

University of Michigan Law School

## University of Michigan Law School Scholarship Repository

---

Articles

Faculty Scholarship

---

2020

### Mapping the Iceberg: The Impact of Data Sources on the Study of District Courts

Christina L. Boyd

*University of Georgia*

Pauline T. Kim

*Washington University School of Law*

Margo Schlanger

*University of Michigan Law School, mschlan@umich.edu*

Available at: <https://repository.law.umich.edu/articles/2185>

Follow this and additional works at: <https://repository.law.umich.edu/articles>



Part of the [Courts Commons](#), [Judges Commons](#), and the [Litigation Commons](#)

---

#### Recommended Citation

Bpyd, Christina L. "Mapping the Iceberg: The Impact of Data Sources on the Study of District Courts."  
Pauline T. Kim and Margo Schlanger, co-authors. *J. Empirical Legal Stud.* 17, no. 3 (2020): 466-92.

This Article is brought to you for free and open access by the Faculty Scholarship at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Articles by an authorized administrator of University of Michigan Law School Scholarship Repository. For more information, please contact [mlaw.repository@umich.edu](mailto:mlaw.repository@umich.edu).



# Mapping the Iceberg: The Impact of Data Sources on the Study of District Courts

*Christina L. Boyd, Pauline T. Kim\**, and Margo Schlanger

Three decades ago, Siegelman and Donohue aptly characterized research about courts and litigation that relied only on published opinions as “studying the iceberg from its tip.” They implored researchers to view published district court opinions “with greater sensitivity to the ways in which such cases are unrepresentative of all cases”. The dynamic, multistage nature of trial court litigation makes a focus solely on published opinions particularly ill-suited to the study of federal district courts. Expanded electronic access to court documents now allows more precise analysis of the ways in which published cases are unrepresentative and what differences that makes for conclusions about the work of district courts. Heeding Siegelman and Donohue’s admonition, this study seeks to map the iceberg, exploring the extent to which the visible part misrepresents what lies below the surface. Using a supplemented version of the Kim, Schlanger, and Martin *EEOC Litigation Project* data, this article examines the varying extent to which cases and judicial activity are visible in the several data sources commonly used by district court researchers. More specifically, we analyze how the work of federal district courts looks different depending on whether research relies on published opinions, on opinions available on Westlaw or Lexis (both “published” and “unpublished”), or on more comprehensive data available on PACER (Public Access to Court Electronic Documents). Our results reveal vast variation in visibility of cases and motions, depending on the data source used. We also demonstrate that these differences in case and motion visibility can affect the results of empirical analyses relating to, for example, the success rates of litigants and whether the party of the appointing president affects judicial behavior. Our findings mean that utilizing docket sheets, now available electronically, to gather data will often be required to draw accurate conclusions about the nature of district court litigation and the behavior of district court judges.

## I. INTRODUCTION

Federal district courts are incredibly important actors in the implementation of federal law. Today, these courts receive well over 350,000 new civil and criminal cases per year,

---

\*Address correspondence to Pauline Kim, Daniel Noyes Kirby Professor of Law, Washington University School of Law, Campus Box 1120, One Brookings Drive, St. Louis, MO 63130; email: kim@wustl.edu. Boyd is Associate Professor, Department of Political Science, University of Georgia; Schlanger is the Wade H. and Dores M. McCree Collegiate Professor of Law, University of Michigan. We are grateful for the generous support of the William W. Cook Endowment of the University of Michigan, the National Science Foundation (SES-0718831), and to the following research assistants for their work on this project: Adam Rutkowski, Emma Brunner, Vander Copeland, Estefania Edens, Katherine Feagin, Young Jeon, Savannah Lawson, Caitlyn Kinard and Jordan McGill.

compared to just 50,000 matters in the federal courts of appeals and fewer than 80 merits cases at the U.S. Supreme Court annually (Administrative Office of the U.S. Courts 2018; Spaeth et al. 2019). Because of this unequal distribution of cases across the judicial hierarchy, federal district judges, who constitute over three-quarters of authorized federal Article III judgeships (Judicial Business 2018), have been referred to as “the workhorses of the federal judiciary” (Abraham 1998). For most federal cases, the pursuit of justice not only begins in the district court but ends there, too (Carp & Wheeler 1972:361).

But while the importance of district courts can hardly be disputed, empirical scholars of district courts face high hurdles studying them in an appropriate and comprehensive way. As Hoffman et al. argue, “empirical work about [federal] trial courts is more expensive, more time-consuming, and more uncertain than one might imagine” (2007:727). And as Levin notes, “studying the district courts in a systematic way is difficult—more difficult than studying federal appellate courts and far more difficult than studying the Supreme Court” (2008:981). What makes good empirical district court research so difficult? Much of the difficulty stems from the dynamic, multistaged nature of litigation at the trial court level. Kim et al. explain:

[A] district judge may rule in a single case on multiple occasions and on different types of questions, only a few of which could be dispositive but all of which affect the case’s progress and ultimate outcome. Moreover, because many of the judge’s actions are taken in response to motions by the parties, there is no determinate sequence in which pretrial litigation events occur. Rather, how a case proceeds depends on the choices made by the parties—what motions are filed by whom and how discovery unfolds (2009:85).

The bulk of the prior literature has identified district court cases for study through the opinions published, in print, in official sources like the *Federal Supplement* or *Federal Rules Decisions* (e.g., Rowland et al. 1984; Rowland & Carp 1996; Schultz & Petterson 1992; Segal 2000; Walker & Barrow 1985; Winkler 2006). More recent district court empirical scholarship often also includes “unpublished” opinions that are available through legal databases like Westlaw and Lexis (e.g., Banks & Tauber 2014; Sen 2015)<sup>1</sup>.

Does studying published and unpublished opinions available in commercial legal databases accurately capture the work of the federal district courts? We have multiple reasons to suspect that the answer is no. One study (Schlanger & Lieberman 2006:163–64) found 28,000 district court opinions in 2004 that were accessible on Westlaw, even though over 10 times that many cases were terminated in the district courts during the

---

<sup>1</sup>For years, some district court scholars have used the Administrative Office of the U.S. Courts and Federal Judicial Center’s Integrated Data Base (see Eisenberg 2004; Eisenberg & Schlanger 2005). These data provide case-level information on every federal district court case dating back to 1970 that includes “names of the parties, the subject matter category and the jurisdictional basis of the case, the case’s origin in the district as original or removed or transferred, the amount demanded, the dates of filing and termination in the district court or the court of appeals, the procedural stage of the case at termination, the procedural method of disposition, and, if the court entered judgment or reached decision, who prevailed” (Clermont et al. 2003). The FJC has recently made these data far more accessible by posting them in close to real time and with caption and docket number information at <https://www.fjc.gov/research/idb>.

same period. Schlanger and Lieberman (2006) also reported a highly skewed geographical variance:

[W]hile over half the districts (58 of the 94) produced opinions at a rate of 5% or less of terminations, the others were more diverse, ranging from 6–10% to 40–45% and even (for two outliers) 60–80% .... Most notably, the Northern District of Illinois, the Southern District of New York, and the Northern District of Texas between them terminated 9.2% of all federal district cases in 2004, but produced 33.4% of all opinions.

Similarly, Hoffman et al.'s (2007) deep dive into how often federal district court cases produced Westlaw- and Lexis-available opinions indicated that only 3 percent of the judge actions in their four-district study led to a published or unpublished opinion in one of the legal databases. As to why this is the case, Boyd argued that the reasons range “from the frivolous nature of some lawsuits to the limited judicial activity required in some cases ... to the presence of judicial discretion in engaging in opinion writing behavior. For these judges, many case activities requiring judicial intervention can be resolved orally or through short, publicly unavailable orders, meaning that writing an opinion is voluntary and, at the extreme, perhaps even superfluous” (2015:255).

Whether an unpublished opinion is included in Westlaw or Lexis is likely to be driven by factors such as whether a judge provides written justification for her decision on a motion, whether that written decision is classified as an opinion or merely an order, whether the court has made the opinion electronically available, and when it was decided (see Boyd 2015; Hoffman et al. 2007; Kim et al. 2009). For example, McCuskey's (2016) in-depth examination of nearly 200 federal question jurisdiction removal cases in the Northern District of Illinois and Eastern District of Virginia reveals that a substantial number of district court decisions are not visible on Westlaw or Lexis. While some of these non-visible decisions are unreasoned orders, others are fully reasoned opinions—what McCuskey calls submerged precedent.

Which district court opinions are published in the *Federal Supplement* or *Federal Rules Decisions* involves an additional nonrandom selection process. Olson explained in her study of Minnesota district court publication tendencies from 1982 to 1984, “[t]he choice of which and how many judicial opinions to send up [to the reporter system] is left up to the individual district judges with only minimal policy guidance” (1992:786). Other research on district court publication practices indicates that key factors include judge preference, local court culture, judicial promotion potential, case salience and importance, and the publishers' rules, among others (McCuskey 2016; Morriss et al. 2005; Ringquist & Emmert 1999; Rowland & Carp 1996; Siegelman & Donohue 1990; Songer 1988; Swenson 2004; Taha 2004).

In short, some types of outcomes, cases, judges, and courts are more likely to appear in print or on Westlaw and Lexis (whether as published and unpublished opinions) than others. This, in turn, means that the district court cases identified through a search of published *Federal Supplement* and *Federal Rules Decisions* opinions or through a Westlaw or Lexis search of published and unpublished opinions are probably unrepresentative of the broader population of district court cases. If “published opinions are not representative of all opinions; opinions are not representative of all district court

decisions; and adjudicated cases are not representative of all filed cases” (Kim et al. 2009:97), then research designs examining published or unpublished opinions will often lead to skewed empirical results.

Of course, we are not the first to recognize the limitations of some of the district court research data sources. As Hoffman et al. put it, it is stark “how little trial court work is explained through written opinions” (2007:727). Siegelman and Donohue, in their extensive study of the publication of opinions in employment discrimination cases, argued that the “body of [district court] cases with published opinions needs to be viewed with greater sensitivity to the ways in which such cases are unrepresentative of all cases” (1990:1137).<sup>2</sup> Ringquist and Emmert argued that “in examining published cases alone, we are investigating an unrepresentative subset of decisions which in turn may seriously compromise the validity of conclusions regarding judicial behavior” (1999:33). But while multiple scholars have pointed out the potential impact of incomplete data, the existing literature on the impact of the data source choice is limited. Some studies are dated or examined only a small number of cases or districts (Hoffman et al. 2007; McCuskey 2016; Olson 1992; Swenson 2004). Others rely on appeals to identify district court cases (Songer 1988) or omit motion-level activity (Ringquist & Emmert 1999; Siegelman & Donohue 1990). One admirable study used dockets and court documents (often obtained from court clerks’ offices), but had a narrow judge ideology focus (Keele et al. 2009). These limitations are completely understandable, of course, given constraints of cost and limited accessibility to court documents that hampered district court research in the past.

But it is now possible to do better. In the three decades since the publication of the Siegelman and Donohue (1990) study, arguably still the most comprehensive treatment of this topic, advances in electronic docketing and data availability have opened up significant new opportunities for systematic study of district court litigation. Nevertheless, much recent work continues to restrict the number of districts studied, neglect motion-level activity, or examine only materials available on Westlaw/Lexis. This article explores the impact of those choices in research design when more comprehensive data are available.

In what follows, we provide what we believe is the most comprehensive examination to date of the implications of research design for quantitative empirical examination of federal district courts. In Section II, we detail our data, beginning with the Kim, Schlanger, and Martin (2013) *EEOC Litigation Project* data. As we document, for this project we supplemented that comprehensive set of EEOC cases by coding the visibility of each case and each motion across four different data sources: published opinions (*Federal Supplement* and *Federal Rules Decisions*); Westlaw or Lexis legal databases (including both published and unpublished opinions); case documents available electronically through PACER; and PACER docket sheets. In Section III we turn to an in-depth examination of the data, describing differences in the visibility of cases and court activity in each data

---

<sup>2</sup>Siegelman and Donohue define “published opinions” as those available on Lexis rather than our stricter (and more familiar) definition that includes only opinions available in official printed reporters.

source and examining how those differences vary across districts and time. In Section IV we tease out the effects of the visibility differences observed in Section III. Specifically, we examine when and the degree to which potential conclusions about case and motion winners, case termination method, monetary and injunctive relief obtained, and the influence of judge party, sex, and race on motion decisions are affected by the data source. Our results demonstrate that findings about district court cases, motions, and judges are often quite different if based on more selective sources compared to near-comprehensive sources. In Section V we conclude by exploring the broader implications for empirical district court research.

## II. OUR APPROACH: DATA AND RESEARCH DESIGN

### A. EEOC Litigation Project Data

We utilize the Kim et al. (2013) *EEOC Litigation Project* dataset,<sup>3</sup> which samples employment discrimination lawsuits initiated by the EEOC in federal district courts over a 10-year period ending in 2006.<sup>4</sup> Starting with a complete list of all lawsuits brought by the EEOC during that period, Kim et al. (2013) drew a stratified sample.<sup>5</sup> The final dataset includes 2,316 cases.<sup>6</sup> Unless otherwise noted, statistics reported below include appropriate sampling weights. For each case in the dataset, PACER was used to retrieve docket sheets and (where available) relevant underlying case documents such as complaints, orders, opinions, and consent decrees.

The *EEOC Litigation Project* data include information regarding the parties, judges, allegations, and details about how cases were finally resolved (e.g., consent decree, court judgment, or jury verdict), as well as any relief obtained (monetary or injunctive). The data also include information about every significant motion<sup>7</sup> filed in the case—including

---

<sup>3</sup>The *EEOC Litigation Project* data and documentation are available at <http://eeoclitigation.wustl.edu/>.

<sup>4</sup>The data include cases filed in fiscal years 1997 through 2006—namely, cases filed from October 1, 1996 through September 30, 2006.

<sup>5</sup>The resulting data include all cases classified by the EEOC as intended to benefit more than one employee and all cases concluded via a contested court order (whether an opinion or a verdict). The data also include a random sample of all remaining EEOC cases.

<sup>6</sup>All cases without a district court resolution by April 22, 2008 were coded with “ongoing” as the case outcome.

<sup>7</sup>“Significant motion” is broadly defined and includes the following motion types: motion to dismiss (involuntary, including 12(b)(6), 12(b)(7), for jurisdiction or venue), motion for judgment on the pleadings, motion for summary judgment, motion to alter/amend judgment, motion for judgment as a matter of law, motion to dismiss—voluntary, motion to intervene by complainant or EEOC, motion for default, motion for preliminary injunction, motion for injunction, motion for joinder, motion for severance, motion for remittitur/additur, motion for new trial, motion to consolidate, motion or notice to relate cases, and substantive discovery motion. Substantive discovery motions include matters like motions to compel and motions for protective orders but do not include motions involving more administrative matters such as the timing of discovery or numerical limits. Similarly, motions regarding administrative matters like the addition or substitution of counsel were omitted.

the dates it was filed and resolved, whether the motion was opposed, the outcome, and the deciding judge.<sup>8</sup> These data provide a comprehensive picture of district court activity in the cases sampled.

### B. Data Extension

The *EEOC Litigation Project* data capture every significant motion that occurred in any of the cases in the sample. To study the visibility of these case activities in other data sources, we supplemented the data as follows. For every motion within the *EEOC Litigation Project* data (i.e., every significant motion), we determined whether the resolution of the motion is published in the *Federal Supplement* or *Federal Rules Decisions Reporters*<sup>9</sup> or is available as an unpublished decision in Westlaw,<sup>10</sup> LexisNexis,<sup>11</sup> or both. The Westlaw and LexisNexis searches used to extend the *EEOC Litigation Project* data were conducted between 2016 and 2019.

We also coded whether the resolution of each motion, whether in the form of an order or opinion, is electronically available on PACER, by looking for a live hyperlink on the docket sheet.

With these extensions, the data allow us to study the representativeness of district court research conducted with today's electronic research tools.

### C. Important Groundwork: Categories, Definitions, and Levels of Analyses

This section describes our study's key concepts and levels of analyses to lay the groundwork for the discussion and data analyses that follow.

When considering to what extent the activities of district courts are visible to researchers, there are two distinct levels of analysis—the motion level and the case level. Which one should be used depends on the research question posed. For example, if a researcher wants to know how judges make decisions, it may be important to capture how they decide individual motions. On the other hand, if a researcher is concerned with plaintiffs' success rates in litigation, case-level outcomes will be more appropriate. Our

---

<sup>8</sup>After extensive training, law student and lawyer coders captured detailed information about litigation activity in each case from the docket and available case documents. To ensure data reliability, an expert coder checked and approved all coding.

<sup>9</sup>To ensure the inclusion of all relevant opinions and orders across sources, we separately retrieved citations from Westlaw and Lexis.

<sup>10</sup>In Westlaw, coders were required to search for relevant opinions in each of three different ways before concluding that no citation (published or unpublished) was available in the database: (1) by case name for the federal district courts source, filtered by jurisdiction if needed, (2) by the dockets source, searching by state and docket number, and (3) by court source, searching by docket number and party name.

<sup>11</sup>In Lexis, coders limited their searches to "All US District Courts" and then performed multiple searches for a combination of "equal employment" (and other variations of EEOC), the state of the district court, the defendant's name (multiple variants as needed), and the docket number.

data include 2,316 cases, but many more motions—that is, case activity that requires some type of response from a judge.<sup>12</sup> A given case may have no, one, or many motions.

The majority of our analysis that follows focuses on motions. Each motion is categorized according to its availability in different data sources. If the outcome of a motion is available in a particular data source, we say that it is “visible” at that level.

The four levels of visibility are:

1. *Published Opinion*. The opinion or order deciding the motion is published in an official reporter. These official reporters include the *Federal Supplement* (F. Supp., F. Supp. 2d, F. Supp. 3d) and the *Federal Rules Decisions* (F.R.D.).<sup>13</sup>
2. *Westlaw or Lexis Opinion*. The motion opinion or order is available on Westlaw and/or Lexis, whether officially published or unpublished.
3. *PACER Hyperlinked*. An opinion or order falls within this category when the PACER docket sheet includes a hyperlink to an electronic version of the opinion/order. Some of these rulings are conclusory; others are equivalent to what McCuskey calls submerged precedent—that is, “reasoned opinions available only on court dockets, and not on the Westlaw and Lexis commercial databases” (2016:516).
4. *PACER Docket Sheet*. The PACER docket sheet always provides a written record that an opinion and/or order resolved a motion. In many instances, the docket sheet text describes in detail the type of motion, the date filed and resolved, the motion’s outcome, and the judge who resolved the motion. At times, the docket sheet text also includes further details about why the judge reached that decision.

Generally speaking, each level of motion visibility is encompassed or nested within the levels below it.<sup>14</sup> Figure 1 illustrates this relationship. If a motion produced an opinion published in the *Federal Supplement*, it will generally also be available in Westlaw and Lexis databases, hyperlinked on PACER, and listed on the PACER docket sheet. Thus,

---

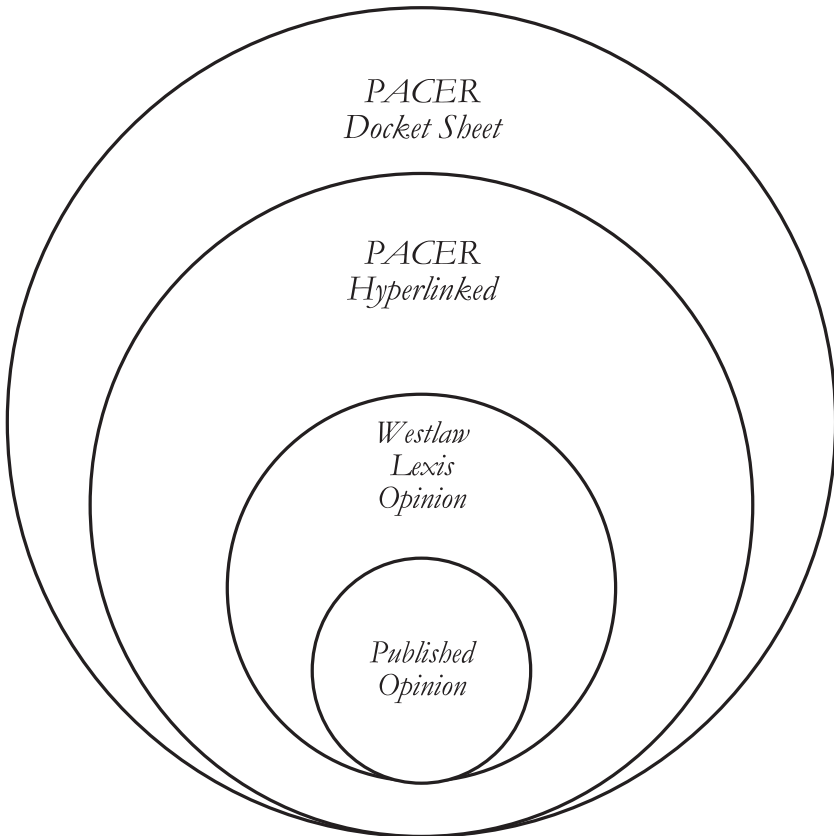
<sup>12</sup>Unless otherwise noted, each observation in the motion-level analyses is based on a unique underlying motion. This is true even if two (or more) motions were resolved by a single court order or opinion. Note, however, that we do not code multiple court orders/opinions resolving the same motion as multiple observations. It is a frequent occurrence for there to be two separate docket sheet entries for the resolution of a motion, with one entering, for example, the judge’s memorandum opinion and the other entering the accompanying order on that motion resolution.

<sup>13</sup>We also have two observations with published opinions in the Bankruptcy Reporter (B.R.). We have coded these as published.

<sup>14</sup>For 0.73 percent of the motion opinions, an opinion is available on Lexis but not on Westlaw, and 0.6 percent of the motion opinions are available on Westlaw but not Lexis. Under 3 percent of motion resolutions are available as published or unpublished opinions through Westlaw or Lexis but are not electronically available as hyperlinked documents on PACER. This circumstance is almost entirely limited to the earlier years of the data when district court dockets were not yet required to link electronic versions of case outcome documents.



Figure 1: Illustration of the four levels of motion visibility.



the four categories are cumulative. The last category, PACER docket sheet, encompasses all motions in the dataset.

At times, a case-level analysis will be more appropriate—for example, when asking about case outcome and termination method, the type of discrimination alleged, or the resulting injunctive and monetary relief. To conduct case-level analyses, we again capture whether a given case is visible at each of the four levels based on whether any of the case documents are available from that source. The difference between the case-level and motion-level analyses is that a case is coded as “available,” and therefore visible, if *any* case document (not only documents relating to a particular motion) appears in that data source. Some cases have multiple opinions and orders, and often these are inconsistently visible. For example, a case may have three available case documents: a published (*F. Supp.*) opinion resolving a dispositive motion, an unpublished opinion on a discovery motion available on Lexis and Westlaw, and an order documenting the in-court

resolution of a motion available through a hyperlink on the PACER docket sheet. For purposes of determining visibility at the case level, we use the highest (most prominent) publication or availability source for any of the case's documents. In the example just set out, the case would be recorded as being visible in a published source. This coding decision reflects the fact that researchers can learn of a case's existence based on just one document. By its nature, our coding approach is very generous toward finding cases to be available in the more selective data sources.

During our analyses that follow, we also refer to two other sources of data for identifying district court activity: the Carp-Manning U.S. District Court Database (Carp & Manning 2016) and appellate court data. The Carp and Manning Database contains over 100,000 federal district court judge decisions from 1927 to 2012 and is the data source for numerous political science studies on the U.S. district courts (e.g., Rowland et al. 1984; Rowland & Carp 1996). The Carp-Manning data are selected from *Federal Supplement* decisions, focusing only on those decisions that contain a liberal-conservative dimension, have a clear-cut winner or loser without division across multiple issues, and were decided by an Article III district judge rather than by a magistrate judge (Carp & Manning 2016). We merged the Carp-Manning data with the *EEOC Litigation Project* data via the *Federal Supplement* citations.

We also consider district court cases that can be identified through the presence of circuit court appeals. Doing so allows comparison with research designs like Songer (1988) and Randazzo (2008). To determine which cases could be identified through appellate data, we merge our data with the Federal Judicial Center's Integrated Database for Appellate Court Terminations, thereby obtaining information on whether there was an appeal in a district court case<sup>15</sup> and, if so, how that appeal terminated.

### III. DIFFERENCES EMERGE: DESCRIPTIVE DATA

In this section, we consider descriptive differences in the visibility of cases and court decisions on motions, depending on the data source used.

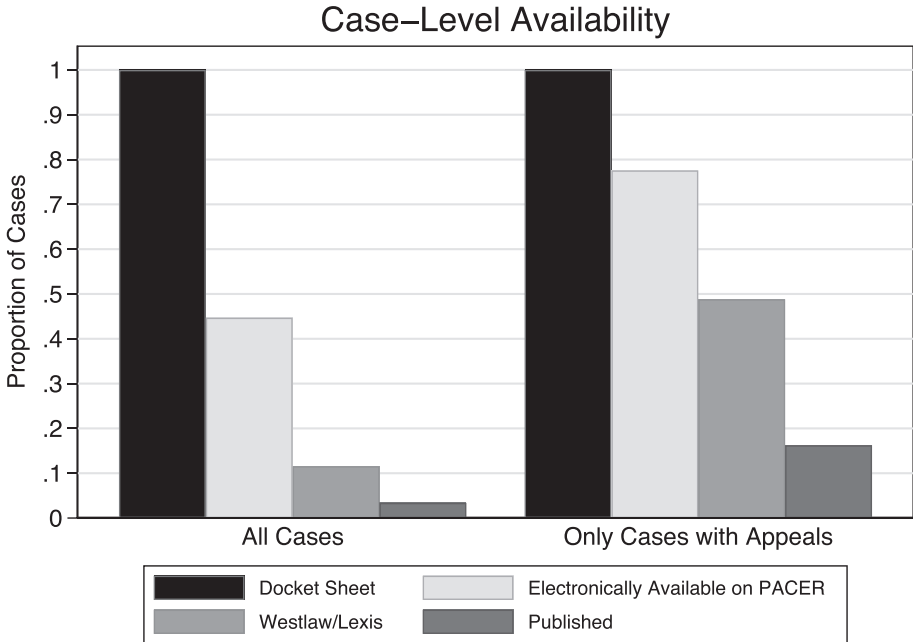
#### A. Variations in Case Visibility

We start by examining variations in the visibility of cases across different data sources. This case-level focus is analogous to many prior district court studies, including Siegelman and Donohue (1990). Case-level analysis allows us to examine case visibility regardless of whether the case involved any motions or resulted in any court-produced case documents. Cases without motions or documents leave no record in traditional opinion searches of *F. Supp.*, Lexis, and Westlaw and have no hyperlinked court documents available on PACER. These zero motions cases, by definition, do not appear in the

---

<sup>15</sup>We define appeals in the most generous way possible to include all instances where there is a notice of appeal.

Figure 2: Visibility of cases by data sources; all cases in data and cases with an appeal.



motion-level data we discuss below or in any study relying on court opinions (published and unpublished), but they do exist, and docket sheets provide information about them.

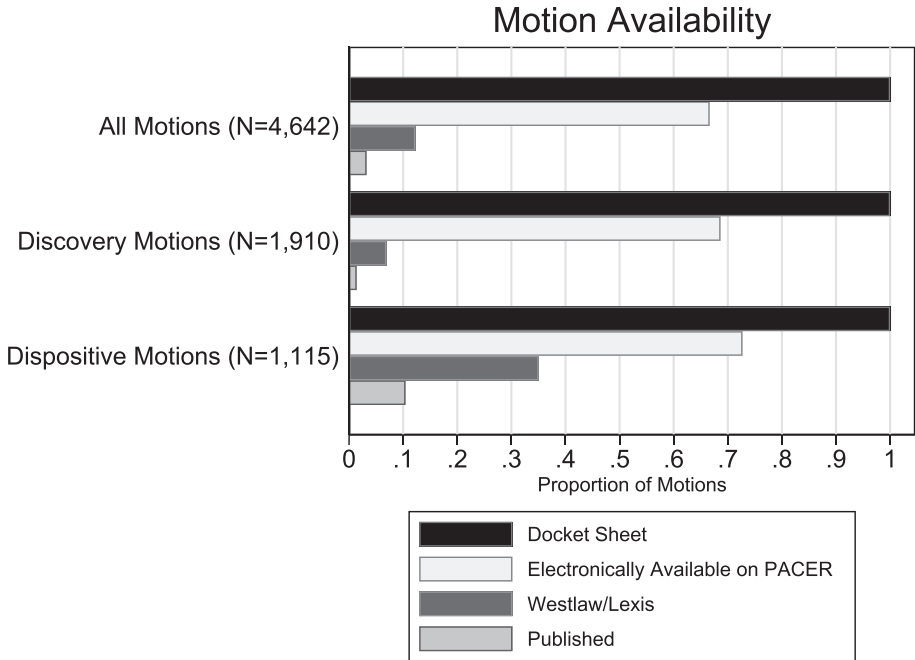
Figure 2 depicts the visibility of cases across data sources. As it shows, for the *EEOC Litigation Project* data, information is available for all cases in PACER docket sheets. While PACER docket sheets record basic information about the case, additional information is available in about 45 percent of the cases by examining docket-hyperlinked case documents. If a researcher studied district court cases by searching all Lexis and Westlaw opinions, only about 10 percent of cases would be identified, and only about 2 percent of cases would be visible if a researcher restricted herself to published reporters.

Focusing only on cases that eventually have an appeal—6 percent of the cases in our data—the differences in visibility are less pronounced. Still, less than 50 percent of the district court cases with appeals are visible through Westlaw and Lexis and only 16 percent are officially published.

### B. Variations in Motion Visibility

One of the advantages of the *EEOC Litigation Project's* data is that they permit examination not just of individual EEOC cases but also more in-depth inquiries into the dynamic motion activity within individual cases. The data include a great deal of motion

Figure 3: Visibility of motions by data source for all motions, discovery motions, and dispositive motions.



activity—a total of 4,642 significant motions (the equivalent of 6,154 population motions once case sampling weights are accounted for).<sup>16</sup> Cases have an average of 5.42 significant motions.

Figure 3 depicts the visibility of motion activity across different data sources, for all motions (a combination of discovery, dispositive, and others) and for only discovery motions and only dispositive motions.<sup>17</sup> While every motion is visible on PACER docket sheets, many are invisible in the other data sources. For all types of motions, the documents (orders or opinions) resolving the motions are electronically available on PACER through hyperlinks around 70 percent of the time. In published reporters and Lexis/Westlaw opinion databases, motion visibility is both lower and more varied.

<sup>16</sup>See footnote 7 for full details on what constitutes a significant motion in the data.

<sup>17</sup>We define the following *EEOC Litigation Project* motions as dispositive motions: 12(b)(6) Motion to Dismiss, Motion to Dismiss—Involuntary, Motion for Judgment on the Pleadings, Motion for Summary Judgment, and Motion for Judgment as a Matter of Law.

Perhaps not surprisingly, dispositive motions are more visible in both published sources or Westlaw/Lexis opinion searches than are discovery motions or all types of motions. Still, only 35 percent of the dispositive motions in the population of the *EEOC Litigation Project's* data are discoverable via Westlaw or Lexis opinion searches. For all motions, that number is 12 percent, and for discovery motions, it is just under 7 percent. Relying on published sources (*F. Supp.* and *F.R.D.*) yields even fewer of the motions—just over 10 percent of the dispositive motions, 3 percent of all motions, and 1 percent of discovery motions.

Overall, then, the descriptive results in Figure 3 indicate that only a small slice of motion activity in district court cases is visible from opinion searches of Westlaw or Lexis. This is so even for dispositive motions. In the next two subsections we explore how the visibility of motion activity varies across districts and over time.

### C. District Court Variation in Motion Visibility

The *EEOC Litigation Project* data include motion activity in cases in 91 of the 94 federal district courts.<sup>18</sup> Visibility of this activity varies widely by district, as can be seen in Figures 4a and 4b. Figure 4a presents the visibility of all motion activity by data source for the 15 district courts with the highest number of motions in the data. Compare districts like E.D. Missouri, N.D. Texas, and E.D. Pennsylvania to districts like E.D. Michigan, W.D. Washington, and W.D. Texas. In the former districts, nearly 20 percent of motion outcome documents are available through Westlaw and Lexis (whether as published or unpublished opinions). By contrast, in the latter districts, there is only about a 1 or 2 percent visibility rate in those easy-to-access Westlaw and Lexis sources.

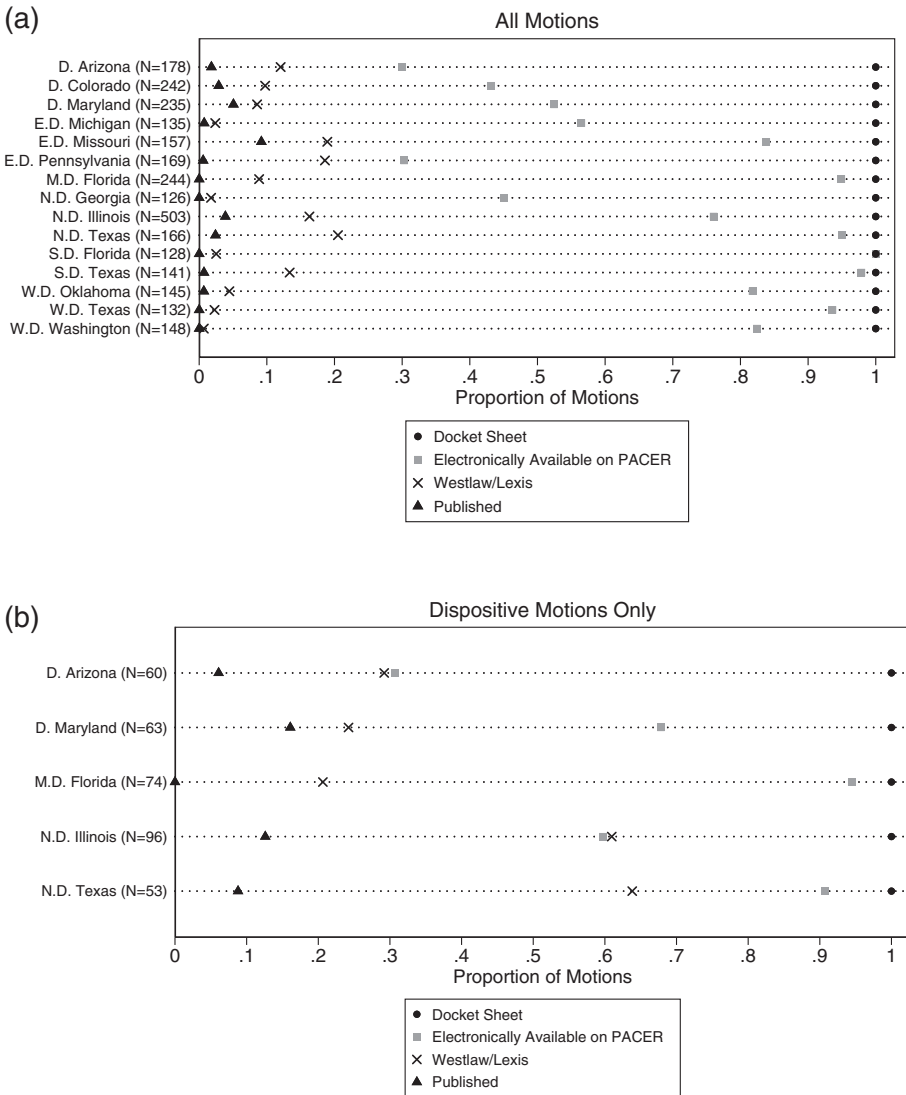
Figure 4b presents similar information, but this time solely for dispositive motions. Because of small numbers, we highlight just the five district courts with the most motions in the data. Variability in visibility is again the norm. For example, in the N.D. Illinois and N.D. Texas, the outcomes of dispositive motions are available through Westlaw and Lexis around 60 percent of the time, whereas in the M.D. Florida, these documents are available for only 20 percent of dispositive motions.

Given prior research finding differences in when and why district court judges submit cases for publication, including based on local legal culture (Ringquist & Emmert 1999; Siegelman & Donohue 1990), variation is not surprising. Nonetheless, the magnitude of the visibility differences across districts is jarring. For researchers designing a data-collection project from Westlaw and Lexis, these differences matter—a topic to which we return shortly.

---

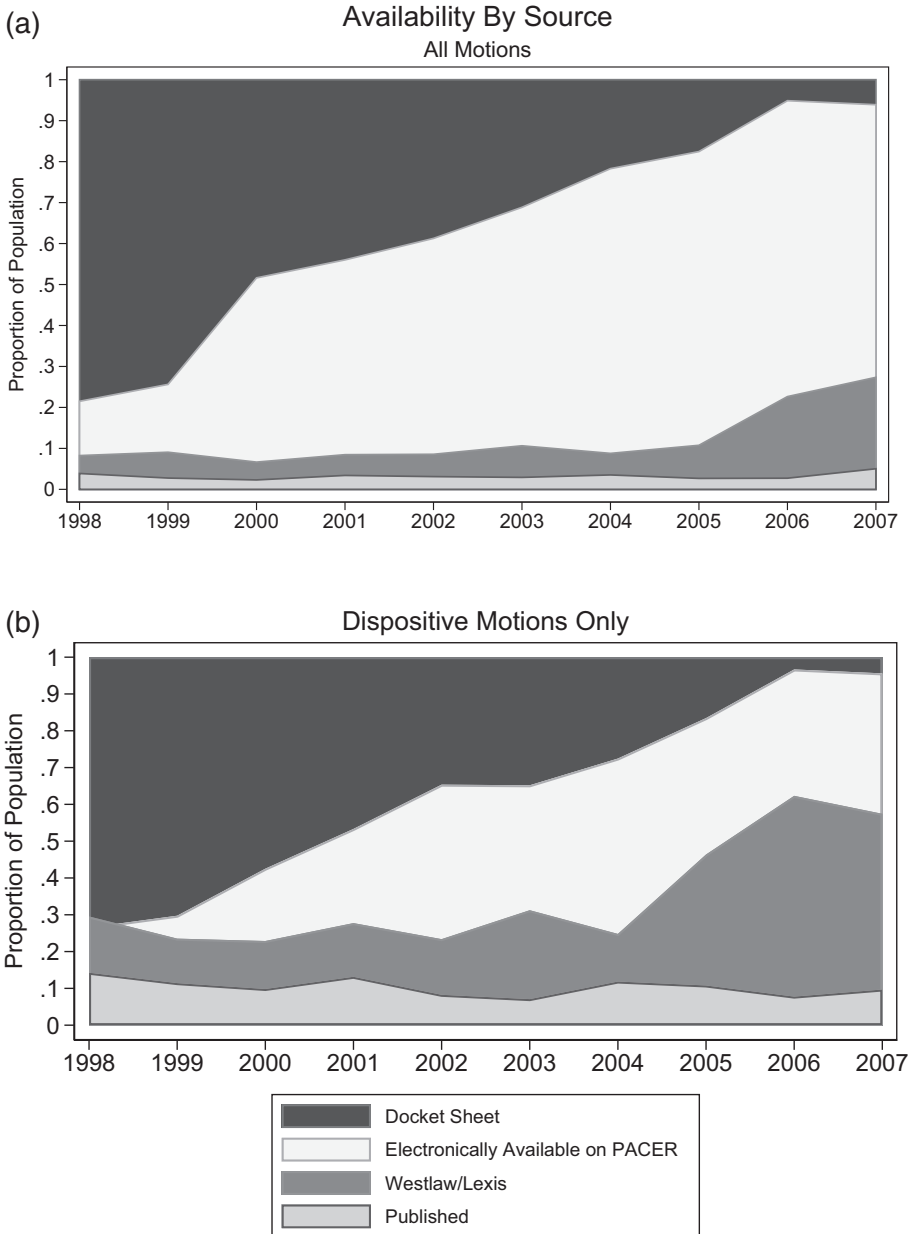
<sup>18</sup>The District Court for the Virgin Islands is the only district court without any cases in the data. In addition, there are no cases in the data with any motions present for the District of Vermont and the Southern District of West Virginia.

Figure 4: (a) Visibility of all motion types by data source for district courts with greatest number of motions. (b) Visibility of dispositive motions by data source for district courts with greatest number of motions.



NOTE: Proportions and Ns are calculated using the sampling weights.

Figure 5: (a) Visibility of motions by data source from 1998–2007, all motions. (b) Visibility of motions by data source from 1998–2007, dispositive motions only.



#### D. Variation in Motion Visibility Based on Decision Date

Figure 5 examines differences in the visibility of motions across the years of our data. The figure plots the overall visibility of motions resolved from 1998 to 2007 in different data sources, for all motions and for dispositive motions.<sup>19</sup> Like Figure 4, Figure 5 highlights enormous variation in visibility by source. It adds notable differences depending on the date of decision.

For decisions issued prior to 2000, under 30 percent of dispositive motions (and 25 percent of all motions) are visible in any non-PACER docket sheet format. By the later years of our data, by contrast, there are very few instances where the only record of a motion's outcome was the docket sheet (solid black shading in Figure 5); for decisions issued after 2005, nearly 90 percent of all motions—and almost 100 percent of dispositive motions—have an opinion or order electronically accessible via PACER hyperlinks. This uptick in electronic document availability in the later years of our data coincides with the gradual implementation of electronic filing in the district courts starting in the late 1990s (U.S. Courts 2013) and the passage of the E-Government Act of 2002 (Kim et al. 2009).

Figure 5 also shows interesting variation in the visibility of motions on Lexis and/or Westlaw. While the overall proportions are low for decisions issued in the early years of our period (below 10 percent for all motions through 2005 (Figure 5a); at or below 30 percent for dispositive motions through 2004 (Figure 5b)), the proportions increase substantially for the later years. This is particularly noticeable for dispositive motions; for this slice of data, which drives the overall results, the percentage of opinions decided after 2004 that are available from Westlaw or Lexis jumps, rising to about 60 percent for decisions in 2006. Presumably, the increase in availability of opinions via PACER hyperlinks in turn has enabled Westlaw and Lexis to expand with little effort the number of unpublished district court decisions they offer for these later time periods.

In contrast to the expansion of electronically available information for cases decided in later years, there is little change in the proportion of published opinions. Throughout our study period, from 1998 to 2007, only about 3 percent of all motions and about 10 percent of dispositive motions are visible in published sources. This lack of variation likely results from invariant selection criteria. Since well before 1998, West Publishers has published reporters like the *Federal Supplement* and has informed judges and courts how and when to submit opinions for publication.

While motion activity during the last full year of our period is much more visible on Lexis and Westlaw (published and unpublished) than for earlier years, nearly 50 percent of the dispositive motion outcomes are visible only via PACER sources (Figure 5b), and this is true for over 70 percent of all motion outcomes (Figure 5a).

Clearly, there is increased availability of court documents for the later years in our time period. Our data do not allow us to observe the trend line for decisions rendered after 2007. Nevertheless, caution regarding research design remains warranted. McCuskey

---

<sup>19</sup>Figure 5 tracks motion resolutions in our dataset, with the x-axis indicating the year the motion was resolved. We exclude 1997 because our dataset begins with cases filed in fiscal year 1997, and few of those cases had time to proceed to a resolution of any motions by the end of 1997.



noted that Westlaw appears to pull its unpublished opinions exclusively from the PACER “Written Opinion” collection (2016:529). This is likely true for Lexis as well. However, “[n]one of the court orders and not every opinion are tagged as an Opinion” (McCuskey 2016:528), and—while we have not systematically tallied such omissions—we have come across some opinions so tagged that nonetheless did not make it into Westlaw’s unpublished opinion collection. In correspondence with Westlaw staff about unpublished opinions, we were told: “Ultimately, ... some decisions not deemed by the courts to be written opinions are included in our case collection, while some decisions deemed to be written opinions are not” (Kenny 2020). In short, we are confident that Lexis and Westlaw have not closed the PACER gap in their unpublished offerings for more recent decisions; even today, no orders and some opinions do not make it into the PACER “Written Opinions” database, and so some motion outcomes will continue to be missing from ordinary Lexis and Westlaw searches.

#### IV. DO DIFFERENCES IN CASE AND MOTION VISIBILITY MATTER?

Informed by the evident differences in case and motion visibility across data sources, we can assess how these differences may affect conclusions often drawn about district court cases, participants, and overall litigation processes. To do this, we examine: (1) case and motion outcomes, the termination method for cases, and amounts and types of monetary and injunctive relief obtained; and (2) judicial behavior.

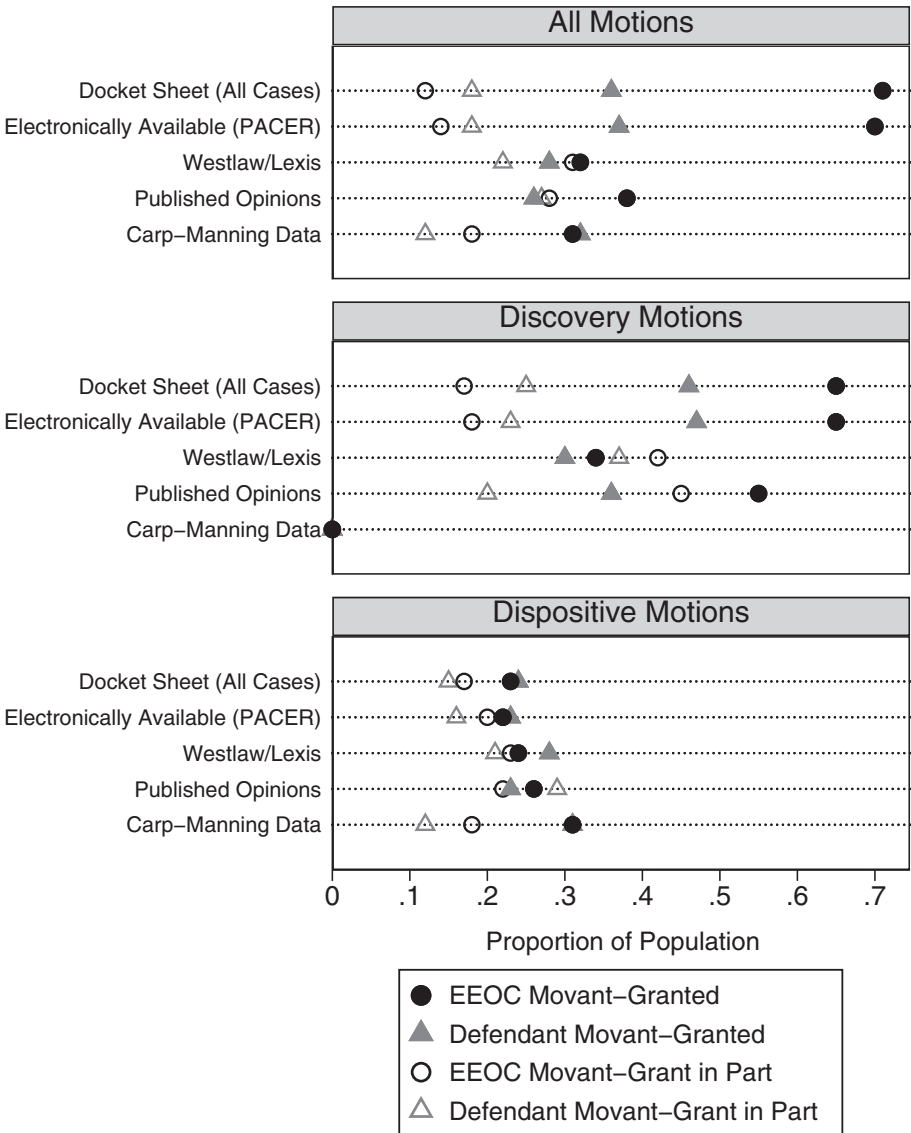
We conduct several analyses below and set out how the results depend on the data source.

##### *A. Likelihood of Winning Individual Motions*

We begin by looking closely at motion win rates. Figure 6 plots the proportion, with sampling weights, of motion outcomes (granted or granted in part/denied in part) based on the identity of the movant (EEOC or defendant) and the type of motion (all types, just discovery motions, and just dispositive motions). Each panel compares these observed outcomes across data sources. Figure 6 illustrates that the observed success rate for motions made by the EEOC turns on data source. This is particularly true for all motions and discovery motions, where the EEOC’s total success rate as a movant is substantially underestimated—by 0.38—if analysis is limited to published and unpublished opinions available on Lexis and Westlaw. In addition, while partial grants of motions are far less common in the data than full grants or denials, when they do happen, a record of their presence varies quite a bit by data source. When the EEOC is the movant, analysis of Lexis/Westlaw-available orders actually over-reports partial successes, again particularly in the all motions and discovery motions subsets. A similar pattern exists for the motions where the defendant is the movant, although the differences are much less extreme.

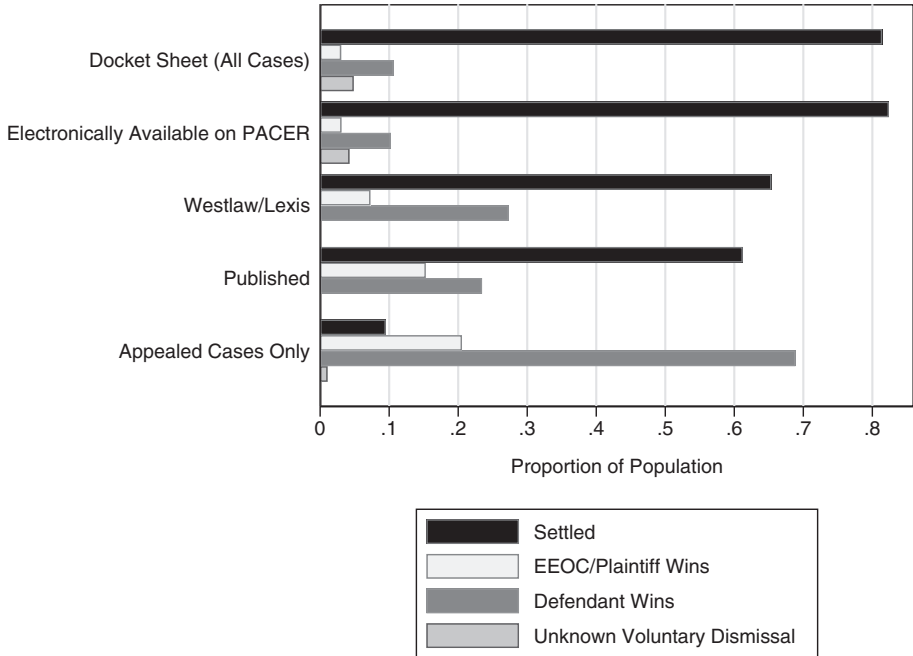
For just dispositive motions, the skew based on different data sources is smaller. There are still some differences of note—for example, outcomes observed in published

Figure 6: Observed motion win rates by data source.



opinions overestimate defendant wins-in-part by about 0.15 relative to the PACER-based sources. However, the differences do not vary as much as for all motions or for discovery motions only. So while Figure 3 demonstrates that *F. Supp.*, Westlaw, and Lexis did not

Figure 7: Case outcomes by data source.



pick up dispositive motions at a very high rate, those motions they did carry turned out, in our data, to be fairly representative of the broader population of those motions, at least as to movants' success rate.

Notably, however, motion outcomes in the Carp-Manning database are less consistent with the comprehensive PACER data. In the Carp-Manning data, dispositive motions (both by the EEOC and defendants) appear to be granted more often than in the full case population. Discovery motions are virtually nonexistent in the Carp-Manning data; indeed, we found only one that matched our PACER docket sheet data. This skew is almost certainly due to the Carp-Manning data's exclusion of unpublished opinions and opinions authored by magistrate judges.

### B. Case Outcomes

We now turn from motion-level to case-level analyses to see whether observed case outcomes differ depending on the data source. In doing so, we return to our publication-generous definition of a case's visibility: we deem a case visible at the most prominent level that any one or more of its motions is available. Figure 7 plots case outcomes by data source. The figure divides the data into four types of case outcome: defendant win,

EEOC win, settlement (often via consent decree), and unknown voluntary dismissal (where documentation is not available to determine definitively whether the case was settled or whether the plaintiff voluntarily abandoned the case).<sup>20</sup> The proportion of cases falling in these four types of case outcome are plotted for each visibility level.

Unsurprisingly, the vast majority of EEOC district court cases are resolved through a negotiated outcome (Schlanger & Kim 2014:1568). However, the estimated proportion of settlements varies drastically based on the source consulted. In the PACER sources (docket sheets only or docket with electronically-linked documents), over 80 percent of the cases settle. However, this number drops to just 60–65 percent in published and combined published and unpublished sources. It falls even further, to just below 10 percent of the cases, when examining only appealed cases.

Unknown voluntary dismissals, which make up about 5 percent of the population, are also significantly underrepresented in the appealed, published, and unpublished data. Indeed, there are almost none of these cases in these three sources. While these lower-than-PACER numbers for settlements and unknown voluntary dismissals are likely not surprising given the nature of published and unpublished opinions and appealed cases, the implications for the litigation representativeness of the resulting data are noteworthy.

Settlements and voluntary dismissals are underrepresented in the unpublished, published, and appealed sources since they have a low likelihood of yielding a written opinion. By contrast, plaintiff and defendant victories are overrepresented in these same sources. For example, of Westlaw/Lexis-opinion-search cases, 30 percent are defendant victories, compared to just over 10 percent in the total population, and about 8 percent are EEOC victories, compared to around 2 percent in the population. In the appealed-only cases, nearly 70 percent were defendant victories in the district court and 20 percent were EEOC/complainant wins.

The relative rates of defendant and EEOC wins also vary depending on the data source. Looking at just cases with party victors (i.e., cases that did not settle or have a voluntary dismissal), the defendants win around 80 percent of these cases if we examine the docket sheets or Westlaw/Lexis visible cases, but just 61 percent of the cases visible in published sources.

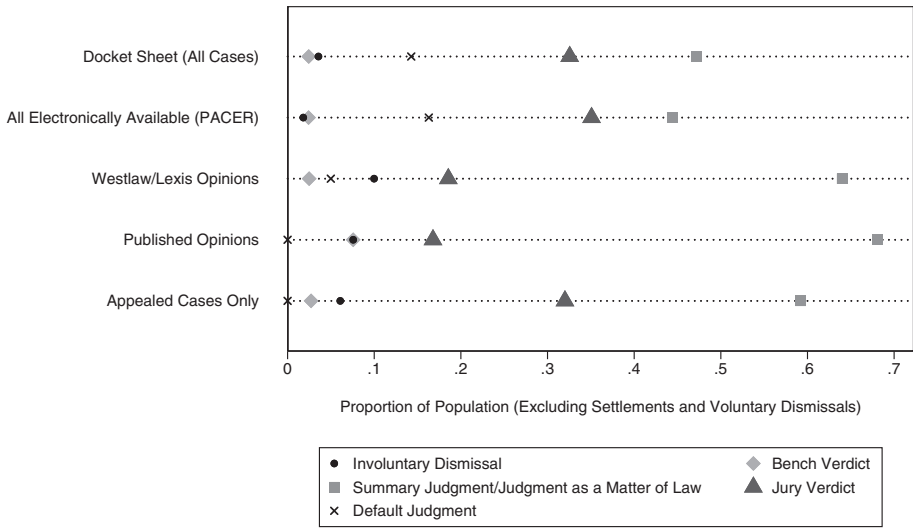
What about variation in case termination method based on data source? We have discussed this already as it relates to settlement, but what about the other ways that cases terminate? Figure 8 plots the distribution of termination methods for cases resolving by court action (i.e., excluding settlements or voluntary dismissals) by data source. For all cases (docket sheets), granted motions for summary judgment and for judgment as a matter of law make up around 47 percent of the cases, and jury verdicts account for nearly 33 percent of the observations.

However, the story changes considerably in published cases and Westlaw/Lexis visible cases, even after excluding settled lawsuits. Granted motions for summary judgment or judgment as a matter of law jump to around 64 percent of the cases. Jury verdicts

---

<sup>20</sup>Coding for case outcome type comes from the *EEOC Litigation Project* data variable Final Resolution Type.

Figure 8: Observed case termination methods by data source.



make up just under 19 percent of the cases with some presence, published or unpublished, on Westlaw or Lexis. The distribution looks very similar for just the published cases. Judge opinion writing patterns (with more opinions in granted summary judgment motions cases and fewer in jury trials) once again likely explain the varying proportions of disposition types observed in different data sources.

For district court cases with an eventual appeal, the case termination methods fall in the middle ground. Granted motions for summary judgment or judgment as a matter of law account for about 59 percent of the appealed cases, and jury verdicts make up over 32 percent.

To round out the discussion of case outcomes, Table 1 reports details on the monetary and injunctive relief obtained, across visibility level. We calculate the monetary and injunctive relief numbers only for those cases where there is a settlement or EEOC/

Table 1: Monetary and Injunctive Relief Obtained by Data Source

	Average Monetary Relief	Injunctive Relief Obtained?		
		Yes	No	Unknown
Docket sheet (all cases)	\$221,868	0.68	0.02	0.30
Electronically available on PACER	\$175,290	0.88	0.02	0.10
Westlaw & Lexis cases	\$574,708	0.78	0.04	0.18
Published opinions	\$284,903	0.71	0.09	0.21
Appealed cases only	\$431,529	0.56	0.27	0.17

complainant case-level victory—that is, only those outcomes where some relief is most plausible. As with the other facets of case outcomes, there is substantial variation here across visibility level. The table's descriptive statistics reveal that the source used to develop a study case set can significantly affect the conclusions reached about the resulting relief. In our dataset, the Westlaw/Lexis cases' average monetary relief was over \$350,000 per case more than the average for the full PACER docket sheet case set. Similarly, Westlaw/Lexis cases have a far higher rate of awarded injunctive relief than the full population.

### *C. Conclusions About Judicial Behavior*

It is clear from the analyses above that choice of data source substantially affects both case and motion outcome estimates. Does it also affect modeling of district court judicial decision making? In appellate court research, there is strong evidence that judicial characteristics like political preferences, gender, and race affect behavior, at least in certain case areas (Boyd et al. 2010; Haire & Moyer 2014; Hettinger et al. 2006; Kastellec 2013; Scherer 2004). There is also evidence that the strength of the effect varies depending on publication status of opinions. (e.g., Carlson et al. 2020). For district courts, however, research results examining judge identity are much more mixed (compare, e.g., Rowland & Carp 1996 and Boyd 2016 to Ashenfelter et al. 1995). We suspect that differing sources of data used by scholars help explain the hodgepodge outcomes in district court judicial behavior research, with certain datasets overrepresenting cases—for example, politically or socially charged cases—where judicial characteristics like sex, race, or party of the appointing president, are more likely to matter.

To assess whether data source can skew research about the effects of judicial characteristics on district court decisions, Table 2 provides the results of multivariate regression analyses of district judge decisions on dispositive motions in our EEOC litigation data. By focusing on just dispositive motions rather than a broader array of motions, we center on the decisions most likely to be affected by judicial preferences or characteristics.

The dependent variable in both models is who wins the motion; plaintiff (either the EEOC or the complainant) wins are coded as 1; defendant wins are coded as 0.<sup>21</sup> The regression models include three main independent variables: judge party (based on the party of the appointing president; Republican = 1, Democrat = 0),<sup>22</sup> judge sex (female = 1, male = 0), and judge race (black = 1, white = 0; other races excluded). We control for the

---

<sup>21</sup>Partial motion victories are excluded from the analysis because victory cannot be assigned.

<sup>22</sup>Because our judge party variable is based on the party of the appointing president, our analysis excludes Article I magistrate judges, who serve and hear cases in the federal district courts alongside Article III district judges. Magistrate judges are appointed by the sitting district judges of their courts. Although some prior district court research uses a proxy for magistrate judges based on the party affiliations of those sitting district judges at the time of the magistrate judge's appointment to the district court (Boyd & Hoffman 2010; Boyd & Sievert 2013), we worry that the imprecision of that measure could be problematic in the context of this study.

Table 2: Logistic Regression of Whether the EEOC Wins a Dispositive Motion

	<i>Motions Coded from PACER Docket Sheets</i>		<i>Motions Coded from Westlaw/Lexis</i>	
	<i>Logistic Regression</i>	<i>Substantive Effect</i>	<i>Logistic Regression</i>	<i>Substantive Effect</i>
Republican-appointed judge	-0.099 (0.19)	NS	-0.644* (0.30)	-0.11
Female judge	0.532* (0.26)	+0.09	-0.022 (0.43)	NS
Black judge	0.399 (0.30)	NS	1.075* (0.44)	+0.17
Defendant movant	1.936* (0.25)	+0.45	1.724* (0.42)	+0.41
More than one complainant	0.223 (0.18)	NS	0.890* (0.31)	+0.21
Number of docket entries	0.000 (0.00)	NS	0.004 (0.00)	NS
Constant	-1.104* (0.28)		-1.523* (0.47)	
Observations	737		274	

\* $p < 0.05$ .

NOTES: Substantive effect is the difference in the predicted probability of an EEOC win when the independent variable's value is changed. NS = difference between values is not statistically significant.

identity of the movant (defendant = 1, plaintiff = 0). We also account for case complexity with two proxy variables: (1) whether the case has a single complainant or more than one complainant and (2) the number of entries on the case's docket sheet.

Table 2 reports the results of two identical logistic regression models for the two case groups of interest here: the full PACER docket sheet population and the subset visible via Lexis/Westlaw opinion search.<sup>23</sup> Where substantive effects are statistically significant, we also report them. These are the differences in the predicted probability of a pro-plaintiff outcome based on a change in the relevant independent variable.

The differences across the two datasets are noteworthy. Most important are the starkly distinct results for the judicial party variable. Based on just the Lexis/Westlaw visible motions—that is, a sample similar to that in many prior studies—judge party matters in judicial decision making about employment discrimination dispositive motions. The Lexis/Westlaw results indicate that Republican-appointed federal district judges are about 0.11 less likely than are Democrat-appointed district judges to rule for the plaintiff's position. However, using the full PACER data, we observe no statistical difference in

<sup>23</sup>Because of a too small number of observations, we cannot model just the published data, the appeals-only data, or the Carp-Manning data. Accordingly, we simplify the analysis by focusing exclusively on PACER data and the combined published and unpublished data available to researchers searching Lexis and Westlaw. Our regression results for the Westlaw/Lexis model remain largely unchanged when we focus on just dispositive motions decided in 2006, 2007, and early 2008. Although this robustness check is based on a small number of observations ( $N = 114$ ), it does provide some evidence that the increase in the inclusiveness of Westlaw and Lexis data over time that we observe in Figure 5b does not affect our basic observation that conclusions about judicial behavior are likely to change depending on the data sources used for analysis.

the voting behavior of Republican and Democrat judges on these dispositive motions. In other words, the results indicate that party of appointing president predicts district judge behavior on dispositive motions but only in a small, nonrandom slice of decisions. It may be that the presence of party effects in this set of cases—Lexis- and Westlaw-available opinions—is where it matters, precisely because these cases are more visible and are therefore likely to influence the future course of the law. Nevertheless, the fact that these effects do not appear in the full dataset suggests caution in extrapolating the results of studies limited to these sources to the full range of district court judge behavior. Given the number of studies examining and finding that party and ideology affect judicial behavior, the differing results for the two models reported in Table 2 are noteworthy even if perhaps not wholly surprising.

We also see differences in judge gender and race results. Modeling the effect of the *Female Judge* variable on the likelihood of a pro-plaintiff vote using just Lexis/Westlaw data, there is no significant difference in how male and female judges rule on dispositive motions. However, there is a difference in how black judges rule; they are 0.16 more likely to rule in favor of the plaintiff's position. Switching to the full PACER data shifts both results. Using full data, female judges are about 0.09 more likely than male judges to support the EEOC or private plaintiffs' positions in dispositive motions. For black judges, the differential disappears. While further investigation is needed to fully understand why this variation exists and whether it holds for different types of disputes and periods of time, the important broader point is that findings about what predicts judicial behavior depend on the data source.

## V. DISCUSSION

Because district court litigation is complex, dynamic, and multistaged, and because of the low rates at which Westlaw and Lexis include opinions and orders, the research design strategies used for appellate court research—that is, relying on published (and perhaps unpublished) opinions—are often inappropriate.

We set out to explore the ramifications of data source choice in empirical research about federal district courts. Our analyses demonstrate great variation in the visibility of cases and case motions based on the source researchers consult, with traditional published reporters providing just a small slice of cases and motions, while databases like Westlaw and Lexis allow access to some but not all unpublished opinions in addition to published opinions. As our district-by-district examination reveals, Westlaw/Lexis searches yield almost no opinions at all for some districts (including E.D. Michigan, N.D. Georgia, W.D. Texas, W.D. Washington, and about 20 other districts). Researchers using traditional search methods would unknowingly virtually eliminate these districts from their studies even though many relevant cases (and motions) might have been brought in these district courts. Our findings buttress a point made by McCuskey: in empirical study of federal district court cases and decisions, “it matters *what* you search for *and where* you look” (2016:522).



Differences in visibility can affect, often substantially, the reliability of estimates of motion outcomes, case resolutions, and judicial behavior. Importantly, settlements are much less frequent in non-PACER sources. In our data, Westlaw and Lexis over-report defendant case victories and under-report successful EEOC motions. They also over-report the amount of monetary relief and the granting of injunctive relief. Finally, our data demonstrate that conclusions about the influence of judge race, sex, and party of the appointing president on judicial decision making depend greatly on the source of data used.

The nature and size of the differences one observes between district court data sources are likely to vary based on the issue areas studied. Because employment discrimination litigation is distinct from many other types of civil litigation and the EEOC is not a typical plaintiff, the direction and magnitude of the differences we observe may not apply in other areas of litigation. Nevertheless, regardless of issue area, a nonrandom process determines which cases are selected into sources like the *Federal Supplement*, Westlaw, and Lexis. As a result, our broader point holds regardless of issue area: there is significant variation in the types of cases and motion activity that are visible across data sources, and those differences will often lead to differing conclusions about those cases and about judicial behavior.

Our results counsel adoption of a docket-based approach to district court data collection for many empirical district court studies. This type of research design—what Hoffman et al. call “docketology”—provides “a way to study litigation that informs readers about the real content of legal rights and obligations ... It enables us to quantify the realities of litigation” (2007:731). It was once true that “published opinions are all we have” (Eisenberg & Johnson 1991:1164) in district court research, but no longer. Docket-sheet-based district court research may not be as easy as collecting a case list from Westlaw or Lexis using traditional search terms or using a pre-collected appellate court dataset, but it has become reasonably easy. Not only has the “development of the electronic filing system in federal courts and the current availability of dockets and litigation documents through PACER ... significantly reduced the obstacles to obtaining detailed information about the work of the district courts” (Kim et al. 2009:112), but in addition, other docket-sheet-based sources are emerging, including Bloomberg Law’s database, govinfo from the Government Publishing Office (GPO) (formerly FDSys), RECAP,<sup>24</sup> and more docket sheet content on Westlaw and Lexis.<sup>25</sup> Our mapping of the differences between the opinions and cases visible in different data sources should give researchers pause. If a district court study relies only on Westlaw or Lexis opinions, whether to describe

---

<sup>24</sup>RECAP is “a free digital archive of federal district court and bankruptcy case documents developed in 2008 by the Center for Information Technology Policy at Princeton University. RECAP’s repository is sourced through Internet users of PACER (Public Access to Court Electronic Records), the federal judiciary’s pay service for accessing electronic court records” (Boyd et al. 2013:257).

<sup>25</sup>McCuskey (2016) provides helpful details on sources like Bloomberg Law and GPO and their coverage of PACER docket sheet data.

outcomes or to study judicial decisions, the risks of unrepresentativeness should be taken into account when describing findings or drawing conclusions.

Yes, “[d]ata are good” (Clermont & Eisenberg 2002:154) but more data are better. For empirical research about federal district courts, more comprehensive data are now electronically available, and this study shows that leveraging those data can make a difference in our understanding of the work of these important courts.

## REFERENCES

- Abraham, Henry J. (1998) *The Judicial Process*. 7th ed. New York: Oxford Univ. Press.
- Administrative Office of the U.S. Courts. (2018) “Statistical Tables for the Federal Judiciary.” Available at <<https://www.uscourts.gov/statistics-reports/statistical-tables-federal-judiciary-december-2018>>.
- Ashenfelter, Orley, Theodore Eisenberg, & Stewart J. Schwab (1995) “Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes,” 24(2) *J. of Legal Studies* 257.
- Banks, Christopher P., & Steven Tauber (2014) “U.S. District Court Decision-Making in USA PATRIOT Act Cases After September 11,” 35(2) *Justice System J* 139.
- Boyd, Christina L. (2015) “Opinion Writing in the Federal District Courts,” 36(3) *Justice System J* 254.
- (2016) “Representation on the Courts? The Effects of Trial Judges’ Sex and Race,” 69(4) *Political Research Q* 788.
- Boyd, Christina L., Epstein Lee, & Andrew D. Martin (2010) “Untangling the Causal Effects of Sex on Judging,” 54(2) *American J. of Political Science* 389.
- Boyd, Christina L., & David A. Hoffman (2010) “Disputing Limited Liability,” 104(3) *Northwestern U. Law Rev* 853.
- Boyd, Christina L., David A. Hoffman, Zoran Obradovic, & Kosta Ristovski (2013) “Building a Taxonomy of Litigation: Clusters of Causes of Action in Federal Complaints,” 10(2) *J. of Empirical Legal Studies* 253.
- Boyd, Christina L., & Jacqueline M. Sievert (2013) “Unaccountable Justice? The Decision Making of Magistrate Judges in the Federal District Courts,” 34(3) *Justice System J* 363.
- Carlson, Keith, Michael A. Livermore, & Daniel N. Rockmore (2020) “The Problem of Data Bias in the Pool of Published U.S. Appellate Court Opinions,” 17(2) *J. of Empirical Legal Studies* 224.
- Carp, Robert A. & Kenneth L. Manning. (2016). “U.S. District Court Database.” 2016 version n=110977. Available at <<http://districtcourtdatabase.org>>
- Carp, Robert A., & Russell Wheeler (1972) “Sink or Swim: The Socialization of a Federal District Judge,” 21(1) *J. of Public Law* 359.
- Clermont, Kevin M., & Theodore Eisenberg (2002) “Litigation Realities,” 88 *Cornell Law Rev* 119.
- Clermont, Kevin M., Theodore Eisenberg, & Stewart J. Schwab (2003) “How Employment-Discrimination Plaintiffs Fare in the Federal Courts of Appeals,” 7(2) *Employee Rights & Employment Policy J* 547.
- Eisenberg, Theodore (2004) “Appeal Rates and Outcomes in Tried and Nontried Cases: Further Exploration of Anti-Plaintiff Appellate Outcomes,” 1(3) *J. of Empirical Legal Studies* 659.
- Eisenberg, Theodore, & Sheri Lynn Johnson (1991) “Effects of Intent: Do We Know How Legal Standards Work,” 76(6) *Cornell Law Rev* 1151.
- Eisenberg, Theodore, & Margo Schlanger (2005) “The Reliability of the Administrative Office of the U.S. Courts Database: An Initial Empirical Analysis,” 78(5) *Notre Dame Law Rev* 1455.
- Haire, Susan B., & Laura P. Moyer (2014) *Diversity Matters: Judicial Policy Making in the U.S. Courts of Appeals*. Charlottesville, VA: Univ. of Virginia Press.
- Hettinger, Virginia A., Stefanie A. Lindquist, & Wendy L. Martinek (2006) *Judging on a Collegial Court*. Charlottesville, VA: Univ. of Virginia Press.
- Hoffman, David A., Alan Izenman, & Jeffrey Lidicker (2007) “Docketology, District Courts, and Doctrine,” 85(4) *Washington Univ. Law Rev* 681.

- Judicial Business. (2018). "Status of Article III Judgeships." *United States Courts*. Available at <<https://www.uscourts.gov/statistics-reports/status-article-iii-judgeships-judicial-business-2018>>
- Kastellec, Jonathan P. (2013) "Racial Diversity and Judicial Influence on Appellate Courts," 57(1) *American J. of Political Science* 167.
- Keele, M. Denise, Robert W. Malmsheimer, Donald W. Floyd, & Lianjun Zhang (2009) "An Analysis of Ideological Effects in Published Versus Unpublished Judicial Opinions," 6(1) *J. of Empirical Legal Studies* 213.
- Kenny, Amanda, Product Developer, Thompson Reuters. Email to Margo Schlanger (June 9, 2020), on file with authors.
- Kim, Pauline T., Margo Schlanger, Christina L. Boyd, & Andrew D. Martin (2009) "How Should We Study District Judge Decision-Making," 29 *Washington University Journal of Law and Policy* 83.
- Kim, Pauline T., Margo Schlanger, & Andrew D. Martin (2013) "EEOC Litigation Project," Washington University in St. Louis. Available at <http://eoclitigation.wustl.edu/>
- Levin, Hillel Y. (2008) "Making the Law: Unpublication in the District Courts," 53(5) *Villanova Law Rev.* 973.
- McCuskey, Elizabeth Y. (2016) "Submerged Precedent," 16 *Nevada Law J.* 515.
- Morriss, Andrew P., Michael Heise, & Gregory Sisk (2005) "Signaling and Precedent in Federal District Court Opinions," 13 *Supreme Court Economic Rev* 63.
- Olson, Susan M. (1992) "Studying Federal District Courts Through Published Cases: A Research Note," 15(3) *Justice System J.* 782.
- Randazzo, Kirk A. (2008) "Strategic Anticipation and the Hierarchy of Justice in the U.S. District Courts," 36(5) *American Politics Research* 669.
- Ringquist, Evan J., & Craig E. Emmert (1999) "Judicial Policymaking in Published and Unpublished Decisions: The Case of Environmental Civil Litigation," 52(1) *Political Research Q* 7.
- Rowland, C. K., & Robert A. Carp (1996) *Politics and Judgment in Federal District Courts*. Lawrence, KS: Univ. Press of Kansas.
- Rowland, C. K., Robert A. Carp, & Ronald A. Stidham (1984) "Judges' Policy Choices and the Value Basis of Judicial Appointments: A Comparison of Support for Criminal Defendants Among Nixon, Johnson, and Kennedy Appointees to the Federal District Courts," 46(3) *J. of Politics* 886.
- Scherer, Nancy (2004) "Blacks on the Bench," 119(4) *Political Science Q* 655.
- Schlanger, Margo, & Pauline T. Kim (2014) "The Equal Employment Opportunity Commission and Structural Reform of the American Workplace," 91 *Washington Univ. Law Rev.* 1519.
- Schlanger, Margo, & Denise Lieberman (2006) "Using Court Records for Research, Teaching, and Policymaking: The Civil Rights Litigation Clearinghouse," 75 *U.M.K.C. Law Rev* 153.
- Schultz, Vicki, & Stephen Petterson (1992) "Race, Gender, Work, and Choice: An Empirical Study of the Lack of Interest Defense in Title VII Cases Challenging Job Segregation," 59 *Univ. of Chicago Law Rev.* 1073.
- Segal, Jennifer A. (2000) "Representative Decision Making on the Federal Bench: Clinton's District Court Appointees," 53(1) *Political Research Q* 137.
- Sen, Maya (2015) "Is Justice Really Blind? Race and Reversal in US Courts," 44(1) *J. of Legal Studies* 187.
- Siegelman, Peter, & John J. Donohue III (1990) "Studying the Iceberg from its Tip: A Comparison of Published and Unpublished Employment Discrimination Cases," 24(5) *Law & Society Rev.* 1133.
- Songer, Donald R. (1988) "Nonpublication in the United States District Courts: Official Criteria Versus Inferences from Appellate Review," 27(1) *Justice System J.* 28.
- Spaeth, Harold J., Lee Epstein, Andrew D. Martin, Jeffrey A. Segal, Theodore J. Ruger, & Sara C. Benesh. (2019) "Supreme Court Database, Version 2019 Release 01." Available at <<http://Supremecourtdatabase.org>>
- Swenson, Karen (2004) "Federal District Court Judges and the Decision to Publish," 25(2) *Justice System J.* 121.
- Taha, Ahmed E. (2004) "Publish or Paris? Evidence of How Judges Allocate Their Time," 6 *American Law & Economics Rev* 1.

- U.S. Courts. (2013). "25 Years Later, PACER, Electronic Filing Continue to Change Courts." *United States Courts*. December 9. Available at <<https://www.uscourts.gov/news/2013/12/09/25-years-later-pacer-electronic-filing-continue-change-courts>>
- Walker, Thomas G., & Deborah J. Barrow (1985) "The Diversification of the Federal Bench: Policy and Process Ramifications," 47(2) *J. of Politics* 596.
- Winkler, Adam (2006) "Fatal in Theory and Strict in Fact: An Empirical Analysis of Strict Scrutiny in the Federal Courts," 59 *Vanderbilt Law Rev* 793.