Estimating the Effect of Leisure on Judicial Performance

Tom S. Clark, Benjamin G. Engst, and Jeffrey K. Staton

ABSTRACT

Past research suggests that natural preferences for leisure influence the ways in which federal judges carry out their work. We consider the extent to which incentives for leisure reduce the speed with which judges work and the quality of their output. We take advantage of a natural experiment caused by an annual sporting event that creates differential distractions across judges. Using a difference-in-differences design, among federal courts of appeals judges we show that a judge's alma mater's participation in the National Collegiate Athletic Association Men's Basketball Tournament both slows the rate at which opinions are drafted and ultimately undermines the opinions' quality, even accounting for the additional time judges spend writing them. The findings suggest that incentives for leisure influence important normative concerns for swift and high-quality justice.

1. INTRODUCTION

One of the most significant lessons of the social science of law and courts during the 20th century might be summarized as follows: judges are people too. A decades-long debate about the role of ideology in judicial decision-making was primarily about whether, how much, and under what conditions judges reference their own worldviews when exercising their authority (see, for example, Segal and Spaeth 2002; Epstein and Knight 1998; Maltzman, Spriggs, and Wahlbeck 2000). Though ideological models provide useful summaries in many contexts, more broadly, judges have come to be understood as motivated by largely the same set

TOM S. CLARK is the Charles Howard Candler Professor of Political Science at Emory University. BENJAMIN G. ENGST is a Research Fellow at the Collaborative Research Center SFB 884 at the University of Mannheim. JEFFREY K. STATON is a Professor of Political Science at Emory University. We thank Chuck Cameron, Adam Glynn, Thomas Gschwend, Kosuke Imai, John Kastellec, and Royce de R. Barondes for helpful suggestions and guidance.

[*Journal of Legal Studies*, vol. 47 (June 2018)]
 0047-2530/2018/4702-0022$10.00

of concerns that motivate individuals in the workplace generally, including reputations, salaries, career advancement, and, of course, personal satisfaction (Epstein, Landes, and Posner 2013; Bainbridge and Gulati 2002; Ash and MacLeod 2015). Critically, like all workers, judges must balance these typical career-related interests against a preference for leisure (see, for example, Posner 1993). In this paper, we ask whether an increase in the marginal utility of time spent on leisure activities causes a decrease not only in the rate at which judges work but also in the quality of work begun during periods of distraction.

That judges, or anyone for that matter, would put in fewer hours on projects begun in periods when they are naturally drawn to nonwork activity should not be surprising or necessarily alarming. There are two reasons why understanding the effects of incentives for leisure (hereafter, leisure incentives) is important. The first is that leisure incentives might produce delays in the administration of justice. Although it is certainly possible to devote more hours to a project tomorrow when fewer hours are devoted to it today to ensure that justice is carried out in a timely fashion, the constant flow of cases in most court systems makes it possible that leisure incentives can result in delays that would be considered inappropriate. A second, arguably more important, reason is that leisure incentives may undermine the quality of justice. As we discuss below, expectations about the effects of leisure on work rate would seem fairly straightforward: judges should devote less time to managing cases when they are distracted by personal concerns. The key question with respect to the rate at which judges work is whether leisure incentives ultimately produce delays in the resolution of cases. With regard to the effect of leisure on the quality of opinions, given extant theoretical claims, our expectations are less clear. For example, judges facing a near-constant inflow of cases might devote less time to projects when leisure is particularly attractive and fail to fully offset this reduction in effort at a later date, when their interest in leisure is less pressing. In this sense, leisure incentives would undermine quality. On the other hand, when faced with an increased desire for leisure, judges might merely opt to slow down their work, simply taking longer to resolve cases, working harder to catch up when they are not otherwise distracted, and maintaining quality at the same level. Leisure would surely result in delays, but at the cost of quality control. And finally, by slowing the rate of work, judges might even produce more high-quality work than they would have produced at a higher work rate, given that they would be taking longer to consider alternative options.

Each of these possible strategies implies different empirical patterns in the relationship between an incentive for leisure and the judicial work product, but they also highlight the importance of some very basic empirical questions. Do leisure incentives produce delays? Do leisure incentives undermine or enhance the quality of judicial work? The importance of these questions follows from the normative questions that they invite. If we observe a delay that we believe is caused by leisure, should we be alarmed? It is hard to answer that question if we do not know whether leisure also influences quality, as we might be willing to suffer some delays to sustain quality. Our tolerance for delay may be different if we also believe that leisure incentives undermine the quality of judges' resolutions of cases. Our paper provides an empirical basis for answering these questions.

Our study offers several contributions to research on the role of leisure in judicial behavior. Prior studies suggest a number of ways in which the preference for leisure influences judicial behavior (for example, Cohen 1991; Klein and Hume 2003; Posner 1993; Bainbridge and Gulati 2002). A significant challenge to evaluating expectations related to leisure incentives is that these incentives are fundamentally latent. For this reason, scholars have largely adopted caseload as a proxy (Epstein, Landes, and Posner 2013). Heavy caseloads raise the possibility that judges may not be able to devote as much time to family, friends, or pastimes as would be desired. Judges who are busier at work confront particularly strong incentives to find ways to protect their leisure time. Conceptualized in this way, Cohen (1991) finds that US federal district court judges with heavy caseloads were more likely to find unconstitutional the sentencing guidelines developed by the US Sentencing Commission, ostensibly because they believed that the guidelines would undermine plea bargaining and result in a greater number of trials. Cohen (1992) also finds that busy district court judges issued increasingly punitive sentences in antitrust criminal cases to defendants who pleaded not guilty, which suggests an effort to incentivize plea bargaining as a way to reduce work in time-consuming trials. Similarly, Epstein, Landes, and Posner (2013) find that judges with higher caseloads were more likely to take advantage of the Supreme Court's decision in *Ashcroft v. Iqbal* (556 U.S. 662, 129 S. Ct. 1937 [2009]), which made it easier to dismiss discrimination lawsuits aimed at high-level government officials.

Our first contribution is a focus on quality. Promoting more plea bargaining or dismissing more lawsuits will surely save judges' time at trial, but whether judges should promote more or less plea bargaining or admit more lawsuits to produce high-quality justice is unclear. Our goal is to develop a method for evaluating not only the way that leisure incentives influence the rate at which judges work but also the quality of their work. Second, we focus on the behavior of judges on the US Courts of Appeals. The majority of prior research examines judges on district courts, where, as many scholars argue, the incentives to protect leisure time are heightened. Still, relative to the Supreme Court, leisure incentives may still play a role at least at the margin for appellate court judges (Drahozal 1998). Furthermore, because appellate court judges set precedent for relatively large jurisdictions, it is particularly important to know if and how their preferences for leisure influence quality.

Third, existing measurement approaches, which rely on differences in the caseloads across judicial districts, identify the effects of the context in which judges operate. They identify the effect of increasing incentives to protect leisure time at the judge level under the assumption that, on average, judges in busier districts have greater pressures on their leisure. While we believe that this is a reasonable assumption, our research design attempts to get a judge-specific measure of leisure incentives.

Fourth, the credibility of the causal claims in existing studies depends on the assumption that differences in caseloads across districts reflect exogenous sources of variation in their analyses of judges' behavior. While we do not take a strong position on whether this is a valid assumption, it is worth considering that different types of people, with fundamental differences in their preferences for how to trade off work and leisure, might be attracted to jobs in different districts. Two possibilities seem plausible. On the one hand, busy districts might be particularly attractive to candidates with weak preferences for leisure, and, if so, existing studies might underestimate the effect of caseload. On the other hand, busy districts might be more likely to attract candidates who are more willing to compromise legal principles to protect their leisure time, in which case existing studies would overestimate the effect of caseload. Our approach, which relies on a natural experiment analyzed using difference-in-differences estimation, rules out by design these forms of sample selection bias, which will result in more credible empirical claims.[1]

1. As we develop below, this is immediately obvious if variation in caseload is largely cross-sectional and invariant over time across districts; however, our design is also robust to variation in time.

Our research design relies on a natural experiment introduced by the annual National Collegiate Athletic Association (NCAA) Men's Basketball Tournament, one of the most popular, attention-getting sporting tournaments in the United States, and one that creates opportunities for leisure even among people who do not pay much attention to basketball. The tournament is credited with substantial drops in productivity in the for-profit private sector (Challenger, Grey, & Christmas 2015), so there is reason to believe that, if judges behave in many of the same ways as workers generally, the tournament will serve as a useful source of variation in our context. We collect data on the teams participating in the NCAA tournament each year and match judges to the institutions from which they received their undergraduate degrees. Using a difference-in-differences design, we examine the effect of a judge's alma mater participating in the tournament and show that the effect of having her team in the tournament is to delay the time it takes her to write opinions for cases heard during the tournament and to lower the quality of the opinion.

Although we find that the tournament represents a useful opportunity to evaluate the effects of leisure, we are not interested in learning about the effects of a basketball tournament per se. The tournament represents a window through which we can evaluate leisure incentives. We certainly believe that the effects are interesting, but their real value derives from their ability to suggest the effects of other sources of leisure, some of which will impose stronger and more frequent forces on judges.

In Section 2, we outline the theoretical framework underlying our expectations regarding judicial leisure and opinion writing. Sections 3 and 4 describe our empirical strategy and evaluate the effect of a judge's alma mater's participating in the tournament on the amount of time it takes her to write an opinion. Section 5 evaluates the effect of the team's participation on the quality of the opinions the judge writes. Finally, Section 6 offers concluding remarks about the consequences of judicial insulation from accountability in light of the evidence and extant theory about judicial leisure.

2. THEORETICAL BACKGROUND

We assume that judges care about their professional reputations among a potentially diverse group of observers (Baum 2009; Miceli and Coşgel 1994; Posner 1993). When processing and resolving cases, judges are evaluated on a number of elements of their decision-making. Among those elements, two are of particular importance: the quality of their decisions and the speed or efficiency with which they resolve cases (Garoupa and Ginsburg 2009; Miller 2004). The job of writing a high-quality opinion takes time (Choi, Gulati, and Posner 2012; Gulati and McCauliff 1998), and, given that reputations are related to the quality of opinions, time is a precious resource for federal judges. Unfortunately, appellate court judges experience a more or less constant flow of cases that must ultimately be resolved, and there is a general concern for avoiding too large a backlog (Reinhardt 1993). Crucially, as Ash and MacLeod (2015) point out, a judge's incentive to pursue leisure interacts with her ability to manage which cases she hears and the flow of judicial work she encounters. Others have similarly shown that judges are susceptible to nonjudicial distractions that can prime their perspectives and even shape how they vote (for example, Berdejó and Chen 2017). One of the primary goals of our empirical analysis below is to exploit the NCAA tournament as a shock to judges' leisure incentives.

According to the Administrative Office of the US Court's Federal Judicial Caseload Statistics report, as of 2014, roughly 300–500 cases per judge were filed in the US courts of appeals, which is roughly the rate at which the judges terminated cases. However, there was also a backlog in each circuit of roughly 200 cases per judge.[2] A great deal of public, political, and scholarly attention has been paid to the rate of cases' resolution and backlogs on the courts. For example, the Judicial Conference routinely requests additional judgeships to deal with caseload problems (Judicial Conference 2015). The Administrative Office of the US Courts is required to provide an annual report of statistical information, which always includes an accounting of caseloads and backlogs, identifying particular circuits that are lagging behind (28 U.S.C. sec. 604[a][2]). This information is commonly reported by media outlets (see Palazzolo 2015), and particular judges are sometimes shamed (see Carvajal 1995). And critically, increasingly caseloads have potential implications for judicial quality. Consider Judge Stephen Reinhardt's plea to Congress, which develops the consequences of an increased caseload.

> Simply put, our federal court system is too small for the job. We seem to assume that judges can perform the same quality of work regardless of the number of cases they are assigned. That simply is not correct. Most of us are now working to maximum capacity. As a result, when our caseload increases, we inevitably

2. US Courts, Federal Judicial Caseload Statistics (https://www.uscourts.gov/statistics-reports/analysis-reports/federal-judicial-caseload-statistics).

pay less attention to individual cases. . . . Those who believe we are doing the same quality work that we did in the past are simply fooling themselves. We adopt more and more procedures for "expediting" cases, procedures that ensure that individual cases will get less attention. (Reinhardt 1993, p. 52)

On the surface it appears that the goals of writing high-quality opinions and ensuring that backlogs are kept at a reasonable level may be in tension. If judges confront an increase in leisure incentives, their work rate should drop. Similarly, we might expect a reduction in quality as well. The question is whether this logic is clear on a more careful consideration.

2.1. Judges as Nonprofit Workers

Posner (1993) claims that because judges (like all others) desire leisure, the rate and quality of their work will suffer when they are distracted. Crucial for Posner's claim is that judges are best thought of as analogous to nonprofit workers. Workers in the for-profit sector are commonly judged on the quality of their work, where quality is easily observable. For those workers, increased time devoted to leisure must be offset by spending increased time on work so that quality can be kept constant. A failure to adjust results in a less competitive product and potential loss of employment. At the very least, it means a loss in reputation. In contrast, Posner claims that the quality of work produced by federal judges is relatively unobservable. Judges are not paid by the number of cases they resolve, the quality of their decisions, or the number of hours they devote to resolving a dispute.[3] Of course, the judicial work product is not truly unobservable. The public, politicians, the media, and scholars all frequently comment on both judicial quality and efficiency. The crucial point, however, is that the quality of judicial work is relatively harder to discern than the amount of work judges do. Backlogs and the length of time cases take to reach conclusion are readily apparent to anyone with even a passing familiarity with the judiciary. Quality is far harder to measure.

If quality and work rate are not equally observable, increasing leisure

3. Of course, a market for law could, in principle, result in better quality. Early common-law courts were based on a model of judges being paid by the volume of cases they resolved. However, it is notable that such systems also tended to have multiple judicial systems, which created competition for business among judges and thereby undermined the incentives judges had to churn through cases without regard for the quality of their work (see, generally, Blatcher 1978).

incentives should decrease the work rate of a judge who is distracted by personal interests and the quality of judicial outputs she produces while distracted. This follows because since quality is less observable than backlog, to avoid a backlog, decreases in effort need not be offset by increases in effort at a later date (or dates). Instead, the level of effort can be kept constant, and quality can be sacrificed to avoid a backlog.

2.2. Alternative Arguments

There are several plausible alternative arguments. Here we consider three, which emphasize the selection mechanism for federal judges, the management of workflow, and the judge's staff.

2.2.1. Judicial Section. Judges are selected via a relatively robust searching process. They are recruited typically from among lawyers with considerable and laudable histories of work (see, for example, Epstein et al. 2005; Savchak et al. 2006). Through this process, it may be possible to select individuals with a strong taste for high-quality judicial work and thus completely offset the concerns for quality created by leisure incentives. It may be either that federal judges are not commonly distracted by leisure or that, if they are, they behave like for-profit workers whose product quality is observable; that is, they offset temporary reductions of effort in the present with temporary increases in the future. Indeed, recent research suggests that judges have an intrinsic desire to write high-quality opinions, even while constraints such as time pressures inhibit their ability to pursue that goal (Ash and MacLeod 2015). If this kind of selection process holds, the empirical implication is that while we may observe a decrease in work rate when leisure incentives increase, we should not observe a decrease in quality.

2.2.2. Managing Workflow. Judges might manage the trade-off between backlog and quality in one of two ways, which would suggest different empirical implications. First, consider the ability to smooth out reductions in effort over time. It may be possible that leisure undermines quality, but by smoothing it is difficult or impossible to observe it (see, for example, Clark and Strauss 2010). For example, suppose there is a judge with a constant flow of one incoming case per day. Now, suppose that when using the appropriate amount of effort, the judge can complete only one case per day. If on any given day she chooses to devote additional time to leisure, she will have a problem: she can either delay resolving the case and allocate some of the next day's time to the current case or apply suboptimal effort to resolving cases. If she delays resolving the case, she creates a backlog in her resolution of cases. Either this delay will perpetuate over time, because she will then have to push off future cases accordingly, or it will cause her to decrease the quality of future decisions she makes. If she decides to apply suboptimal quality to her work, either she can simply make a lower-quality decision on the current case or she can balance delays and quality, spreading out the reduction in quality over time, marginally decreasing the quality with which she decides the current and future cases. Assuming that all judges experience some form of personal distraction at some point in a year, if this alternative approach is taken, then it will be impossible to observe an effect on quality, as all judges will have similarly smoothed out the effects of their leisure incentives over the course of a given time period. Thus, again, whereas we might expect to observe a reduction in work rate associated with a particular source of leisure, this kind of quality management implies that it will be highly difficult to detect a decrease in quality.

A second possibility is that, by slowing down their work rate, leisure incentives give judges more time to contemplate solutions to legal problems. With an increased window of time to work on an opinion, the effort needed to write the same high-quality opinion may be less than it would have been with a smaller window. In this sense, there may be no trade-off between backlog and quality. For sure, judges would work less when distracted by leisure, but quality would not be harmed. Indeed, it might be improved.

2.2.3. The Role of Clerks. Federal courts of appeals judges do not work alone. They manage a staff typically consisting of a secretary, a team of clerks, and sometimes externs or volunteer law students. It is well documented that clerks play a number of important roles in the work that judges do, from the development of bench memos and the preparation of nonlegal materials like speeches to the drafting and editing of judicial opinions (Peppers, Giles, and Tainer-Parkins 2014). Indeed, 98 percent of courts of appeals judges in Peppers, Giles, and Tainer-Parkins (2014) report using clerks to conduct legal research, 88 percent use clerks to develop bench memos, and 95 percent use clerks to draft opinions.

To understand whether and how clerks might influence the effect of a judge's leisure incentives, it is useful to summarize how clerks are typically used. Gulati and Posner (2016) suggest that federal courts of appeals judges operate under one of three basic management models. One, which is increasingly rare, is that of the authoring judge. An authoring judge does what the model suggests: she writes her own opinions. The clerks of an authoring judge do not typically prepare bench memos; however, they provide important assistance to the judge in the form of legal research and editing.

The second and most common model is that of the editing judge. The clerks of editing judges, when assigned to particular cases, commonly prepare bench memos and attend oral arguments when possible. Judges meet with the clerks following the post-oral-argument conference, and the clerks are then typically expected to draft the initial opinion. The judge reviews the opinion and may ask for additional edits. Although the editing judge's clerk is involved in many aspects of the development of the opinion, it is important to stress that the judge guides the process. Gulati and Posner (2016, p. 484) write, "A distinction worth noting at this point between the editing judges and the authoring judge . . . is that the editing judges made clear to us that they specify outcomes to their clerks and then tell them to explain and justify that outcome in the opinion draft. Few clerks are bold enough to come to the judge and tell him that the arguments in favor of that outcome are simply not good enough and therefore the judge should change his vote."

A final model is that of the hierarchical or delegating judge. The delegating judge manages a staff led by a head clerk, sometimes one of the clerks whom the judge hires in a year but more often a permanent clerk. The head clerk is responsible for managing the other clerks, who largely relate to the judge through the head clerk. As is the case with the editing-judge model, delegating judges allow clerks to develop draft opinions—the judge primarily serves as an editor, again, with the additional layer of hierarchy provided by the head clerk.

If clerks are really in charge of the production of opinions, the judge's distractions due to leisure need not have any effect on her work rate or quality, unless her clerks' leisure interests happen to coincide commonly with hers. Of course, although clerks are important in the process by which judges produce opinions, it is hard to imagine a theory of the judicial hierarchy in which clerks are fully in charge. Judges are ultimately responsible for their opinions. That said, because clerks take on a good deal of the work involved in producing an opinion, typical management structures for judicial staff may attenuate the effect of a judge's increase in leisure incentives. If this is true, it will be more difficult to observe the effect of a judge's preferences for leisure than it would be absent the role of clerks.

A second implication of the role of clerks lies in the interpretation of the findings we report. How might we interpret findings showing that judges work at lower rates and produce lower-quality work when they are distracted by leisure? For one, it may be most appropriate to interpret the effects as dealing with entire chambers rather than judges. If judicial opinions are largely produced by teams, then we might attribute a negative effect of a judge's leisure on work rate to her team. Furthermore, the mechanisms connecting leisure to outputs are likely to differ subtly across management models. For example, the effect on an authoring judge will be more directly related to the judge's own writing, whereas the effect under the other two models will relate more to the judge's ability to edit well. When a judge is distracted, final products may more validly reflect the work of clerks, and that could account for a decrease in quality.

Finally, without carefully measuring the management style of every judge in our sample (we do not), the effects we estimate are best thought of as average effects across the three models. It is not clear ex ante which management model is likely to be more affected by distraction due to leisure. It is possible that the effects will be stronger on authoring judges because they control more aspects of the final product. Yet it is also possible that the writing process sharpens the attention of a judge in ways that editing does not; if that is true, the effects will be stronger among the editing and delegating judges.

2.3. Summary

As these features of the judicial process illustrate, it is not clear what to expect from the observation that judges occasionally face a heightened marginal benefit from leisure. We might expect judges to simply slow down their decision-making, accepting a backlog in order to maintain high-quality work. Alternatively, we might expect judges to sacrifice the quality of their opinions in order to keep pace with their constant flow of cases. Or we might expect a smoother trade-off between the two dimensions of their decision-making. Unfortunately, the theoretical ambiguity has not been allayed by powerful empirical strategies. Because the theoretical model turns on a difficult- or impossible-to-observe factor—the marginal utility associated with leisure—scholars have had to rely at best on crude proxies for judicial motivations to work, such as career interests. Our primary goal is to offer a powerful design to estimate the causal effect of a judge's leisure incentives.

3. ANALYZING THE EFFECTS OF JUDICIAL LEISURE

To study the effect of a marginal increase in the utility that judges place on leisure, we identify a salient feature of American culture that leads to distraction from the daily workflow by drawing a great deal of public attention but that affects workers differentially each year. These features make the event measurable compared with private events that are less frequently reported. In particular, we examine how judges on the US Courts of Appeals are affected by the annual NCAA Division I Men's Basketball Tournament. Importantly, given the nature of the event and the activities that commonly take place around it, the NCAA tournament is likely to influence not only individuals who typically enjoy watching college basketball (or even sports) but a large proportion of the US workforce.

The NCAA tournament takes place in late February and March of each year and ends no later than very early April. The number of teams competing in the tournament has varied over time, but between 1985 and 2000 the number was fixed at 64. (We limit our empirical analysis to this time period.) The tournament is played in a series of single-elimination rounds. There are several methods by which a team can qualify for participation in the tournament, and, once qualified, the teams are seeded into a bracket and divided into four regions. The tournament begins midday on a Thursday, and the first two rounds are played continuously through the following Sunday. The third round (the Sweet Sixteen) is played in the evenings of the tournament's second Thursday and Friday, with fourth-round games being played over the weekend. The tournament's semifinals and finals (the Final Four) are played Saturday of the third weekend and the subsequent Monday night.

The NCAA tournament is a very popular sporting event. The average rating for an NCAA tournament game is historically between a 6 and 7, which is comparable to the CBS Evening News. An estimated average of 28.3 million viewers were tuned in to watch the 2015 championship game between Duke University and the University of Wisconsin. Moreover, according to the NCAA, 350 million social impressions of the tournament were shared on Facebook and Twitter, and 17.8 million hours of live video were watched online in 2015 (NCAA 2015). This highlights the large societal impact of this sporting event. In addition to the tournament's games, it is common for workers to take part in betting pools in which individuals fill out brackets and success is a function of how well participants predict outcomes throughout the tournament.[4] For example,

4. In surveys conducted by CareerBuilder.com, during recent years between 15 and 20 percent of respondents report participating in such pools (CareerBuilder 2015).

in a widely publicized survey, Challenger, Grey, & Christmas (2015) estimate that 60 million American workers were expected to participate in a March Madness pool in 2015. Importantly, individuals who do not commonly follow basketball or even sports at all often participate in office pools.

The effect of the NCAA tournament on workplace productivity in the United States has been documented. Assuming that each of the individuals estimated to participate in a pool spends just 1 hour of work time following the NCAA tournament, the cost to employers would be over $1.9 billion (Challenger, Grey, & Christmas 2015). (Indeed, Challenger, Grey, & Christmas [2015] find that 56 percent of respondents indicated they would spend at least 1 hour of work time on the March Madness pool.) According to *Fortune,* people spent a collective 664 million hours watching television broadcasts of NCAA tournament games (Vandermey 2015). Similarly, Clotfelter (2012) finds that downloads of academic articles through JSTOR at university libraries drops sharply (about 6 percent) during the first week of the NCAA tournament. These estimates dovetail with those from other sporting megaevents. For example, Lozano (2011) finds that there is a considerable reduction in the hours individuals work during the World Cup. However, and critically, he finds that the effect is concentrated among salaried workers rather than hourly workers. The judges we study are, in many ways, even more immune to the professional pressures that differentiate salaried and hourly workers.

Crucially, attention to the NCAA tournament is almost surely a function of whether one's preferred team is participating in the tournament. When one's alma mater is selected to play in the tournament, the team receives considerable national media attention. Moreover, a team's success is linked to other teams' performance, which creates an incentive for fans to watch games involving competing teams. Consistent with this claim, Clotfelter (2012) also shows that the effects during later weeks in the tournament are most pronounced when the victor is a surprise.[5]

The tournament can influence individuals who are not watching games, commenting on games on social media, or participating in office pools. Because media coverage of colleges and universities is highly correlated with whether the institution is in the tournament, a person whose school makes it is likely to be exposed to much more information about her school than usual. In fact, depending on the school, she may see

5. Clotfelter (2012) takes this as evidence that individuals rationally anticipate their teams' participation and so adjust their work schedules accordingly. As a consequence, only those surprised by their team's success should have a noticeable decrease in productivity.

national coverage of her alma mater only during this period when her school is playing. Increases in the salience of her school can serve as reminders of past friendships, past mentors, and a variety of forms of nostalgia. It can encourage somewhat time-consuming searching on social networking sites, reconnections, and a variety of potential social activities ultimately unrelated to the tournament. Importantly, when schools advertise during games, they present information about the academic and social aspects of their institutions, largely setting aside athletic factors that sell themselves through a team's participation. This type of advertising is especially appealing to non–sports fans. Thus, even the casual fan or the non–sports fan who happens to tune in for a few minutes is reminded of broad features of her alma mater. In all of these ways, the tournament represents a particularly useful way of evaluating productivity under a form of personal distraction that can influence people with diverse preferences for leisure.

Our design relies on the exogeneity of a judge's alma mater with respect to its appearance at NCAA tournaments years after the judge has graduated. To estimate the causal effect of tournament-induced distraction on judges' performance, we use a difference-in-differences design. We also assess the causal effect of the tournament on the quality of the opinion a judge writes. We show that opinions written by judges with teams in the tournament are more negatively cited than other opinions. We again employ a difference-in-differences estimator to account for other factors driving that relationship. We now turn to our two empirical analyses.

4. THE EFFECT OF THE TOURNAMENT ON DELAYS IN DECISIONS

Our first analysis considers the effect of judges' leisure incentives on the speed with which they decide cases. We examine how interest in the NCAA tournament affects the amount of time the author of an opinion takes to issue the opinion after the case has been heard. Our expectation is that the NCAA tournament will differentially affect judges whose alma maters are participating in the tournament that year by increasing the time it takes for them to prepare and publish their opinions.

4.1. The Data

To test our expectation, we require data on the timing of judicial decisions, teams' participation in the NCAA tournament, the judges partici-

Table 1. Summary Statistics for Key Variables in All Analyses

	N	Mean	SD	Min	Max
Days until Publication	8,818	101.4	57.6	1	249
Tournament Author	8,818	.04		0	1
Any Judge in Tournament	8,818	.78		0	1
March Madness$_{t[i]}$	8,818	.20		0	1
Negative Citations	1,424	2.65	3.42	0	25
Negative Citations (without Distinguishing Citations)	1,424	.91	1.92	0	18
Positive Citations	1,437	77.56	109.80	0	875
Total Citations	1,429	57.98	94.59	0	973

pating in and authoring each opinion, and judges' alma maters. Table 1 summarizes the data.

4.1.1. Case Data. We collect the text of all decisions included in the *Federal Reporter,* from volume 797 of the second volume (F.2d) through volume 529 of the third volume (F.3d). This includes all decisions from 1985 through 2005. We obtain the texts from Bulk.Resource.org, which provides the text of all judicial opinions formatted in HTML. We limit our attention, though, to cases decided between 1985 and 2000, when the number of teams competing in the NCAA tournament was fixed at 64. We then extract from each HTML file the date of oral argument (if provided in the header for the opinion), the date of decision, the names of the judges hearing the case, and the case's citation.[6] In the HTML code, the line with the dates of argument and decision is tagged, which makes it easy to extract that text. We then write a regular expression to parse it into the date of oral argument and the date of decision. In many instances, oral argument is not given in a case or is not reported. In some instances, courts report the date a case was submitted to the court. In even rarer instances, the date a case was decided is not formatted in a standard way and so is not easily extracted from the HTML code. This process yields 9,309 cases for 1985–2000 for which we have the date of argument and the date of the decision. Our variable of interest, Time to Decision$_i$, is the number of days between oral argument and opinion in case i. When either piece of information is unavailable from the *Federal*

6. Commonly used databases, such as the US Appeals Courts Database (Songer 1999), do not contain the dates of oral argument.

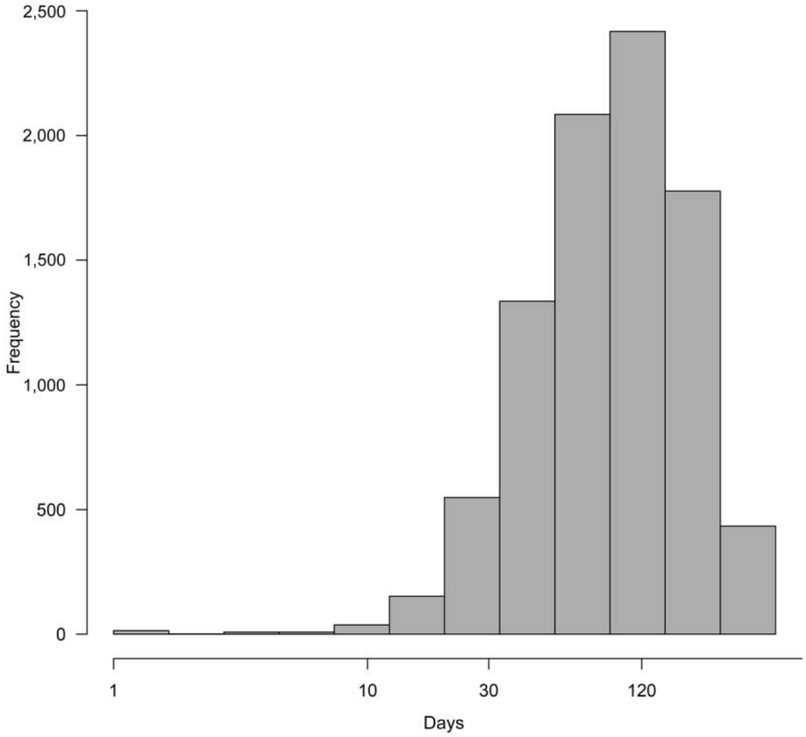

Figure 1. Distribution of days from oral argument to decision, 1985–2000. Figure shows the distribution of (logged) days from oral argument to the decision in a case. Note that the figure shows the natural log of the number of days, whereas the values reported on the *x*-axis are on the linear scale.

Reporter file, we code this variable as missing. Figure 1 shows the distribution of the logged number of days from oral argument to the decision.[7]

To identify the judges on the panel hearing a case, we also write a regular expression. Fortunately, the *Federal Reporter* has standardized how it identifies the judges hearing a case, by reporting "before" in the header, followed by the judges' names in capital letters. We search the headers of the decision for a line matching those parameters, focusing only on cases in which three judges hear the case, and extract the three names. When we cannot identify three judges hearing the case, or when a decision is decided by more than three judges (for example, en banc decisions), we

7. A total of 99 percent of the cases are decided in less than 300 days. For the cases that took longer than 300 days to decide, the average time to decision is 413 days. We exclude these cases because they are such extreme outliers that they could unduly influence our analysis.

exclude the case from our data. The case citation for each decision is also specially tagged in the HTML code, and we extract that citation from the raw code.

4.1.2. Judge Data. To identify the alma mater of each judge on a panel, we rely on the Federal Judicial Center's Biographical Directory of Article III Federal Judges, which is available for download and updated daily.[8] The database contains extensive biographical information for every Article III judge in US history, including the postsecondary schools each judge attended. From these data, we extract all judges serving after 1985 (the start of our data) and match judges by their last names, circuits, and dates of service (a judge's dates of service must include the date a case was decided). There are a small number of instances of judges having multiple last-name matches in the biographical data, which we handle manually, matching on first names when the last name does not uniquely identify the judge.

4.1.3. National Collegiate Athletic Association Tournament Data. To identify the teams playing in the NCAA Men's Basketball Tournament, we reference the *Wikipedia* pages for each year's tournament. The *Wikipedia* entries contain tables of teams playing in each regional division of the tournament. Using the XML package (Lang 2013) for R, we extract the tables and create a database of the 64 teams playing in each tournament.

We are able to identify all of the variables of interest for 9,309 cases. The majority of missing data occur when cases are decided per curiam, which means that no author is identified. Other instances occur when judges' names are misspelled or ambiguous, which makes it impossible to match them with their records in the Federal Judicial Center's data. Furthermore, in 26 cases our data indicate that a decision took less than a day. Because this is unlikely, we exclude these cases, and in addition we exclude 465 cases that took more than 300 days, which occurred during March Madness in 2 years. This leaves us with 8,818 cases to analyze.

We code the variable $\text{Tournament}_{a[i]y[i]}$ as one if the author in case i has an alma mater participating in the NCAA tournament in the year case i was heard and zero otherwise. We code the variable $\text{March Madness}_{t[i]}$ as one if case i is heard (has oral argument) during February or March. We choose this coding because work on those cases is likely to be disrupted by distraction due to the NCAA tournament. Cases heard during Febru-

8. Federal Judicial Center, Biographical Directory of Article III Federal Judges, 1789–Present (https://www.fjc.gov/history/judges).

Table 2. Distribution of Cases by National Collegiate Athletic Association Tournament Alma Mater Separated by Authorship Status

	Nontournament Months	Tournament Months
Nontournament Author	6,723 (76.24)	1,738 (19.71)
Tournament Author	273 (3.10)	84 (.95)

Note. Rows distinguish panels with Tournament Authors from those panels with Nontournament Authors. Columns distinguish cases heard during February and March from those heard during other months. Each cell shows the raw number of cases heard by each pair of conditions. Percentages are in parentheses.

ary are in the process of having opinions written during the tournament, which begins in the middle of March, as are cases heard during March. Cases heard during January are likely to be (at least nearly) completed by the time of the NCAA tournament.[9]

In our data, 357 decisions were authored by a judge whose alma mater participated in the NCAA tournament during the year the case was heard. We call these judges tournament authors, while we call the judges authoring the remaining 8,461 cases nontournament authors. (Note that a given judge can be both a tournament author and a nontournament author, as she is a tournament author only in the years her alma mater was in the tournament.) Table 2 compares cases decided during the tournament with those decided in other months by authorship status.

4.2. Identification Strategy

To estimate the causal effect of a judge's alma mater participating in the tournament, we rely on a difference-in-differences design. In particular, we consider the difference between the time from oral argument to decision for cases that were heard in months other than February and March (nontournament months) with a tournament author and with a nontournament author. We then consider that difference for cases heard during February and March (tournament months). The difference between the two differences is our quantity of interest. This difference-in-differences design captures the expectation that tournament authors take longer to write a decision than nontournament authors during February or March.

9. The findings we report below are robust to relaxing these coding rules in sensible ways. We discuss our robustness checks as we present the results.

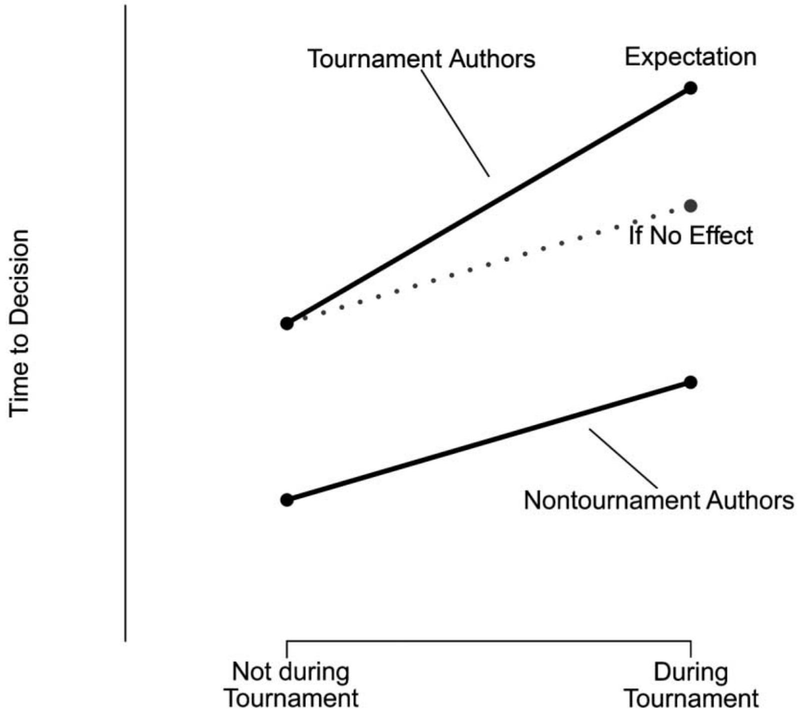

Figure 2. Illustration of hypothetical common trends and our theoretical expectation

By focusing on how the difference between the two types of judges changes in February and March, we can account for other underlying differences between the groups. These should be constant over time.

This design is particularly important in this setting. As described above, there have been many studies of whether labor markets and firms decrease in productivity during major sporting events such as the NCAA tournament. However, if we simply ask whether the judiciary slows down, then we cannot evaluate the normative and theoretical issues at hand—whether individual judges behave differently as they place increased weight on dimensions of their utility other than work, such as leisure.

The key identifying assumption involves common trends. The assumption of common trends holds that any differences between the two groups (here, tournament authors and nontournament authors) will be the same across the two periods of observation (here, cases during tournament months and cases during the rest of the year) if there is no treatment effect. Figure 2 summarizes the common-trends assumption and how our hypothesis relates. The lower black line shows the hypothetical difference between the time to decision for cases heard in months other than February and March and cases heard during February and March by authoring judges without a team in the NCAA tournament. The dotted line shows our assumption for authoring judges with a team in the tournament. If there is no causal effect of having your team in the tournament, then those judges' behavior should change between tournament months and nontournament months in the same ways as the behavior of other judges. Note that the model does not assume no difference for nontournament authors in the two time periods. Rather, it assumes only that the differences between the two groups across the time periods will be similar but for the causal effect of having a team in the tournament. The upper black line shows, by contrast, our expectation that the tournament authors will exhibit a larger delay in deciding cases during February and March, relative to other months, than nontournament authors. Note, as well, that we do not assume a strict increase in delay—just that any change in delay will be toward a longer delay among tournament authors.

4.3. The Empirical Model

Figure 3 compares the time to make a decision during tournament months and nontournament months. The black lines show the distribution of the time to decision for cases not heard during the tournament, and the black triangles show the means for those distributions. In both panels, the mean for these distributions is 99 days. In other words, both tournament authors and nontournament authors write their opinions, on average, in 99 days when a case is not heard during the tournament. This picture changes during the NCAA tournament. The gray lines show the distribution of the time to decision for cases heard during the tournament, and the gray triangles show the means for those distributions. Nontournament authors are faster in writing decisions, at 110 days on average, compared with tournament authors, who take 120 days on average. Hence, in nontournament months it takes nontournament authors and tournament authors equally long to make decisions. While both groups of authors slow down during tournament months, tournament authors are on average slower compared with their colleagues.

We model Time to Decision$_i$ as a function of whether a panel judge's alma mater is in the tournament the year the case was argued, whether the case was heard during tournament months, and the interaction of the two variables. We estimate the model by coding Tournament$_{a[i]y[i]}$ in two

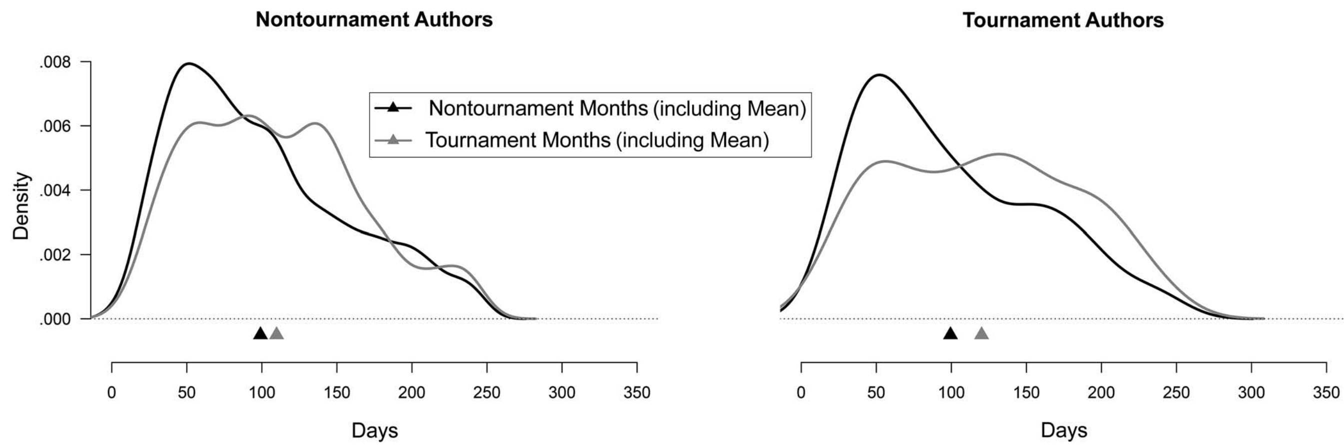

Figure 3. Distribution of days to a decision in month other than February and March (*black lines*) and during the NCAA Tournament (*gray lines*) including mean values (*triangles*) separated by Tournament Authors and Nontournament Authors. The figure highlights only the major parts of the heavily right skewed distribution of days to a decision.

ways—if the alma mater of the opinion's author is in the tournament and if any judge on the panel has an alma mater in the tournament.[10] This helps to capture the notion that peer effects may extend to the collective workflow among judges on a panel and slow down the opinion-writing process. Our expectation is that the interaction will have a positive relationship to Time to Decision$_i$, as judges from tournament teams work more slowly during the tournament months. We employ a linear difference-in-differences model. Formally, we assume that

$$\begin{aligned}\text{Time to Decision}_i = \alpha + \beta_1\text{Tournament}_{a[i]y[i]} + \beta_2\text{March Madness}_{t[i]} \\ + \beta_3\text{Tournament}_{a[i]y[i]} \times \text{March Madness}_{t[i]} + \varepsilon_i,\end{aligned} \tag{1}$$

where ε_i is a normally distributed random variable with a mean of 0 and a standard deviation σ_ε. The model include both year fixed effects and author fixed effects. Standard errors are clustered at the level of the unique combination of three judges deciding each case (that is, the panel) to avoid any downward bias in uncertainty that might result from different numbers of observations from the circuits. We also consider alternative specifications that include covariates for demographic features of the panel; details of other specifications are reported in the Appendix. We estimate each model in a frequentist framework. We also consider estimating the models using a negative binomial specification rather than a Poisson model. The substantive findings we report hold in these alternative models. The primary results are reported in Table 3.

The results show that we estimate a positive and statistically significant effect of the interaction of Tournament$_{a[i]y[i]}$ and March Madness$_{t[i]}$

10. One might also propose a third option, which is to calculate the proportion of the judges who have alma maters in the tournament in a given year. The only possible values for the mean of Tournament Author are 0, .33, .67, and 1. (These correspond to the situations in which zero, one, two, and three of the judges have alma maters in the tournament, respectively.) There are no instances in which the variable takes on a value of 1; that is, there are no observations in which all three judges have alma maters in the tournament. Furthermore, in the data, there are only 70 instances in which two of three judges have alma maters in the tournament. Of those, only eight cases occur during the tournament. Therefore, for only eight of the 8,818 observations (less than .1 percent of the data) could the interactive term capturing the difference-in-differences effect be coded differently from the Tournament Judge variable. (The mean time to decision among those eight observations is 138.38 days, which is in line with the mean for observations for which any judge is the author and the case is heard during the tournament—132.19 days. These two means are not statistically distinguishable.)

Table 3. Empirical Models of Number of Days between Oral Argument and Decision on Federal Court of Appeals, 1985–2000

	Author Only		Any Judge	
	Year Fixed Effects	Author and Year Fixed Effects	Year Fixed Effects	Author and Year Fixed Effects
$\text{Tournament}_{a[i]y[i]}$	−.59	−1.23	12.17**	8.89**
	(4.57)	(4.61)	(3.17)	(1.97)
$\text{March Madness}_{t[i]}$	18.06**	16.37**	10.96*	12.45**
	(2.51)	(2.61)	(5.12)	(4.00)
$\text{Tournament}_{a[i]y[i]} \times \text{March Madness}_{t[i]}$	13.23*	7.75	20.57**	11.64*
	(4.42)	(5.93)	(6.90)	(4.67)
Intercept	102.04**	80.24**	83.90**	43.61**
	(5.24)	(1.20)	(6.61)	(1.97)
R^2	.09	.33	.84	.90

Note. Cell entries are ordinary least squares regression coefficients, with clustered standard errors in parentheses. $N = 8{,}818$.
* $p \leq .05$.
** $p \leq .01$.

and the time it takes to render a decision across three of our four specifications. This means that the difference in the time it takes to render a decision between judges with teams in the tournament and judges without teams in the tournament increases for cases heard during the tournament. Importantly, the difference-in-differences design accounts for any effect of either the tournament itself or being from a school in the tournament on the speed with which judges render decisions. This estimate captures only the effect of the case being heard during the tournament on judges from tournament teams. In the models with fixed effects for the author's identity and the year of the case, the estimate is identified from variation in individual judges' tournament status. Furthermore, in the final model, we see that the effects we detect are consistent when we recode the data to measure whether any judge on a panel had an alma mater participating in the tournament.

To illustrate the effect of the NCAA tournament on tournament authors, Figure 4 summarizes the mean expected days to make a decision during the tournament and in other months. The points show point estimates, and the black bars show 95 percent confidence intervals. The effects are separated by tournament and nontournament authors, over-

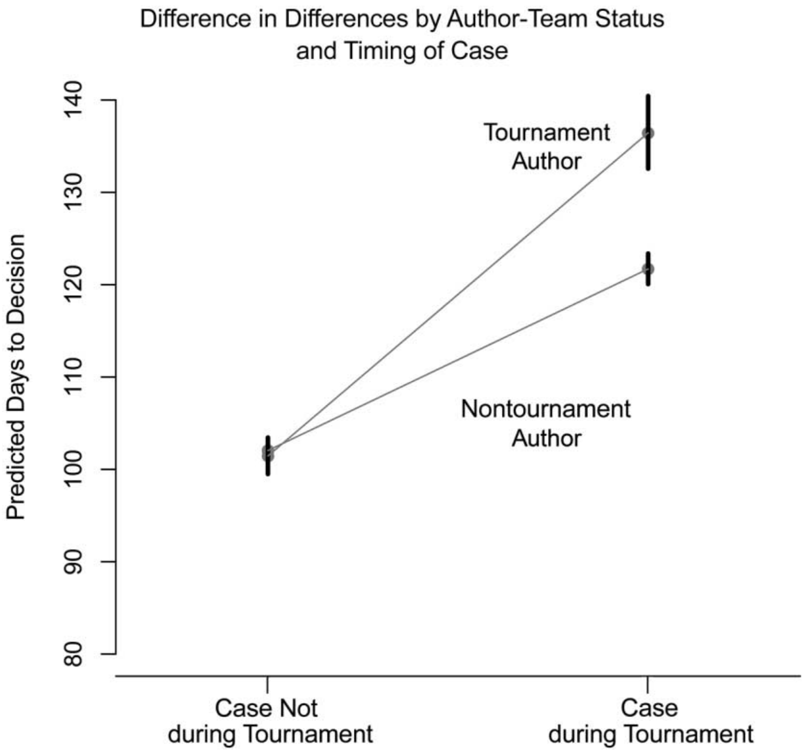

Figure 4. Mean expected days (including 95 percent confidence intervals) tournament and nontournament authors need to write an opinion during the tournament compared to nontournament month.

whelmingly supporting our hypothesis. It takes both kinds of judges roughly 102 days, on average, to produce opinions for cases heard during nontournament months. During the tournament, that number increases to 120 days, on average, for nontournament authors; however, it increases to just over 133 days for tournament authors. That difference—between 120 and 133 days—is our estimate of the causal effect of having an alma mater's team participating in the tournament. Thus, the effect of having one's team in the tournament is to delay the publication of an opinion by 21 days.

What is more, these findings are fairly robust to alternative coding schemes for our key variables. For example, if we code only cases heard during March as being cases decided during the NCAA tournament, we still estimate a positive, statistically significant interactive effect

($\hat{\beta}_3 = 14.0$, SE $= 3.10$). By contrast, if we conduct a placebo test and code cases heard during any given month, we find no effect on months other than the tournament months. Moreover, we still find a positive, though substantively smaller, effect if we include cases heard during January with cases heard in February and March ($\hat{\beta}_3 = 15.3$, SE $= 2.65$). This is consistent with our expectation discussed above that the causal effect of the NCAA tournament is strongest for cases heard during February and March, as those are the cases for which the mechanism—distraction by the tournament affecting effort on opinion writing—is at work. We report the full results of these tests in the Appendix.

4.4. Threats to Inference

4.4.1. Can Judges Avoid Their Work when Tempted with Leisure? One might worry that when a panel includes a judge whose alma mater is in the tournament, her colleagues would (in the collegial spirit) not ask her to take on the responsibility for drafting the majority opinion. If this is true, then the tournament treatment would not be applied as suspected, which would threaten the causal inference we draw. There are, however, theoretical reasons to doubt this possibility. In general, panels of judges sit for a fixed period of time, hearing a (sometimes alleged) randomly assigned set of cases. Because of the temporary nature of the panel, there is a strong norm of equity in the workload among the judges, and oftentimes the opinions are assigned before the cases are heard. If this norm is binding, for whatever reason, then the possibility that the tournament does not affect the panels as suspected is mitigated. Furthermore, relative to some other kinds of personal distractions, including those that are difficult and serious (for example, the illness of a family member), the NCAA tournament is unlikely to be the kind of personal distraction that would warrant a relaxation of typical work norms.

What is more, there is empirical evidence that judges do not get to shirk from their authorship responsibilities when their teams are playing in the tournament. Using our data, we estimate a conditional logit model in which for each case there is a choice of which judge on the panel to assign the opinion, subject to the constraint that only one judge can be assigned. The conditional logit model allows us to include covariates that vary at the level of the alternatives rather than the choice itself. Here we have whether each judge's alma mater is in that year's tournament, which

Table 4. Predictors of Whether a Judge is Selected to Author an Opinion

	$\hat{\beta}$
$\text{Tournament}_{a[i]y[i]}$	.16**
	(.03)
$\text{March Madness}_{t[i]}$	−.01
	(1.41)
$\text{Tournament}_{a[i]y[i]} \times \text{March Madness}_{t[i]}$	.03
	(.07)

Note. Results are from a conditional logit model in which the choice is among the three judges hearing a case, and the predictors are each judge's alma mater's participation in the tournament, whether the case is heard during the tournament, and the interaction of the two. Standard errors are in parentheses. $N =$ 84,879 observations and 24,923 cases.
** $p \leq .01$.

varies at the level of the judge (alternative) rather than at the level of the panel (the choice to whom to assign the opinion).

Let

$$A_{ai} = \exp(\beta_{1a}\text{Tournament}_{a[i]y[i]} + \beta_{2a}\text{March Madness}_{t[i]} + \beta_{3a}\text{Tournament}_{a[i]y[i]} \times \text{March Madness}_{t[i]}).$$

Formally, we assume the probability that judge a is selected to author the opinion in case i as follows:

$$\Pr(\text{Author}_i = a \mid \boldsymbol{M}_i) = \frac{A_{ai}}{\sum_{k=1}^{3} A_{j[i[k]]i}}, \tag{2}$$

where $j[i[k]]$ is the kth judge hearing case i, Author_i is the identity of the author of the majority opinion in case i, and $\boldsymbol{M}_i$ is a matrix of covariates.

We estimate this model using all panels for which the opinion's author is specified (the decision is not per curiam or coauthored) and for which we have clearly matched the author to the panel. This results in 24,923 cases. The results from estimating model (2) are reported in Table 4. The evidence is clear: the interactive effect—that is, our estimate of β_{3i}—is substantively very close to 0 and statistically indistinguishable from 0. Taken in conjunction, the organization of the courts, the norms of equity in opinion writing, and the empirical results indicate that judges are not able to strategically avoid opinion-writing responsibilities during the

Table 5. Characteristics of Judges

	Tournament Judges		Nontournament Judges	
	Mean	SD	Mean	SD
Birth year	1936	17.1	1935	15.9
Proportion male	.68	.5	.69	.7
Proportion white	.94	.2	.82	.4
American Bar Association rating	1.27	.8	1.43	.7
Senate voice vote	.85	.4	.93	.3
Proportion Republican	.30	.5	.34	.5

Note. Comparison of judges' characteristics between judges with tournament teams authoring opinions during the tournament (treatment group) and other judges (control group).

tournament and therefore are unable to avoid the treatment effect. These two empirical tests show that the effect of a judge's team participating in the tournament is associated with a sharp increase in the amount of time she takes to issue an opinion, and that effect cannot be attributed to strategic avoidance of writing opinions.

4.4.2. Are Judges from Tournament-Participating Alma Maters Different? One might also be concerned that judges who attended schools that often participate in the NCAA tournament—big state schools and other athletic powerhouses—are different in other characteristics that might exacerbate the slowdown we observe in February and March for reasons unrelated to leisure. We consider a handful of demographic characteristics of our judges across those who have a team in the tournament (tournament judges) and those who do not (nontournament judges). Table 5 summarizes these results. Across virtually all characteristics, the two groups are identical. The average judge in each group was born in 1935. Men constitute 68–69 percent of the observations. In the tournament and nontournament judge groups, 94 percent and 82 percent are white, respectively, which suggests a slight, though statistically insignificant difference. We also consider the American Bar Association (ABA) rating, which is a 4-point scale rating how qualified the judge is for the position when nominated. (Some judges serve in multiple federal courts and so have ABA ratings for each nomination. We use the ABA rating from the first post, whether it was the court of appeals or another court; using the rating from the court of appeals confirmation process yields identical results.) The statistics indicate, again, that the two groups are identical. In addition, we consider whether the Senate confirmed the judge on a voice vote as a proxy, perhaps, for less-well-quantified metrics of the judge's temperament, political divisiveness, or quality. We find among the tournament and nontournament judges similar rates of voice votes, 85 percent and 93 percent, respectively, a difference that is not statistically significant. Finally, we consider the proportion of judges who are Republicans, as opposed to Democrats, and find that there is no statistical difference between the two groups.

These data suggest that there are no significant differences between the two groups of judges that might give rise to a spurious correlation in the difference-in-differences design and therefore that we have adequate balance in the kinds of judges in each of our categories. That said, it is critical to notice that these demographic features—that is, age, gender, race, perceived qualifications, political support, and ideological orientation—do not vary over time, much less across the periods we compare (tournament months and nontournament months). A notable feature of the difference-in-differences design is the ability to account for time-invariant heterogeneity in potentially confounding features. To the extent that one worries about baseline differences between the groups, the difference-in-differences design addresses the concern.

4.4.3. Are Published Opinions Special? Another concern is that we focus only on published opinions. Our design assumes that judges cannot offset their leisure incentives by changing their workload in the set of unpublished (that is, less-important) decisions. Suppose their workloads in unpublished cases shifts, which decreases their workloads on these less-important cases. This would allow judges to offset the consequences of distraction due to leisure, biasing against the result we find. Insofar as we constrain the implications of our findings to published opinions, we restrict our findings to the cases that will be read by the legal community and most likely to influence future law.

As described above, there are multiple dimensions along which judicial work can be evaluated. A marginal increase in the desire for leisure may involve simply trading off, for example, the amount of time it takes to reach a decision in order to maintain the quality of the judicial product.

4.4.4. Is It Really the Clerks? As we describe in Section 2.2.3, it may be appropriate to interpret the effects we find as reflecting the work of a judge's entire staff. But one might wonder if there is something about the clerks that the judge hires that might explain the outcome. This is a particular concern here in that clerks may change over the course of the year, potentially in ways that confound the inference we have drawn. A natural concern is that the clerks' distraction causes the delay. For this to be the source of confounding, at the very least, it must be that federal courts of appeals judges are particularly likely to hire clerks who attended their own undergraduate institutions. Judges do take considerable interest in the quality of a potential clerk's education; however, the typical concern is with a candidate's law school. Peppers, Giles, and Tainer-Parkins (2014) report that 91 percent of the judges in their sample mentioned law school ranking as a factor that they consider when hiring clerks. Fully 66 percent suggested that law school ranking was the most important factor. In contrast, only 44 percent of judges in the sample thought to mention the quality of a candidate's undergraduate institution as a possible relevant factor, and only 5 percent reported that it was the most important factor. It seems unlikely that this possible source of confounding is problematic.

We consider the availability of data on judges' and clerks' alma maters. To identify the undergraduate institutions of the clerks working for federal courts of appeals judges, we consulted the *Judicial Yellow Book* (Leadership Directories 1996, 1997). We began with the fall 1996 volume and also considered the spring 1997 volume.[11] Across all circuits, including the federal circuit, we identified data on 664 clerks and their judges. Educational information is commonly reported for judges (judge's undergraduate institution, 95 percent; law school, 99 percent). Consistent with Peppers, Giles, and Tainer-Parkins (2014), we find that although the law school attended by clerks is reported more often than not (69 percent), information about their undergraduate institution is routinely ignored: of the 664 clerks in the *Yellow Book* during this period, there is undergraduate information for only 53 clerks (8 percent). Even among those clerks, there is very little evidence that judges and clerks commonly have the same alma maters. Indeed, there are only four cases of a match between the undergraduate institutions of clerks and their judges (8 percent). In every case, the alma mater is Harvard University, a school that was never in the NCAA tournament during the period of our study. Likewise, a

11. After speaking with other scholars and reviewing the results of the first two volumes, we concluded that further research—that is, beyond spring 1997—was unnecessary.

judge's undergraduate institution and her clerk's law school are rarely the same (8 percent).

Finally, the difference-in-differences design with year and judge fixed effects accounts for judge-specific clerk characteristics, assuming that tournament authors do not disproportionately assign excess work to their clerks during the tournament season. If this is ultimately what happens, this process creates a bias against finding a difference in the differences between tournament authors and nontournament authors.

5. THE EFFECT OF THE TOURNAMENT ON JUDICIAL QUALITY

As we outlined above, there are multiple ways in which judges might respond to heightened leisure incentives. One possibility might be to simply slow down their rate of resolution of cases, taking longer to write their opinions, in which case the preceding result is all we might expect to find. On the other hand, we might also expect judges to sacrifice the quality of their decisions to avoid large effects of their delays on case backlogs. Or, alternatively, we might expect a smoother balancing of timing and quality of decisions. In this section, we investigate the consequences of the tournament for judicial opinions' quality and consider these multiple causal pathways—a direct effect of the tournament on quality and the possibility that taking longer to resolve cases mediates the deleterious effects of judicial distraction on quality.

5.1. The Data

How can we best study the quality of judicial opinions? The concept is notoriously elusive, and a veritable cottage industry has developed around the goal of measuring it. An increasingly common metric is to use citation patterns to proxy for quality (see, for example, Posner 2000; Choi and Gulati 2007; Ash and MacLeod 2015). Citations to opinions come in many forms: some simply relate a case to past cases on factual grounds, some describe past cases to provide doctrinal context, some praise past cases, and some criticize past cases. Here we rely on the number of negative citations to an opinion as a proxy for judicial quality (or, really, lack thereof). We prefer to use negative citations because, as contrasted with positive citations, there is little ambiguity in what is a negative, adverse citation. By contrast, positive citations often include a hodgepodge of forms of citation, many elements of which are likely not driven by judi-

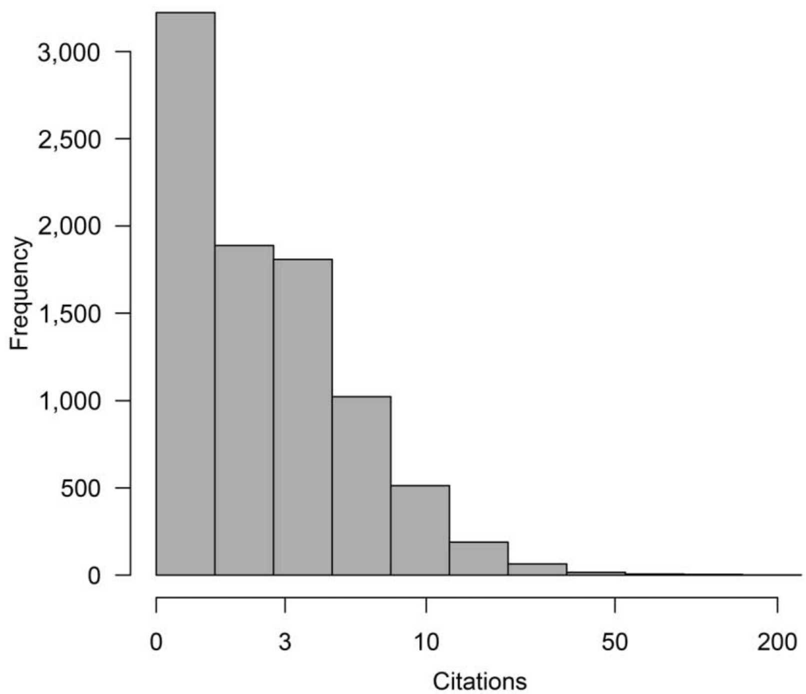

Figure 5. Distribution of negative citations to precedent cases

cial quality. In conjunction with the data collected for the above analyses, data on citations allow us to evaluate the effect of the tournament on the quality of judicial opinions.

To collect data on citations, we performed a KeyCite search on a sample of all courts of appeals decisions, which uses Westlaw's databases to identify every subsequent case citing the case we search (including unpublished decisions). Westlaw returns a report that divides all subsequent citations into discrete, mutually exclusive categories. We then identified cases from our sample of KeyCite reports that are also in our data and for which we have the full set of other relevant covariates—the judges on the panel, the author of the opinion, and the author's alma mater. Most important for our purposes is the category "negative cases," which are cases that cite the case adversely. This category is a relatively narrowly defined classification: a subsequent case criticizes, distinguishes, overrules, or otherwise refers to the case negatively. We are able to identify the total number of negative citations for 5,449 cases. However, only 1,442 of

Table 6. Empirical Models of Numbers of Citations to Federal Court of Appeals Decisions, 1985–2000

	Negative Citations	Negative Citations without Distinguishing Citations	Positive Citations	All Citations
$\text{Tournament}_{a[i]y[i]}$	−.41	−.04	−3.64	−2.46
	(.33)	(.19)	(11.79)	(10.39)
$\text{March Madness}_{t[i]}$	−.46^{+}	−.20	6.42	32.96**
	(.27)	(.15)	(9.10)	(7.94)
$\text{Tournament}_{a[i]y[i]}$ × March $\text{Madness}_{t[i]}$	1.35^{+}	1.00*	−12.48	−39.64^{+}
	(.74)	(.42)	(24.68)	(21.74)
Akaike information criterion	.01	.01	.28	.25
N	1,424	1,424	1,437	1,429

Note. Cell entries are ordinary least squares regression coefficients, with standard errors in parentheses.
$^{+}$ $p \leq .1$.
* $p \leq .05$.
** $p \leq .01$.

those are cases for which we have a complete set of the other covariates (an identifiable author, the date of argument, and the date of decision).[12] For each case, we code our primary variable of interest, Negative Citations_i, as the total number of negative citations for each case, which is summarized in Figure 5.

5.2. Identification Strategy and Empirical Model

As in the preceding analyses, the appropriate research design here is the difference-in-differences design. We model the same empirical model, but now our dependent variable is the number of negative citations an opinion receives rather than the number of days it takes to produce an opinion. We estimate

$$\begin{aligned}\text{Negative Citations}_i = {} & \gamma_0 + \gamma_1\text{Tournament}_{a[i]y[i]} + \gamma_2\text{March Madness}_{t[i]} \\ & + \gamma_3\text{Tournament Author}_{a[i]y[i]} \times \text{March Madness}_{t[i]} \\ & + \xi^T \boldsymbol{X}_i + \varepsilon_i,\end{aligned} \tag{3}$$

where, as before, ε is a normally distributed random variable with a

12. As we discuss in Section 6, this is but one possible measure of an opinion's quality. However, because a negative treatment is such a significant and rare occurrence, we believe that it is a particularly powerful measure of quality.

mean of 0 and a standard deviation σ_ε. For each model, we exclude a small number of observations for which the number of citations is an extreme outlier.[13]

We report the results of estimating this model in the first column in Table 6. We find a positive correlation between the interaction of Tournament$_{a[i]y[i]}$ and March Madness$_{t[i]}$ and the number of negative citations an opinion receives. These results indicate that, as with the time it takes to write a decision, the quality of the work product is affected by a judge's distraction during March Madness. Substantively, these results indicate that in a nontournament month, an opinion written by a judge whose alma mater is not playing in the tournament will receive, in expectation, 1.9 negative citations. Opinions written during the tournament by the same judge receive, in expectation, 2.8 negative citations, a 50 percent increase. Again, most of our results hold if we employ a negative binomial specification. The result in the model for positive citations does not.

It is important to note, of course, that these effects are not limited to the metric of quality we selected. In Table 6, we also report estimates of the model using alternative metrics of quality. First, we separate citation types. In particular, we examine the number of negative citations, excluding distinguishing citations, which are often not critical of an opinion but simply set the opinion aside as not on point. We also consider the total number of positive citations and the total number of all citations (negative, positive, and neutral). If the tournament has an effect on the quality of opinions, we should expect to see a positive effect on negative citations and a negative effect on positive citations. The expectation concerning all citations is more ambiguous, though one might expect a lower-quality opinion to be cited less often, overall.

The results from the models using alternative metrics corroborate the finding from the first model. The tournament has a positive effect on the number of negative citations, even when we exclude distinguishing citations. The tournament has a negative effect on both the number of positive citations and the total number of citations. Taken together, these results suggest the tournament causes a drop in the impact or quality of a published opinion.

One might be worried that other factors, in particular the length of

13. For the model of negative citations, we exclude observations with 50 or more negative citations, both with and without distinguishing citations (the top 1 percent in the data). For the models of positive and total citations, we exclude observations with more than 1,000 positive citations (the top .5 percent in the data).

time to write a decision, affect the rate of negative citation. In general, it is inappropriate to include the time to decision as a covariate, as citations occur after treatment with the March Madness variables and therefore including the time to decision introduces bias. However, when we estimate such a model, we still find a positive effect of the Tournament interaction and the number of negative citations.

Furthermore, one might be concerned that the time it takes to write an opinion mediates all of the effect on negative citations—that there is no direct effect of distraction on the rate of negative citation. Judges simply take longer to write their opinions (as we saw above), and that affects the rate of negative citation. The challenge for an empirical analysis is to assess whether any effect of the tournament on the quality of opinions can be offset by taking longer to complete an opinion. As we described above, judges do not necessarily face strict deadlines for their decisions, so if they want to avoid the deleterious effects of a marginal increase in leisure utility on the quality of their decisions, they could always delay their work and take the time necessary to write decisions of sufficient quality. As we saw above, a desire for leisure certainly has an effect on the time it takes a judge to do her work. At the same time, the caseload pressure that judges face suggests that there is a limit to how long they can delay, so it is possible that they cannot fully offset the effect of distraction through delay.

6. DISCUSSION AND CONCLUSION

The findings we report here have implications for at least three areas in which the literature on judging has grown in recent years. First, our findings speak to the growing literature, with roots in the account of the judicial process in Posner (1993), on the various incentives judges face. Second, our analysis and findings provide groundwork for the analysis of two components of the judicial work product—speed and quality—and highlight the value of articulating the range of options judges have when deciding how to respond to changes in their incentives. Third, our findings show how empirical approaches to studying judicial leisure, which are often deployed at the trial court level, can be used to study appellate judging, where the theoretical framework that dominates the literature often applies.

6.1. A Broader Model of Judging

The literature on judicial behavior has evolved considerably over the last century. Whereas debates once centered on whether and precisely how ideological beliefs influence judicial work, ideology is now understood to be just one of many forces that operate on judicial behavior. As we have come to think about judges as motivated by general career concerns, various claims have been made about how judges, as laborers in a market, change their work behavior when they increasingly value leisure time (see, for example, Posner 1993). A central challenge to the ability to advance research on the role of leisure is empirical. How do we measure a judge's interest in leisure, let alone variation in leisure incentives in a given judge over time or across judges at a given point in time? Our research design offers an opportunity for assessing the consequences of heightened incentives for leisure. The NCAA Men's Basketball Tournament is one of the most popular sporting events in the country every year, and reports often demonstrate a large effect of the tournament on workplace productivity in the private sector. By leveraging the selection of teams into the tournament exogenous of individual judges' workloads, our difference-in-differences design shows that when judges' alma maters appear in the tournament, they divert effort away from judging, as predicted by the literature on judges in the labor market. Crucially, the causal identification of our research design is stronger than in past studies because we are able to leverage exogenous shocks to judges' leisure interests that vary over time and across individual judges.

Of course, our analysis is in many ways only a single step forward. Our empirical leverage on judicial leisure is limited—we exploit exogenous variation in a judge's leisure incentive from the NCAA tournament. Surely judges face much more significant and common increases to their leisure incentives, such as when a child is to marry, at the death of a loved one, or because of other personal interests. A richer investigation of the effects of leisure would employ designs that can exploit exogenous variation in other, potentially more powerful, shocks to leisure incentives.

Furthermore, what we cannot address, and what the two strands of literature need, is a comprehensive model that directly evaluates how judges balance their myriad interests. Epstein, Landes, and Posner (2013) lay out many interests a judge might have and evaluate many of them empirically. However, what remains elusive is a comprehensive theoretical model that identifies how those incentives interact and is subject to empirical evaluation. While we do not propose a theoretical innovation, we expect that the empirical strategy we identify will contribute to ongoing efforts to elaborate a more holistic model of judging.

6.2. Are the Effects We Find Important?

By synthesizing existing claims about the many forces that influence judicial behavior, we have shown that the existing literature unequivocally predicts that judges work more slowly when confronted with a heightened interest in leisure but that the literature makes inconclusive predictions with respect to the quality of the work judges produce. Moreover, our empirical strategy provides analytic leverage for both components. We are naturally led to ask what all of this means. Are the effects we observe important?

Unfortunately, there is no definitive answer to this question. People will have different opinions about how much of a delay in the resolution of cases is tolerable. And people will differ over whether a particular delay is tolerable in light of the effect of leisure on quality. That said, we are confident that we can place some structure on our normative evaluations. If delay is the cost of quality, we are in a position to ask how much of a loss in quality we would be willing to sustain to ensure that cases are not delayed when judges are distracted. If we suppose that this kind of trade-off exists, then it becomes easier to evaluate the substantive significance of our findings. What we find is that changes in leisure incentives result in both delays in the completion of cases and a reduction in quality. As far as we can tell, judges are not fully trading off one harm for another but appear to be balancing the harms simultaneously. More to the point, if we continue to assume that judges want to resolve cases without delay and to write high-quality opinions, we are led to conclude that the delays we are observing are likely shorter than they would be if judges were willing to compromise even more on the quality dimension. Similarly, the reduction in quality that we observe is lower than it would be if judges were willing to eliminate delays. In evaluating the substantive significance of particular findings, it is critical to keep this trade-off in mind.

With respect to work rate, we find that when a judge's alma mater participates in the tournament, she takes longer to write the opinions for which she is responsible. In particular, during nontournament months, there is no difference in the speeds with which judges with teams and judges without teams in that year's tournament work. However, during tournament months, tournament judges take between 11 and 20 days longer to complete their opinions. The 95 percent confidence intervals include effects as long as 37 days.[14] The absolute magnitude of the estimated effect may not seem particularly long, but it is commensurate with what one might reasonably expect given the nature of the variation we exploit. Were we to have analytic leverage on other, more powerful shocks to a judge's leisure incentives, we might find more substantively meaningful effects. The insight from our analysis is instead the credible claim of a causal effect of leisure on the judicial work product.

With respect to the quality of their work, we find that judges' opinions do suffer when the authors face heightened incentives for leisure. In particular, the effect of an author of an opinion having her alma mater in the tournament is to increase the number of negative citations to the opinion. Of course, this is but one possible measure of an opinion's quality, but the finding is important. Negative citations are rare, so the finding that judges' leisure incentives cause an increase in negative treatment of an opinion is particularly striking. Moreover, as noted, the extant theoretical accounts of judicial responses to leisure incentives is equivocal on the implications for the quality of judges' work. Judges might slow down to maintain their work quality, or they might trade off on both dimensions, balancing a deterioration of quality and speed. We find evidence suggesting the latter. The substantive magnitude of these effects is less clear, as it can be hard to know what to interpret from a given change in citations to an opinion.

6.3. Expanding the Model to Appellate Judges

One way in which we might make such headway relates to a third aspect of our analysis—considering the different positions in which various judges find themselves. As described above, much of the extant empirical literature focuses on trial judges. However, one might suspect that the theoretical incentives identified in the literature more naturally apply to appellate judges. In particular, judges on the US Courts of Appeals face little risk of review by their only superior court—the Supreme Court. Moreover, their more prestigious positions insulate them somewhat from the kinds of concerns about career and reputation that might influence judges in lower courts. A comprehensive theoretical model of the judicial process would examine the different judicial positions to assess how in-

14. The confidence intervals for the interaction term in the four models reported in Table 3 are (3.39, 21.07), (−4.11, 19.61), (4.48, 36.67), and (2.30, 20.98).

stitutional circumstances condition a judge's responsiveness to incentives and willingness to trade off components of the judicial work product.

More important, the greater stature and consequences of appellate judging—especially in written opinions—suggests larger implications for our findings than those in the context of lower courts. Delays in the resolution of important questions of law have significant implications for the rule of law and justice. Deteriorating quality of appellate opinions can affect the law in many cases and even the extent to which the Supreme Court must intervene with law making in the circuits. Moreover, we might expect that the way judges respond to leisure incentives can influence the behavior of judges in lower courts, given that most models of hierarchy and oversight predict that subordinates' behavior is conditioned by the preferences of their superiors. It is again critical to remember that the effects we identify are modest. Still, it also bears repeating that the treatment is relatively weak. Future work should look for stronger sources of variation in leisure incentives.

Taken together, the various implications of our study—building on literature that develops a broader model of judging, incorporating multiple components of the judicial work product, and expanding empirical studies to appellate courts—provide a step forward in the study of judicial incentives and performance. At the same time, we recognize that our findings are just that—only a step. We anticipate that future research will build on our findings and research design as the literature moves toward a more comprehensive model of judging.

APPENDIX: RESULTS OF PLACEBO TESTS

First, we replicate our analysis as a placebo test, coding each month of the year outside the tournament as the period of the tournament. We then reestimate equation (2) with the key variable, March Madness$_{t[i]}$, coded, in turn, according to the placebo. We include fixed effects for the author of the opinion and for the year the case was decided. Standard errors are clustered on the level of the panel of judges that decides the case. We report the ordinary least squares estimates of the difference-in-differences estimates, given by the key term—Tournament$_{a[i]y[i]}$ × March Madness$_{t[i]}$—in Table A1. We report results from a model that codes the panel as having a Tournament Judge when any member of the panel has an alma mater participating in the tournament. The alternative coding yields similar results, again with none of the difference-in-differences estimates statistically distinguishable from 0. We do not report the other estimated coefficients for reasons of

Table A1. Results of Placebo Test

Month	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
January	9.86**	2.96	23.84
	(2.92)	(10.00)	(12.60)
April	11.21**	15.36+	7.33
	(2.31)	(9.21)	(8.38)
May	13.84**	6.05	−13.93
	(1.78)	(12.26)	(14.68)
June	8.51*	−31.28**	11.53
	(4.26)	(8.65)	(9.98)
July	11.03**	−31.94	13.13
	(4.10)	(30.72)	(21.16)
August	11.79**	3.82	−9.30
	(4.23)	(10.95)	(13.24)
September	11.13**	−46.95	−12.00
	(4.09)	(55.10)	(35.59)
October	13.79**	8.50	−16.85
	(4.22)	(9.34)	(8.52)
November	12.70	11.16	−21.51
	(4.16)	(10.22)	(16.68)
December	11.92**	3.96	−9.74
	(4.22)	(11.84)	(19.50)

Note. Results are from 10 separate ordinary least squares models. Three main coefficients are reported. The column with $\hat{\beta}_3$ gives the difference-in-differences estimate. In each model, we code the indicated month as the National Collegiate Athletic Association tournament and estimate the model from equation (2). For reference, in the true data, the estimated coefficient associated with the difference-in-difference estimate is $\hat{\beta}_3 = 11.64$, SE $= 4.67$. Standard errors are in parentheses.
$^{+}p \leq .1$.
$^{*}p \leq .05$.
$^{**}p \leq .01$.

space. The key finding is that none of these coefficients are statistically distinguishable from 0.

REFERENCES

Ash, Elliott, and W. Bentley MacLeod. 2015. Intrinsic Motivation in Public Service: Theory and Evidence from State Supreme Courts. *Journal of Law and Economics* 58:863–913.

Bainbridge, Stephen M., and Mitu Gulati. 2002. How Do Judges Maximize? (The Same Way Everybody Else Does—Boundedly): Rules of Thumb in Securities Fraud Opinions. *Emory Law Journal* 51:83–151.

Baum, Lawrence. 2009. *Judges and Their Audiences: A Perspective on Judicial Behavior*. Princeton, NJ: Princeton University Press.

Berdejó, Carlos, and Daniel L. Chen. 2017. Electoral Cycles among US Courts of Appeals Judges. *Journal of Law and Economics* 60:479–96.

Blatcher, Marjorie. 1978. *The Court of King's Bench, 1450–1550: A Study in Self-Help*. London: Athlone Press.

CareerBuilder. 2015. More Workers Participating in College Basketball Tourney Office Pools, Finds CareerBuilder Survey. Press release. March 18. https://www.careerbuilder.com/share/aboutus/pressreleasesdetail.aspx?sd=3%2f17%2f2015&id=pr874&ed=12%2f31%2f2015.

Carvajal, Doreen. 1995. Awaiting Judgment: A Special Report. *New York Times,* April 17. https://www.nytimes.com/1995/04/17/nyregion/awaiting-judgment-special-report-new-york-s-clogged-us-courts-delaying-civil.html.

Challenger, Grey, & Christmas. 2015. Office Pools Cost Millions, but Boost Morale. Press release. Chicago, March 11. http://www.challengergray.com/download/file/fid/215.

Choi, Stephen J., and Mitu Gulati. 2007. Ranking Judges according to Citation Bias (as a Means to Reduce Bias). *Notre Dame Law Review* 82:1279–1309.

Choi, Stephen J., Mitu Gulati, and Eric A. Posner. 2012. What Do Federal District Judges Want? An Analysis of Publications, Citations, and Reversals. *Journal of Law, Economics, and Organization* 28:518–49.

Clark, Tom S., and Aaron B. Strauss. 2010. The Implications of High Court Docket Control for Resource Allocation and Legal Efficiency. *Journal of Theoretical Politics* 22:247–68.

Clotfelter, Charles. 2012. March Madness and Productivity: The Real Story. *Harvard Business Review,* March 15. https://hbr.org/2012/03/the-march-madness-really-destr/.

Cohen, Mark A. 1991. Explaining Judicial Behavior or What's "Unconstitutional" about the Sentencing Commission? *Journal of Law, Economics, and Organization* 7:183–99.

———. 1992. The Motives of Judges: Empirical Evidence from Antitrust Sentencing. *International Review of Law and Economics* 12:13–30.

Drahozal, Christopher R. 1998. Judicial Incentives and the Appeals Process. *SMU Law Review* 51:469–503.

Epstein, Lee, and Jack Knight. 1998. *The Choices Justices Make*. Washington, DC: CQ Press.

Epstein, Lee, William M. Landes, and Richard A. Posner. 2013. *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice*. Cambridge, MA: Harvard University Press.

Epstein, Lee, Jeffrey A. Segal, Nancy Staudt, and René Lindstädt. 2005. The Role of Qualifications in the Confirmation of Nominees to the U.S. Supreme Court. *Florida State University Law Review* 32:1145–73.

Garoupa, Nuno, and Tom Ginsburg. 2009. Reputation, Information, and the Organization of the Judiciary. *Journal of Comparative Law* 4:228–56.

Gulati, Mitu, and C. M. A. McCauliff. 1998. On *Not* Making Law. *Law and Contemporary Problems* 61(3):157–227.

Gulati, Mitu, and Richard A. Posner. 2016. The Management of Staff by Federal Court of Appeals Judges. *Vanderbilt Law Review* 69:479–97.

Judicial Conference. 2015. Judicial Conference Judgeship Recommendations, March 2015. http://www.uscourts.gov/file/document/2015-judicial-conference-judgeship-recommendations.

Klein, David E., and Robert J. Hume. 2003. Fear of Reversal as an Explanation of Lower Court Compliance. *Law and Society Review* 37:579–606.

Lang, Duncan Temple. 2013. XML: Tools for Parsing and Generating XML within R and S-Plus (computer file). R package version 3.98-1.1. Davis: University of California Department of Statistics. http://CRAN.R-project.org/package=XML.

Leadership Directories. 1996. *Judicial Yellow Book: Who's Who in Federal and State Courts*. Vol. 2, no. 1. New York: Leadership Directories, Inc.

———. 1997. *Judicial Yellow Book: Who's Who in Federal and State Courts*. Vol. 2, no. 2. New York: Leadership Directories, Inc.

Lozano, Fernando A. 2011. The Flexibility of the Workweek in the United States: Evidence from the FIFA World Cup. *Economic Inquiry* 49:512–29.

Maltzman, Forrest, James F. Spriggs II, and Paul J. Wahlbeck. 2000. *Crafting Law on the Supreme Court: The Collegial Game*. New York: Cambridge University Press.

Miceli, Thomas J., and Metin M. Coşgel. 1994. Reputation and Judicial Decision-Making. *Journal of Economic Behavior and Organization* 23:31–51.

Miller, Geoffrey P. 2004. Bad Judges. *Texas Law Review* 83:431–87.

NCAA (National Collegiate Athletic Association). 2015. 2015 Tourney Most-Watched in 22 Years. *NCAA.com.* April 7. https://www.ncaa.com/news/ncaa/article/2015-04-07/2015-ncaa-tournament-has-highest-average-viewership-22-years.

Palazzolo, Joe. 2015. In Federal Courts, the Civil Cases Pile Up. *Wall Street Journal,* April 6. https://www.wsj.com/articles/in-federal-courts-civil-cases-pile-up-1428343746.

Peppers, Todd C., Micheal W. Giles, and Bridget Tainer-Parkins. 2014. Surgeons or Scribes? The Role of United States Court of Appeals Law Clerks in "Appellate Triage." *Marquette Law Review* 98:313–32.

Posner, Richard A. 1993. What Do Judges and Justices Maximize? (The Same Thing Everybody Else Does). *Supreme Court Economic Review* 3:1–41.

———. 2000. Is the Ninth Circuit Too Large? A Statistical Study of Judicial Quality. *Journal of Legal Studies* 29:711–19.

Reinhardt, Stephen. 1993. A Plea to Save the Federal Courts: Too Few Judges, Too Many Cases. *ABA Journal* 79:52–54.

Savchak, Elisha Carol, Thomas G. Hansford, Donald R. Songer, Kenneth L. Manning, and Robert A. Carp. 2006. Taking It to the Next Level: The Elevation of District Court Judges to the U.S. Courts of Appeals. *American Journal of Political Science* 50:478–93.

Segal, Jeffrey A., and Harold J. Spaeth. 2002. *The Supreme Court and the Attitudinal Model Revisited*. New York: Cambridge University Press.

Songer, Donald R. 1999. The United States Courts of Appeals Databases, 1925–1996 (computer file). Columbia, SC: Don Songer Project. http://www.songerproject.org/us-courts-of-appeals-databases.html.

Vandermey, Anne. 2015. This Is How Much Time We Spent Watching March Madness. *Fortune,* March 6. http://fortune.com/2015/03/06/march-madness/.