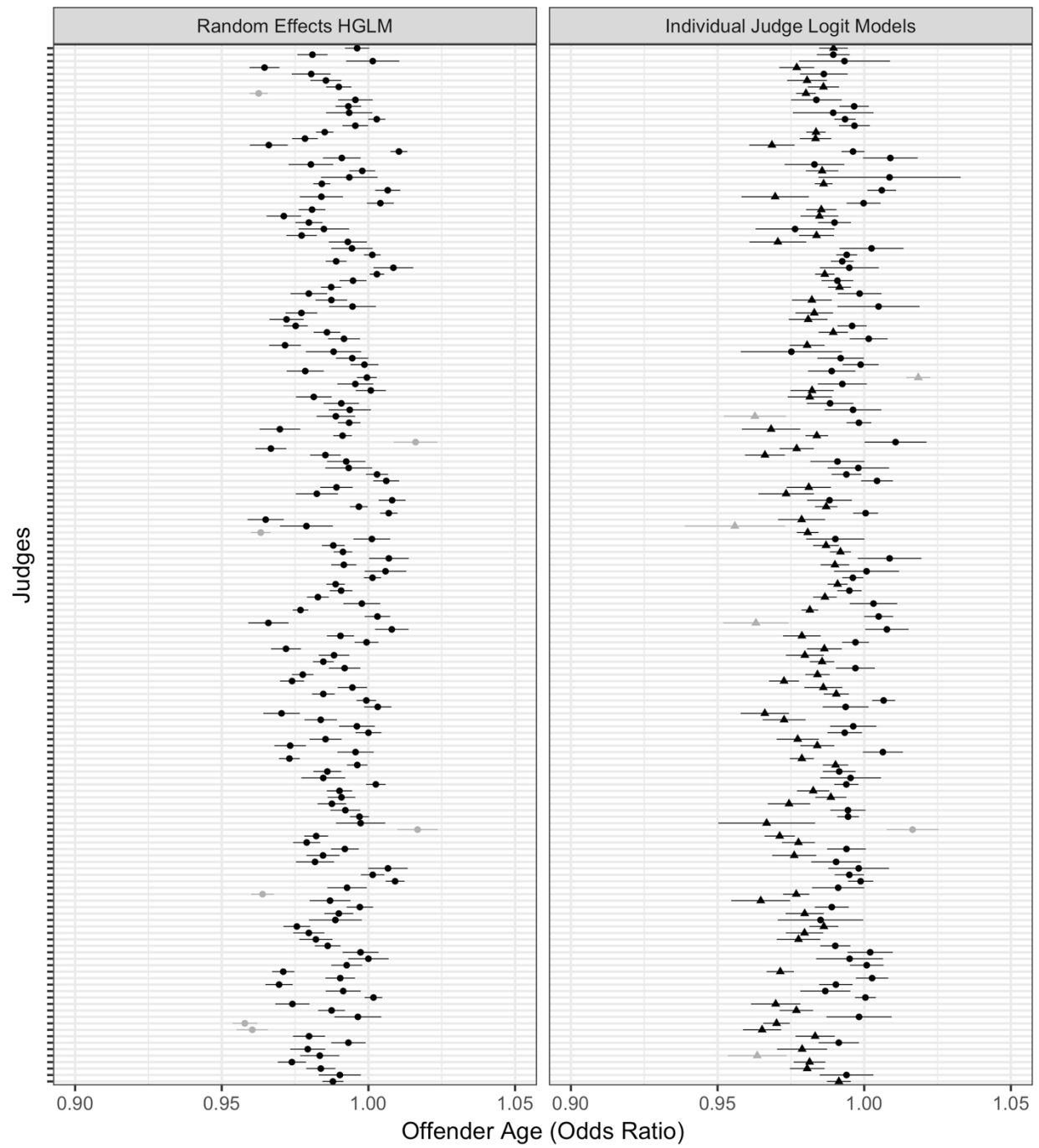




Figure 3. HGLM and Logit Effects for Age



incarceration for younger offenders are limited to 80 of the 161 judges, and one judge is more likely to incarcerate older offenders.

The findings for female offenders are presented in Figure 4, and the HGLM estimates show substantial pooling when compared to the judge logit models. Odds ratios range from 0.55 to .90 (standard errors from 0.04 to 0.09) in the random effects panel, and 0.19 to 1.48 (standard errors from 0.05 to .72) in the logit models. Nearly all of the judges' values in the HGLM fall in the 95 percent predicted range of 0.53 to 0.90, while a number of judges are outside the 0.33 to 0.97 range provided by the logit models. In terms of statistically significant effects, 112 of the 161 judges are less likely to incarcerate female offenders than males.

Race effects can be found in Figure 5, and from a pooling standpoint, appear contrary to previous results. Odds ratio ranges associated with offender race are much larger than what was found in the analyses above, spanning 0.52 to 4.09 (standard errors 0.50 to 0.67) in the HGLM, and 0.42 to 5.49 (standard errors 0.08 to 3.15) in the logit models. Ninety-five percent of the values are predicted to fall between 0.58 and 3.20 and 0.93 and 2.09 in the HGLM and logit models, respectively. Whereas analysis of the other sentencing predictors showed less variation in the multilevel analysis compared to the judge logit models, the race models show an opposite pattern. The difference is due to several judges' multilevel estimates for race moving farther, as opposed to closer, to the overall mean of 1.36. Further, the individual judge logit model p-values suggest that 90 of the 161 judges are more likely to incarcerate black offenders compared to whites.

The final incarceration models provide findings concerning mode of conviction (Figure 6). Odds ratios for the HGLM are between 0.51 and 4.65 (standard errors ranging from 0.10 to 1.22), and 0.32 and 8.14 (standard errors ranging from 0.10 to 5.78) for the logit models. The

Figure 4. HGLM and Logit Effects for Female Offenders

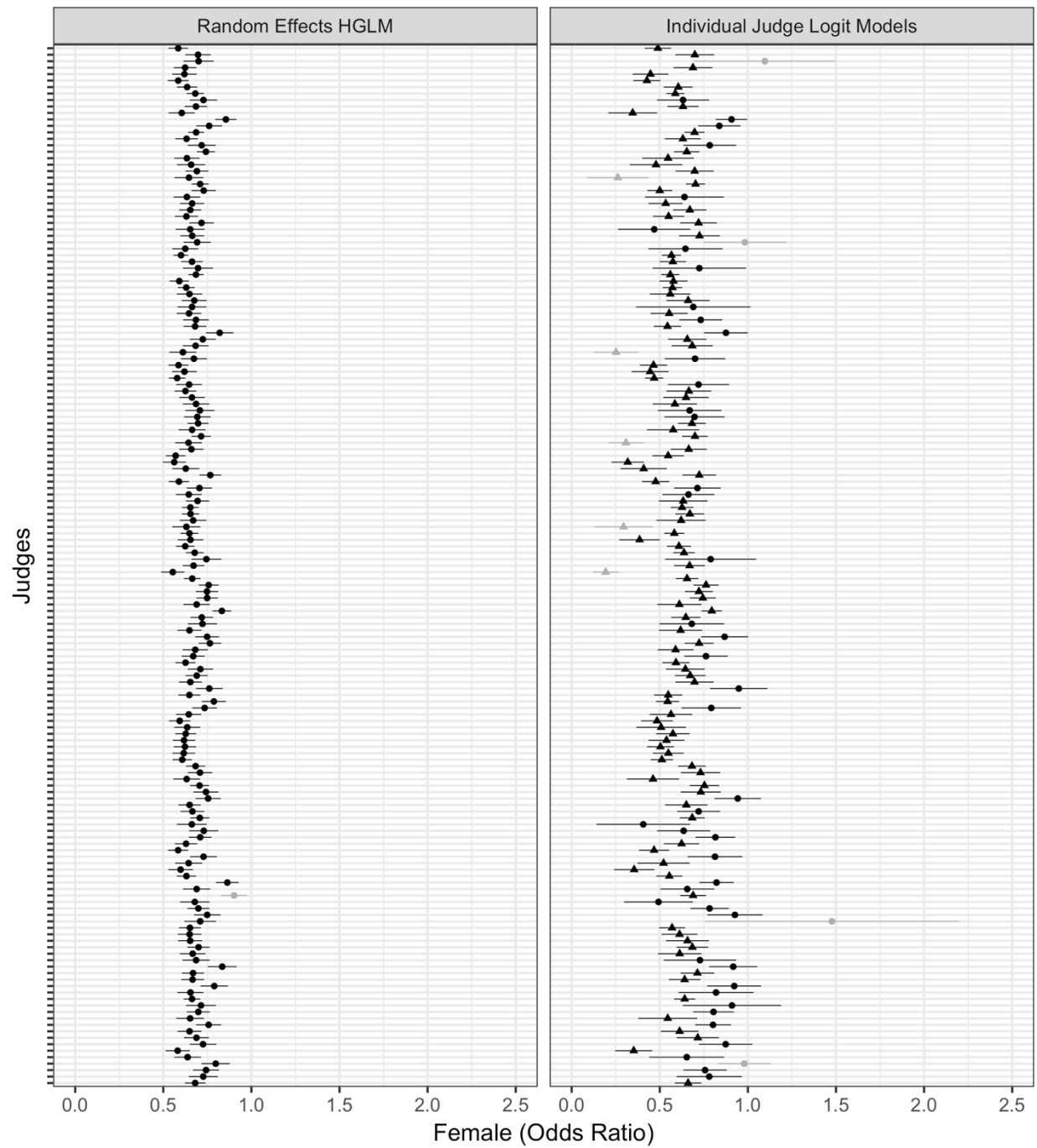


Figure 5. HGLM and Logit Effects for Black Offenders

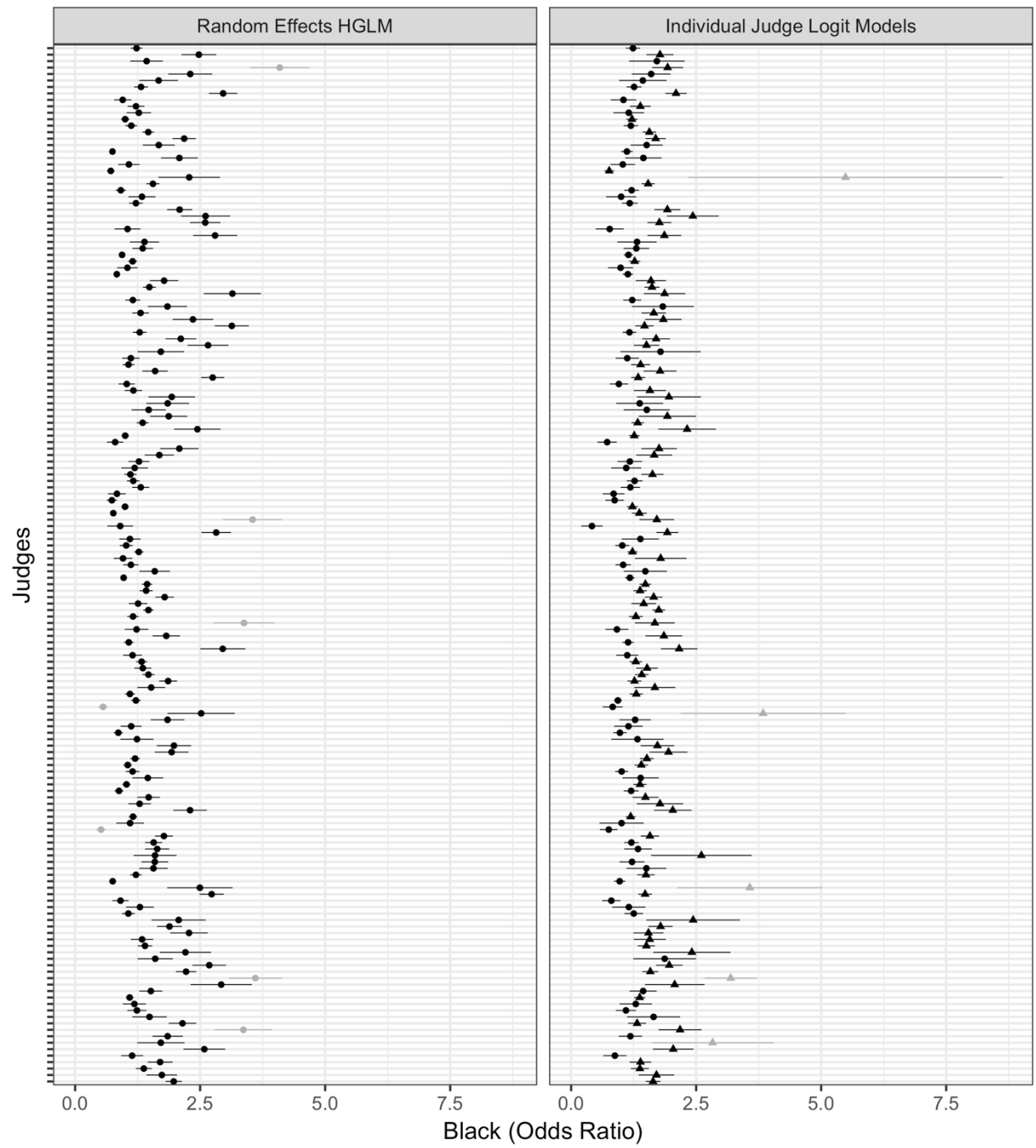
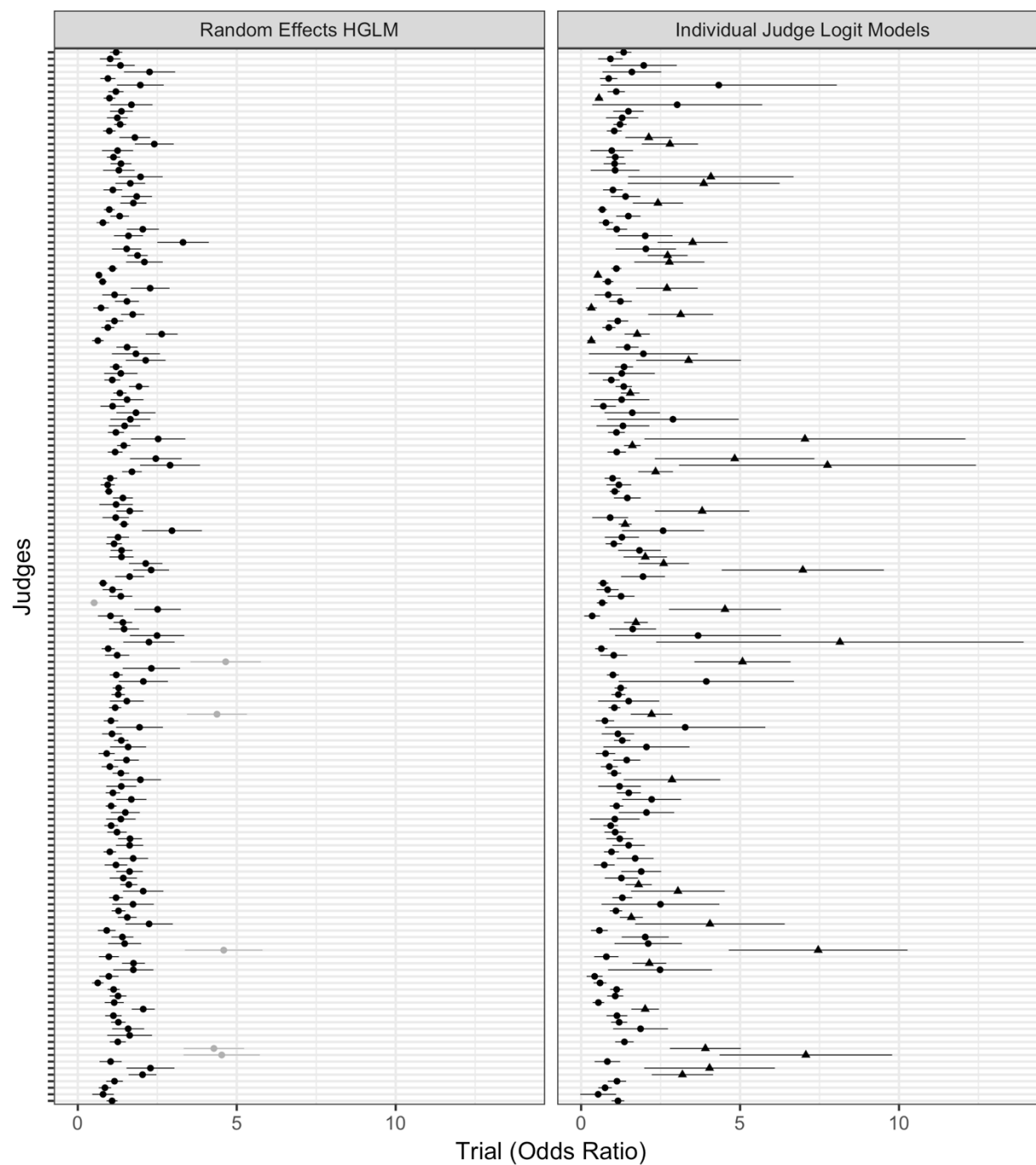


Figure 6. HGLM and Logit Effects for Trial Convictions



judge logit plot excludes three judges whose standard errors were extremely high (i.e., over 20) and likely unreliable; due to pooling, these judges' estimates are included in the HGLM and were shrunk to the overall mean of 1.36. Results also show that 95 percent of the predicted values lie between 0.53 and 3.50 in the HGLM, and 0.40 and 3.36 in the logit models. Comparing the panels, the HGLM estimates show substantial pooling. This suggests that the small number of offenders convicted after trial produce less reliable estimates across a large number of judges, resulting in more shrinkage to the overall mean. Finally, despite the fixed effect trial estimate in Table 4 being statistically significant, only 45 of 161 judges are more likely to incarcerate offenders convicted after trial as opposed to those entering a guilty plea.

### *Sentence Length Models*

The next set of findings provide information for legal and extralegal effects on the length of the sentence imposed. Figure 7 shows the percent increase for a one-unit increase in offense severity for the sentence length models. Percent increases range from 22 to 82 (standard errors from 0.01 to 0.03) in the LMM, and 11 to 98 (standard errors 0.01 to 0.12) in the judge OLS models. Ninety-five percent of the predicted values lie between 27 and 61 percent in the LMM, and 23 and 63 percent across the OLS models. Similar to offense severity findings in the incarceration models, slightly less variation is observed in the LMM estimates than the OLS values. P-values from each judges' individual OLS model indicate that offense severity significantly affects the length of the sentence imposed for all 161 judges. For prior record (Figure 8), the LMM findings show a one-unit increase in prior record score results in an increase in sentence length ranging from 11 to 32 percent across the judges (standard errors 0.01 to 0.03), and 95 percent of the values are predicted to fall between 11 and 31 percent. In the OLS models, prior record effects span between seven and 32 percent, with slightly higher standard

Figure 7. LMM and OLS Effects for Offense Severity

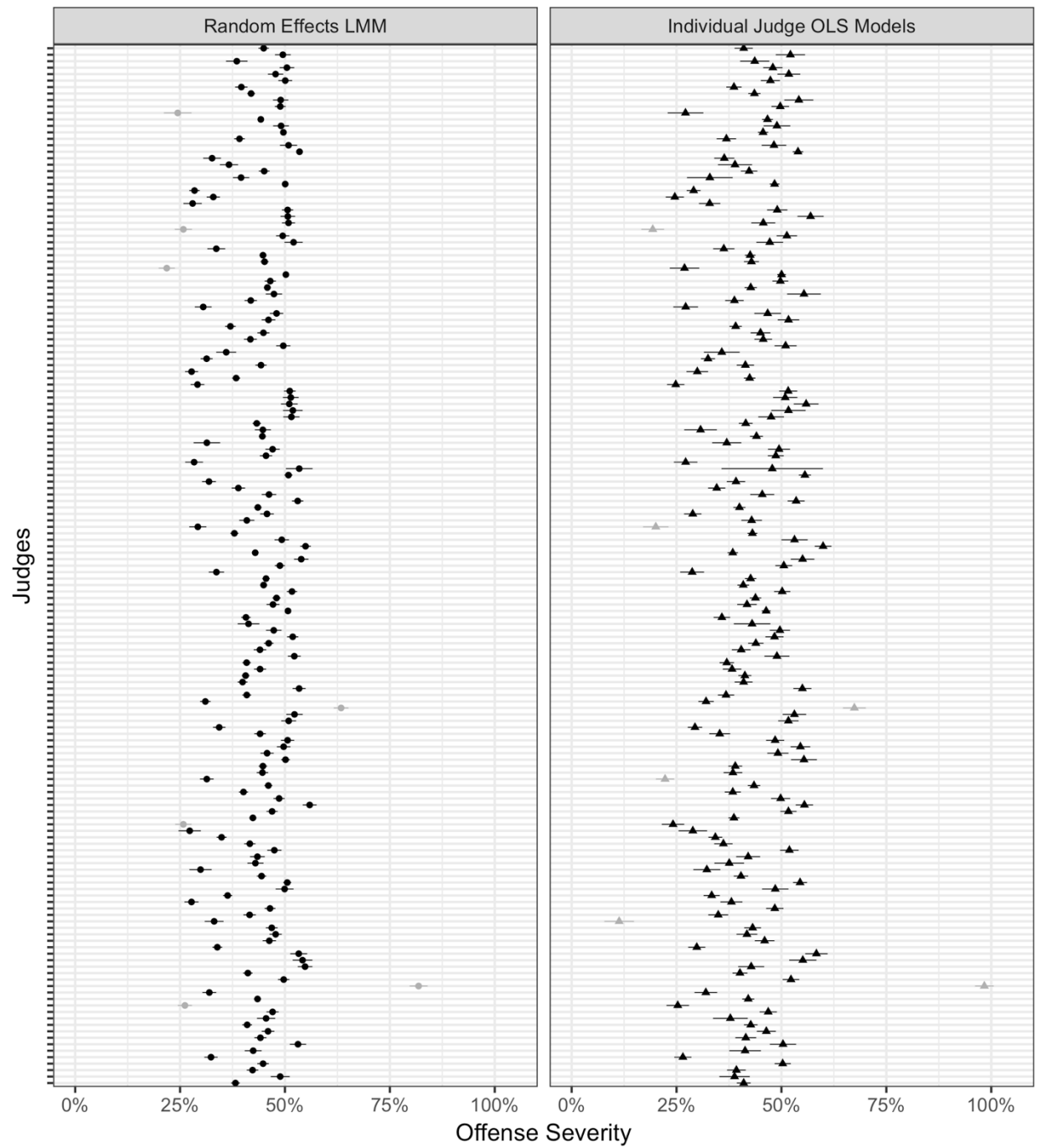
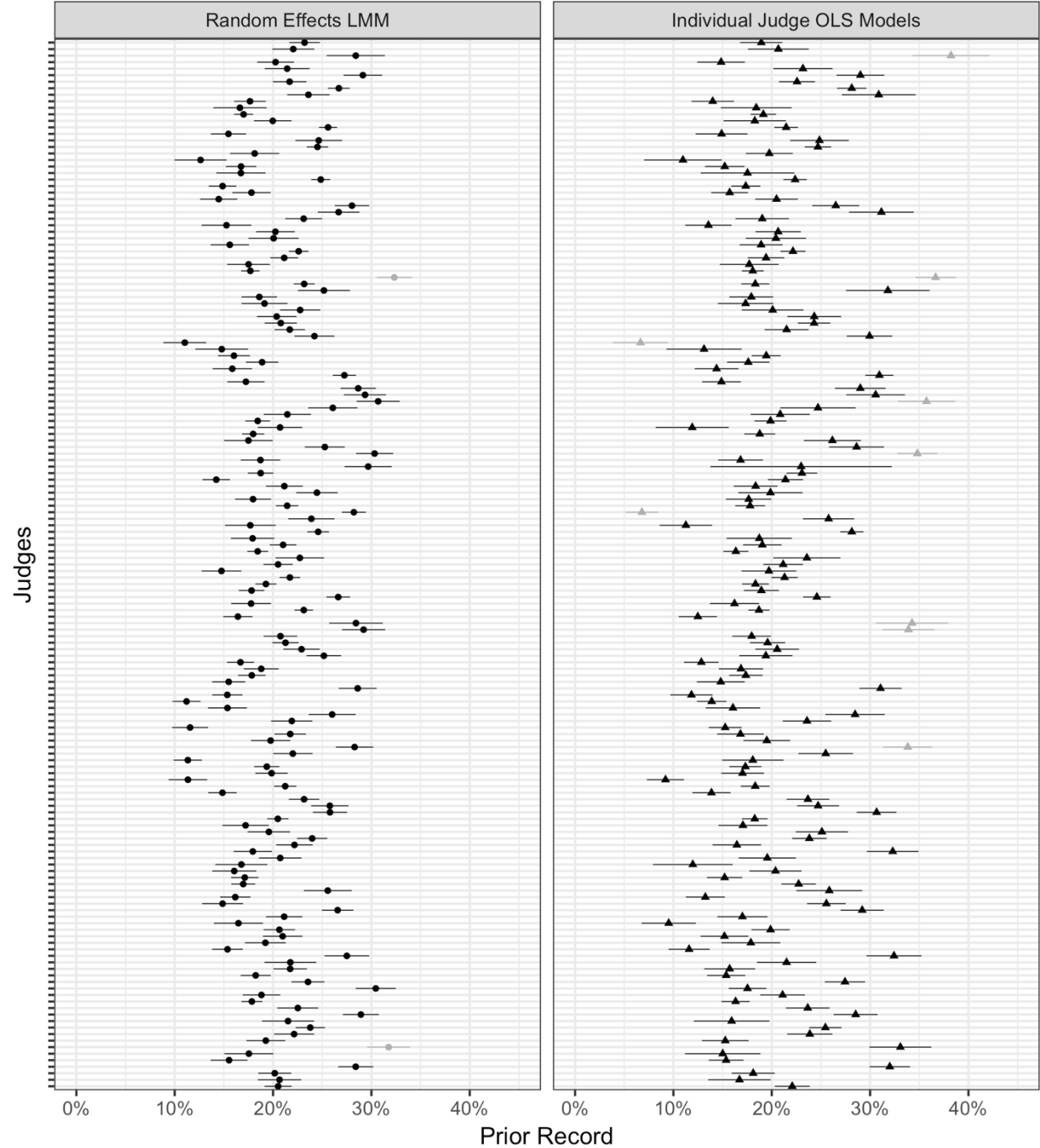


Figure 8. LMM and OLS Effects for Prior Record





errors overall (0.01 to 0.09). Predicted values for 95 percent of the judges range from eight to 33 percent. Similar to offense severity, the LMM estimates show some pooling to the overall mean of 21 percent, and the OLS models reveal that prior criminal history results in a significant increase in sentence length across all 161 judges.

Figures 9 through 12 examine differences between the LMM and OLS sentence length models for extralegal factors. Concerning offender age (Figure 9), the LLM shows a range of a two percent decrease in sentence length to a one percent increase (standard errors 0.00), while the OLS models range from a two percent decrease to a three percent increase (standard errors from 0.00 to 0.01). Predicted values for 95 percent of the judges lie between negative one percent and one percent for both the LMM and the OLS models. As expected, the LMM estimates are shrunk closer to the overall mean of zero, and statistically significant effects from the OLS models (39 judges with  $p < .05$ ) are less consistent than what was found for offense severity and prior record. In addition, whereas nearly all judges in the logit models were more likely to incarcerate younger offenders, statistically significant effects from the OLS models show that 33 of the 39 judges impose longer sentences for older offenders.

As shown in Figure 10, estimates for female offenders range from a 52 percent decrease in sentence length to a 21 percent increase (standard errors from 0.04 to 0.11) for the LMM, and a 69 percent decrease to a 25 percent increase (standard errors 0.05 to 0.40) in sentence length for the judge OLS models. Estimates in the LMM model are pooled closer to the model mean of an 11 percent decrease in sentence length, and predicted values for 95 percent of the judges fall between negative 35 percent and 13 percent. For the OLS models, predicted estimates range from negative 44 to 20 percent. The judge OLS models indicate only two judges show significant

Figure 9. LMM and OLS Effects for Age

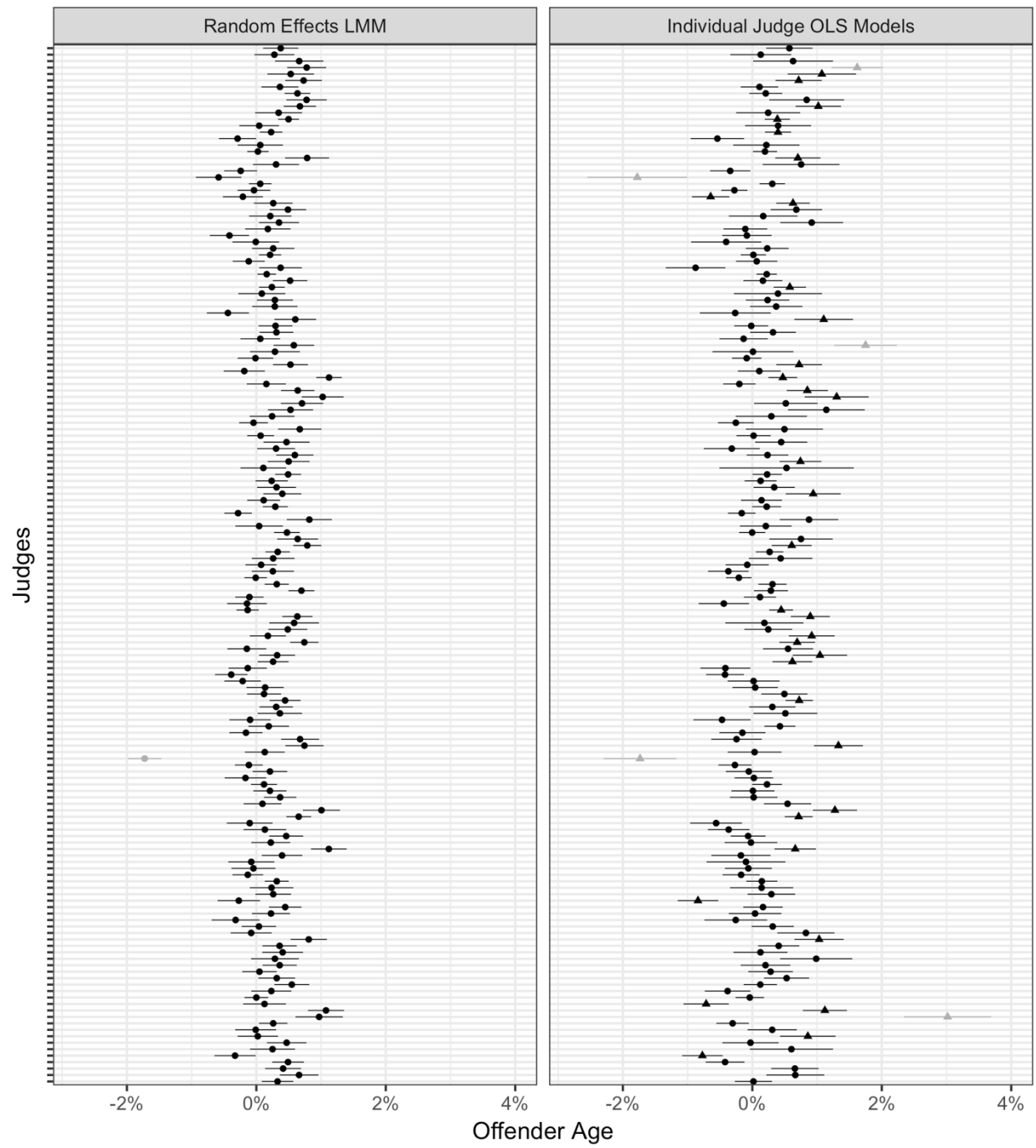
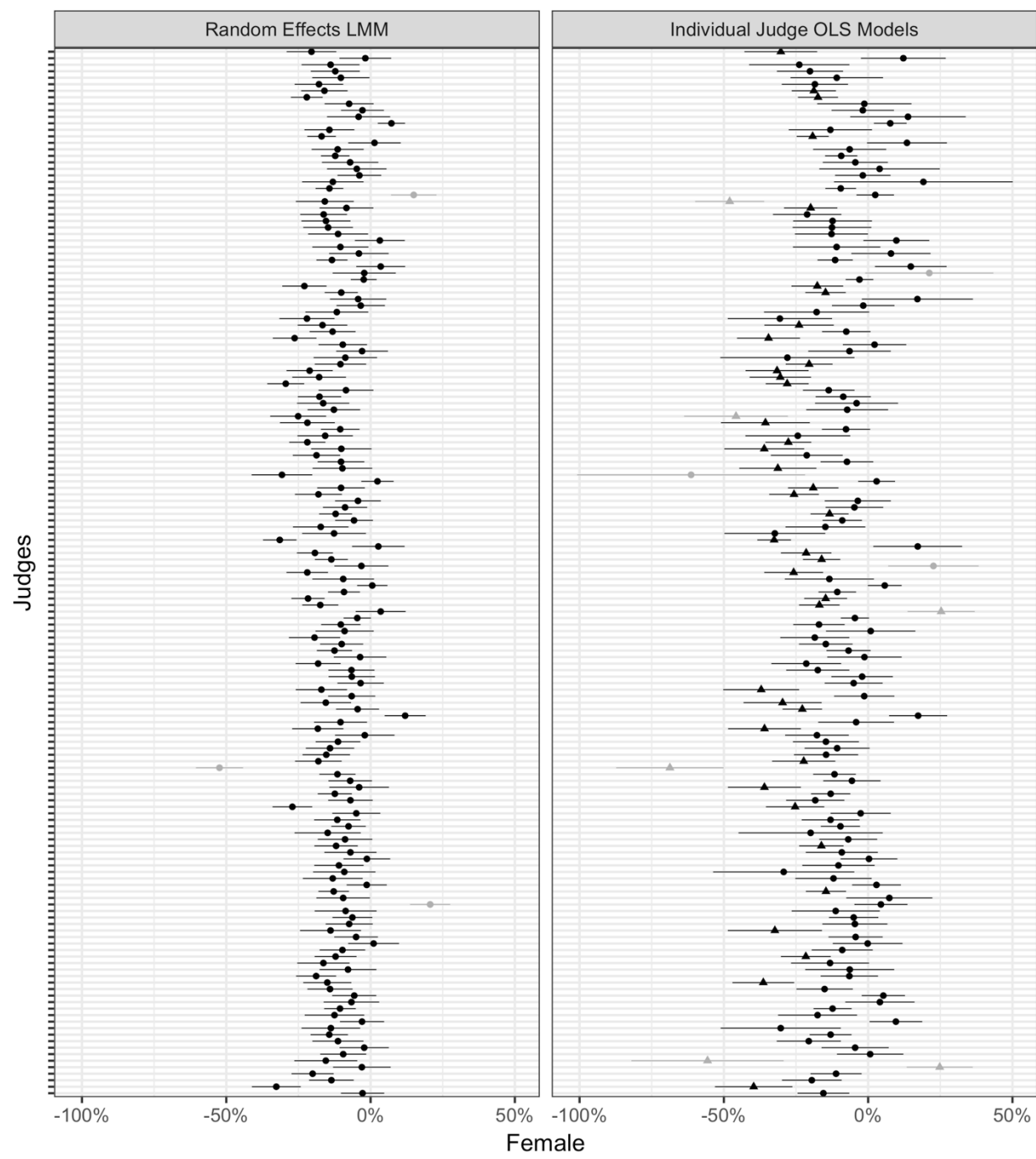


Figure 10. LMM and OLS Effects for Female Offenders



effects for sentencing females to longer periods of incarceration, while the remaining 43 of the 161 judges impose shorter sentences for female offenders.

Offender race effects are displayed in Figure 11. Values range from negative 37 percent to 25 percent (standard errors 0.04 to 0.11) in the LMM and negative 55 percent and 31 percent (standard errors 0.05 to 0.28) in the OLS models. The lower end of the predicted values for 95 percent of the judges is the same in both the LMM and OLS models (negative 27 percent), and the upper end is 24 percent and 25 percent in the LMM and OLS models, respectively. Once again, the LMM estimates show some shrinkage to the model mean of negative two percent. Among the 24 of 161 judges who exhibit statistically significant effects in the OLS models, 14 judges impose shorter sentences for black offenders, and 10 sentence black offenders to longer periods of incarceration.

The final figure for the sentence length models shows effects associated with trial convictions (Figure 12). Judge percents range from negative five to 75 (standard errors from 0.07 to 0.17) in the LMM, and negative 33 to 113 (standard errors 0.05 to 0.84) in the OLS models. Substantial pooling occurs around the LMM mean of 22 percent, and predicted values for the judges fall between negative 16 and 59 percent in the LMM, and negative 29 and 71 percent across the OLS models. Though no judges sentence offenders convicted after trial to shorter sentences, the statistically significant effects from the OLS models show that 55 of the 161 judges impose longer sentences for offenders convicted after trial.

Figure 11. LMM and OLS Effects for Black Offenders

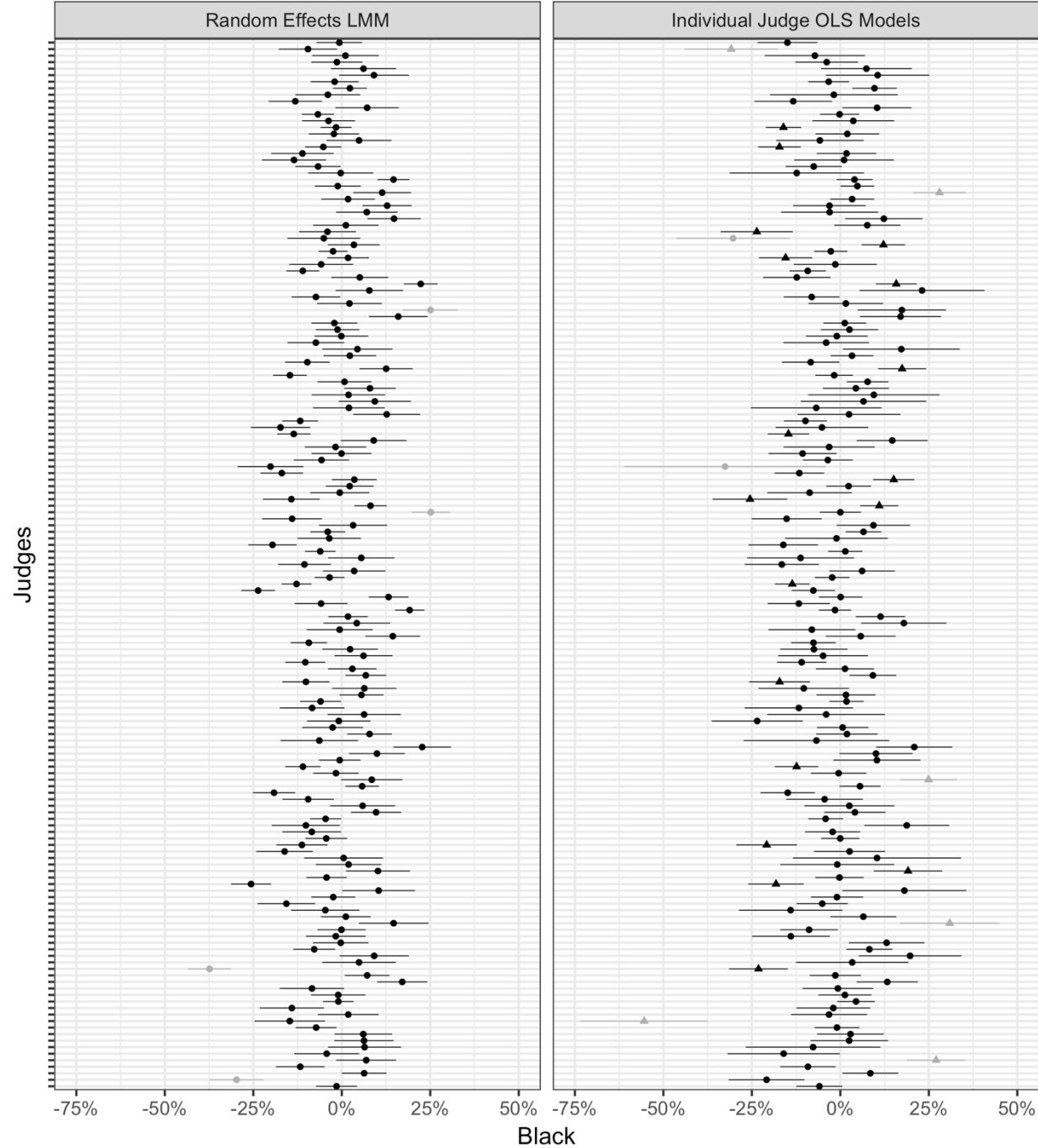
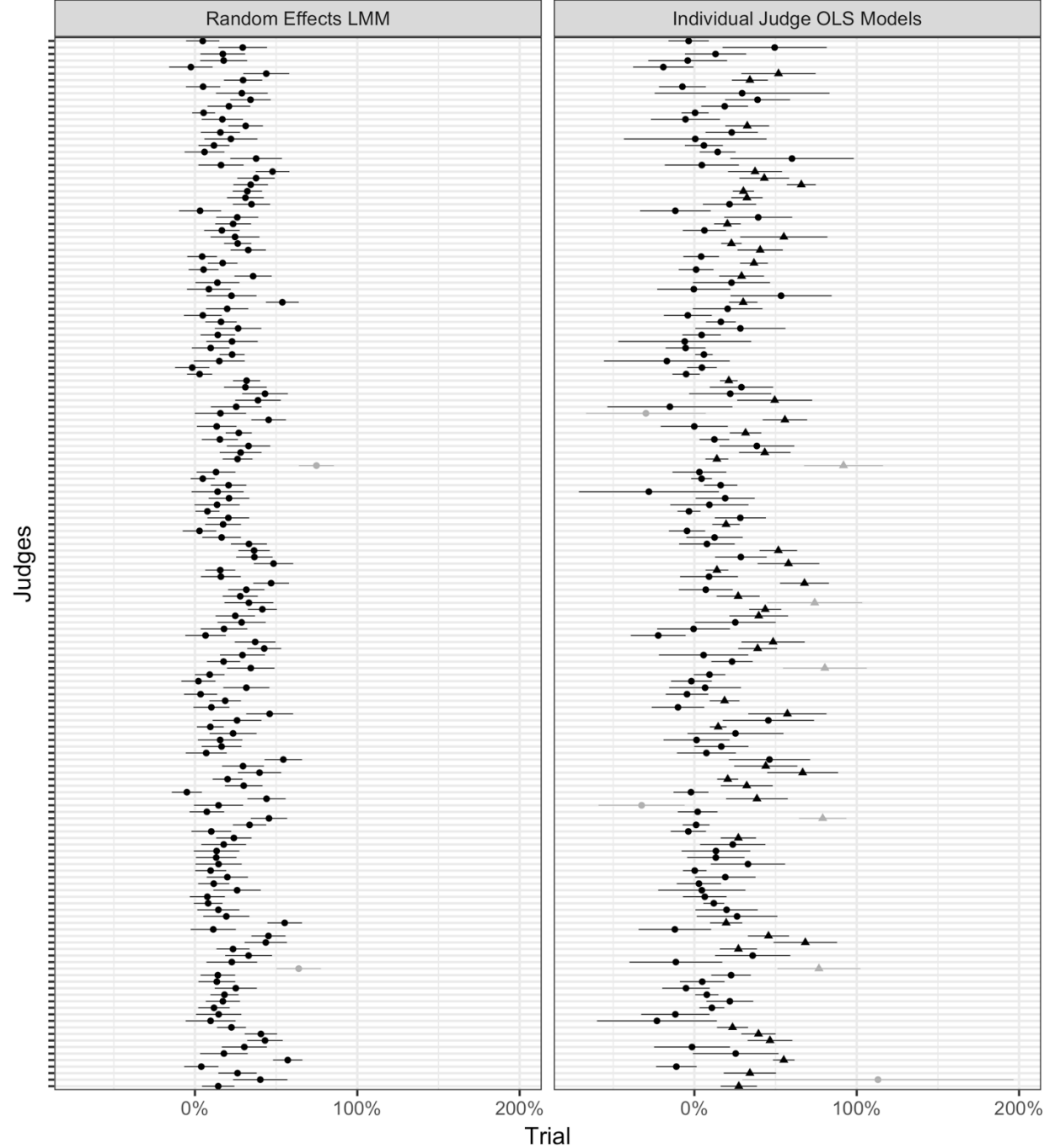


Figure 12. LMM and OLS Effects for Trial Convictions



Before moving to the third research question, examination of the ways in which judges consider offender and case characteristics for the incarceration decision and the length of the sentence imposed revealed additional information that is worth noting. In the logit and OLS models, nearly all judges showed statically significant effects for offense severity and prior record, indicating that increases in these legally relevant variables increased the odds of incarceration and the length of the sentence imposed. Figures 13 through 16 provide side by side comparisons of significant and non-significant effects from the logit and OLS models for offender age, gender, race, and mode of conviction. The results suggest that some judges consider these extralegal factors for both sentencing outcomes, others when deciding whether to incarcerate but not when determining the appropriate sentence length (and vice versa), and still others who do not consider these factors in either decision.

Concerning offender age (Figure 13), most judges (65 of 161) are nonsignificant for age effects in either decision, followed by 57 judges who consider age in both the decision to incarcerate and the length of the sentence imposed. Figure 14 displays effects for female offenders and indicates that only 32 of the 161 judges consider gender in both decisions, 80 show significant effects for incarceration alone, and 13 for just the sentence length decision. Significant race effects (Figure 15) are most prevalent when deciding whether to incarcerate (77 of 161 judges), followed by 60 judges who do not consider race a significant factor when determining sentence severity. Finally, for just over half of the judges (84 of 161), whether offenders enter a guilty plea or take their case to trial seems to have no bearing on either the decision to incarcerate or sentence length (Figure 16). Remaining comparisons can be found in Table 5, but overall these findings suggest individual judges' consideration of extralegal factors is conditioned by the sentencing outcome.

Figure 13. Logit and OLS Effects for Age

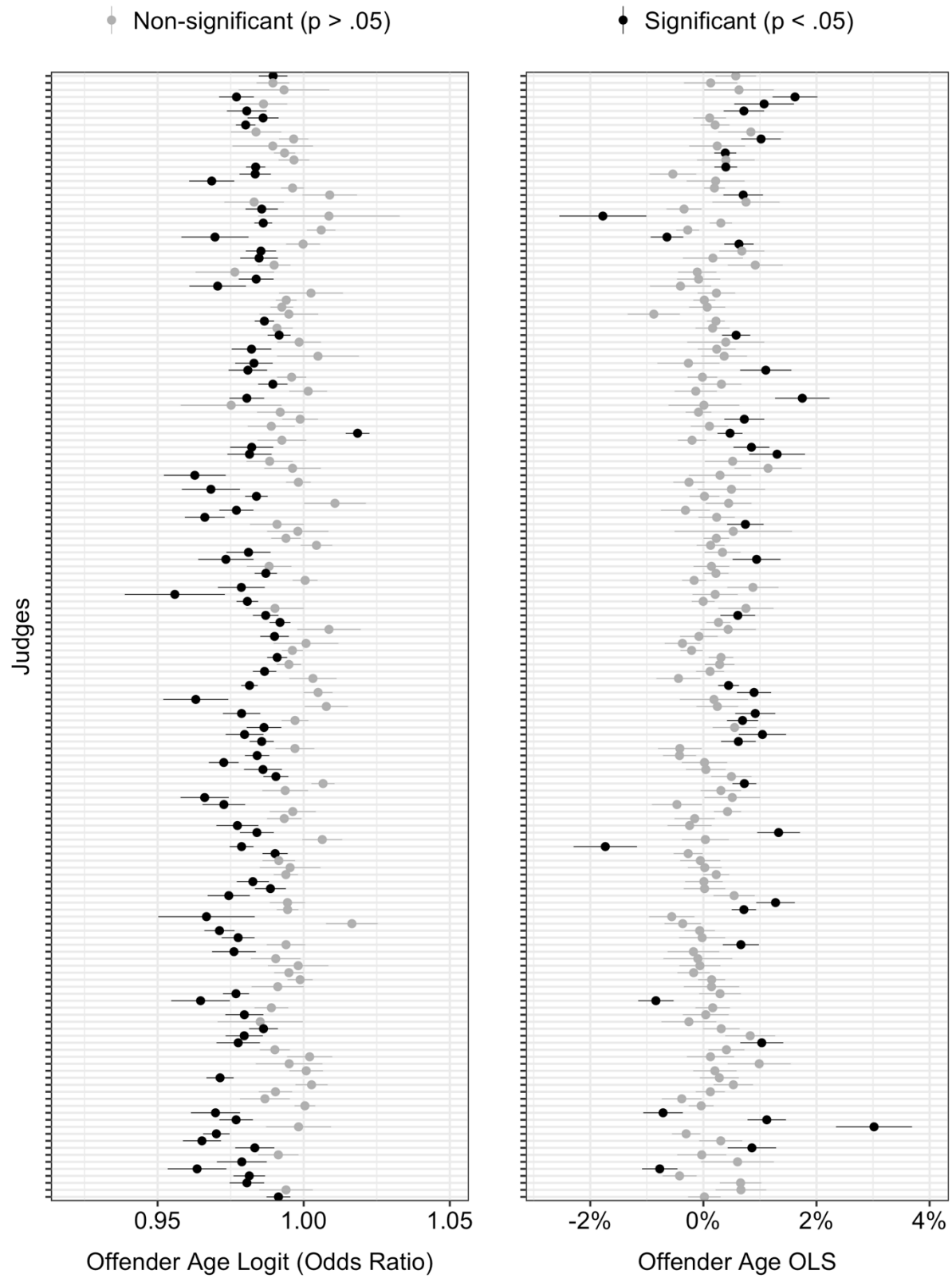




Figure 14. Logit and OLS Effects for Female Offenders

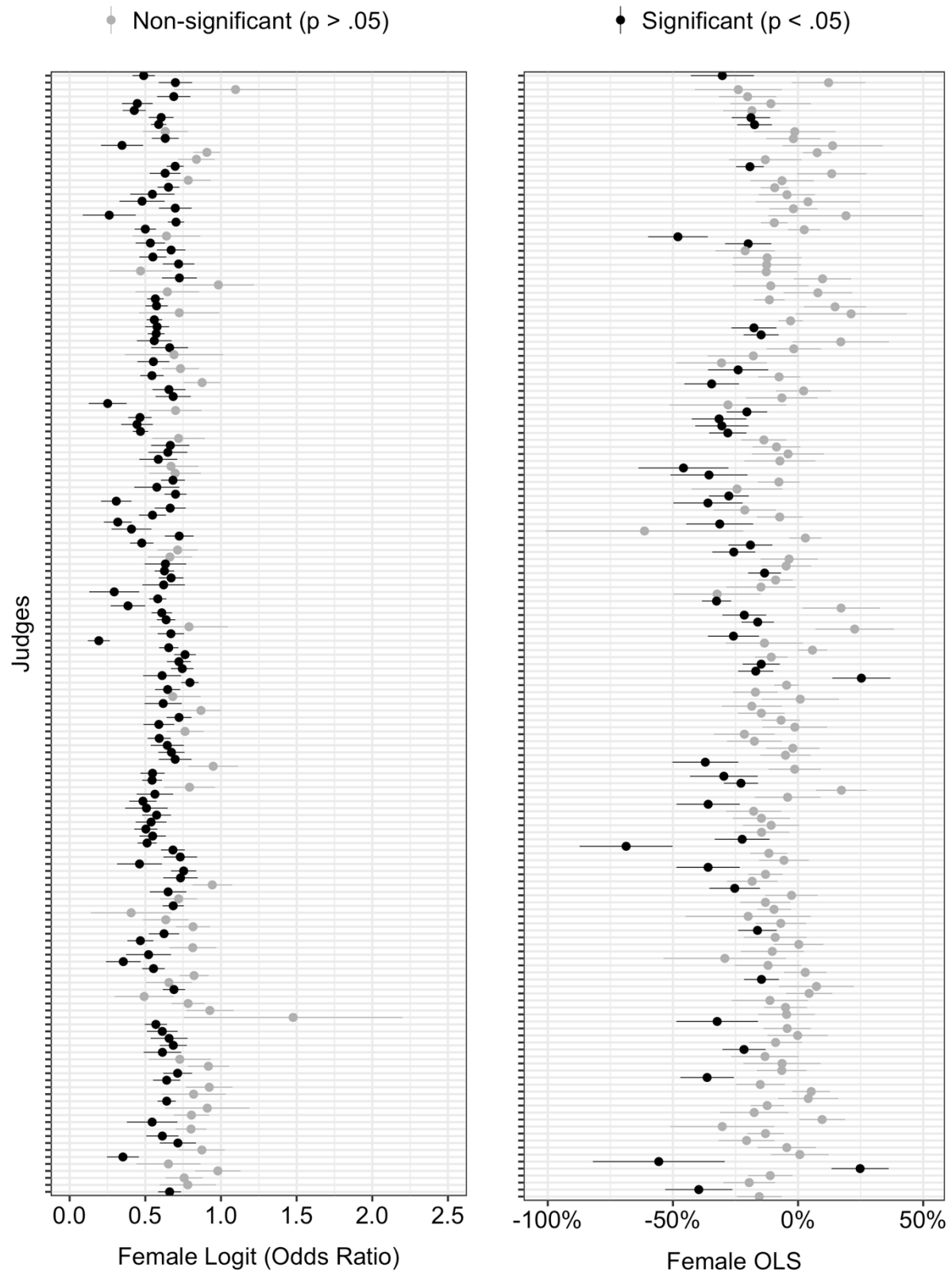


Figure 15. Logit and OLS Effects for Black Offenders

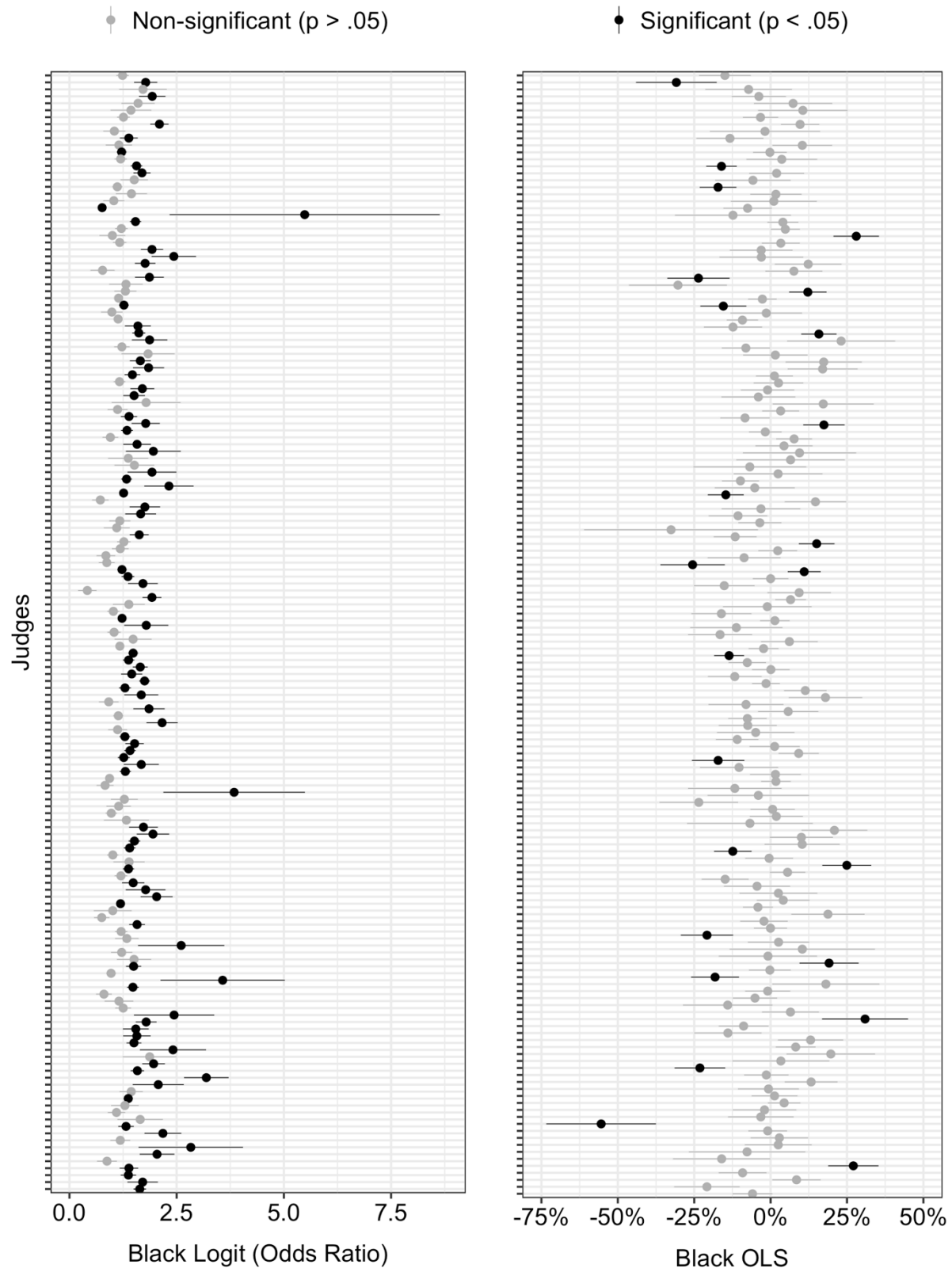


Figure 16. Logit and OLS Effects for Trial Convictions

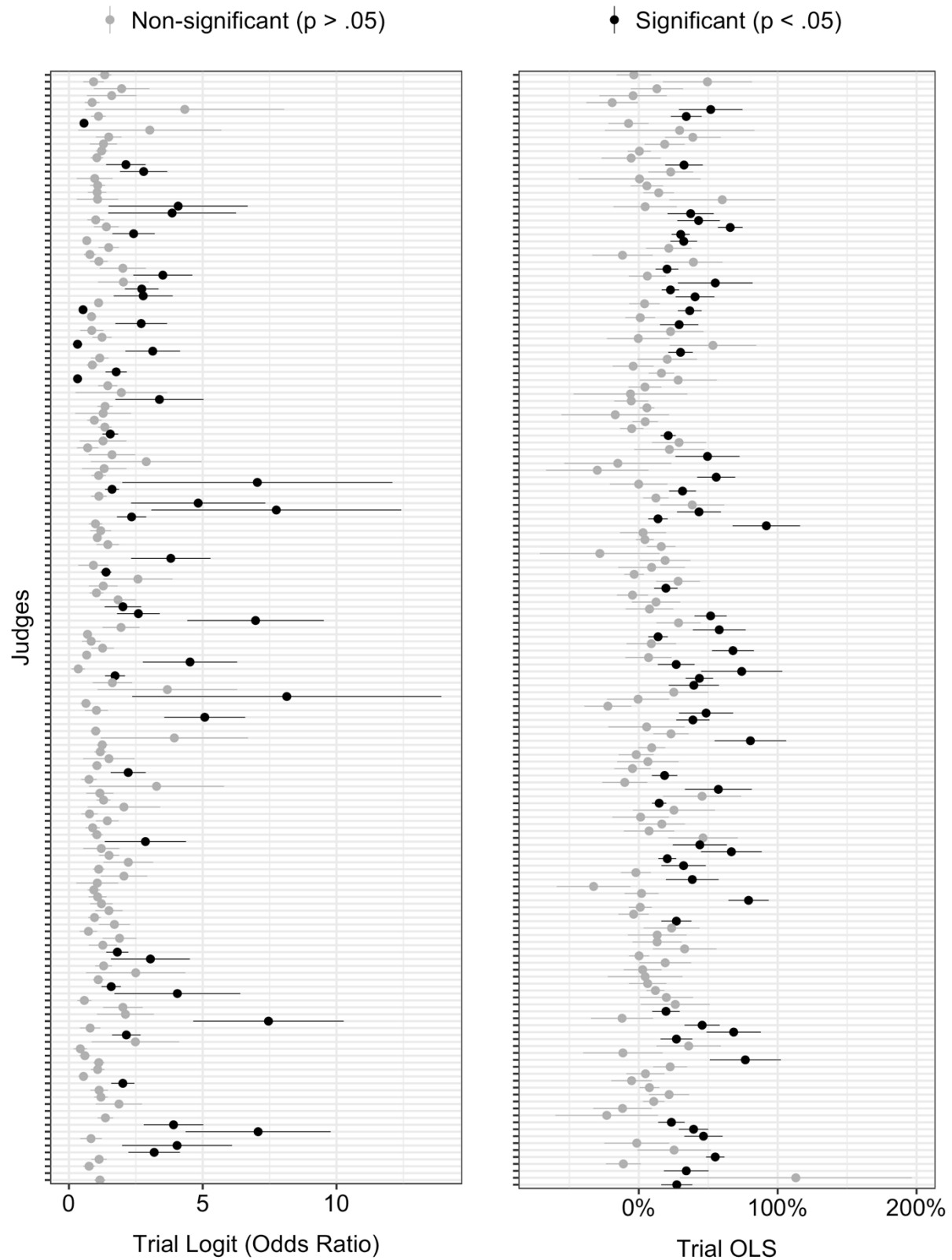


Table 5. Number of Judges with Significant Effects for Sentencing Outcomes  
(N = 161)

<i>Sentencing Outcome</i>	<u>Age</u>	<u>Female</u>	<u>Black</u>	<u>Trial</u>
Incarceration and Sentence Length	24	32	13	23
Incarceration Only	57	80	77	22
Sentence Length Only	15	13	11	32
Neither Incarceration Nor Sentence Length	65	36	60	84

#### *Analysis of Judges within Court Communities*

The final portion of the analysis examined the extent to which judges in the same courthouses exhibit similar sentencing patterns. Integrating the focal concerns and court community perspective suggests that sentencing decisions may be influenced by the court community in which punishment decisions occur, but differences in court actor autonomy in large, medium, and small courts may condition this relationship. Specifically, court communities in large courts have been characterized as diffuse, based in part on a high degree of autonomy between members of the courtroom workgroup, whereas the close working relationship developed among small court actors is expected to limit autonomy (Eisenstein, Flemming, & Nardulli, 1988; Jacob, 1997). Court actor autonomy in medium courts is expected to fall somewhere in between large and small courts. Figures 17 through 22 provide results from select courts. The plots display individual judge logit (odds ratios and standard errors) and OLS (percents and standard errors) findings for legal and extralegal factors. Estimates in black represent coefficients with a p-value of <.05, while those in gray are above this threshold.

Significant legal and extralegal effects for the remaining courts are included in Tables 6 through 8.

Figure 17 provides findings for the largest court in the sample (Large Court 2). The results indicate that while legal factors are the primary determinants of punishment, substantial variation exists in the role extralegal factors play in sentencing outcomes. Nearly all 29 judges are associated with increasing the likelihood of incarceration and sentence length as offense severity and prior record increase. Concerning extralegal factors, very few judges consider offender age in the incarceration decision (five of 29) and when determining sentence length (eight of 29). In addition, while five of the judges are associated with a very slight increase in sentence length for younger offenders, the other three sentence older offenders to longer sentences. Significant findings are more prevalent for gender, with just over half of the judges being less likely to incarcerate female offenders. Fewer judges consider gender in the sentence length decision (11 of 29), and one judge imposes longer sentences for female offenders compared to males. Similar to age, very few judges are significant for race effects, though race in this court appears to play a larger role in determining the sentence length than whether to incarcerate black offenders. Finally, 13 of 29 and 18 of 29 judges are more likely to incarcerate and impose longer sentences for offenders convicted after trial compared to those who enter a guilty plea, respectively. With the exception of gender effects for the incarceration decision, more judges exhibit significant effects associated with trial convictions than any other extralegal variable.

Table 6 provides information for all of the large courts. The table includes the percent of judges with significant effects for the key predictors of sentencing across both punishment decisions. Findings show significant effects associated with offense severity and prior criminal

Figure 17. Individual Judge Effects in Large Court 2

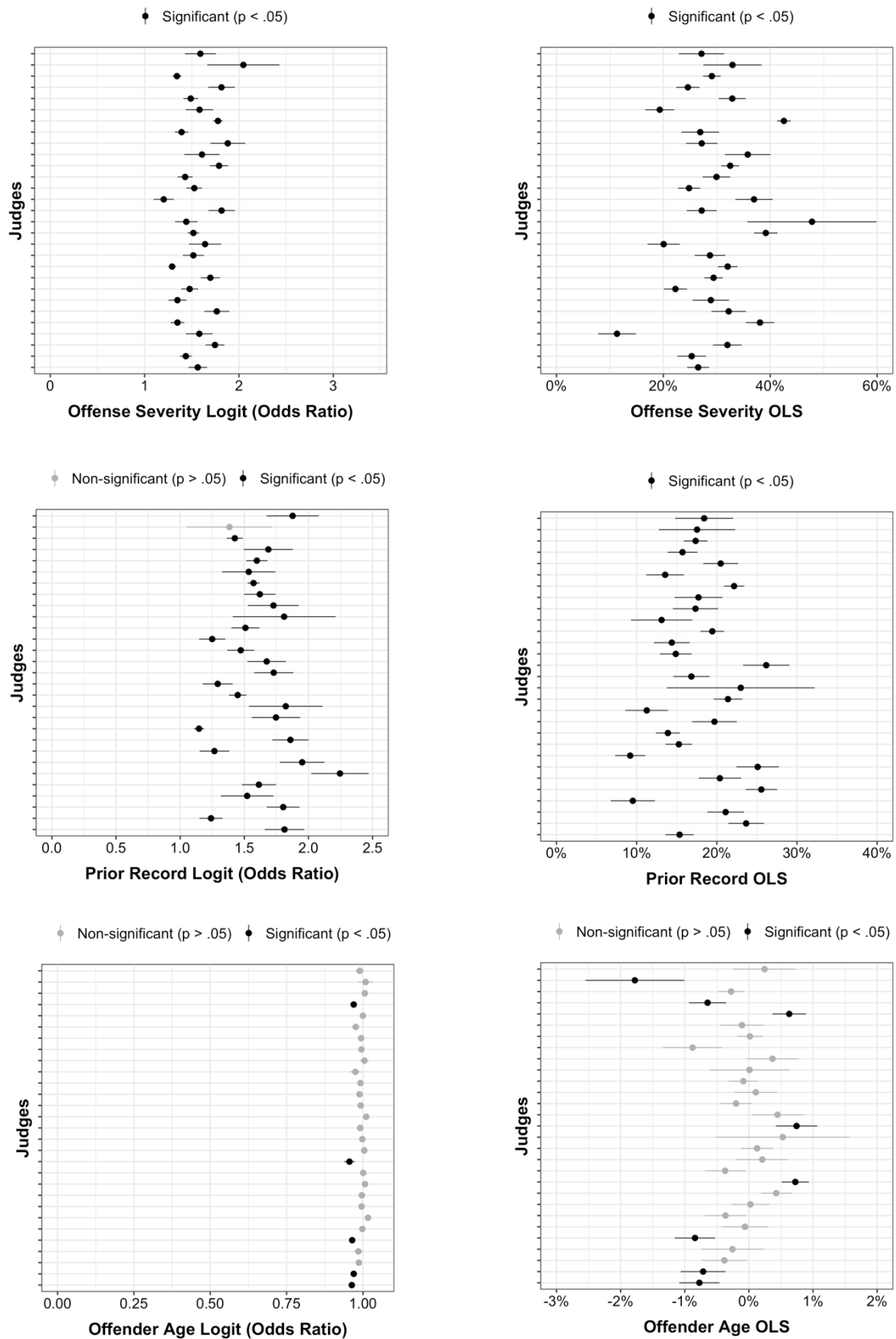


Figure 17 (cont'd).

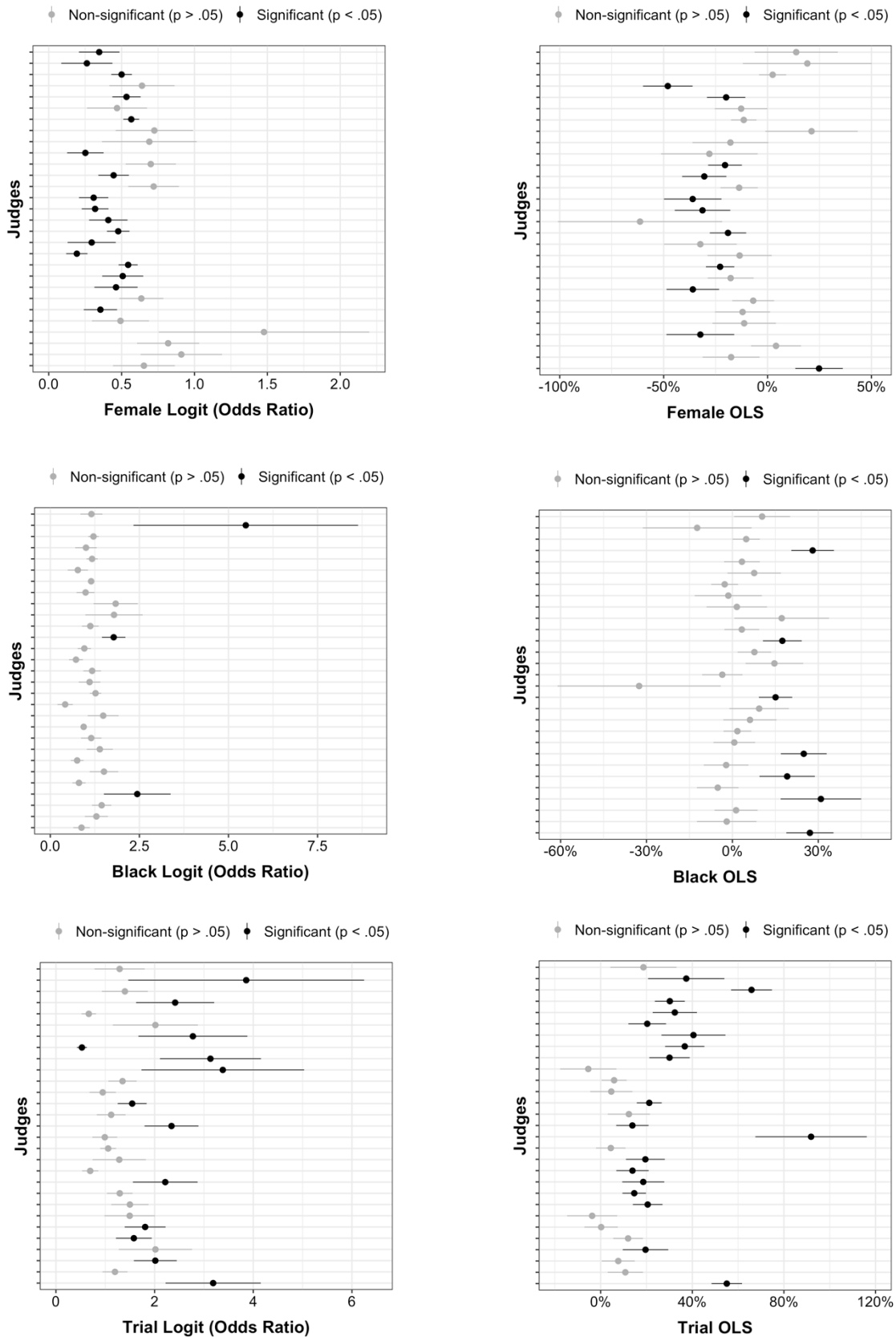


Table 6. Percent of Large Court Judges with Significant Effects

<u>Court</u>	<u>Offense Severity</u>	<u>Prior Record</u>	<u>Age</u>	<u>Female</u>	<u>Black</u>	<u>Trial</u>
<i>Large Court 1 (N=19)</i>						
Incarceration	89%	100%	74%	95%	74%	26%
Sentence Length	100%	100%	5%	32%	26%	16%
<i>Large Court 2 (N=29)</i>						
Incarceration	100%	97%	17%	59%	10%	45%
Sentence Length	100%	100%	28%	38%	24%	62%
<i>Large Court 3 (N=11)</i>						
Incarceration	100%	100%	36%	91%	64%	36%
Sentence Length	100%	100%	9%	18%	18%	27%
<i>Large Court 4 (N=16)</i>						
Incarceration	100%	100%	31%	75%	69%	19%
Sentence Length	100%	100%	19%	31%	13%	69%



history for nearly all judges in these courts. Conversely, judges in these large court communities vary more in significant effects associated with age, gender, race, and mode of conviction. Further, with the exception of trial convictions in Large Court 4, significant extralegal effects are more prevalent for the incarceration decision than the sentence length decision.

The next set of findings provides information on judges' sentencing patterns in medium sized courts, where court actor autonomy exists but to a lesser degree than found in large courts. Figure 18 provides findings for Medium Court 3, and shows highly consistent sentencing patterns among the judges in this court. Specifically, both judges increase punishment severity as offense severity and prior record increases, and sentence female offenders more leniently than males. In addition, both judges are less likely to incarcerate older offenders, are more likely to sentence blacks to jail or prison when compared to whites, and exhibit non-significant effects associated with age, race, and trial and the sentence length decision. The only exception is trial convictions in the incarceration models, where one judge is more likely to incarcerate offenders convicted after trial and the other is not.

In contrast, Figure 19 provides results from another medium court where individual judges exhibit substantial differences across extralegal effects. In Medium Court 6, while eight of the 10 judges are less likely to incarcerate older offenders, only half of the judges consider age in the sentence length decision. Notably, the latter increase sentence length as age increases, and the four judges exhibiting significant effects for both outcomes are more lenient on older offenders for the incarceration decision, but more punitive when determining the appropriate sentence length. Seventy percent of judges are less likely to incarcerate female offenders, and 40 percent impose shorter sentences for females than males. Additional inconsistency is found concerning race effects, where 60 percent of judges are more likely to incarcerate blacks than

Figure 18. Individual Judge Effects in Medium Court 3

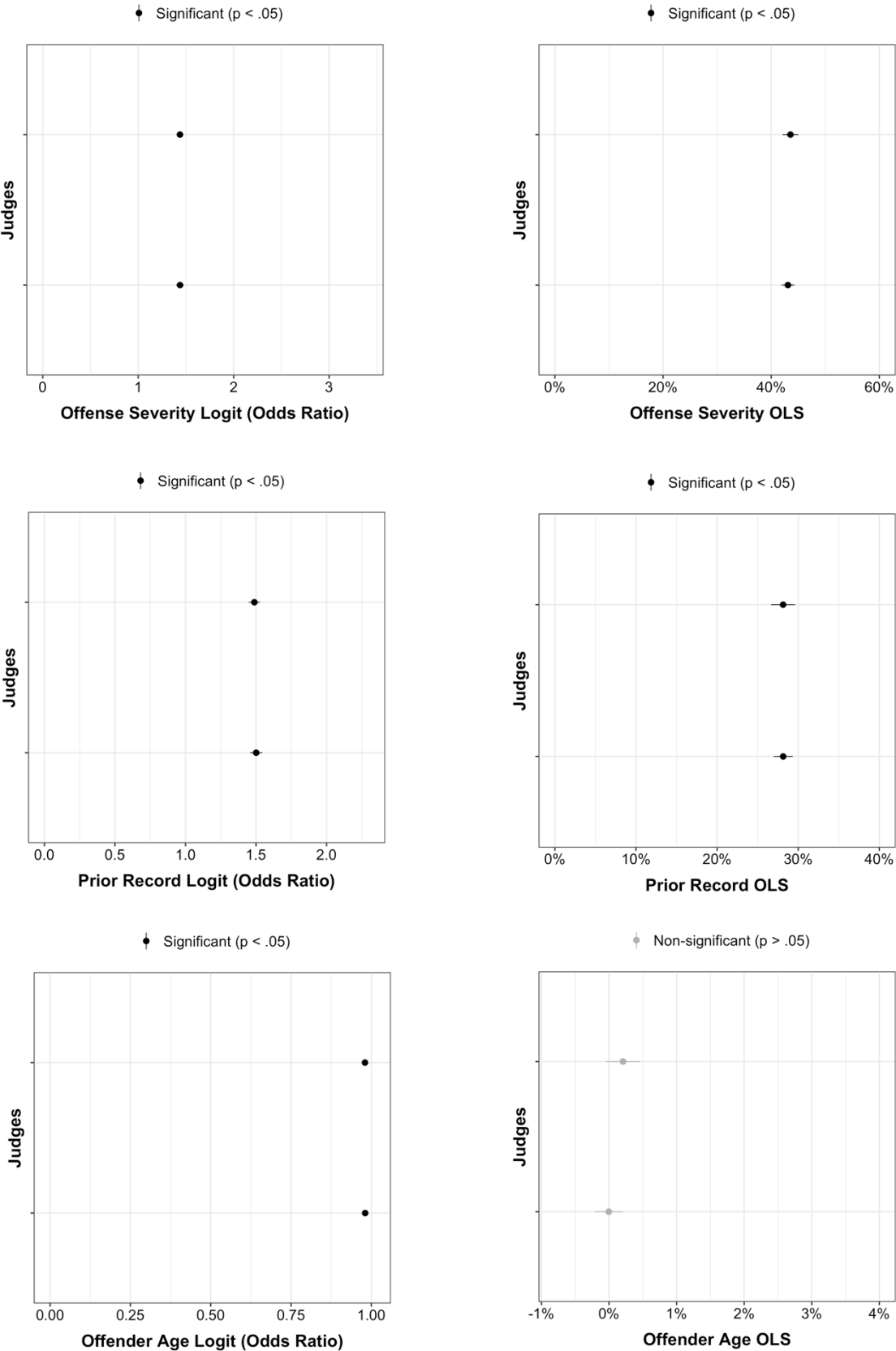


Figure 18 (cont'd).

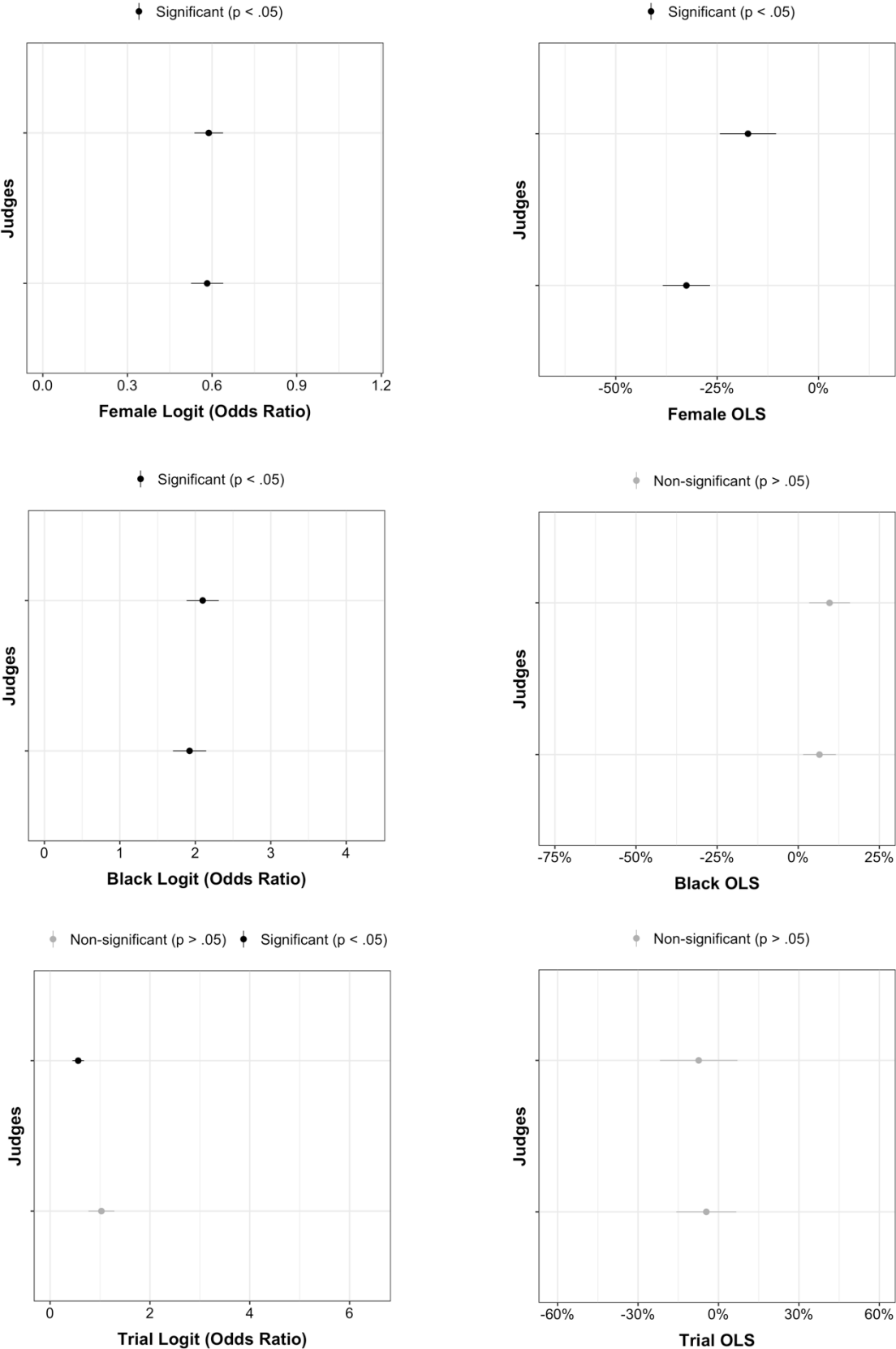


Figure 19. Individual Judge Effects in Medium Court 6

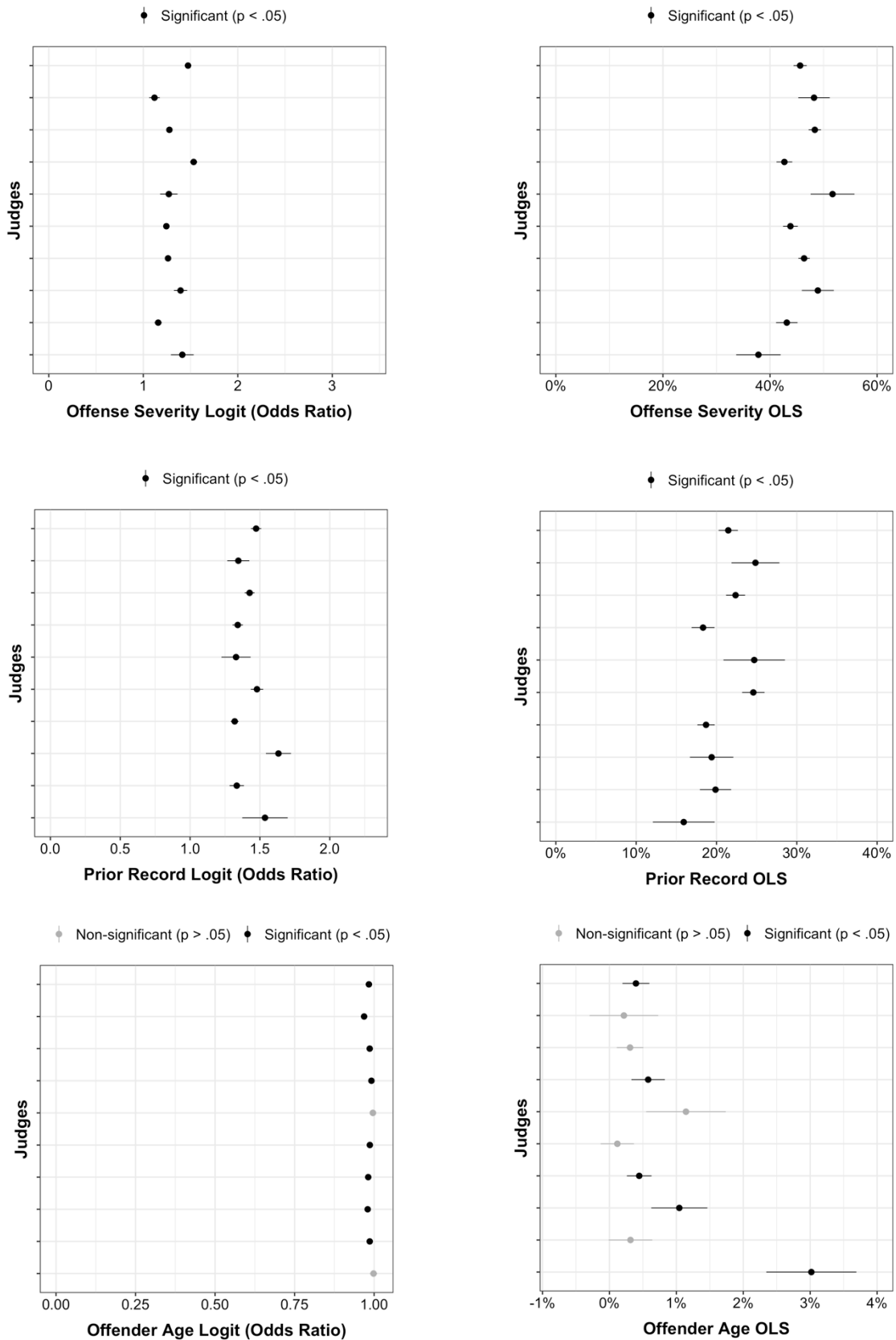
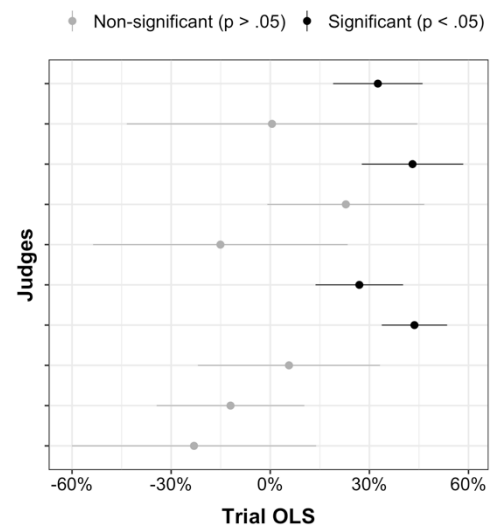
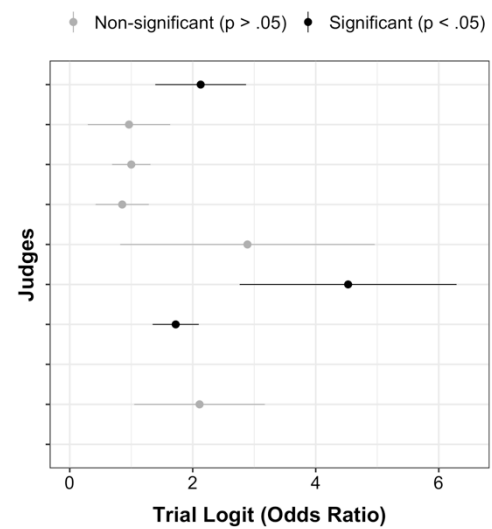
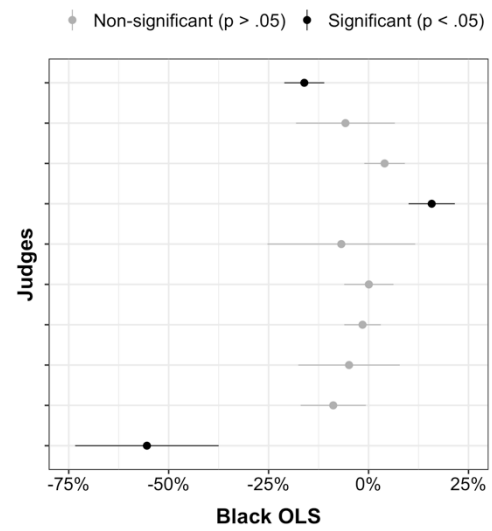
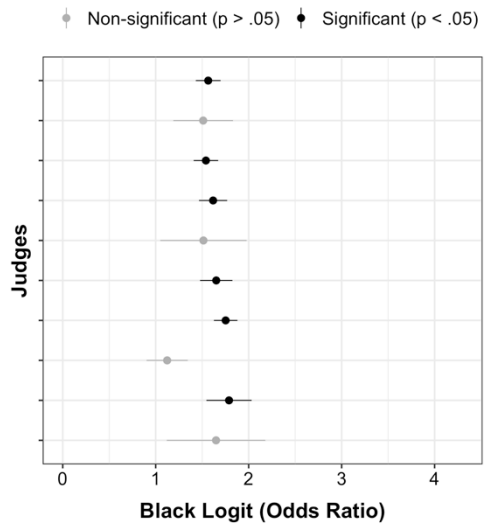
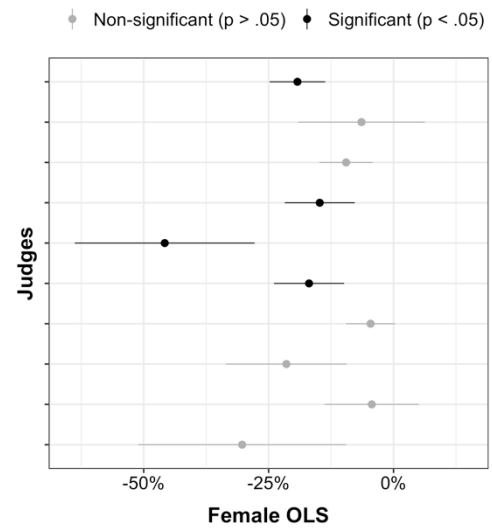
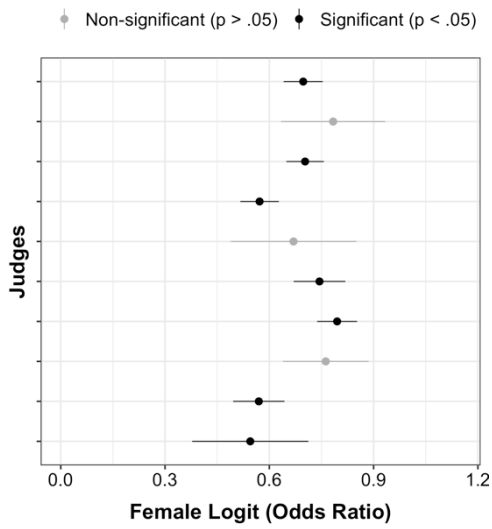


Figure 19 (cont'd).



whites, and 30 percent consider race in the sentence length decision. However, findings from the sentence length models show judges in this court sentence black offenders in different ways, with two judges imposing shorter sentences and one sentencing black offenders to longer periods of incarceration. Finally, 30 percent and 40 percent of judges increase the odds of incarceration and sentence length (respectively) for offenders convicted after trial, while the remaining judges show non-significant effects.<sup>15</sup>

Table 7 provides information for judges in all of the medium courts. Overall, judges in these courts exhibit similar patterns to those found in the large courts. The vast majority of judges in these courts show significant effects for offense severity and prior criminal history. In contrast, significant findings associated with age, gender, race, and mode of conviction vary substantially across these judges, and extralegal factors are generally more influential for the incarceration decision as opposed to length of the sentence imposed.

The last set of results provides information on judges in small courts. Court actors in these court communities work closely with one another, which is likely to limit individual autonomy. Findings for three of the 11 small courts conformed to these expectations, showing consistent sentencing patterns among judges. As expected, judges in these small courts show significant effects for legal factors across both outcomes. In Small Court 5 (Figure 20), both judges are associated with a decrease in the odds of incarceration as offender age increases, and are more likely to incarcerate blacks compared to whites. With the exception of mode of conviction, where one judge imposes longer sentences for offenders convicted after trial, no other extralegal predictors are significant for either judge. Similarly, in Small Court 9 (Figure 21), the only difference between the judges is that one judge is less likely to incarcerate female

---

<sup>15</sup> An additional judge exhibits significant effects for trial convictions in the incarceration decision, but was excluded from the 30 percent due to an extremely high standard error.

Table 7. Percent of Medium Court Judges with Significant Effects

<u>Court</u>	<u>Offense Severity</u>	<u>Prior Record</u>	<u>Age</u>	<u>Female</u>	<u>Black</u>	<u>Trial</u>
<i>Medium Court 1 (N=7)</i>						
Incarceration	100%	100%	43%	86%	71%	29%
Sentence Length	100%	100%	29%	14%	0%	0%
<i>Medium Court 2 (N=10)</i>						
Incarceration	100%	100%	10%	50%	30%	20%
Sentence Length	100%	100%	10%	10%	30%	0%
<i>Medium Court 3 (N=2)</i>						
Incarceration	100%	100%	100%	100%	100%	50%
Sentence Length	100%	100%	0%	100%	0%	0%
<i>Medium Court 4 (N=6)</i>						
Incarceration	100%	100%	83%	67%	67%	17%
Sentence Length	100%	100%	33%	67%	0%	50%
<i>Medium Court 5 (N=9)</i>						
Incarceration	100%	89%	33%	67%	89%	0%
Sentence Length	100%	100%	44%	22%	0%	0%
<i>Medium Court 6 (N=10)</i>						
Incarceration	100%	100%	80%	70%	60%	30%
Sentence Length	100%	100%	50%	40%	30%	40%

Figure 20. Individual Judge Effects in Small Court 5

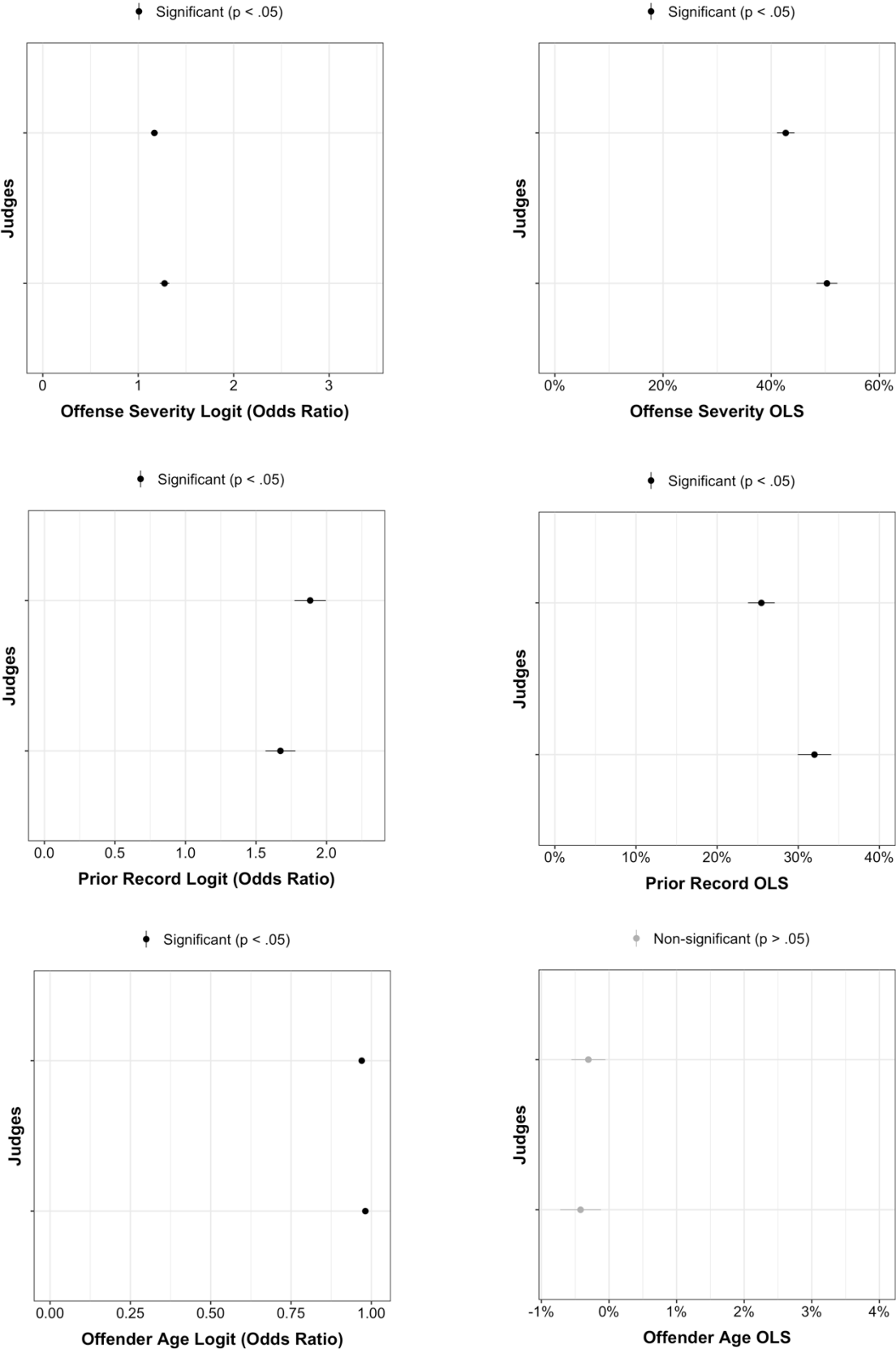




Figure 20 (cont'd).

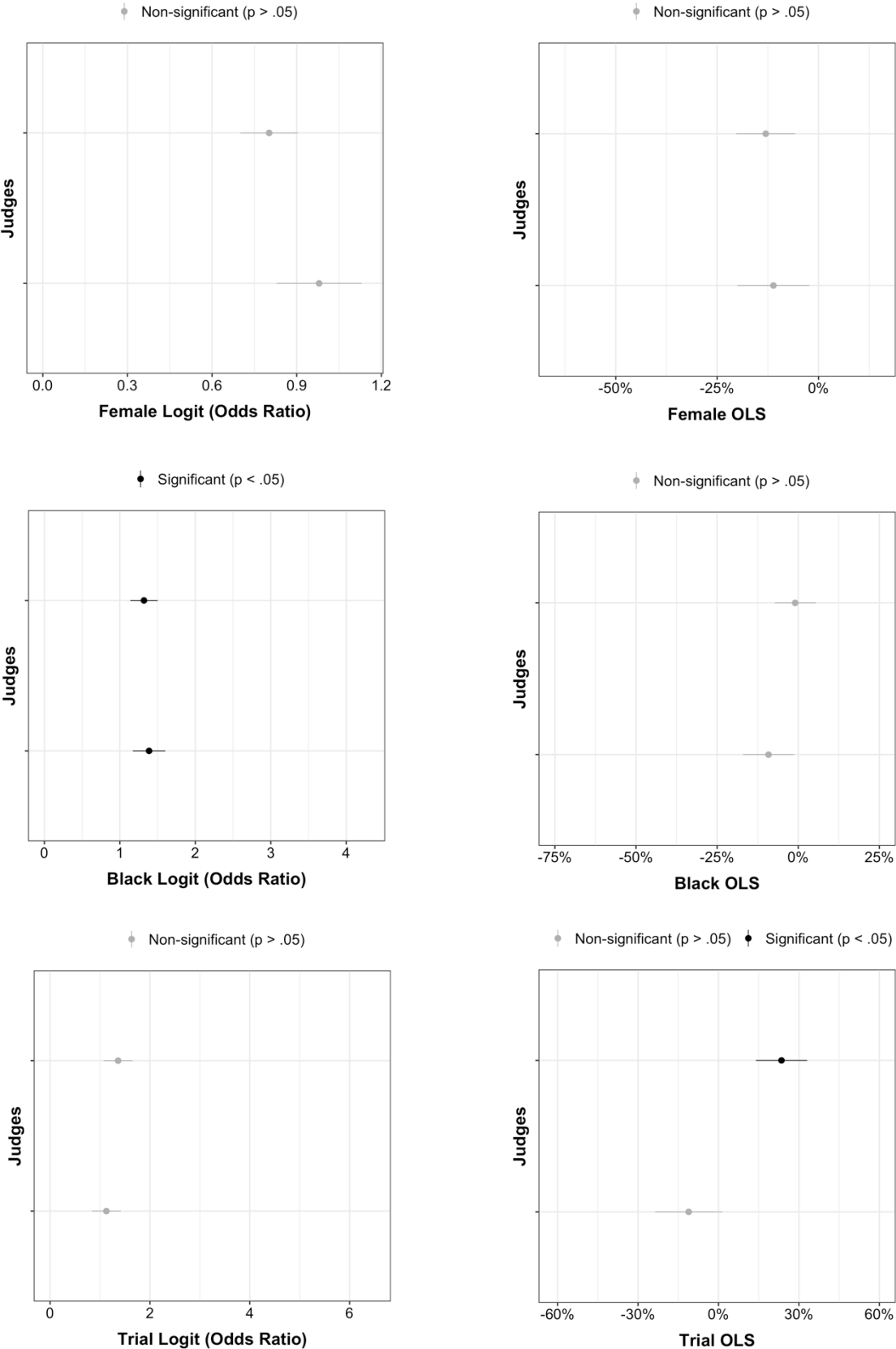


Figure 21. Individual Judge Effects in Small Court 9

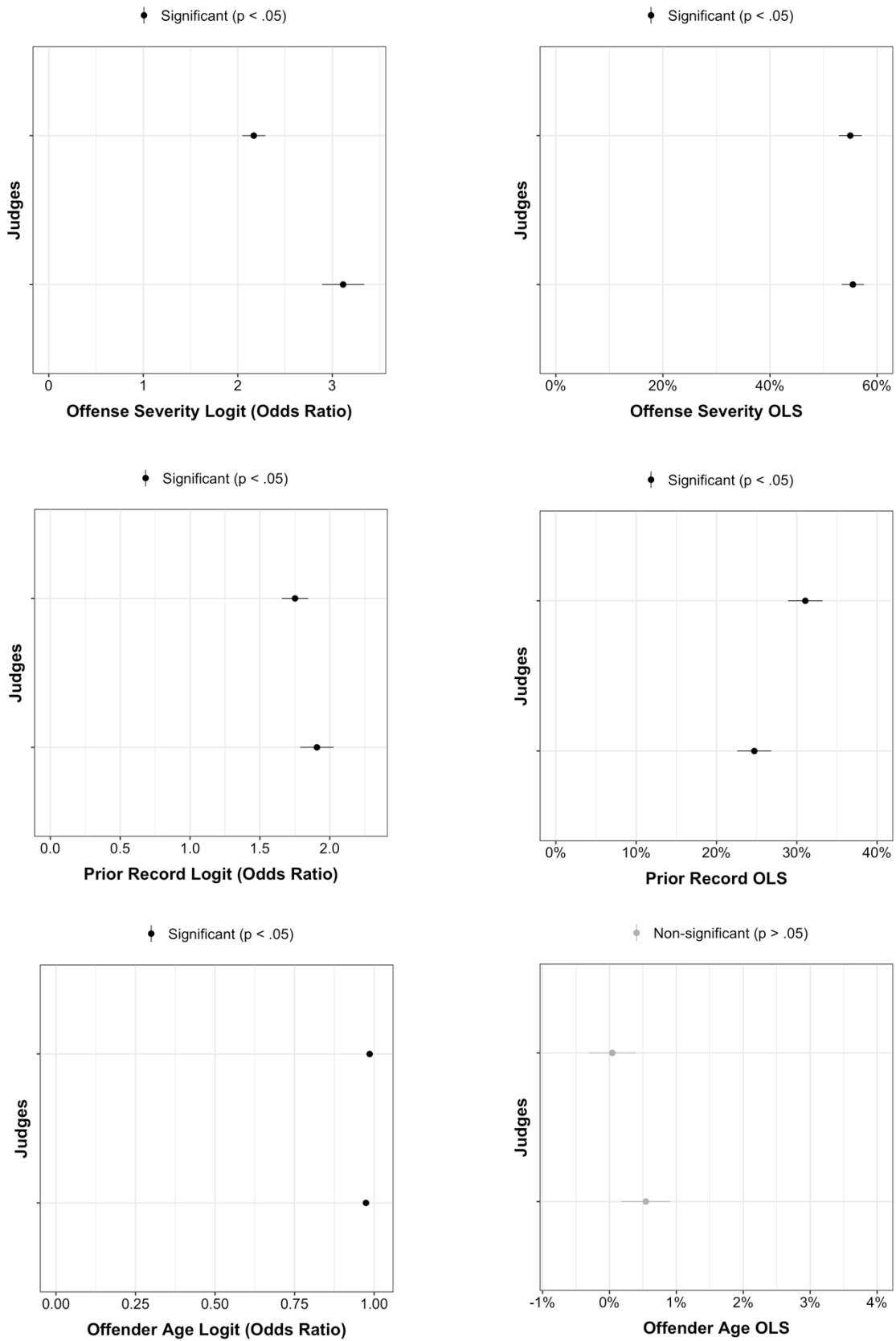
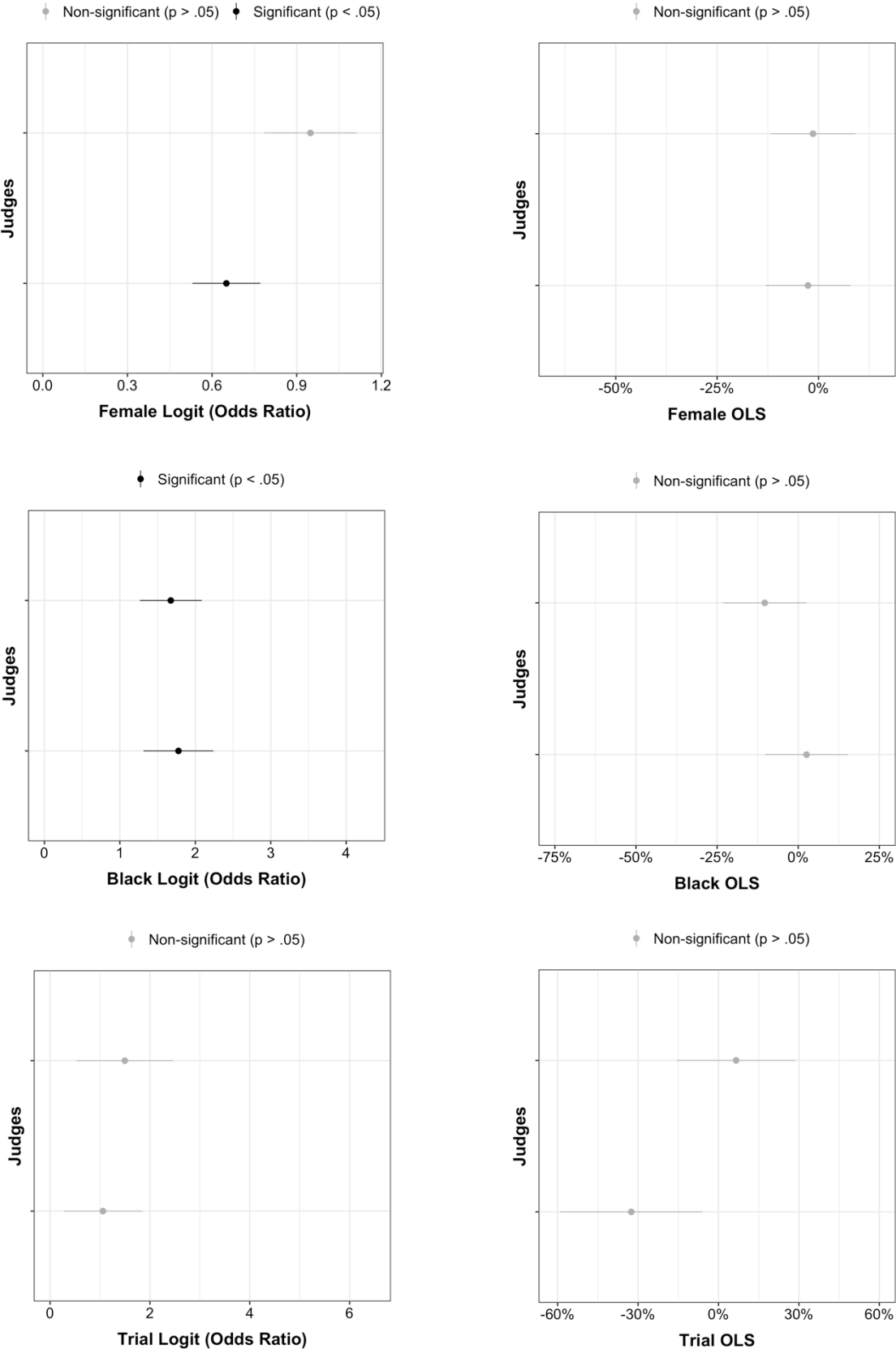


Figure 21 (cont'd).



offenders, while the other is not. Effects associated with the legal factors for both outcomes and age and race in the incarceration models are significant and in the same direction, and the remaining extralegal factors are non-significant. Judges in Small Court 4 (Figure 22) only differ in terms of offender race and mode of conviction in the incarceration models, though these findings should be interpreted with caution given the large standard errors. However, for the sentence length models, only legal factors influence these judges' decisions.

Finally, Table 8 includes findings for judges in all of the small courts. In contrast to judges in large and medium courts (with the exception of Medium Court 3), there are a few pockets of consistency among judges in small courts. For example, in Small Court 3, judges are consistent in terms of significant effects for offense severity, prior record, age, and mode of conviction in the incarceration models, and for all predictors in the sentence length models. In addition, though judges in Small Courts 10 and 11 exhibit differences for the incarceration decision, similar sentencing patterns are found for the sentence length models. Judges differ in race effects only in Small Court 10, and mode of conviction in Small Court 11.

Figure 22. Individual Judge Effects in Small Court 4

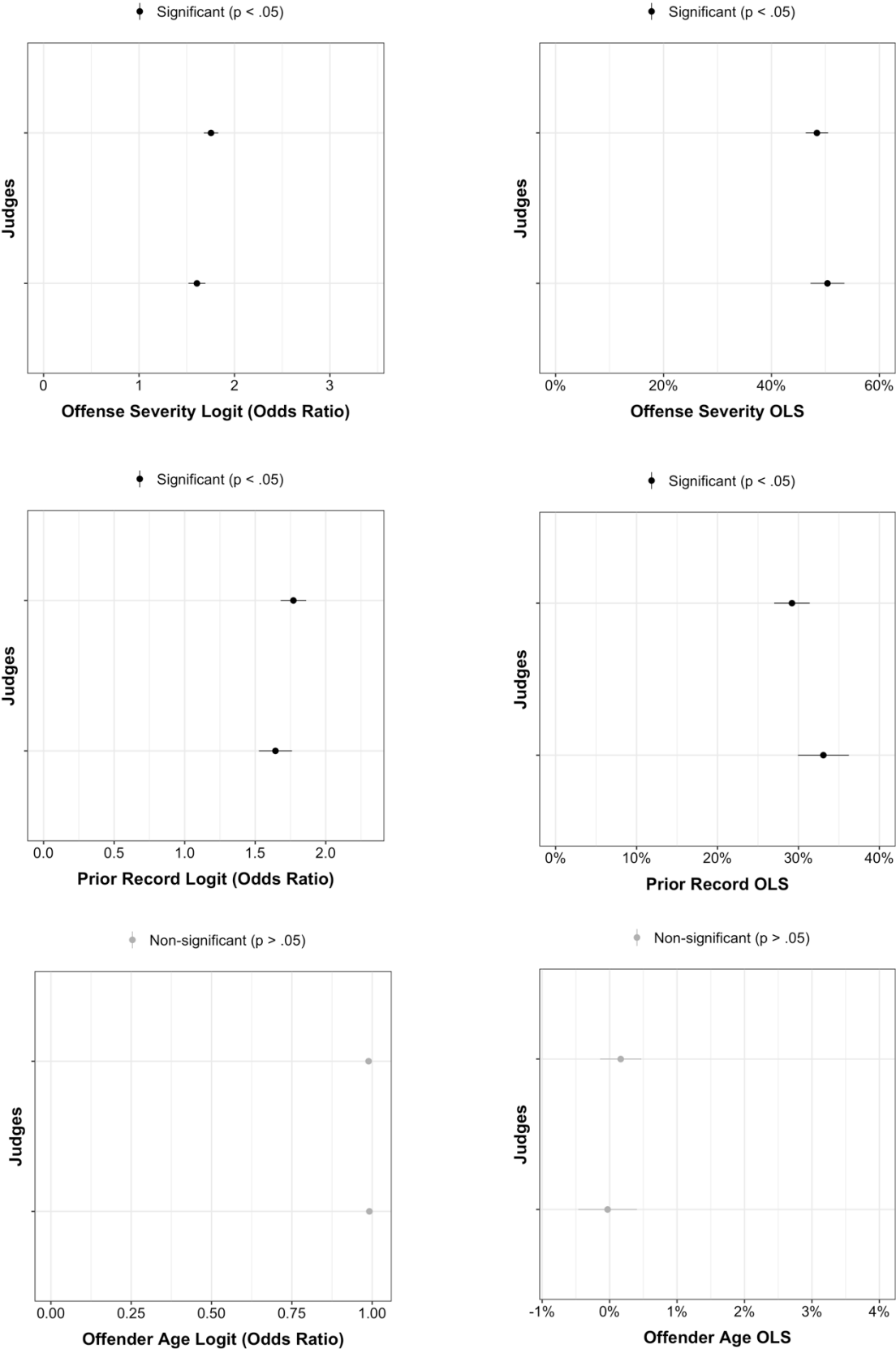


Figure 22 (cont'd).

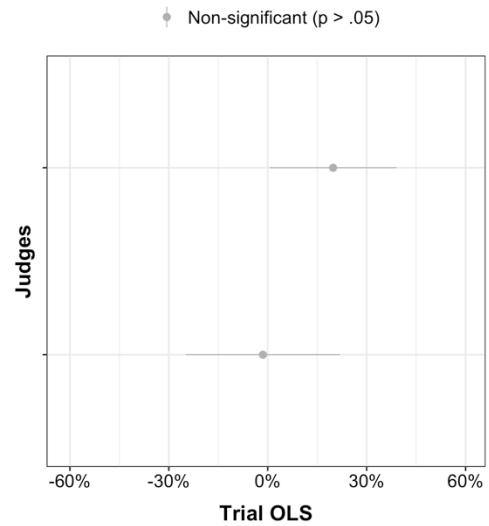
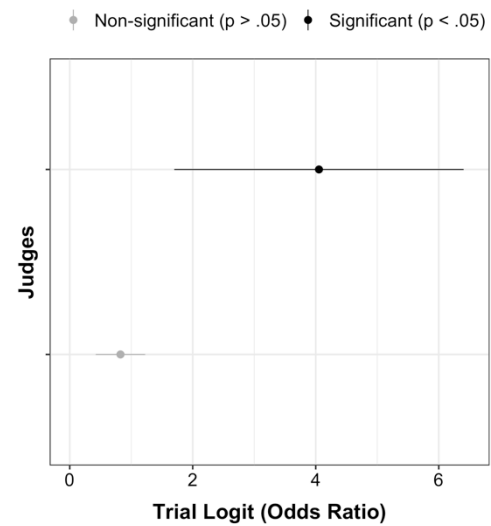
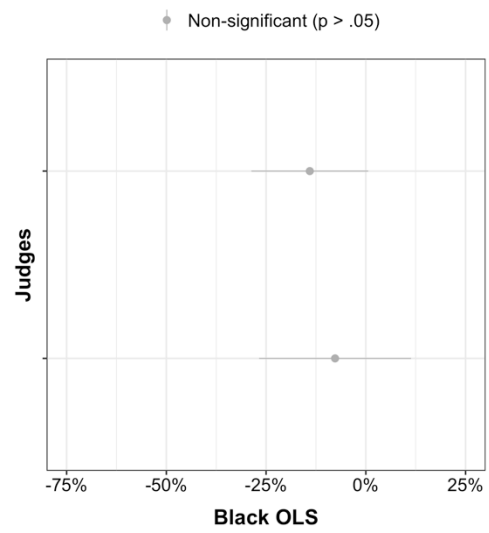
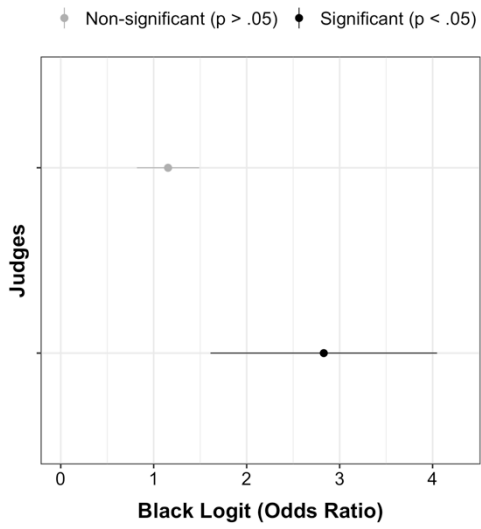
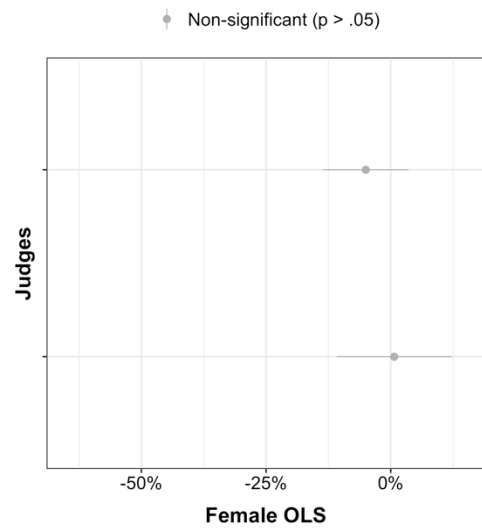
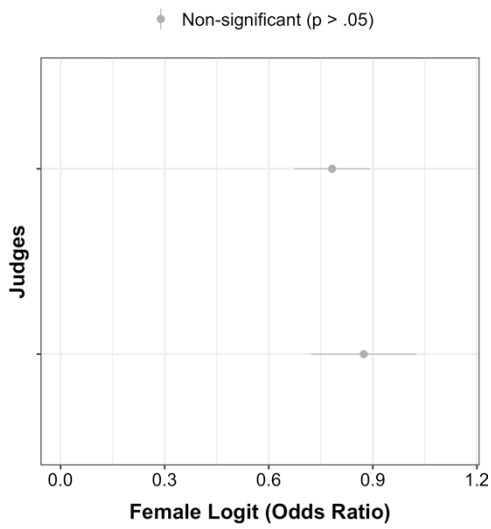


Table 8. Percent of Small Court Judges with Significant Effects

<u>Court</u>	<u>Offense Severity</u>	<u>Prior Record</u>	<u>Age</u>	<u>Female</u>	<u>Black</u>	<u>Trial</u>
<i>Small Court 1 (N=5)</i>						
Incarceration	80%	100%	80%	80%	60%	20%
Sentence Length	100%	100%	40%	60%	0%	40%
<i>Small Court 2 (N=4)</i>						
Incarceration	100%	100%	50%	75%	50%	75%
Sentence Length	100%	100%	25%	25%	0%	50%
<i>Small Court 3 (N=2)</i>						
Incarceration	100%	100%	100%	50%	50%	0%
Sentence Length	100%	100%	0%	0%	0%	0%
<i>Small Court 4 (N=2)</i>						
Incarceration	100%	100%	0%	0%	50%	50%
Sentence Length	100%	100%	0%	0%	0%	0%
<i>Small Court 5 (N=2)</i>						
Incarceration	100%	100%	100%	0%	100%	0%
Sentence Length	100%	100%	0%	0%	0%	50%
<i>Small Court 6 (N=6)</i>						
Incarceration	100%	100%	50%	83%	67%	0%
Sentence Length	100%	100%	33%	0%	0%	50%

Table 8 (cont'd).

<u>Court</u>	<u>Offense Severity</u>	<u>Prior Record</u>	<u>Age</u>	<u>Female</u>	<u>Black</u>	<u>Trial</u>
<i>Small Court 7 (N=5)</i>						
Incarceration	100%	100%	80%	20%	40%	0%
Sentence Length	100%	100%	20%	40%	0%	20%
<i>Small Court 8 (N=5)</i>						
Incarceration	100%	100%	100%	100%	80%	100%
Sentence Length	100%	100%	20%	0%	20%	80%
<i>Small Court 9 (N=2)</i>						
Incarceration	100%	100%	100%	50%	100%	0%
Sentence Length	100%	100%	0%	0%	0%	0%
<i>Small Court 10 (N=4)</i>						
Incarceration	100%	100%	50%	75%	100%	0%
Sentence Length	100%	100%	0%	0%	25%	0%
<i>Small Court 11 (N=4)</i>						
Incarceration	100%	100%	75%	75%	50%	0%
Sentence Length	100%	100%	100%	0%	0%	25%



## CHAPTER 5: DISCUSSION

This chapter begins by briefly reviewing the purpose of the current inquiry, including the research questions and hypotheses examined in the analyses. Next, an overview of the findings is presented and theoretical and methodological implications are discussed. The final portion of this chapter provides policy implications, limitations, and directions for future research.

### *The Current Inquiry*

This study explored three research questions to advance knowledge of interjudge disparity and judicial sentencing patterns within court communities. The first research question involved multilevel analysis of all judges in the sample to examine legal and extralegal effects on the decision to incarcerate offenders and the length of the sentence imposed, and whether these effects varied significantly across judges. Drawing on extant research (e.g., Johnson, 2006; Kramer & Ulmer, 2009), increases in offense severity and prior record were expected to increase sentence severity. Younger offenders, males, black offenders, and offenders convicted after trial were expected to be punished more harshly than older offenders, females, whites, and those who entered a guilty plea, respectively.

The second research question employed individual judge logistic and OLS regression models to assess judges' contributions to legal and extralegal disparities found in the multilevel analysis. Given the sentencing guidelines' use of offense severity and prior criminal history in determining the appropriate punishment, these legal factors were expected to significantly affect most individual judges' sentencing decisions. However, judicial consideration of extralegal factors is influenced, at least in part, by individual judges' subjective decision-making. Thus,

significant effects associated with extralegal predictors were expected to vary across judges more so than effects associated with legal factors.

The final research question examined the sentencing patterns of judges in the same court communities. Prior work suggests that court communities develop distinctive case processing and sentencing norms that may influence punishment outcomes (Eisenstein, Flemming, & Nardulli, 1988; Kramer & Ulmer, 2009; Ulmer, 1997), but differences in court actor autonomy in small, medium, and large courts may condition this relationship. Consequently, similar sentencing patterns may be most prevalent in small courts where court actor autonomy is restricted; on the other hand, high levels of autonomy associated with large courts may result in more diverse sentencing patterns among large court judges.

### *Theoretical and Methodological Implications*

*Multilevel Analysis of Judge Variation.* Extant research on interjudge disparity using multilevel analysis consistently shows both legal and extralegal factors influence sentence severity (Anderson & Spohn, 2010; Johnson, 2006; Wooldredge, 2010). Using this approach, the current work finds a one-unit increase in offense severity and prior record increase the likelihood of incarceration and the length of the sentence imposed, and female offenders are treated more leniently than male offenders. Prior research also indicates that younger offenders, black offenders, and those convicted after trial are more likely to be sentenced to jail or prison and receive longer sentences than older offenders, whites, and offenders pleading guilty, respectively (e.g., Kramer & Ulmer, 2009; Ulmer & Johnson, 2004). The multilevel analysis offers support for judges imposing a trial penalty, but findings associated with offender age and race produced mixed results. More specifically, while younger offenders and black offenders are more likely to be incarcerated, age has almost no effect on the length of the sentence imposed, and black

offenders receive shorter sentences than whites. Differences in these results may be attributable to using data from different time periods and/or over a longer period of time, or may be unique to the sample of judges selected for this analysis. Ultimately, these findings offer support for hypotheses one, three, and five, and partial support for hypotheses two and four.

Research employing multilevel analysis also shows that effects associated with legal and extralegal factors differ across judges (Anderson & Spohn, 2010; Johnson, 2006; Wooldredge, 2010). In line with hypothesis six, the current research indicates that effects associated with offense severity, prior record, age, gender, race, and mode of conviction vary significantly across judges. Overall, the results from this portion of the analysis provide support for the notion that judges rely on legal and extralegal factors when assessing the focal concerns, and effects associated with these offender and case characteristics vary across judges.

*Individual Analysis of Judge Variation.* To explore this variation across judges in more detail, the second research question employed individual judge logistic and OLS regression models to assess judges' contributions to legal and extralegal disparities found in the overall models. As expected, nearly all judges mete out harsher punishments for offenders committing more serious crimes and those with lengthier criminal records, whereas significant extralegal effects were less consistent. Further, consistent with the limited prior work on individual judges (Wooldredge, 2010), comparing findings from the overall models with the judge logit and OLS models indicates variation at the judge level is masked when using multilevel analysis.

Specifically, very few individual judge effects from the logit and OLS models were in line with the fixed effects estimates for the legal and extralegal factors found in the multilevel models. In addition, statistically significant extralegal effects vary widely across individual judges, despite age, race, gender, and mode of conviction being significant in the multilevel fixed

effects. This should be expected to some degree, given the overall estimates are just the predicted values across all judges (Hox, 2010). Still, the overall estimates are generally the focus of much of the existing sentencing literature (Anderson & Spohn, 2010; Johnson, 2006; Wooldredge, 2010), and have played a substantial role in developing theories that explain the differential treatment of similarly situated offenders (i.e., those convicted of the same crimes, with similar criminal histories). Results from the random effects portion of the multilevel analyses provide more information about individual judge effects for a given variable, and offer three primary findings with methodological and theoretical implications.

First, as discussed previously, the individual judge estimates in the multilevel analysis are weighted averages that take into account group information and the overall model mean, and less reliable estimates are shrunk closer to the model average. Consequently, values from multilevel models are biased, but also more precise. Yet, this is not always the case. Though the majority of the findings from the current work show shrinkage towards the mean, results from the multilevel and individual judge logit incarceration models for offender race revealed a number of judges' odds ratios in the multilevel model increased (when compared to the logit models) and were pushed farther away, as opposed to closer to, the overall model mean. Burnham (2017) notes that while shrinkage estimators are optimal for the overall set of model parameters (i.e., they minimize the mean squared error for all groups), they may not be for all individual parameters. Consequently, individual estimates may be "incorrectly shrunk" or move in the wrong direction (Burnham, 2017: 20; see also Lipsky et al., 2011), resulting in misleading estimates about how judges consider offender and case characteristics in sentencing decisions.

The second finding indicates that even when shrinkage estimators are operating as expected, the random effects estimates are of limited value when the purpose of the research is to

assess individual judge decision-making. For example, the multilevel and logit incarceration models (and to a lesser extent the sentence length models) for gender show the differences in values for female offenders in the logit models are reduced to a substantially tighter grouping around the overall mean in the multilevel model. As such, the random effects allow for drawing general conclusions about leniency for female offenders, but understate the extent of the variation across judges. More noticeable, however, are the shrinkage effects for the mode of conviction variable in the incarceration models. Strong pooling should be expected, since shrinkage estimators are designed to obtain the most precise estimates, and the individual models show a number of judges with large effects and high standard errors. Yet, comparing the results from the different methodological approaches raises questions about how mode of conviction should be used in sentencing research, and what can be interpreted from findings associated with offenders convicted after trial. Including mode of conviction as a predictor of punishment severity is ubiquitous in sentencing research, and findings consistently show that trial convictions result in harsher punishment (e.g., Dixon, 1995; Johnson, 2003, 2006; for reviews, see Ulmer, 2012; Ulmer & Bradley, 2006). However, to the extent these results are driven by unreliable estimates and/or a small number of judges who impose extremely high penalties, as the individual models suggest, drawing general conclusions from multilevel models about how judges sentence offenders convicted after trial should be reconsidered.

The third finding highlights an additional limitation when relying exclusively on multilevel analyses to examine interjudge disparity. Multilevel models are designed to obtain the most precise estimate for each group, but may not be appropriate when predictors are hypothesized to produce effects for some groups, but not others (Gelman & Hill, 2016). This is particularly important for sentencing research on individual judges because sentencing theories

are not only concerned with whether judges vary in effects associated with offender and case characteristics, but also whether these factors matter in punishment decisions. Specifically, focal concerns suggests that the influence of legal and extralegal factors on punishment decisions is likely to differ based on judges' subjective assessments of blameworthiness and community threat. Findings from extant studies have generated broad conclusions about sentencing predictors, such as females receiving more lenient sentences than males, and blacks being punished more harshly than whites. Yet, the current work finds substantial differences in significant effects associated with extralegal factors in the individual judge analyses, indicating that more variation exists among certain sentencing predictors than previously understood. These include effects for age (80 of 161 judges  $p < .05$ ), gender (112 of 161), race (90 of 161), and mode of conviction (45 of 161) for the incarceration decision, and even fewer significant effects for the sentence length decision (with the exception of mode of conviction). Ultimately, these findings suggest that extralegal factors matter in sentencing outcomes, but the ways in which they influence punishment severity is conditioned by the individualized nature of the judicial decision-making process. As such, analysis of individual judges provides new insights about the key predictors of sentencing over what is traditionally found using multilevel analysis, and offer important implications for theory.

Similar to findings from the overall model, the results from the individual judge models offer support for the focal concerns perspective. Judges rely primarily on legal factors when assessing offender blameworthiness and community threat, but also engage in subjective decision-making based on attributions associated with extralegal characteristics when considering the focal concerns. Yet, given that few individual judges are in line with the overall multilevel model estimates, the pooling of estimates in the random effects, and that individual

judges' significant effects vary in ways that are not represented in the overall model or the random group effects, the results suggest that focal concerns may be more appropriately tested using separate judge models. In particular, individual analyses would provide a better understanding of whether and how judges consider the key predictors of sentencing when assessing the focal concerns (see also Wooldredge, 2010).

However, findings from the analytic strategies used in the current work highlight additional problems with the focal concerns perspective. Additional analyses (not shown) indicate that some judges are associated with significant effects for legal factors only. Others are significant for legal factors and some extralegal factors, and still other judges exhibit significant effects for all legal and extralegal variables. Moreover, the influence of extralegal factors is conditioned by the sentencing outcome (e.g., incarceration, sentence length). In some sense, all of these judges sentencing decisions can be explained by the focal concerns perspective. This is because focal concerns recognizes that judges will vary in their assessments of the focal concerns, and the factors used to assess them. For some judges, protection of the community may be assessed based on prior criminal history, while others might consider the nature of the offense (e.g., violent versus property crime). Others may consider one or both of these legal factors in addition to offender characteristics such as race and gender if they view certain offenders as posing a greater threat to the community.

In another sense, the variation found in the individual judge models concerning significant effects suggests this perspective may be too parsimonious because it cannot *explain* these patterns in sentencing decisions. Focal concerns outlines broad concepts associated with punishment decisions; blameworthiness, protection of the community, and practical constraints. The perspective also notes legal and extralegal factors that may be associated with assessing the

focal concerns, but acknowledges that judges are likely to vary in how they consider legal and extralegal factors in relation to the focal concerns. It does not, however, provide a clear indication of the specific factors that influence each of the focal concerns, or why certain factors may be relevant for the incarceration decision but not the sentence length decision (and vice versa), which limits testing this perspective using existing data and quantitative methods (see Hartley, Maddan, & Spohn, 2007). Moreover, the perspective relies heavily on judges' perceptions and subjective decision-making, but provides little information about how judges develop their sentencing philosophies. Additional theoretical perspectives that may address this gap are discussed in the current work's section on directions for future research.

*Sentencing Within Court Communities.* The third research question examined judges grouped by court to assess whether and how judges in the same court communities consider legal and extralegal factors in the decision to incarcerate offenders. The current study hypothesized that less autonomy among small court judges would result in similar sentencing patterns among these judges. Differences in sentencing patterns were expected to increase in medium courts, and the largest variation was predicted in large courts where autonomy is highest.

Overall, findings concerning the relationship between court size and sentencing patterns provide limited support for the current work's hypotheses. As expected, judges in large courts exhibit substantial variation in terms of significant extralegal effects. These findings are in line with limited prior work that suggests while certain aspects of case processing are tightly coupled in large courts (e.g., docket management, courtroom assignment), judges exercise discretion in punishment decisions in ways they feel promotes justice (Jacob, 1997).

More similarities in judges' sentencing patterns were expected in medium-sized courts, but the findings show a mix of patterns. While individual judge legal and extralegal effects are



near identical in one medium court, judges in the remaining five medium courts look similar to those found in the large courts; that is, judges vary substantially in effects associated with offender age, gender, race, and mode of conviction. Notably, the medium court where judges exhibit very similar sentencing patterns has a total of 11 authorized judgeships, but only two judges handle the majority of the criminal caseload (roughly 90 percent of cases). Thus, though the court community is categorized as medium, it may operate in ways that are more reflective of small court communities.

With prior research suggesting that small courts are composed of very few court actors who work closely with one another (Eisenstein, Flemming, & Nardulli, 1988; Ulmer, 1997), the current work expected judges in small courts to exhibit similar sentencing patterns than found in large and medium courts. Findings offer some support for this expectation, with judges in three of the 11 small courts exhibiting relatively similar sentencing patterns in terms of statistically significance effects associated with offender and case characteristics. Three additional courts show pockets of consistency across judges for most extralegal effects, offense severity and/or prior record, but much of this is limited to the sentence length decision.

Though the results concerning the conditioning effect of court size on individual judges' sentencing patterns garnered limited support, the findings overall are consistent with the focal concerns and court community perspective. Similar to focal concerns theory alone, this is because current theorizing about the court community influence on sentencing has not been well defined. Recall that integrating these perspectives suggests that differences in judicial consideration of legal and extralegal factors in sentencing decisions can be explained in part by the distinctive case processing and sentencing norms present in the court community in which punishment decisions occur. However, scholars also note that the court community influence is

likely dependent on the presence of, and adherence to, shared norms within these communities (Ulmer, 2012). As such, similarities between individual judges' sentencing decisions in the same court communities can be interpreted as consistent with the focal concerns and court community perspectives, but differences in judges' sentencing patterns can as well.

Overall, the findings from this portion of the analysis suggest court size is too broad of a measure to explain the relationship between court communities and sentencing decisions. Though extant qualitative work provides clear evidence of court actors developing working relationships and case processing strategies, it stops short of explaining how these court community elements affect sentencing. Thus, the current research highlights a need for additional theoretical development to explain the ways in which court communities influence sentencing decisions.

### *Implications for Policy*

The current work has implications for sentencing law and policy. The sentencing of criminal offenders is a fundamental mechanism of formal social control in society, and disparity in punishment raises questions about the legitimacy of legal institutions (Reitz, 1998; Tonry, 1996). Perceived illegitimacy in the application of criminal sanctions may have a significant impact on crime rates, the deterrent capacity of the criminal justice system, race relations, and the generation and reproduction of social inequalities (Anderson, 1999; Gottschalk, 2008; Klinger, 1994; LaFree, 1998; Russell, 1998; Ruth & Reitz, 2003; Tyler, 1990; Western, 2006).

Sentencing guidelines were developed to increase uniformity in punishment and reduce unwarranted disparity (Kramer & Scirica, 1986). Since the present study does not compare sentencing decisions pre- and post-guidelines, it is unclear whether the guidelines have achieved their intended goals. However, findings suggest that extralegal disparity associated with age,

gender, race, and mode of conviction continues to exist under structured sentencing systems, and the ways in which these factors influence sentencing varies across judges. As a result, the probability of incarceration and the sentence length imposed for two similar offenders may be significantly different depending on the judge who sentences them.

Despite the existence of legal and extralegal disparities, sentencing guidelines offer a compromise between eliminating judicial discretion entirely and sentencing bounded by only wide ranging statutory minimums and maximums. The former would include policies such as mandatory minimum sentencing provisions, which have been widely criticized as unduly harsh (Tonry, 1996). Further, some research suggests mandatory minimums simply shift sentencing discretion to other court actors, such as prosecutors (Tonry, 1996). Sentencing bounded by only statutory minimums and maximums would grant judges nearly unfettered discretion, which would almost guarantee disparate treatment of similar offenders.

Still, policy changes may be necessary to achieve more uniformity in criminal sanctions. These changes may include training for judges to ensure that sentencing is based primarily on the guidelines rather than extralegal criteria, as well as having judges provide some explanation of their reasons for imposing the selected sentence. In addition, since Pennsylvania's guidelines allow for more judicial discretion than any other state operating under a guidelines system (Kramer & Ulmer, 2009), the current work may signify a need for stricter appellate review standards. Finally, the Pennsylvania Commission on Sentencing's (PCS) Annual Reports are currently limited to descriptive analyses of offender and case characteristics. If the PCS is concerned with guideline compliance and accountability, more rigorous examinations of individual judges' sentencing decisions are necessary.

### *Limitations*

Though the current research offers significant theoretical and methodological contributions to the sentencing literature, like any research several limitations exist. As noted by others who have used the PCS data (Johnson, 2005, 2006; Kramer & Ulmer, 2009; Ulmer & Kramer, 2004), the data do not include information on charging decisions, bail outcomes, offender socioeconomic status, and victim characteristics, all of which may predict variation in punishment severity (e.g., Baumer, 2010). Further, some research shows that offender characteristics interact to produce greater disparity than found when only exploring direct effects alone (e.g., Doerner & Demuth, 2010; Spohn & Holleran, 2000; Steffensmeier, Ulmer, & Kramer, 1998). As such, it is possible that effects associated with extralegal factors from the individual judge models would be more prevalent if age, gender, and race, were examined in combination.

In addition, the analyses are limited to a sample of judges from large, medium, and small courts in Pennsylvania. Consequently, findings are only generalizable to these judges in these courts. It is possible that research on individual judges in other courts in Pennsylvania, as well as judges in other states, would produce different findings.

Similar to other studies that have applied the focal concerns and court community perspectives (Johnson, 2006; Kramer & Ulmer, 2009; Ulmer & Johnson, 2004), the present work lacks direct measures of judicial sentencing philosophies, as well as information about judges' perceptions associated with offender and case characteristics. In addition, the current work does not include measures of court community features, such as information about other courtroom actors (e.g., prosecutors, defense counsel), workgroup relationships, case processing strategies, and sentencing norms. Thus, interpretations of workgroup autonomy, judges' sentencing

patterns, and the role the court community plays in influencing punishment outcomes only serve as inferences.

However, the current work's use of multilevel modeling is generally consistent with the way in which other studies have examined interjudge variation and court communities (e.g., Anderson & Spohn, 2010; Johnson, 2006). More importantly, the present study is the first to apply these perspectives to analyze individual judges' sentencing patterns from a relatively large sample of courts differing in size. As such, it offers a substantial contribution in terms of identifying individual judge variation, and further advances knowledge of judges' sentencing patterns within court communities. Thus, despite these limitations, this work provides a number of avenues for future research.

#### *Directions for Future Research*

Given most theories of sentencing recognize individual differences in the ways judges consider legal and extralegal factors in sentencing decisions, multilevel models have become increasingly popular in sentencing research. However, results from the current research suggest that future studies should consider using separate judge models to gain a better understanding of variation across judges, to examine extreme cases and patterns in the data, and to assess whether and how judges consider extralegal factors, which are likely to influence some judges' decisions, but not others. This is not to say that multilevel analyses are inappropriate for sentencing research, but they are likely better suited for some research questions over others. Multilevel models may be beneficial when examining effects of sentencing predictors that theory and empirical research suggest are highly influential for all judges, such as offense severity and prior criminal history. These factors are consistently associated with significantly affecting punishment severity, and the shrinkage estimators used in multilevel analysis may provide

precise estimates for judges with varying sample sizes. Multilevel analysis may also be useful for drawing general conclusions about variation for other sentencing predictors, particularly when some judges sentence a small number of offenders. Yet, for testing theories that predict effects associated with offender and case characteristics are likely to vary based on judges' subjective decision-making, individual judge models offer some advantages. Though using separate regression models is dependent on having access to judge information and sample size, this kind of research would complement the larger body of sentencing literature that has focused mostly on multilevel analyses.

To move beyond describing whether and how judges consider offender and case characteristics, future work would benefit from additional theoretical integration and development to better understand *why* certain legal and extralegal factors influence sentencing outcomes. For example, future research might take an organizational view of case processing, which suggests that individual judges' sentencing patterns are influenced by the cases they handle over time. According to Emerson (1983: 425), "the individual case provides an adequate unit of analysis only if social control agents themselves examine and dispose of cases as discrete units, treating each on its own merits independently ... of other cases." What is more likely, he argued, is that individual cases are not treated independently, but rather viewed in connection with the agents' overall flow of cases (Emerson, 1983). In the context of sentencing, judges who handle a larger number of violent cases over time may become desensitized; as a result, these judges may sentence violent offenders less harshly than judges who encounter violent cases less often (Johnson, 2006). Additional work indicates that attributions associated with extralegal factors may also be influenced by the overall flow of cases (Maynard-Moody & Musheno, 2003). Thus, examining individual judges' sentencing decisions in relation to their caseload may further

understanding of why judges consider legal and extralegal factors in different ways. Future research in this area may include trajectory analysis to examine judges' caseloads and changes in sentencing patterns over time, and qualitative work to gain an in-depth understanding of how judges' overall flow of cases affects sentencing decisions.

In addition, more qualitative research is needed to better understand the relationship between court communities and sentencing decisions. Research with judges should explore whether judges are aware of their colleagues' sentencing decisions, and the extent to which those decisions influence their own. Additional work is also needed to assess the prosecutor's role in punishment decisions, and particular attention should be devoted to the courtroom workgroup's approach to prosecutor recommended sentences as part of plea agreements. Limited work suggests this varies across courts (Eisenstein, Flemming, & Nardulli, 1988; Ulmer, 1997), and it is an important component to understanding the root causes of differences in sentencing outcomes.

More generally, the findings from the current work highlight the need for more research at the individual court actor level (see also Ulmer, 2012). Current theories of sentencing view punishment decision-making as an individualized process, where judges and potentially other court actors assess offender blameworthiness and community threat in their own ways. The factors that influence these decisions, as well as the weight afforded to these factors, is likely to vary across decision-makers. In addition, contextual theories concerning court influences recognize that court communities are unique, and they develop their own distinctive case processing strategies and sentencing norms. Yet, much of the extant research has taken these theories and tested them at the aggregate level, using large datasets, and pooling cases across all judges in a jurisdiction or a state. Though research at the individual judge level may limit

generalizability, it has the potential to further refine current sentencing perspectives and advance theoretical development.



## REFERENCES

## REFERENCES

- Albonetti, C. A. (1991). An integration of theories to explain judicial discretion. *Social Problems*, 38, 247-66.
- Albonetti, C. A. (1997). Sentencing under the federal sentencing guidelines: Effects of defendant characteristics, guilty pleas, and departures on sentence outcomes for drug offenses, 1991-1992. *Law and Society Review*, 31, 789-822.
- Albonetti, C. A. (2002). The joint conditioning effects of defendant's gender and ethnicity on length of imprisonment under federal sentencing guidelines for drug trafficking-manufacturing offenders. *Journal of Gender, Race, and Justice*, 6, 39-60.
- Allison, P. D. (2001). *Missing data*. Thousand Oaks, CA: Sage.
- Anderson, D. A. (1999). The aggregate burden of crime. *Journal of Law and Economics*, 42(2), 611- 642.
- Anderson, A. L., & Spohn, C. (2010). Lawlessness in the federal sentencing process: A test for uniformity and consistency in sentencing outcomes. *Justice Quarterly*, 27(3), 362-393.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R., Singmann, H., ... Green, P. (2016). *lme4: Linear Mixed-Effects Models using 'Eigen' and S4*. R package version 1.1-12. <https://cran.r-project.org/web/packages/lme4/index.html>
- Baumer, E. P. (2013). Reassessing and redirecting research on race and sentencing. *Justice Quarterly*, 30(2), 231-261.
- Britt, C. L. (2000). Social context and racial disparities in punishment decisions. *Justice Quarterly*, 17(4), 707-732.
- Britt, C. L. (2009). Modeling the distribution of sentence length decisions under a guidelines system: An application of quantile regression. *Journal of Quantitative Criminology*, 25(4), 341-370.
- Burnham, K. P. (2017). Appendix D: Variance components and random effects models in MARK. In E. G. Cooch & G. C. White (Eds.), *Program MARK: A gentle introduction* (D1-D45). Retrieved from [http://www.phidot.org/software/mark/docs/book/pdf/app\\_4.pdf](http://www.phidot.org/software/mark/docs/book/pdf/app_4.pdf)
- Casella, G., & Berger, R. L. (1990). *Statistical inference*. Pacific Grove, CA: Wadsworth.
- Chambliss, W. J., & Seidman, R. B. (1982). *Law, order, and power*. Reading, Massachusetts: Addison-Wesley.

- Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Wunder, J., & De Rosario, H. (2015). *pwr: Basic functions for power analysis*. R package version 1.1-3. <https://cran.r-project.org/web/packages/pwr/index.html>
- Chiricos, T. G., & Waldo, G. P. (1975). Socioeconomic status and criminal sentencing: An empirical assessment of a conflict proposition. *American Sociological Review*, 40, 753-772.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- de Leeuw, J., & Kreft, I. G. (1995). Questioning multilevel models. *Journal of Educational and Behavioral Statistics*, 20(2), 171-189.
- Dixon, J. (1995). The organizational context of criminal sentencing. *American Journal of Sociology*, 100, 1157-1198.
- Doerner, J. K., & Demuth, S. (2010). The independent and joint effects of race/ethnicity, gender, age on sentencing outcomes in U.S. federal courts. *Justice Quarterly*, 27(1), 1-27.
- Eisenstein, J., Flemming, R. B., & Nardulli, P. F. (1988). *The contours of justice: Communities and their courts*. Boston, MA: Little, Brown.
- Eisenstein, J., & Jacob, H. (1977). *Felony justice: An organizational analysis of criminal courts*. Boston, MA: Little, Brown.
- Emerson, R. M. (1983). Holistic effects in social control decision-making. *Law & Society Review*, 17(3), 425-455.
- Fitz-Gibbon, C. T. (1991). Multilevel modeling in an indicator system. In S. W. Raudenbush & J. D. Willms (Eds.), *Schools, classrooms, and pupils: International Studies of Schooling from a multilevel perspective*. San Diego, CA: Academic Press, Inc.
- Fitz-Gibbon, C. T. (1996). *Monitoring education: Indicators, quality and effectiveness*. London, UK: Continuum.
- Gelman, A., & Hill, J. (2016). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.
- Gottschalk, M. (2008). Hiding in plain sight: American politics and the carceral state. *Annual Review of Political Science*, 11, 235-260.
- Hagan, J. (1974). Extra-legal attributes and criminal sentencing: An assessment of a sociological viewpoint. *Law and Society Review*, 8, 357-383.

- Hagan, J. (1989). Why is there so little criminal justice theory? Neglected macro- and micro-level links between organization and power. *Journal of Research in Crime and Delinquency*, 26(2), 116-135.
- Hartley, R. D., Maddan, S., & Spohn, C. (2007). Concerning conceptualization and operationalization: Sentencing data and the focal concerns perspective—a research note. *The Southwest Journal of Criminal Justice*, 4(1), 58-78.
- Hauser, W., & Peck, J. H. (2017). The intersection of crime seriousness, discretion, and race: A test of the liberation hypothesis. *Justice Quarterly*, 34(1), 166-192.
- Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, 45, 1-47.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and Applications*. New York, NY: Routledge.
- Jacob, H. (1997). The governance of trial judges. *Law and Society Review*, 31(1), 3-30.
- Johnson, B. D. (2003). Racial and ethnic disparities in sentencing departures across modes of conviction. *Criminology*, 41(2), 449-489.
- Johnson, B. D. (2005). Contextual disparities in guidelines departures: courtroom social context, guidelines compliance, and extralegal disparities in criminal sentencing. *Criminology*, 43(3), 761-796.
- Johnson, B. D. (2006). The multilevel context of criminal sentencing: Integrating judge and county level influences in the study of courtroom decision making. *Criminology*, 44, 259-298.
- Johnson, B. D. (2014). Judges on trial: A reexamination of judicial race and gender effects across modes of conviction. *Criminal Justice Policy Review*, 25(2), 159-184.
- Johnson, B. D., Ulmer, J., & Kramer, J. (2008). The social context of guideline circumvention: The case of federal district courts. *Criminology*, 46, 711-783.
- Kautt, P. M. (2002). Location, location, location: Interdistrict and intercircuit variation in sentencing outcomes for federal drug-trafficking offenses. *Justice Quarterly*, 19(4), 633-671.
- Kim, B., Spohn, C., & Hedberg, E. C. (2015). Federal sentencing as a complex and collaborative process: Judges, prosecutors, judge-prosecutor dyads, and disparity in sentencing. *Criminology*, 53(4), 597-623.
- Kleck, G. (1981). Racial discrimination in criminal sentencing: A critical evaluation of the evidence with additional evidence on the death penalty. *American Sociological Review*, 46(6), 783-805.

- Kleck, G. (1985). Life support for ailing hypotheses: Modes of summarizing the evidence for racial discrimination in sentencing. *Law and Human Behavior*, 9, 271-85.
- Klinger, D. A. (1994). Demeanor or crime? Why "hostile" citizens are more likely to be arrested. *Criminology*, 32, 475-493.
- Kramer, J., & Scirica, A. (1986). Complex policy choices: The Pennsylvania commission on sentencing. *Federal Probation*, 50, 15-23.
- Kramer, J. H., & Ulmer, J. T. (2009). *Sentencing guidelines: Lessons from Pennsylvania*. Boulder, CO: Lynne Rienner.
- Kreft, I. G., & Yoon, B. (1994). Are multilevel techniques necessary? An attempt at demystification. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. Retrieved from <http://files.eric.ed.gov/fulltext/ED371033.pdf>
- LaFree, G. (1998). *Losing legitimacy: Street crime and the decline of social institutions in America*. Boulder: Westview Press.
- Levin, M. A. (1977). *Urban politics and the criminal courts*. Chicago, IL: University of Chicago Press.
- Lipsky, A. M., Gausche-Hill, M., Vienna, M., & Lewis, R. J. (2011). The importance of "shrinkage" in subgroup analyses. *Annals of Emergency Medicine*, 55(6), 544-552.
- Lizotte, A. (1978). Extra-legal factors in Chicago's criminal courts: Testing the conflict model of criminal justice. *Social Problems*, 25, 564-580.
- Maynard-Moody, S., & Musheno, M. (2003). *Cops, teachers, counselors: Stories from the front lines of public service*. Ann Arbor, MI: University of Michigan Press.
- MacKenzie, D. L. (2001). Corrections and sentencing in the 21st century: evidence-based corrections and sentencing. *Prison Journal*, 81, 3-17.
- Miethe, T. D., Moore, C. A. (1985). Socioeconomic disparities under determinate sentencing systems: A comparison of preguideline and postguideline practices in Minnesota. *Criminology*, 23, 337-363.
- Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for equality of regression coefficients. *Criminology*, 36(4), 859-866.
- Pennsylvania Commission on Sentencing (PCS). (n.d). *Sentencing Guidelines Manuals*. Retrieved from <http://pcs.la.psu.edu/guidelines/sentencing/sentencing-guidelines-and-implementation-manuals>

- Pratt, T. (1998). Race and sentencing: A meta-analysis of conflicting empirical research results. *Journal of Criminal Justice*, 26(6), 513-523.
- R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Reitz, K. (1998). *Sentencing*. In M. Tonry (Ed.), *The handbook of crime and punishment* (pp. 543-546). New York: Oxford University Press.
- Russell, K. (1998). *The color of crime: Racial hoaxes, white fear, black protectionism, police harassment, and other macro-aggressions*. New York: New York University Press.
- Ruth, H. S., & Reitz, K. (2003). *The challenge of crime: Rethinking our response*. Cambridge: Harvard University Press.
- Savelsberg, J. (1992). Law that does not fit society: Sentencing guidelines as a neoclassical reaction to the dilemmas of substantivized law. *American Journal of Sociology*, 97, 1346-1381.
- Snijders, T.A., & Bosker, R.J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Spohn, C. (2000). Thirty years of sentencing reform: The quest for a racially neutral sentencing process. *Criminal Justice: The National Institute of Justice Journal*, 3, 427-501.
- Spohn, C., & Holleran, D. (2000). The imprisonment penalty paid by young, unemployed black and Hispanic male offenders. *Criminology*, 38(1), 281-306.
- Steenbergen, M. R., Jones, B. S. (2002). Modeling multilevel data structures. *American Journal of Political Science*, 46(1), 218-237.
- Steffensmeier, D., & Demuth, S. (2001). Ethnicity and judges' sentencing decision: Hispanic-black-white comparisons. *Criminology*, 39(1), 145-178.
- Steffensmeier, D., & Demuth, S. (2006). Does gender modify the effects of race-ethnicity on criminal sanctioning? Sentences for male and female white, black, and Hispanic defendants. *Journal of Quantitative Criminology*, 22, 241-261.
- Steffensmeier, D., Ulmer, J. T., & Kramer, J. H. (1998). The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, black, and male. *Criminology*, 36(4), 763- 98.
- Tate, R. L. (2004). A cautionary note on shrinkage estimates of school and teacher effects. *Florida Journal of Educational Research*, 42, 1-21.

- Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. New York, NY: Falmer Press.
- Tillyer, R., Hartley, R. D., & Ward, J. T. (2015). Differential treatment of female defendants: Does criminal history moderate the effect of gender on sentence length in federal narcotics cases? *Criminal Justice and Behavior*, 42, 703-721.
- Tonry, M. (1996). *Sentencing matters*. New York, NY: Oxford University Press.
- Tyler, T. R. (1990). *Why people obey the law*. New Haven: Yale University Press.
- Ulmer, J. T. (1997). *Social worlds of sentencing: Court communities under sentencing guidelines*. Albany, NY: State University of New York Press.
- Ulmer, J. T. (2012). Recent developments and new directions in sentencing research. *Justice Quarterly*, 29(1), 1-39.
- Ulmer, J. T., & Bradley, M. S. (2006). Variation in trial penalties among serious violent offenses. *Criminology*, 44(3), 631-670.
- Ulmer, J. T. & Johnson, B. D. (2004). Sentencing in context: A multilevel analysis. *Criminology*, 42(1), 137-175.
- Ulmer, J. T., & Kramer, J. H. (1996). Court communities under sentencing guidelines: Dilemmas of formal rationality and sentencing disparity. *Criminology*, 34, 383-408.
- Western, B. (2006). *Punishment and inequality in America*. New York: Russell Sage Foundation.
- Wheeler, S., Weisburd, D., & Bode, N. (1982). Sentencing the white-collar offender: Rhetoric and reality. *American Sociological Review*, 47, 641-659.
- Wilbanks, W. (1987). *The myth of a racist criminal justice system*. Monterey, CA: Brooks/Cole.
- Willms, J. D. (1992). *Monitoring school performance: A guide for educators*. Washington, DC: Falmer Press.
- Wolfe, S. E., Pyrooz, D. C., Spohn, C. (2011). Unraveling the effect of offender citizenship status on federal sentencing outcomes. *Social Science Research*, 40, 349-362.
- Wooldredge, J. (2010). Judges' unequal contributions to extralegal disparities in imprisonment. *Criminology*, 48(2), 539-567.
- Zatz, M. (1987). The changing forms of racial/ethnic bias in sentencing. *Journal of Research in Crime and Delinquency*, 24(1), 69-92.
- Zatz, M. (2000). The convergence of race, ethnicity, gender, and class on court decisionmaking:

Looking toward the 21st century. *Criminal Justice: The National Institute of Justice Journal*, 3, 503-552.