MANDATORY RETIREMENT FOR JUDGES IMPROVED PERFORMANCE ON U.S. STATE SUPREME COURTS

Elliott Ash
W. Bentley MacLeod

Mandatory Retirement for Judges Improved Performance on U.S. State Supreme Courts
Elliott Ash and W. Bentley MacLeod
NBER Working Paper No. 28025
October 2020, Revised January 2023
JEL No. D02,J26,J41,J44,K0,K4

ABSTRACT

Mandatory retirement for judges has often been considered as a policy response to an aging judicial workforce. This paper provides empirical evidence on how mandatory retirement influences judge performance using a wave of such reforms in U.S. state supreme courts as a natural experiment. We find that introducing mandatory retirement improves court performance as measured by output (number of published opinions) and legal influence (number of forward citations to those opinions). While older judges are cited less often than younger judges, the effect of mandatory retirement on performance is much larger than what would be expected from a change in the age distribution. We find some evidence that the additional effect is due to selective attrition and that the presence of older judges reduces the performance of younger judges.

Elliott Ash
ETH Zurich
IFW E47.1
Zurich 8044
Switzerland
ashe@ethz.ch

W. Bentley MacLeod
Department of Economics
Princeton University
214 Robertson Hall, 20 Prospect St.
Princeton, NJ 08544-1013
and NBER
wbmacleod@wbmacleod.net

1   Introduction

The goal of this paper is to provide empirical evidence on how mandatory retirement rules influence work performance. The empirical setting is state supreme courts, the state-level equivalent to the federal-level U.S. Supreme Court. State supreme court judges are tasked with reviewing lower-court decisions and setting new legal rules in a common-law milieu. Researching, deciding, and justifying the law is a high-skilled constellation of tasks requiring expertise and professionalism (see Posner, 2008), comparable in technicality to physicians and scientists. Moreover, judges work alongside peer judges and supervise teams of clerks and other staff, meaning the job also entails significant social and managerial skills.

We measure judge performance using a database of all decisions in state supreme courts for the years 1947-1994, combined with citation measures from 1947-2012. Because these opinions comprise the near-entirety of a judge's work product, they can be used to produce cleaner measures of performance than is possible in other high-skill domains. More specifically, we measure work performance as the number of times a judge is cited (positively) by future judges in a given year. While forward citations do not identify a "correct" decision, they do index the degree to which a new legal interpretation or clarification influences or helps future peer judges.[1] Hence, citations provide a measure of performance that does not require a normative evaluation of the rule applied.

Beyond the performance measurements, state supreme courts have a number of desirable features for our research objectives. First, the job of a judge has not changed substantially over time, nor does it change over the course of the career. Relative to other high-skill professions, such as medicine or management (Choudhry et al., 2005; Bloom and Reenen, 2007; Bloom et al., 2012), the skills relevant for good judging are relatively constant and thus historical evidence is relevant for current policy. Second, the workload for judges is held constant; it cannot be influenced by judges themselves and does not vary according to age or experience. Third, compensation does not vary between judges, nor is it contingent on work performance (see Landes and Posner, 2009). Fourth, state supreme court judges have relatively strong tenure protections with little chance of termination (Kritzer, 2011). Fifth and finally, they are at the top of their profession without many further opportunities for promotion.

Appellate judging is a technically and professionally demanding career, at least on par with physicians, scientists, and managers. Our results will therefore be informative about

---

[1]See Posner (2008) for a seminal discussion of the work of judges and what citations mean to them. Choi et al. (2008) and Choi et al. (2010) discuss in detail why citations are used to measure judicial performance. In particular, decisions by appellate courts are law, and thus they constrain future decisions. A citation to a judge's decision is a direct measure of their performance in their role of rule maker.

how retirement policies would influence work productivity in these other high-skill professions. Beyond that, state supreme court judges are themselves an elite and powerful group. They have the authority to review not just the decisions of lower courts, but also legislation passed by state assemblies. Important common-law rules – such as contracts, real property, and torts – are made, applied, and distinguished in state supreme court decisions. These decisions have the force of binding precedent for all courts in a state and can even influence law in other states through a shared legal discourse and persuasive precedent.

Given the legal and social impacts, judge performance is an important policy objective. In particular, the potential decline in performance due to aging may warrant a policy response. According to Posner (1995), "it is well-known within professional circles that some federal judges, including Supreme Court Justices, have continued to sit long after their judicial performance has become severely compromised by age-related disabilities" (p. 3). Recent news stories have highlighted anecdotal evidence of old age interfering with work performance in U.S. federal courts, where judges have life tenure and can stay on as long as they like. These accounts include some examples of older judges with dementia symptoms continuing with their work.[2] In addition, the risk of a death on the court increases as the set of working judges gets older, a disruptive event exemplified by the 2020 death of Justice Ruth Bader Ginsburg.

In response to these concerns, many states have introduced maximum age constraints for full-time judges. Table 1 reports these rules and their reforms for the 1947-1994 period. In 1947 (the first year of our judge performance panel), 17 states had a mandatory retirement rule. By 1994 (the last year in the panel), an additional 14 states had adopted mandatory retirement. Besides the extensive-margin variation of having a rule or not, there is additional intensive-margin variation in the maximum age: 70, 72, or 75.[3]

Our empirical strategy uses the introduction of mandatory retirement rules for judges as a natural experiment. We ask how introducing mandatory retirement affects court performance in a differences-in-differences regression framework. Court fixed effects adjust for time-invariant characteristics by court, while year fixed effects adjust for nationwide trends affecting all courts. Our identification assumption is parallel trends, for which we provide

---

[2]See the 2011 *ProPublica* article, "Life Tenure for Federal Judges Raises Issues of Senility, Dementia," available at https://www.propublica.org/article/life-tenure-for-federal-judges-raises-issues-of-senility-dementia. There is also the issue of uncertainty around health: e.g., "Justice Ginsburg in the Hospital Again", *New York Times*, 29 July 2020, available at https://www.nytimes.com/2020/07/29/us/politics/justice-ginsburg-hospital.html.

[3]Note that, despite the common reference to these rules as requiring "mandatory retirement", these rules do not necessarily entail perfect binary enforcement with a 100% workload the day before turning (e.g.) 70 and a 0% workload the day after. Some of the laws allow judges to finish out the year or term, or they can exempt incumbent judges. Even when the rules apply, judges often return as visiting senior judges with a partial caseload. These additional details are discussed in detail below.

Table 1: Judge Retirement Rules and Reforms by State

A. Status Quo Rules at Period Start (1947)	
Retirement Rule	List of States
No Max Age	AR, CA, DE, GA, ID, KY, ME, MS, MT, ND, NE, NM, NV, OK, RI, TN, WI, WV, VT*
Retirement at Age 70	AK, HI, LA, MD, MA, MI, MO, NH, NJ, NY, OH
Retirement at Age 72	NC, SC
Retirement at Age 75	IL, IN, TX, UT

B. Retirement Rule Changes, 1948-1993		
Retirement Age		List of States (with Year Enacted)
Before	After	
None	70	AL (1973), AZ (1992), CT (1974), FL (1972), MN (1973), PA (1968), VA (1970), WI (1955), WY (1972)
None	72	CO (1962), IA (1965)
None	75	KS (1993), OR (1960), TX (1948), WA (1952)
70	None	WI (1984)

Notes. Initial retirement rules (in 1947, Panel A) and their reforms (Panel B), by state. * Vermont (VT) has mandatory retirement at age 90; we classify it as no mandatory retirement since there are just 2 judges in our entire sample (not in Vermont) who live that long.

evidence using event-study regressions.

Do mandatory retirement rules affect the judge age distribution? We find that introducing a retirement age decreases the age of working judges by 2-4 years. The effect is observed across the age distribution, with the oldest incumbent judges being replaced by new younger judges.

In our main result, we estimate the treatment effect of retirement rules on court performance (as measured by citations to judicial opinions issued by the court). We find a positive and statistically significant effect of mandatory retirement on court performance. The effect is quantitatively large, at about a 25-30% proportional increase in positive citations to the court. The effect is robust to a number of alternative specifications, and we find similar effects using a less restrictive citation measure (all citations, not just positive ones) and a more restrictive one (just citations from other states, where a ruling is persuasive rather than binding precedent). We find that the effect is driven by both quantity (an increase in opinions authored) and influence (citations per opinion).

Why does mandatory retirement improve court performance? We explore a number of mechanisms. First, we rule out that the effect is driven by changes in the composition of cases appealed or overall caseload. Next, we consider the salient possibility that older judges are less productive, so replacing them with younger judges would mechanically increase performance. While performance falls with age, we can show that the aging effect does not fully explain

the mandatory retirement effect. That is, the retirement reform's effect on performance is greater than that expected from the shift in the age distribution. We can partly explain this gap through a selective attrition mechanism, where the higher-performing judges return as visiting judges with a partial caseload after mandatory retirement. Finally, we find that a significant component of the retirement reform effect comes from a within-judge shift, where the younger judges on the court increase their output after the reform. These results point to a quantitatively important *team effect* of aging, in which the presence of older judges slows down the pace of work on the court.

These results add to the significant and active literature on the economics of aging and retirement (e.g. Lumsdaine and Mitchell, 1999). One of the closest papers is Ashenfelter and Card (2002), who find that banning a mandatory retirement age for university faculty significantly increased the share of older academics (see also Ho et al., 2021). Frederiksen and Flaherty Manchester (2019) look at the private sector and find that banning mandatory retirement reduced the share of long-term employment contracts and increased the prevalence of performance pay. Our setting is different in that we observe the *implementation* (rather than removal) of mandatory retirement. Unlike Ashenfelter and Card, we directly observe changes in productivity in addition to changes in worker age. Moreover, our judicial context holds fixed other aspects of the contract (e.g. performance pay), allowing us to isolate productivity effects of a mandatory retirement rule. Finally, the team performance of judges on the court is also directly and jointly observed.

A second related literature is the structural work on retirement choice, which has analyzed worker responses to pensions and other retirement incentives (Gustman and Steinmeier, 1986; Stock and Wise, 1990; Gustman and Steinmeier, 1991, 2005). In political economy, Diermeier et al. (2005) and Keane and Merlo (2010) derive structural estimates of the parameters underlying retirement choices of U.S. congressional representatives. The structural approach allows rich counterfactual exercises but relies on strong assumptions of the underlying choice model. Our approach is different in that we focus on producing causal estimates using natural experiments changing the retirement rule.

Third, our paper adds to the literature on judicial behavior (Epstein et al., 2013; Posner, 2008) and, in particular, the literature on state supreme court judges. In our own previous work using earlier versions of this paper's dataset, we show that state supreme court judges increase citations in response to institutional changes that give them more time and discretion (Ash and MacLeod, 2015). Further, we show that judges selected by technocratic merit commissions or by non-partisan elections receive more citations than their colleagues on the same court selected by partisan elections (Ash and MacLeod, 2021). There are many other empirical papers exploring other dimensions of state supreme courts and the behavior of

these judges.[4] The current paper adds to this literature by analyzing the issue of mandatory retirement.

More generally, these results adds to the emerging literature on how organizational practices influence performance of public-sector professionals. This literature includes Prendergast (2001) and Shi (2009) on police performance, Bloom et al. (2015) on education, and Coviello et al. (2015, 2014) on work organization in the Italian judiciary. We add to this literature by showing that retirement rules work not just at the level of individuals, but also at the level of organizations.

The rest of the paper is organized into the following parts. Section 2 provides some background and describes the data. Section 3 outlines the empirical strategy and identification assumptions. Section 4 reports the main results. Section 5 explores some mechanisms for the effect. Section 6 concludes.

2 Background and Data

2.1 Overview

A key factor in the decision to stop working is the inevitable depreciation in skills with age. Take the case of professional runners: As shown in Appendix Figure A.1 Panel (A), the speed with which runners can complete a 10 kilometer race falls continuously from about age 40, and there is a very steep decrease around age 85. Most jobs in a knowledge economy are not based on physical speed or other singular performance dimension, yet aging still takes its toll: cognitive ability declines continuously starting at age 20 (Appendix Figure A.1 Panel (B)).

Since age-related performance decline takes several years, it may not be clear when to step down. Due to nominal rigidity in wages (see Appendix Figure A.2), wage cuts are not usually an option for older workers. Meanwhile, a predetermined age-based wage cut – even if it would be accepted by workers – would require an accurate estimation of age-related productivity decline (perhaps many years in the future), and would require (inefficiently) identical treatment of workers with highly heterogeneous performance trajectories.[5] As health choices and technologies improve, one could expect even more dramatic variation across workers in longevity and late-life productivity.[6]

---

[4]See Hall and Bonneau (2006); Choi et al. (2008); Shepherd (2009b,a); Choi et al. (2010); Canes-Wrone et al. (2010); Kritzer (2011); Iaryczower et al. (2013); Canes-Wrone et al. (2014); Kritzer (2015); Lim and Snyder (2015); Bonneau and Cann (2015); Kang and Shepherd (2015).

[5]Frederiksen and Flaherty Manchester (2019) provide evidence that firms have historically addressed the problem of varying performance trajectories by keeping base wages low and using more performance pay. This is also consistent with the evidence in Lemieux et al. (2009) that finds an increased use of performance pay over time for highly skilled professionals.

[6]In the United States in 2010, one could expect about 19 more years of life conditional upon reaching 65;

A potential solution to these end-of-career conflicts is the use of mandatory retirement provisions (Lazear, 1979). When parties have agreed in advance to a specified retirement date, then the employee's performance is not being called into question during the termination process. Beyond avoiding such discomfort, another benefit of such a rule is that it helps employees plan in advance their savings. A potential cost is that many individuals are still productive into advanced age and would prefer to continue working in some capacity.

Regardless of the net benefits, mandatory retirement is prohibited (with some exceptions) in the United States under the Age Discrimination Act of 1967.[7] The justification for this prohibition is that mandatory retirement by age is facially discriminatory toward workers above that age. Instead, employers are required to evaluate the performance of the worker so that potentially high-productivity workers are retained. Economic theory provides little guidance on whether mandatory retirement dates should be allowed: with costless renegotiation, mandatory retirement would have no effect or, if it does have an effect, private parties would themselves choose the optimal rule (see Lazear, 1979; Frederiksen and Flaherty Manchester, 2019).

Separate from the question of legality, it is an open empirical question whether mandatory retirement improves work productivity. If performance declines with age, then reducing the average age of workers to some degree would improve productivity. On the other hand, if high-productivity older workers are removed during their prime due to the mandate, that could reduce productivity. In addition, with a looming mandatory retirement date, workers might reduce investment in job-specific skills. To see which effects dominate, empirical evidence is needed.

These factors of job performance and engagement are important determinants of the judge retirement decision, which tends to come later than that for the average U.S. worker (see Appendix Figure A.4). Such lengthy careers reflect in part a high intrinsic professional motivation of the judges (Ash and MacLeod, 2015). One indicative example of this motivation is the fact that a small minority of federal judges take senior status (a reduced caseload at full salary) as soon as it becomes available (Posner 1995, pg. 186; see also Choi et al. 2013). Meanwhile, a related literature in political science has shown evidence of strategic retirement – judges timing their retirement based on the party of the appointing governor or president to influence the political ideology of the successor judge (e.g. Nixon and Haskin, 2000).[8]

---

this number was up from about 14 years in 1960. Appendix Figure A.4 Panel (a) shows the average retirement age by year, which has been stable for men but increasing significantly for women.

[7]The Age Discrimination Act of 1967 is designed "to promote employment of older persons based on their ability rather than age". The text of the law is excerpted in Appendix D.

[8]Retirement choices have also been used for identification of electoral-retention effects. For jurisdictions with competitive reappointment processes, judges planning to retire do not face the same retention-related

2.2 Institutional Context

Our empirical setting is state supreme courts, which are the highest courts in state systems and are analogous to the U.S. Supreme Court for the federal system.[9] The fundamental role of a state supreme court judge is to rule on questions of state law (rather than federal law), arising in cases appealed from lower state courts. Such cases begin when a plaintiff files a lawsuit or a prosecutor indicts a criminal. At trial, facts are litigated and a judge/jury gives a verdict, which the losing party can appeal. If the state has an intermediate appeals court (IAC), that court will then take the case and may affirm, reverse, or modify the trial verdict. After this intermediate court's decision (or after the trial decision when the state does not have an IAC), the ruling can be appealed to the state supreme court, which is the last appeal on matters of state law.[10]

If a state has an IAC, then there is usually discretionary review, where the judges can decide whether to certify an appeal or else to reject without reviewing it. There are some exceptions (e.g., the death penalty and other serious cases), but this varies across states and is not perfectly categorizable. Without an IAC, there generally is mandatory review of appeals from trial courts. However, even with mandatory review, cases can be disposed of without an opinion being reported. We will come back to this issue in Section 5.1 below.

If the state supreme court accepts a case for review, the judges will rehear the case at oral argument and then review the submitted briefs for legal error. Each judge votes whether to affirm or reverse the lower decision. One of the majority judges, with the help of clerk staff, then researches and writes an opinion explaining the decision. In our data we cannot directly disentangle a judge's work from that of their clerks, so any performance effects we observe may be driven in part by changes in how clerks are hired or managed.

A notable feature of state supreme courts is that there is variation in how judges are selected and retained. There are three main systems: partisan elections, non-partisan elections, and merit commissions (where judges are nominated by a panel of senior judges and then appointed by the governor). A rich empirical literature has explored the relevance of these selection systems and shown that they influence judge decisions and quality (Shepherd, 2009b; Choi et al., 2010; Iaryczower et al., 2013; Canes-Wrone et al., 2014; Lim and Snyder, 2015; Kang and Shepherd, 2015; **?**). For the present paper on mandatory retirement, we control for the appointment process as a potential confounder.

---

incentives as judges who intend to stay in office (Shepherd, 2009a; Gordon and Huber, 2007; Shepherd, 2009b).

[9]Much of the existing literature on judges focuses on federal judges (e.g. Choi et al., 2013; Epstein et al., 2013).

[10]In rare cases when federal law (rather than state law) is pivotal, state supreme court decisions can be appealed to the U.S. Supreme Court. In two states (Texas and Oklahoma), there are separate high courts for criminal and civil matters.

Figure 1: Summary Histograms on Judge Age
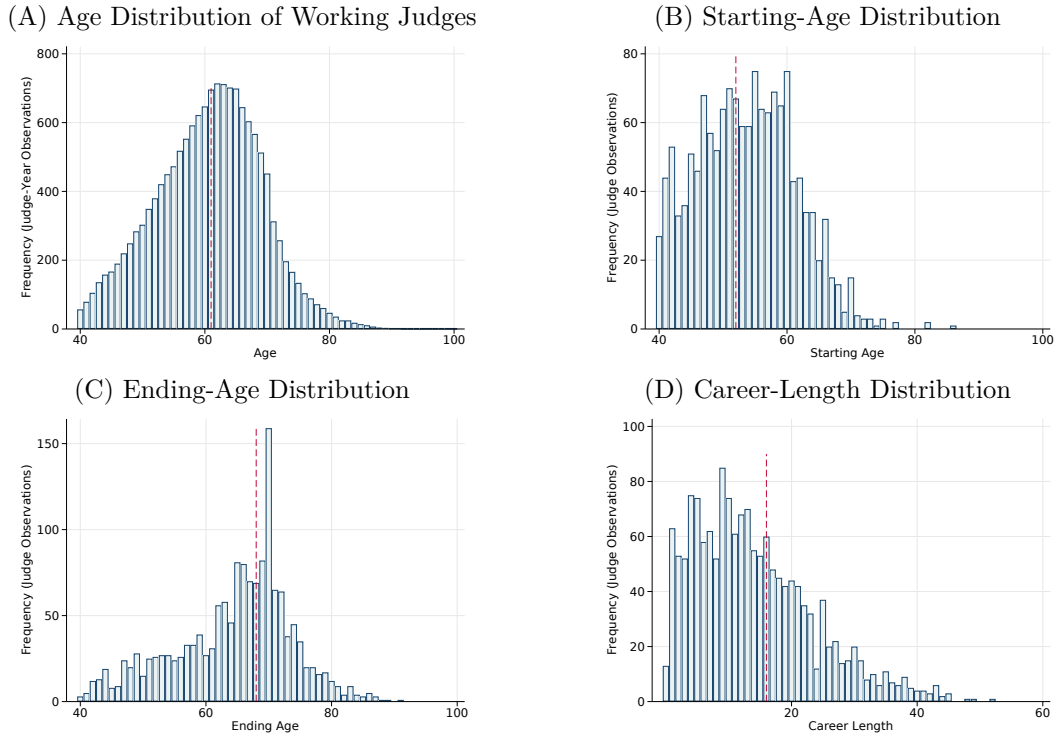
Notes. Distributions of age, start age, end age, and career length, as indicated. Vertical dashed line at median.

2.3 Judge Age and Retirement Decisions

The starting point for data collection is the existing data on state supreme courts from Ash and MacLeod (2015). A team of research assistants collected these data from a range of sources and built biographies for each judge in the sample. The key sources include state court web sites, judge obituaries, and Marquis Who's Who. Items that were unavailable from these sources were obtained through records requests or interviews of state court administration staff.

The dataset includes 1,558 state supreme court judges, for which we were able to obtain birth date information for all but 3 individuals. Beyond birth, we collected date information on judgeship starts, judgeship terminations, and judge deaths. We also collected information on how judgeships ended and previous and subsequent career information.

Figure 1 provides some visual descriptives on the age and retirement decisions of state supreme court judges. These graphs only include states that do not have a mandatory retirement age. First, Panel A shows the age distribution for all state supreme court judges

working between 1947 and 1994. We can see a wide range of ages of active working state supreme court judges. The other panels show the distributions of the starting age (Panel B), ending age (Panel C), and career length (Panel D). Judges tend to start in their position late in life (in their 50s) and work late as well (into their 70s). Appendix Figure A.3 shows the wide distribution over birth cohort decades in the dataset (from the 1860s to the 1950s).

The average retirement age of judges has not changed much over the time period 1947-1994 (Appendix Figure A.4, Panel B). This is somewhat different from the long-run changes in the broader economy which include an aging work force (Appendix Figure A.4, Panel A). The retirement age for judges has been consistently 4-5 years higher than the average worker over this time period. Meanwhile, the proportion of judges working in the private sector after leaving the court has increased over this time period (Appendix Figure A.5).

2.4 Mandatory Retirement Rules

Our second collection of data relates to the rules on mandatory retirement. Table 1 reports these rules and records on their reforms in the 1947-1994 period. As discussed in the introduction, we have 14 reforms to analyze in the empirical part. These include maximum ages of 70, 72, or 75.[11]

In some states, the mandatory retirement laws allow judges to finish out the year or term, or they can exempt incumbent judges. Even when the rules apply, judges often return as visiting senior judges with a partial caseload. That is, the chief justice has the right to keep on a retired judge in what the federal courts call "senior status" or "active retirement status". Appendix Table A.2 presents some information on how these rules work by state. When senior status is available under mandatory retirement, the control rights on the retirement decision switch to the court, which can decide to keep a high-performing judge.

A practical implication of senior status is that we often see judges in our data working past the official retirement age. This can be seen in Appendix Figure A.6, showing the probability of retirement at any given age, separately by the mandatory retirement rule. The blue line, with no mandatory retirement, is relatively smooth, peaking in the early 70s. The red line, with mandatory retirement at age 70, shows big increases for ages 69 and 70. We see corresponding jumps for retirement at 72 (green line) and 75 (purple line). We do see, however, that the rules are not perfectly enforced. Some judges stay on past the mandatory retirement age due to senior judge status.

Since the maximum age rule is not 100% binding, our estimates from the rule changes

---

[11]Our understanding of these laws is that they also apply to lower court appellate judges. Our data on judge performance is from the highest courts, but we will explore this issue in more detail below in Section 5.1.

should be interpreted as intention-to-treat effects.[12] Further, it could be that mandatory retirement causes selective attrition, in the sense that the chief judge invites back the better judges to work as visiting senior judges. We explore this issue in greater detail in Section 5.3.

2.5   Measuring Judge Performance

Our performance measures are constructed from published and unpublished state supreme court opinions for the years 1947 through 1994, obtained (along with some annotated metadata) from Bloomberg Law. The full sample includes 1,126,555 opinions. Of these, the sample of lead opinions with a named author comprises 404,928 observations. The metadata include records on forward citations up until 2012.

As outlined in the introduction, state supreme courts have several unusual features which are convenient for measuring judge performance: the job has not changed much over time, workload is held constant, there is no performance pay, judges have strong tenure, and there is negligible probability of promotion. These features are quite different from other high-skill jobs, such as medicine or science, where technology changes rapidly, professionals choose their workload, there is significant performance pay, strength of tenure varies, and there are opportunities for transfer / promotion.

Perhaps most importantly, in the context of state supreme courts, we can measure court-level and individual performance based on published judicial opinions, which represent the main output of their work. Our preferred measure of court performance is the total number of positive forward citations to a court in a year. Judges in a common-law system cite previous cases that are useful to their decision, therefore citations can be seen (on average) as an evaluation by judicial peers of a decision's quality or usefulness (Posner, 2008; Choi et al., 2010; Epstein et al., 2013; Ash and MacLeod, 2015). Note that this is not a measure of whether the decision is correct or not, which we do not observe. In some cases, citations reflect binding precedent and a judge does not have a choice about whether to cite a particular case or not.[13] But on average, more citations means that a case was more useful to future judges.[14]

---

[12]Results using variation in senior status rules are reported in Appendix Table B.5. We did not find differential impacts of reforms in states with formal senior-status policies. Appendix Figure A.7 shows that with voluntary retirement, judges are more likely to die within a year of leaving office.

[13]State supreme court decisions are binding precedent on the lower courts in the state. For the supreme court itself, those cases are precedents but can be overruled. For a given case, judges often have a number of relevant precedents to cite, and there is discretion in which ones are cited. For robustness, we also show results with out-of-state cites as the outcome. Those are persuasive precedents and would not be the result of self-cites or other local factors.

[14]Legal citations are therefore quite unlike academic citations, which can be included for many reasons, such as being on a similar topic or to curry favor with reviewers. In U.S. states, judges are appointed by the governor, selected by a merit commission, or elected by voters, rather than being selected by incumbent judges.

Using citations to measure judge performance is well-established in the literature (e.g. Choi et al., 2008; Posner, 2008; Choi et al., 2010; Epstein et al., 2013; Ash and MacLeod, 2015). In a discussion of the various influences on judges, Posner (2008) notes that judges care about their reputation as a good judge, and that includes being cited. Like academics, judges can observe the number of citations to their opinions and, as professionals, desire more of them. Consistent with this idea, Ash and MacLeod (2015) find that in response to exogenous reductions in workload, state supreme court judges use the time to produce more highly cited decisions, suggesting that judges personally value citations as a signal of professional performance. Further, relative citation counts across judges within court persist over time, consistent with a judge-specific ability component (Ash and MacLeod, 2021). [15]

Still, empirical analysis of judge performance using citations is not without controversy (e.g. Baker et al., 2009). Citations reflect not just the quality of the decision or opinion, but also the topic or ideological bias. We try to address these concerns by controlling for court and time factors, using multiple measures, and adjusting for legal topics. In our empirical analysis, we are implicitly assuming that it is the quality or productivity factor of citations that is most systematically changing in response to mandatory retirement reforms, rather than other unobservable contributors. Kaheny et al. (2008) find that the behavior of judges can change over their careers, but that judges are stable both in their early and later careers. In our case, appellate court judges are in their later career, and hence our findings are consistent with Kaheny et al.'s results on the stability of late career citation behavior.

Appendix Figure A.8 shows the distribution of the number of positive citations at the case level. We can see a steadily decreasing power-law distribution with a median of about 7 cites per case. While there is a large minority of cases with zero cites, overall the distribution is smoothly and slowly decreasing. Citations are relatively widely distributed, rather than showing up only in a few blockbuster cases.[16]

Citations are annotated as positive, negative, or distinguishing by the data provider. For the baseline, we look only at positive citations, as those are more likely to reflect a high-quality decision that influences future judges. As a more inclusive measure of performance that does not rely on subjective annotations about "positive", we use all cites (including

---

This means that, unlike scientists, choices to cite other particular judges would not be used to strategically influence future hiring or promotion decisions.

[15]If citations were due to other factors, then under random assignment of cases there should be no persistent variation in citations due to a judge fixed effect. Meanwhile, it remains a challenging open question whether or not more highly cited decisions have a positive effect upon measures of social well-being, such as economic growth or inequality.

[16]We show robustness of our results to dropping these blockbuster cases (formally, the top five percent of cases in each court-year by citations).

negative and distinguishing).[17]Appendix Figure A.8 shows that the distribution of citation counts per case is close to log normal, with some extra mass at zero. About half (49 percent) of cases have at least one citation. The average case is cited 13.8 times, of which 11.6 cites are positive.

As more restrictive measures of performance, we use discussion cites (where the case was discussed at length by the citing court) and out-of-state cites (only citations in other jurisdictions). Because state supreme court precedents are not binding in other states, out-of-state citations provide a stronger signal of legal usefulness or influence (Choi et al., 2010). In addition, out-of-state cites exclude self-cites. Finally, while judges might have time to network with colleagues in their own court to attract cites, lobbying factors are less pronounced for out-of-state cites.

The most recent cases in the data are from 1994, and the citations data were collected in 2012. Thus, we can measure a case's legal impact over at least 18 subsequent years. Given that the median delay between case publication and subsequent citation is ten years, we can ensure that most of the variation in case citations is captured. In the empirical distribution one observes a drop off in citation counts around year 18, but it is clear that the bulk of citations occur before then. The variation in the number of years where a case could be cited should not affect our empirical analysis because it will be absorbed by the year fixed effects. Nevertheless, we also show that our results hold when limiting to citations within ten years of the original opinion.

Our measure of productivity is a combination of quantity (number of opinions) and quality (number of citations per opinion). For analyzing the impacts of reforms, such as mandatory retirement, we feel that the impact measure is the most policy-relevant. To decompose the quantity and citations dimensions, we report as additional outcomes the number of opinions authored (quantity), and the number of positive citations per opinion (legal influence). Further, we provide evidence on how the reforms affect the overall caseload (including cases where no opinion was authored). The appendix includes analysis for a range of other outcomes, including measures of total written output in words, case law research, and number of discretionary opinions written. For assessing the relevance of the lower courts, we have an auxiliary dataset with information on the number of opinions per year published by state intermediate appellate courts (IACs), and the number of cites to those IAC opinions.

For all these variables, the baseline measure is the log of one plus the average value for the court in a year. Using the inverse hyperbolic sine transformation rather than the log-one-plus

---

[17]About 84% of citations are annotated as positive. The citation measure available from the data provider includes citations to concurring and dissenting opinions. From annotating a sample of cases, we found that, unlike the U.S. Supreme Court, citations to discretionary opinions are extremely rare in state supreme courts, comprising well under 1% of cites.

specification generates nearly identical estimates. Summary statistics for all of our outcome measures are reported in Appendix Table A.5. The main outcome is the log of positive citations, with additional measures meant to provide additional dimensions of performance. Appendix Figure A.9 shows the distributions for the main outcome variables.

2.6 Case Assignment and Case Characteristics

The citation count for a decision is a joint product of both the type of case and the judge's efforts. For example, cases that review the constitutionality of statutes will generally get more citations than summary habeas denials. When looking at the effects of reforms or aging on case influence, we have to check whether that is driven by changes in the composition of the caseload, rather than changes in a judge's work efforts.

A relevant institutional rule is how cases are assigned to judges, especially when comparing judges to their colleagues. There are three systems for case assignment, collected by the State Supreme Court Project and updated by Christensen et al. (2012).[18] Appendix Table A.4 lists the state supreme courts by rule. Discretionary assignment by the chief justice (the rule at the U.S. Supreme Court) is the minority rule followed in just 15 states. In 13 states, cases are randomly assigned by lottery to authoring judges. In the remaining 22 states, cases are assigned on a rotating system, with cases arbitrarily assigned to judges based on their order on the docket.[19] Christensen et al. (2012) found that in state supreme courts, case characteristics and judge characteristics are correlated even under random/rotation assignment. This is important for interpreting any effects, which could be due to changes in case types.[20]

At the decision level, we have data on the area of law of a case, as well as the related industries of a case. These are coded for each case by the data provider, and there may be up to three legal areas and three related industrial sectors for any particular case. Appendix Table A.3 reports summary tabulations of these characteristics. In the data, we include a vector of dummy variables for each area and sector, equaling one if the case is annotated as that area or sector. Because there are so many of these categories, including separate covariates for every category would almost saturate the dataset. Instead, we construct the

---

[18]See http://www.ruf.rice.edu/~pbrace/statecourt/. These rules were confirmed by Brace and Hall in the early 1990s and late 2000s. We tried to check the rules for earlier years. We could not get comprehensive information, but for those states where we could find information, it comported with the Christensen et al. (2012) information.

[19]There are complex rules across states that affect the rotation. Senior judges have fewer cases. Judges can occasionally recuse themselves. On appeal after remand, the same panel normally reviews a case. There can be exceptions for specialized cases such as those involving the death penalty.

[20]There is also evidence of strategic voting on the U.S. Supreme Court as a function of case characteristics (Johnson et al., 2005). We have a much larger sample of cases in this study, hence by controlling for case characteristics we are adjusting for case-specific strategic behavior and asking how retirement affects performance across many case categories.

first five principal components of this matrix of categorical variables, which explains 65% of the variance.[21] We use these factors as controls, but also look at how they respond to the treatments.

3 Empirical Approach

The empirical approach is generalized differences-in-differences (two-way fixed effects). The main estimating equation is

$$y_{st} = \alpha_s + \alpha_t + \rho M_{st} + X'_{st}\beta + \epsilon_{st} \tag{1}$$

where y_{st} is an annual outcome (e.g., age or log citations) for court s during year t. To control for time-invariant court characteristics that may be correlated with the retirement system and with performance measures like citations, we include court fixed effects α_s. To control for national trends in performance, we include year fixed effects α_t. The term X_{st} includes other covariates, to be described further below. Standard errors are clustered by state to allow correlation in the residuals over time across time in the same state.

The treatment indicator M_{st} equals one for years after introducing mandatory retirement. Fourteen states introduced a mandatory retirement age during the time period of our data (see Table 1). The coefficient ρ measures the corresponding causal effect of interest. Due to the length of the panel, in the baseline specification we estimate effects in an eight-year window before and after the reforms.[22] Formally, X_{st} includes an indicator equaling one for the baseline time window of eight years before and eight years after a change to the retention system. In turn, M_{st} is a dummy for the eight years after the change. Thus, as y_{st} is specified in logs, the estimates can be interpreted as the average proportional change in within-court performance for the eight years after the policy change relative to the eight years before the policy change and can be further differenced relative to the untreated comparison group captured by the year fixed effects. For additional flexibility, we also allow for state-specific treatment windows.[23]

A number of factors can motivate the introduction of a mandatory retirement age. Having older judges on the court could be problematic if they are ill and have to recuse themselves from more cases. With older judges, there is a higher risk of a judge having to step down

---

[21]Using more or fewer components does not change our estimates.

[22]In Appendix Table B.4 we use a six-year window, fourteen-year window, or no window (all years). The main effects are robust across these specifications.

[23]Note that this "window" approach to differences-in-differences is often implemented instead by including indicators for for the years outside the window (rather than inside, as done here). That approach would be equivalent in terms of the identification assumptions and generates identical results in our setting.

unexpectedly due to illness or even to die on the court. The courts have an odd number of judges to avoid deadlocks; having a missing judge increases the likelihood of a deadlock. Under voluntary retirement, there could be issues of inefficient strategic timing in order to influence the political affiliation of the successor. For our purposes, the most important potential driver is the perception that older judges are not delivering sufficiently high decision quality.[24] On the other hand, there have been some recent moves to repeal mandatory retirement rules, so the pressure to change operates in both directions.[25]

Consistent estimation of ρ requires parallel trends between treated states and comparison states (e.g. Bertrand et al., 2004). That is, the comparison states should provide a counterfactual for the trend in the treated states in the absence of the rule change. A potential threat to parallel trends in this setting is that the reforms may be implemented in response to preexisting trends in performance. To assess the parallel-trends assumption and the dynamics of the effect, we use a panel event-study specification. Formally, we estimate

$$y_{st} = \alpha_s + \alpha_t + \sum_{k=-8, k \neq -1}^{8} \rho_k M_{st}^k + X_{st}'\beta + \epsilon_{st} \tag{2}$$

where all of the items are as above, except the singular treatment indicator M_{st} is replaced with a sequence of event-study year indicators M_{st}^k. We let k index the years before and after treatment, with the year before treatment being left out as the comparison year. Then $\hat{\rho}_k$ give the dynamic effects on performance k years before/after the reform.[26] Parallel trends are consistent with $\hat{\rho}_k = 0$ for $k < -1$. For $k >= 0$, $\hat{\rho}_k$ will elucidate the dynamics of the differences-in-differences effect measured by $\hat{\rho}$ from Equation (1). In the reported results, we will test the joint significance of the coefficients $\hat{\rho}_k$ for $k < -1$ and $k \geq 0$ and report the F-test's associated p-value.

We add further covariates to the regressions in robustness checks. First, we allow for preexisting state trends in performance that may be confounded with the reforms by including state-specific linear trends. Further, we include initial-period court characteristics – institutional rules, case types, and judge age distribution – interacted with year fixed effects. These covariates allow for different trends along these different dimensions of state court characteristics. As shown in Appendix Table B.1, these initial-period variables are predictive

---

[24]See Posner (1995, ch.8) and Goldstein (2011) for discussions of the issue with respect to federal judges.

[25]See `http://ncsc.contentdm.oclc.org/cdm/ref/collection/judicial/id/440` for an update on the state situation. We looked for pre-trends in state-level measurable economic and political variables, such as state GDP and state Democrat vote share, and these were flat.

[26]As before, our event-study regressions include state-specific effect windows for the period covered by the event study coefficients. Note that a statistically similar approach is often taken by including indicators for for the years before and after the window. That approach is almost equivalent in terms of the identification assumption and generates identical results in our event study.

15

of whether a state is treated and also of the timing of the reform. However, our results are robust to selecting initial-period covariates via LASSO and interacting them with year fixed effects.

Even when we have evidence of parallel trends based on the event study estimates, exogenous timing of these reforms is a strong assumption. It could be, for example, that legislatures are responding to unpopular decisions made by judges. Or there could be contemporaneous shocks affecting performance trends as well as the demand for such reforms. These points should be taken into account when considering the external validity of our results.

Our setting features a staggered treatment, where different treated states adopt the reforms at different times. As noted in Goodman-Bacon (2018), in this setting the standard two-way fixed-effects estimator might not capture the average treatment effect on the treated due to the presence of previously treated states in the comparison group. We address this issue in two ways. First, we run the diagnostic from De Chaisemartin and d'Haultfoeuille (2020) to check for the presence of negative weights in our estimation sample. We find that there are zero negative weights, which provides reassurance that the two-way fixed-effects regression specification is valid in our setting.[27]

Second, we follow the method in Sant'Anna and Zhao (2020) and implement an alternative stacked-regression specification. For each treated state, we estimate the event study (2) dropping all other treated states, and including in the comparison group only the set of 18 states who had not adopted mandatory retirement before the end of our time period. Hence, no previously treated states are in the implied comparison group for any treatment cohort. As shown in Appendix Figure B.6, these adjusted event study estimates are the same as the results from our baseline model.

4   Main Results

4.1   Effect on Court Performance

Figure 2 shows event-study estimates (Equation 2) with log citations per court per year as the outcome. We use 8 years before the reform, up until 8 years after, as the event window. We can see a clear break and increase in log citations per court per year after treatment. The effect increases over a number of years, peaking at about a 37 percent increase relative to the counterfactual seven years after the reform. Meanwhile, the precise zeros before the reform support our identification assumption of parallel trends. According to a joint significance

---

[27]More specifically, we use the provided stata command *twowayfeweights* applied to our differences-in-differences regression. For any of the arguments for the diagnostic command, we have that zero ATT's receive a negative weight.

16

Figure 2: Event-Study Effect of Retirement Reform on Court Performance
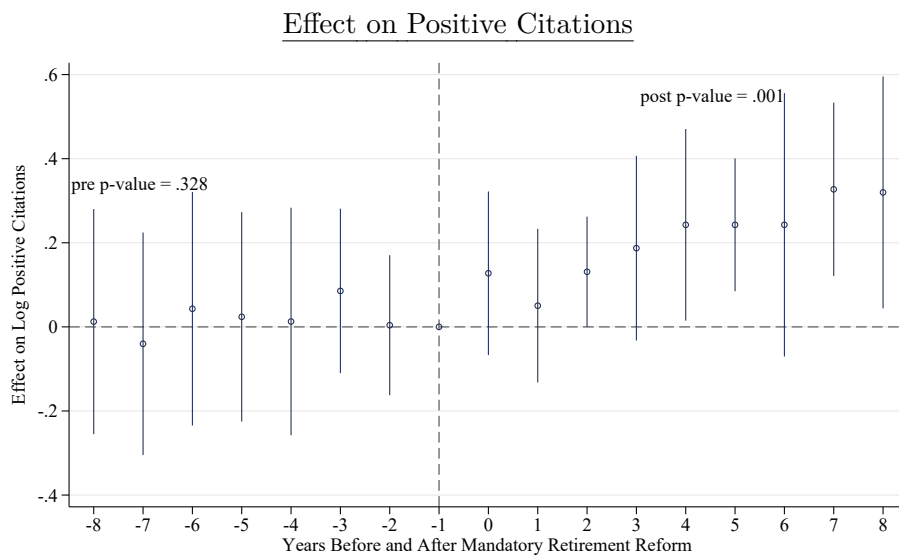
Effect on Positive Citations

Notes. Court performance before and after reforms implementing retirement ages of 70, 72 or 75. Time series is a coefficient plot from the event study regression (2), with coefficients estimated relative to the year before the reform. Regression includes court and year fixed effects. 95% confidence intervals constructed with standard errors clustered by court. Overlaid text reports the p-value from an F-test for the joint significance of the pre-reform indicators (top left) and of the post-reform indicators (top right).

Table 2: Effect of Mandatory Retirement Reform on Positive Citations

	(1)	(2)	(3)	(4)	(5)	(6)
	Effect on Log Positive Cites per Court-Year					
Retirement Reform	0.186*	0.243*	0.257*	0.330**	0.320**	0.279**
	(0.0807)	(0.0999)	(0.109)	(0.112)	(0.132)	(0.101)
Court FE, Year FE	X	X	X	X	X	X
Court Trends/Windows		X	X	X	X	X
Init Court Rules $\times$ Year FE			X	X	X	
Init Case Types $\times$ Year FE				X	X	
Init Age $\times$ Year FE					X	
Lasso $X^0 \times$ Year FE						X
N	2448	2448	2448	2448	2448	2448
R-sq	0.732	0.828	0.845	0.862	0.871	0.853

Notes. Differences-in-differences effect of mandatory retirement reform on log positive citations to a court's opinions in eight years after reform, relative to eight years before reform. Observation is a court and year. "Retirement Reform" is a treatment indicator for the eight years after the introduction of mandatory retirement. Court Treat Windows means court-specific treatment windows (eight years before and after reform). "Init X" $\times$ year FE means initial values are interacted with year. "Init Court Rules" includes a state's 1947 rules for judge selection/retention system, admin office, intermediate appellate court, number of judges, and term length. "Init Case Types" includes a court's 1947 average values for case characteristics (legal area and related industries). "Init Age" includes the initial mean and standard deviation for judge age on the court. "Lasso X^0" includes the set of initial-period covariates selected by LASSO as predictive of treatment or treatment timing (see Appendix Table B.1). Standard errors clustered by court in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

test, the pre-period coefficients are jointly not significant ($p = .328$), while the post-period coefficients are jointly significant ($p = .001$).

The results for the differences-in-differences (DD) regressions for Equation (1) are reported in Table 2. Across a range of specifications, there is a positive and significant effect of introducing a mandatory retirement age on a court's influence, as measured by citations. The result is consistently significant when adding controls for initial court rules, case types, and the age distribution, interacted with year. Further, results are robust to selecting initial-period covariates that are predictive of treatment or treatment timing with LASSO and then interacting those with year fixed effects.[28] Appendix Table B.2 adds an interacted time trend with the treatment indicator before and after the reform. While the pre-trend is statistically insignificant, there is a statistically significant positive trend after the reform, reflecting an increasing effect over time.

[28]See Appendix Table B.1 for the LASSO regressions and selected variables: average judge age, state population, and two of the principal components for case type. Appendix Table B.12 shows identical results using an inverse hyperbolic sine rather than log transformation. Appendix Figure B.2 shows the event study with court-specific trends.

Table 3: Effect of Mandatory Retirement Reform, Other Citation Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Cites in Levels		Within 10 years		All Cites		Discuss Cites		Out-of-State Cites	
Ret. Reform	386.2*	437.2**	0.283**	0.318**	0.186*	0.232**	0.214**	0.211**	0.174+	0.206*
	(192.7)	(181.9)	(0.0954)	(0.121)	(0.0779)	(0.0951)	(0.0687)	(0.0741)	(0.0915)	(0.0828)
Year / Court FE	X	X	X	X	X	X	X	X	X	X
Trends/Windows		X		X		X		X		X
N	2448	2448	2448	2448	2448	2448	2448	2448	2448	2448
R-sq	0.588	0.695	0.744	0.849	0.781	0.861	0.810	0.873	0.791	0.853

Notes. Observation is a court and year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. "Cites in Levels" means the outcome is not logged. "Within 10 years" is the log positive cites within eight years of an opinion. "All Cites" is the log number of all citations (positive, negative, and distinguishing) to a judge in a year. "Discuss Cites" is only the positive cites where the latter judge discussed the cited opinion. "Out-of-State Cites" is the count of number of positive citations from courts in other states. "Positive Cites" is the number of positive cites (in levels). Court Treat Windows means court-specific treatment windows (eight years before and after reform). Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

4.2   Robustness Checks

We undertook an array of checks to assess the robustness of the effect of the retirement reform on court performance. First, Table 3 shows that the effect holds for a number of alternative measures besides positive citations. The result is robust to using levels rather than logs (Columns 1 and 2), and note that the estimated effect in levels (60 positive forward citations) is about 24% of the mean in the eight years before the reforms (245 positive forward citations), similar to the effect in log points from above. The effect is robust when holding the number of years of forward citing cases constant to eight years of an opinion (Columns 3 and 4). We report effects on a more inclusive measure of influence (all cites, not just positive) in Columns 5 and 6. We use a more restrictive measure (discussion cites, where the previous case was specifically discussed and applied) in Columns 7 and 8. Finally, we report differences-in-differences estimates for out-of-state citations (Columns 9 and 10). We see positive effects for all of these alternative measures. The respective event study estimates are reported in Appendix Figure B.1.

Appendix Table B.3 reports effects on even more measures. Columns 1 and 2 show that there is still an effect for forward citations more than ten years after an opinion is published, indicating that there is also an increase in long-term influence. Presumably, judges cannot easily influence long-term citations via social or political actions. In Columns 3 and 4, we see that there is only a marginally significant effect on the proportion of cases by a judge that

has at least one positive citation. This is perhaps unsurprising since an average of 87% of authored cases have at least one cite. Third, Columns 5 and 6 show that the effect is robust and similar when dropping the top 5% of cases by number of cites in each court-year; this means that our effects are not driven by the production of more "blockbuster" cases.[29]

The appendix reports robustness checks along a number of additional margins. Results are robust to using Poisson regression with citation counts, rather than OLS (Appendix Figure B.3). The event study looks similar using only treated states in the sample (Appendix Figure B.4). Appendix Figure B.5 shows robustness to dropping each treated state individually.[30] Appendix Table B.10 shows that a qualitatively similar effect on judge performance is observed when each of the specified maximum judge ages (70, 72, or 75) are analyzed separately. Appendix Figure B.6 shows event-study estimates from the doubly robust regression approach based on Sant'Anna and Zhao (2020), designed to adjust for negative weighting due to staggered treatment (Goodman-Bacon, 2018). The adjusted event study estimates again show a positive effect and are similar to the results from the baseline model.

Next, in Appendix Tables B.8 we add additional time-varying controls to X_{ist} , which are probably "bad controls" or "colliders" in the sense that they could be affected by the retirement reforms. We include controls for the case characteristics, time-varying court rules (election system, number of judges, and government expenditures on the judiciary), and fixed effects for the number of years a judge has been on the court. Finally, we include the lagged dependent variable, which can perform better in panel data models with persistent shocks (Gentzkow et al., 2011; Caughey and Warshaw, 2018). All the results are robust. Further, the controls for the institutional rules – how cases are assigned, how judges are selected, and whether the state has an intermediate appellate court – can be interacted with time fixed effects and the results hold (Appendix Table B.9).

4.3 Effects on Quantity and Case Influence

Posner (1995) suggests that older judges often maintain high citations by reducing quantity in terms of the number of opinions they write. Ash and MacLeod (2015) provide evidence that these are separate choices in a judge's work and that judges tend to care more about citations at the margin. A clearly important question, then, is whether the effect on total citations is driven by total number of opinions, or citations per opinion.

---

[29]We look at additional non-cite outcomes in Appendix Table B.13, and report effects on work output (log number of words written), case-law research, and the rate of being overruled by the U.S. Supreme Court. There are no effects on these additional dimensions. There is an increase in the number of addendum opinions (log of the count of dissents and concurrences). Further, there is an increase in the dissent rate.

[30]Results are also robust to dropping all states that have an electoral reform within five years of the retirement reform.

Figure 3: Reform Effect on Quantity and Case Influence

Effect on Quantity: Log Number of Authored Opinions

Effect on Case Influence: Log Positive Citations per Opinion
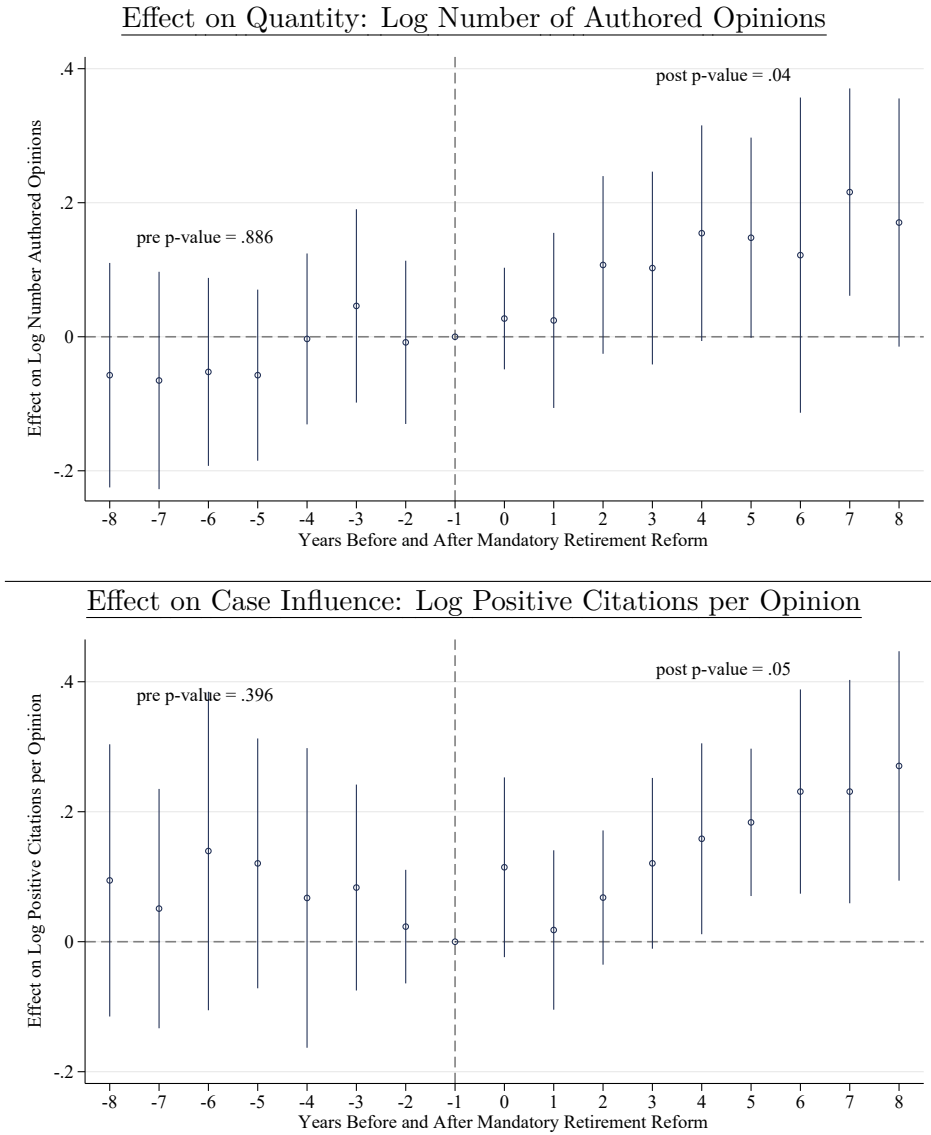
Notes. Court performance effects before and after reforms implementing retirement ages of 70, 72 or 75. Top panel outcome is log number of opinions by a judge in a year; bottom panel is average log positive citations per opinion. Time series is a coefficient plot from the event study regression (2), with coefficients estimated relative to the year before the reform. Regression includes court and year fixed effects. 95% confidence intervals constructed with standard errors clustered by court. Overlaid text reports the p-value from an F-test for the joint significance of the pre-reform indicators (top left) and of the post-reform indicators (top right).

We provide some evidence in this direction by running our analysis with these separated quantity and influence measures. Figure 3 shows the event study estimates. We can see a clear positive effect of the reform on both outcomes. Appendix Table B.11 provides the differences-in-differences estimates. We see that there is an increase in the number of opinions (Columns 1-3), reflecting an increase in quantity. In addition, the point estimates on the effect of cites per opinion are positive (Columns 4-6), even when limiting to out-of-state cites (Columns 7-9). Though, the effect on citations per case is less precisely estimated than the effect on quantity.

5 Analysis of Mechanisms

This section explores the different mechanisms by which maximum age rules could influence the productivity of state supreme courts.

5.1 Changes in the Caseload or Case Characteristics

A first possible mechanism for the effect of the retirement reform on judge performance is that it affects, or is confounded with, the caseload. It could be that the reform states get more appeals, or that the judges choose to review different types of cases post-reform. This section explores those possibilities.

First, we consider confounding changes in the lower courts. Most states have an intermediate appellate court (IAC) that supports the supreme court dealing with appeals. We can rule out that the effect is driven by a confounding establishment of IACs, as the main results are robust to controlling for the existence of those courts, even after interacting with year fixed effects (Appendix Tables B.8 and B.9).

Next, we would like to check for mediating effects on lower courts. In most cases, the mandatory retirement reforms affect IAC judges as well, so it could be that the effects are driven by changes in the effort of the IAC judges. Using the auxiliary dataset on cases in IACs, we regress measures of case output and citations in these courts on the reform treatments. Appendix Table C.1 shows that the reform does not affect the number of IAC opinions, the volume of words in IAC opinions, or citations to IAC opinions.

In Figure 3 above, we showed an effect on the legal output of the treated courts. This could be due to the judges choosing to do more work and publish more opinions, or it could be due to a change in the overall caseload. To check whether the latter is the case, we ran similar regressions but using the total caseload volume as the outcome (including cases where an opinion was not authored). There is no effect on total caseload (Appendix Figure C.1

Table 4: Effect of Retirement Reform, Relevance of Case Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Effect on Log Positive Cites			Share	Case
	Rand Assign		Disc Assign		Case Controls	Crim.Cases	Importance
Retirement Reform	0.169	0.314*	0.222**	0.154*	0.156*	0.0237+	0.0251
	(0.122)	(0.154)	(0.0845)	(0.0536)	(0.0802)	(0.0141)	(0.0505)
Year FE, Court FE	X	X	X	X	X	X	X
Court Trends / Windows		X		X	X	X	X
N	1750	1750	698	698	2448	2448	2448
R-sq	0.719	0.816	0.783	0.874	0.870	0.845	0.711

Observation is a court and year. "Rand assign" means regressions are limited to states with random or rotating assignment of judges to cases. "Disc assign" means regressions are limited to states where the chief judge chooses which judges are assigned to a case. "Case controls" means that the case characteristics are included as controls. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. Court Treat Windows means court-specific treatment windows (eight years before and after reform). "Share Criminal Cases" is the share on criminal law. "Case Importance" is the predicted citations to a case based on case characteristics (legal area and related industries). Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Panel A). That is, the observed quantity effects are driven by the court's choice to author more opinions, conditional on caseload. Further reflecting this point, the number of opinions per case also goes up (Appendix Figure C.1 Panel B).[31]

Now moving to case influence, judicial citations are a combined result of judge input and the importance of a case. We are therefore concerned with whether these effects are driven in part by the types of cases. Table 4 reports some regression results along these lines. First, in Columns 1 through 4 we show that the effect of the reform on legal influence is seen both in states with random case assignment and in states with discretionary case assignment. Appendix Figure B.7 shows the event study effects separately by case assignment rule, which are similar. Column 5 shows the main result is robust to including controls for case type (although note that this is an endogenous control).

To further unpack this dimension of the effect, we look at how the types of cases that the judges review changed after the reform. In Table 4 Column 6, we see that the share of criminal cases increased after the reform, reflecting a change in the composition of the caseload that judges review. However, in Column 7 we do not see a significant change in the importance of opinions, as indicated from the case characteristics. Therefore it appears that

---

[31]Note that the total number of cases, including those without a published opinion, is an imperfect proxy for the total caseload. We do not observe the number of appeals directly. It could be that some, or perhaps many, appeals are refused without any record in our dataset. Hence these results should be taken with some caution.

Table 5: First Stage Effect of Mandatory Retirement Reform on Judge Age

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Court-Level Distributional Statistics				
	Mean Judge Age		Age (Min)	Age (Q25)	Age (Median)	Age (Q75)	Age (Max)
Ret. Reform	-3.112**	-4.053**	-2.850*	-3.315*	-4.329**	-4.843**	-4.309*
	(0.909)	(1.143)	(1.644)	(1.693)	(1.203)	(1.566)	(1.887)
Year/Court FE, Trends	X	X	X	X	X	X	X
Additional Controls		X	X	X	X	X	X
N	2448	2448	2448	2448	2448	2448	2448
R-sq	0.525	0.654	0.569	0.5695	0.615	0.622	0.623

Notes. DD effect of mandatory retirement reform on judge age statistics in eight years after reform, relative to eight years before reform. Observation is a court and year. "Ret. Reform" is a treatment indicator for the eight years after the introduction of mandatory retirement. In Columns 3 through 7, dependent variables are computed at the court-year level. In particular, "Age (Min)" the minimum age, "Age (Q25)" the age at the 25th percentile, "Age (Median)" the median age, "Age (Q75)" the age at the 75th percentile and "Age (Max)" the maximum age. Column 1 includes court/year FE and court trends; the other columns include additional controls. as in Table 2 Column 5. Standard errors clustered by state in parentheses. $+$ p$<$.0.1, * p$<$0.05, ** p$<$0.01.

the portfolio of cases that the court accepts for review is not a major factor in the increase in court citations.

5.2 Aging and Judge Performance

The clearest mechanism for how mandatory retirement reforms could influence output is through the effect on the age distribution in the courts. Notwithstanding the salient examples mentioned in the introduction, the previous statistical studies of how aging influences judge performance have provided only mixed evidence (e.g. Posner, 1995; Smyth and Bhattacharya, 2003; Manning et al., 2004; Teitelbaum, 2006; Dimitrova-Grajzl et al., 2012). Here we show that the reform reduced judge age, and then we ask whether performance tends to decrease with age.

To begin, we substantiate the "first stage" to see if mandatory retirement has its intended effect on judge age. Regression estimates for Equation (1) with age statistics as the outcome are reported in Table 5. Relative to the pre-reform trend, judges after the reform are 3-4 years younger (Columns 1 and 2). Columns 3 through 7 look at the effect on distributional statistics using court-year-level aggregates. We see that the age effect is present across the age distribution. Both the youngest and oldest ends of the distribution shift down. This is consistent with the replacement of the oldest judges with overall younger judges. Presumably, this result is what the reforms were meant to achieve.

24

If the reforms reduce judge age, and judge age reduces performance, that would help explain the positive effect of the reform on performance. To get at the relationship between age and performance, we compare judges of different ages in the same court-year. The benefit of this approach is that judges are evaluated using exactly the same criteria – the number of cases and future citations for a specific setting and period. We can take advantage of the panel structure and control for heterogeneity in a judge's base-line performance. We can then ask if the relative ranking of the judge's performance changes with age.

This approach is implemented with the following specification. Assume a quadratic age model of performance variable y_{ist} for judge i working in court s at year t:

$$y_{ist} = \alpha_{st} + \alpha_i^0 + \gamma_1 A_{ist} + \gamma_2 A_{ist}^2 + X_{ist}'\beta + \epsilon_{ist} \tag{3}$$

where A_{ist} is the age (in years) for judge i in court s at t. Next, α_i^0 measures performance in a judge's first year on the court, providing a baseline value such that our estimates for the effect of age, $\hat{\gamma}_1$ and $\hat{\gamma}_2$, are relative to the individual's baseline. The main source of bias comes from the time-varying changes in the court work environment which are systematically correlated with age. Thus, we include a full set of court-year fixed effects. Therefore, any estimated coefficients are also relative to the court average in each year. Again, standard errors are clustered by state.

X_{ist} includes a number of additional items which we add in follow-up specifications. First, we have cohort fixed effects – indicators for each decade that the judge started on the court. This covariate is meant to rule out mechanical variation due to cohort differences across the time period. In the same vein, we have court-specific linear trends in judge starting cohort: formally, judge starting-year interacted with a court fixed effect. This allows for judges in different states to have a different confounding trend in starting year and performance.

Now we report the main results for variation of judge performance by age. The regression estimates from Equation (3) are reported in Table 6. There is a highly significant negative relationship between age and judge performance as measured by positive citations to a judge's opinions in a year. The estimate does not change that much when adding the judge's first-year as a baseline to the regression (Column 2), or when adding cohort fixed effects and state-specific cohort trends (Column 3). As seen in Column 5, the effect also holds using a rank percentile specification in work performance.

Columns 4 and 6 assume a quadratic model for age. Both of these columns show that the negative linear estimates from the other columns conceal a concave relationship, where the linear term is actually positive. The quadratic term is negative and significant, indicating that the negative relationship between age and performance accelerates at later ages. Taking

Table 6: Judge Age and Judge Performance

	(1)	(2)	(3)	(4)	(5)	(6)
		Log Positive Cites			Rank Percentile Cites	
Judge Age (Years)	-0.00907**	-0.00857**	-0.00734**	0.0302**	-0.00415**	0.0154*
	(0.00167)	(0.00133)	(0.00132)	(0.0106)	(0.000726)	(0.00757)
Age Squared				-0.000316**		-0.000165*
				(0.000094)		(0.000067)
Court-Year FE	X	X	X	X	X	X
First-Year Baseline		X	X	X	X	X
Cohort FE / Trends			X	X	X	X
N	14989	149819	14989	14989	14975	14975
R-sq	0.465	0.529	0.565	0.566	0.089	0.091

Observation is a judge working in a year. "Log Positive Cites" is log of positive cites to a judge in a year. "Rank Percentile Cites" means judges are uniformly distributed between zero and one based on number of positive citations within court-year (0 is lowest, 1 is highest). Court-Year FE is interacted court-year fixed effects. First-Year Baseline means a judge's value for the outcome in their first year on the court is included as a control. Cohort FE means fixed effect for decade that the judge started on the court. Cohort Trends means judge starting-year interacted with court fixed effect. Standard errors clustered by court in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

the quadratic model at face value, the coefficient estimates indicate that judge performance is maximized around age 47 (Column 6) to 48 (Column 4). Given that the median age of working judges is 61, these estimates suggest a negative age-performance relationship for almost all judges.[32]

Note that the reforms also influence the experience of judges as measured by their number of years on the supreme court. While age and experience effects are difficult to disentangle, we have some evidence that age seems to be more important. In Appendix Table C.2, first, we show that first-stage effect of the reform on experience is significantly negative but much smaller in magnitude than the effect for age – about a year and a half on average. Further, the effect of the reform on age holds while controlling for experience, but not vice versa. In Appendix Table C.3, we show that the effect of age on judge-level performance holds when controlling for experience, while the experience effect on performance is no longer significant when controlling for age.[33]

---

[32]Appendix Figure C.2 illustrates the dynamics of the aging effect on citations. The coefficient plots give age-group estimates, summarizing the difference in outcomes between the respective five-year age group and the baseline group of judges who are 44 or younger. The results in Panels A (log positive citations) and B (rank percentile in positive citations) are consistent with the quadratic regression model, showing an initial increase in performance and then a steady decrease starting in the late 40s or early 50s.

[33]To show further that age matters conditioning on experience, we look at how judge performance evolves over the first and last years on the court, but comparing judges by their starting and ending age. We can see the starting-age variation in Appendix Figure C.3. The judges in the different time series have the same

Can these age-performance effects explain the effect of mandatory retirement reform on performance? We take the estimates for the effects of retirement reform on judge age from Table 5 and the estimates for the life-cycle aging effect on citations from Table 6. The simplest approach is to take the reform's average effect on age (decrease by 4 years) and the life-cycle regression's average effect of age on citations (about 1 percent decrease per year). Together, these estimates suggest that the reform would increase cites by about 4 percent due to the aging effect. The estimated effect of the reform is much larger: about 24 percent. This is six times as large as the estimated effect due to changes in aging.

This prediction of 4 percent is too conservative, for at least three reasons. First, the retirement reform has a larger first-stage effect on age for the oldest judges (an almost five-year decrease above the 75th percentile). Given that the imposed retirement ages are at about the 80th percentile of the pre-reform age distributions, the reforms' effect at this spot in the distribution (a five-year decrease) is more realistic. Second, the age effect on citations is quadratic and more negative for older judges. Taking the average age for the treated judges (those above the maximum age at the time of the reforms), the predicted effect of age on citations is 1.6 percent. With these more realistic numbers, the predicted increase in citations due to the reform's age effect is 8 percent. Still, that is one-third the size of the estimated effect of the reform. Third and finally, there is noise in both estimates. For example, the lower 95% confidence bound for the reform effect on citations is about 10 percent, quite close to this predicted effect of 8 percent.

5.3   Selection on Entry and Exit

A more subtle possible mechanism is that mandatory retirement could change the types of judges working on the court. This section explores selection effects on entry and exit.

One possibility is that mandatory retirement attracts better judges to the bench, even conditional on age, through a selection effect. Theoretically, this seems unlikely, given the mandatory retirement reduces the lifetime income from a supreme court judgeship. Still, we test for it empirically following the method from Ash and MacLeod (2021) used to test for

amount of experience (years on the state supreme court), but they start at different ages. As before, we see a generally negative trend in performance over time. Moreover, we can see clearly that younger starting judges begin at a higher influence level than older starting judges that is maintained over time. In turn, Appendix Figure C.4 shows corresponding trends in the last years of a judgeship, separately by the ending age. We see the same pattern: judges that end their career at a younger age have higher end-of-career quality than judges who end their career at an older age. Overall, increased age reduces work performance even holding experience constant.

selection effects of election systems. That is, we estimate

$$y_{ist} = \alpha_{st} + \gamma R_{ist} + X'_{ist}\beta + \epsilon_{ist} \tag{4}$$

where α_{st} includes court-year fixed effects and R_{ist} is an indicator variable for joining the court after the mandatory retirement reform. This regression effectively compares two judges working on the same court at the same time, but who were selected under mandatory retirement versus voluntary retirement. This is a panel event study, but in starting year rather than current year. Appendix Table C.4 shows that, as expected theoretically, there is no effect of starting before/after a reform on current productivity.

Another interesting possibility is that retirement reform has selective effects on exit. That is, after the reform, judges of different quality leave at different times. There is actually an explicit mechanism for this, provided by senior judge status: judges who are above the maximum age cutoff can continue to come back with a partial caseload at the invitation of the chief justice. Presumably, a chief justice interested in a high-performing court will tend to invite back the highest-quality judges. An example of this mechanism is Ho et al. (2021), who show that the removal of mandatory retirement at 70 for law school faculty resulted in older faculty staying longer in the job, and therefore slower integration of more diverse faculty.

To test for this mechanism, we analyze the attrition patterns of judges, as measured by their highest attained age. We estimate

$$y_{ist} = \alpha_{st} + \gamma_1 A^{\mathrm{end}}_{ist} + \gamma_2 M_{st} + \gamma_3 A^{\mathrm{end}}_{ist} M_{st} + X'_{ist}\beta + \epsilon_{ist} \tag{5}$$

where A^{end}_{ist} is a judge's end age and M_{st} is a variable equaling one if a court has mandatory retirement. With the court-year fixed effects absorbed, we evaluate a judge's relative performance compared to the rest of the court, depending on ending age. The term X_{ist} include age fixed effects, which adjusts for variation across the life cycle and the impact of the reforms on the age distribution.

The results are reported in Table 7. Columns 1 through 3 use log positive cites as the outcome, while Columns 4 through 6 use log out-of-state cites as the outcome. In Columns 1 and 4, we see first that in a given court at a given year, judge performance is higher in terms of citations if that judge will end up staying longer. Hence, there is selective attrition by age where higher-cited judges tend to stay longer. In Columns 2 and 5, we see that there is a statistically stronger position selection effect due to attrition under mandatory retirement. That effect is even larger and more significant when allowing for end-age effects to vary across

Table 7: Evidence of Selective Attrition Due to Mandatory Retirement

| | (1) | (2) | (3) | (4) | (5) | (6) |
	Log Positive Cites			Log Out-of-State Cites		
Judge End Age	0.0183**	0.0175**		0.0146**	0.0130**	
	(0.00205)	(0.00212)		(0.00210)	(0.00218)	
End Age × Mand. Retirement		0.00582+	0.00737*		0.0114*	0.0132**
		(0.00319)	(0.00333)		(0.00435)	(0.00444)
Court × Year FE	X	X	X	X	X	X
Age FE	X	X	X	X	X	X
End Age × Year FE			X			X
N	14956	14956	14956	14956	14956	14956
R-sq	0.705	0.705	0.708	0.669	0.670	0.673

Observation is a judge working in a year. Judge end age is a judge's age upon leaving the dataset. Mandatory Retirement is a dummy for courts having mandatory retirement at 70 or 72. Standard errors clustered by state in parentheses. Regressions weighted by inverse career length to treat judges equally. + p<.0.1, * p<0.05, ** p<0.01.

years (Columns 3 and 6).

These regressions provide some evidence for the chief-justice mechanism. Some further supporting evidence is that the number of judges working on a court also tends to increase after the reform (Appendix Table C.5).[34] Under mandatory retirement, the chief justice can improve overall court performance by calling back the highest-performing judges to take on additional work.

5.4 Team Effects of Aging

All of the mechanisms explored so far work a the court-time level or at the judge level, but not at the judge-year level. That is, judge-year fixed effects would not adjust effects due to case characteristics, judge age, or judge selection. To explore these issues further, we revisit the retirement-reform regressions using judge fixed effects instead of court fixed effects. Formally, we use a judge-year dataset and estimate (1) including judge indicators α_i rather than court state indicators α_s. These regressions will summarize the within-judge effect of the reforms, focusing on those judges that stay on the court and are not removed due to old age. If the retirement reform works primarily through composition, the results with judge fixed effects

---

[34]The overall effect is not just driven by the number of judges, as the number of cites per judge also goes up (Appendix Table C.5).

Table 8: Within-Judge Effect of Reform on Log Cites (Judge Fixed Effects)

	(1)	(2)	(3)	(4)	(5)
	Effect on Log Cites per Judge-Year				
Retirement Reform	0.175*	0.173*	0.170*	0.221*	0.216*
	(0.0768)	(0.0726)	(0.0728)	(0.0872)	(0.0868)
Judge FE, Year FE	X	X	X	X	X
Court Trends/Windows		X	X	X	X
Init Court Rules × Year FE			X	X	X
Init Case Types × Year FE				X	X
Init Age × Year FE					X
N	14905	14905	14905	14905	14905
R-sq	0.675	0.683	0.691	0.700	0.707

Observation is a judge working in a year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. Court Treat Windows means court-specific treatment windows (eight years before and after reform). "Init X" × year FE means initial values are interacted with year. "Init Court Rules" includes a state's 1947 rules for judge selection/retention system, admin office, intermediate appellate court, number of judges, and term length. "Init Case Types" includes a court's 1947 average values for case characteristics (legal area and related industries). "Init Age" includes the initial mean and standard deviation for judge age on the court. Standard errors clustered by state in parentheses. + $p<.0.1$, * $p<0.05$, ** $p<0.01$.

would be zero. If the estimates are non-zero, we can reject that composition is the primary driver of how mandatory retirement improves performance.

Table 8 shows the differences-in-differences estimates for the specification with judge fixed effects. There is a positive and significant effect, about half to two-thirds of the magnitude of the full effect measures from Table 2. As before, the estimates are robust to the addition of a number of exogenous state characteristics interacted with year fixed effects. Appendix Figure C.5 shows the corresponding event study estimates, which tell the same story. Appendix C reports additional robustness checks along the same lines as Section 3. Again suggesting that the within-judge effects are positive and significant, yet not as large as the court-wide effects.

In Appendix Table C.7, we look at number of opinions and citations per opinion separately using judge fixed effects. We see that quantity goes up robustly, but not as much as the court-level effect in terms of magnitudes. The per-case effects are positive, yet not as robust across specifications. These results suggest that after the older judges retire, the remaining younger judges take on more work without a reduction in per-case influence.

Taken together, these estimates suggest that the effect of the mandatory retirement reform is driven not just by older judges being replaced. In addition to that, there is a positive effect on the younger judges who are working before and after the reform. That effect is consistent with a team effect of aging workers upon co-workers. The presence of older judges appears

to reduce the productivity of younger judges.

There could be two reasons for this team effect. First, it could be that the presence of older judges imposes unobserved work burdens on younger judges. That is, to maintain a reasonable work product level, the younger judges have to help the older judges with their work. The reforms remove the older judges and remove this burden, so then the younger judges have more time to provide higher performance. The within-judge increase in case citations is consistent with the evidence in Ash and MacLeod (2015) that judges prefer to increase citations at the margin when time pressure is relieved.

A second, perhaps subtler, explanation of this effect is an "equal workload" norm that, combined with the presence of relatively unproductive older judges, imposes a workload cap on younger judges. As we can see in Appendix Figure C.6, all judges are responsible for the same number of authored opinions regardless of age. Thus, as a team, the court would not want to increase the authoring workload above what can be handled by its lowest-performing judges. In particular, before mandatory retirement a court has to maintain a relatively low workload so that the oldest judges can keep up. The effect of the reforms in increasing the number of authored opinions, without increasing the overall caseload, is consistent with this equal-workload explanation. Under an equal workload norm with variation in productivity due to age, the best-performing younger judges have excess work capacity and can expend that on disposing cases that do not require an authored opinion.

6 Conclusion

The goal of this paper has been to measure the effects of retirement policies on court performance. Given that judges have low-powered incentives that do not explicitly link pay to performance, these factors likely have a significant impact on judge behavior. We find that mandatory retirement rules increase the performance of courts as a whole. We have evidence that this is driven in part by replacing older judges with younger ones who have higher performance. There is also a team effect of aging, where older judges reduce the performance of younger colleagues.

Our results are relevant to proposed retirement reforms in the state supreme courts As recently as 2016, Pennsylvanians voted (by a 2 percent ballot initiative margin) to increase the mandatory retirement age for state supreme court judges from 70 to 75. It is likely that other state-level changes in the retirement age will be proposed in the coming years. New York State has had mandatory retirement at 70 since 1869, the same rule that persists in many states. It is unlikely, given changes in health and other technologies, that the same maximum age could be optimal both now and in 1869.

In the case of federal judges, our results are relevant notwithstanding that imposing mandatory retirement is probably unconstitutional (Posner, 1995).[35] While the use of senior status and other incentives can put pressure on older judges (Choi et al., 2013), it is still the case that many federal judges stay on the bench past their prime (e.g. Goldstein, 2011). Our evidence provides more support for proposed reforms (including a constitutional amendment) to mandate retirement for the oldest federal judges. Further, our results suggest that an increased use of senior status to reduce the negative output effect of older judges on younger judges may be appropriate.

What does our evidence say about an optimal retirement age? In the case of judges, our estimates – along with some more structural assumptions – could be used to compute an optimal retirement age (Diermeier et al., 2005; Chetty, 2009). If replacing judges were costless and there were no spillovers between judges, then we could impose mandatory retirement at the age where performance peaks and judges begin to decrease in performance. Yet replacing judges is not costless, and there are certainly spillovers between judges. In addition, there appear to be separate effects of age and experience (Appendix Figure C.3), which might indicate that term limits are a reasonable substitute or complement to a mandatory retirement age. If one could estimate such costs and other factors, a structural approach could deliver a number of useful policy parameters.

Outside of the judiciary, these results may be useful to policymakers seeking to design better retirement policies for other high-skill jobs. When productivity decreases with age, mandatory retirement can increase productivity on average. In particular, the results are useful in an era where an aging workforce is resulting in large structural changes to the economy Acemoglu and Restrepo (2017). Though we have used the mandatory retirement rule change to estimate these effects, it does not imply that mandatory retirement is necessarily the optimal rule. That depends on many factors, including the counterfactual occupations for both the potential retirees.

---

[35]Federal judges "shall hold their offices during good behavior" (U.S. Const. Art. III Sec. 1).

References

Acemoglu, D. and Restrepo, P. (2017). Secular Stagnation? The Effect of Aging on Economic Growth in the Age of Automation. *American Economic Review*, 107(5):174–179.

Ash, E. and MacLeod, W. B. (2015). Intrinsic motivation in public service: Theory and evidence from state supreme courts. *Journal of Law and Economics*, 58(4).

Ash, E. and MacLeod, W. B. (2021). Reducing partisanship in judicial elections can improve judge quality: Evidence from U.S. state supreme courts. *Journal of Public Economics*, 201:104478.

Ashenfelter, O. and Card, D. (2002). Did the elimination of mandatory retirement affect faculty retirement? *American Economic Review*, 92(4):957–980.

Baker, S., Feibelman, A., and Marshall, W. P. (2009). The continuing search for a meaningful model of judicial rankings and why it (unfortunately) matters. *Duke Law Journal*, 58(7):1645–1666.

Ballesteros, S., Nilsson, L.-G., and Lemaire, P. (2009). Ageing, cognition, and neuroscience: An introduction. *European Journal of Cognitive Psychology*, 21(2-3):161–175.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. (2012). Does management matter? evidence from India. *Quarterly Journal of Economics*, 128(1):1–51.

Bloom, N., Lemos, R., Sadun, R., and Van Reenen, J. (2015). Does management matter in schools? *Economic Journal*, 125(125):647–674.

Bloom, N. and Reenen, J. V. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, 122(4):1351–1408.

Bonneau, C. W. and Cann, D. M. (2015). Party identification and vote choice in partisan and nonpartisan elections. *Political Behavior*, 37(1):43–66.

Canes-Wrone, B., Clark, T. S., and Kelly, J. P. (2014). Judicial selection and death penalty decisions. *American Political Science Review*, 108:23–39.

Canes-Wrone, B., Clark, T. S., and Park, J.-K. (2010). Judicial independence and retention elections. *Journal of Law, Economics & Organization*, 28(2):211.

Caughey, D. and Warshaw, C. (2018). Policy preferences and policy change: Dynamic responsiveness in the american states, 1936–2014. *American Political Science Review*, 112(2):249–266.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics*, 1(1):451–488.

Choi, S. J., Gulati, G. M., and Posner, E. A. (2008). Judicial evaluations and information forcing: Ranking state high courts and their judges. *Duke Law Journal*, 58:1313.

Choi, S. J., Gulati, G. M., and Posner, E. A. (2010). Professionals or politicians: The

uncertain empirical case for an elected rather than appointed judiciary. *Journal of Law, Economics, and Organization*, 26(2):290.

Choi, S. J., Gulati, M., and Posner, E. A. (2013). The law and policy of judicial retirement: An empirical study. *The Journal of Legal Studies*, 42(1):pp.111–150.

Choudhry, N. K., Fletcher, R. H., and Soumerai, S. B. (2005). Systematic review: the relationship between clinical experience and quality of health care. *Annals of Internal medicine*, 142(4):260–273.

Christensen, R. K., Szmer, J., and Stritch, J. M. (2012). Race and gender bias in three administrative contexts: Impact on work assignments in state supreme courts. *Journal of Public Administration Research and Theory*.

Coviello, D., Ichino, A., and Persico, N. (2014). Time allocation and task juggling. *American Economic Review*, 104(2):609–23.

Coviello, D., Ichino, A., and Persico, N. (2015). The inefficiency of worker time use. *Journal of the European Economic Association*, 15(5):906–947.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

Diermeier, D., Keane, M., and Merlo, A. (2005). A political economy model of congressional careers. *The American Economic Review*, 95(1):pp. 347–373.

Dimitrova-Grajzl, V., Grajzl, P., Zajc, K., and Sustersic, J. (2012). Judicial incentives and performance at lower courts: Evidence from slovenian judge-level data. *Review of Law & Economics*, 8(1):215–252.

Epstein, L., Landes, W. M., and Posner, R. A. (2013). *The Behavior of Federal Judges.* Harvard University Press.

Frederiksen, A. and Flaherty Manchester, C. (2019). Responding to regulation: The effects of changes in mandatory retirement laws on firm-provided incentives.

Gentzkow, M., Shapiro, J. M., and Sinkinson, M. (2011). The effect of newspaper entry and exit on electoral politics. *American Economic Review*, 101(7):2980–3018.

Goldstein, J. (2011). Life tenure for federal judges raises issues of senility, dementia. *ProPublica.*

Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. Working Paper 25018, National Bureau of Economic Research.

Gordon, S. and Huber, G. (2007). The effect of electoral competitiveness on incumbent behavior. *Quarterly Journal of Political Science*, 2(2):107–138.

Gustman, A. L. and Steinmeier, T. L. (1986). A structural retirement model. *Econometrica*, 54(3):pp. 555–584.

Gustman, A. L. and Steinmeier, T. L. (1991). The Effects of Pensions and Retirement Policies on Retirement in Higher Education. *American Economic Review*, 81(2):111–15.

Gustman, A. L. and Steinmeier, T. L. (2005). The social security early entitlement age in a

structural model of retirement and wealth. *Journal of Public Economics*, 89(2):441 – 463.

Hall, M. and Bonneau, C. (2006). Does quality matter? challengers in state supreme court elections. *American Journal of Political Science*, 50(1):20–33.

Ho, D. E., Mbonu, O., and McDonough, A. (2021). Mandatory retirement and age, race, and gender diversity of university faculties. *American Law and Economics Review*, 23(1):100–136.

Iaryczower, M., Lewis, G., and Shum, M. (2013). To elect or to appoint? bias, information, and responsiveness of bureaucrats and politicians. *Journal of Public Economics*, 97:230–244.

Johnson, T. R., Spriggs, J. F., and Wahlbeck, P. J. (2005). Passing and Strategic Voting on the U.S. Supreme Court. *Law & Society Review*, 39(2):349–378.

Kaheny, E. B., Haire, S. B., and Benesh, S. C. (2008). Change over Tenure: Voting, Variance, and Decision Making on the U.S. Courts of Appeals. *American Journal of Political Science*, 52(3):490–503.

Kang, M. S. and Shepherd, J. M. (2015). Judging judicial elections. *Mich. L. Rev.*, 114:929.

Keane, M. P. and Merlo, A. (2010). Money, political ambition, and the career decisions of politicians. *American Economic Journal: Microeconomics*, 2(3):pp. 186–215.

Kong, Y.-C. and Ravikumar, B. (2012). Earnings growth over a lifetime: not what it used to be. *Regional Economist*, 35:30.

Kritzer, H. M. (2011). Competitiveness in state supreme court elections, 1946–2009. *Journal of Empirical Legal Studies*, 8(2):237–259.

Kritzer, H. M. (2015). *Justices on the Ballot: Continuity and Change in State Supreme Court Elections*. Cambridge University Press.

Landes, W. M. and Posner, R. A. (2009). Rational judical behavior: A statistical study. *The Journal of Legal Analysis*, 1:775–831.

Lazear, E. P. (1979). Why is there mandatory retirement? *Journal of Political Economy*, 87:1261–64.

Lemieux, T., MacLeod, W. B., and Parent, D. (2009). Performance pay and wage inequality. *Quarterly Journal of Economics*, 124(1):1–49.

Lim, C. H. S. and Snyder, J. M. (2015). Is more information always better? party cues and candidate quality in u.s. judicial elections. *Journal of Public Economics*.

Lumsdaine, R. L. and Mitchell, O. S. (1999). Chapter 49 new developments in the economic analysis of retirement. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3, Part C, pages 3261 – 3307. Elsevier.

Manning, K. L., Carroll, B. A., and Carp, R. A. (2004). Does age matter? judicial decision making in age discrimination cases. *Social Science Quarterly*, 85(1):1–18.

Meng, A., Nexø, M. A., and Borg, V. (2017). The impact of retirement on age related cognitive decline–a systematic review. *BMC geriatrics*, 17(1):160.

Munnell, A. H. (2015). The average retirement age–an update. *Center for retirement research*, 1920:1960–1980.

Nixon, D. C. and Haskin, J. D. (2000). Judicial retirement strategies the judge's role in influencing party control of the appellate courts. *American Politics Research*, 28(4):458–489.

Posner, R. (2008). *How Judges Think.* Harvard University Press.

Posner, R. A. (1995). *Aging and old age.* University of Chicago Press.

Prendergast, C. (2001). Selection and oversight in the public sector, with the los angeles police department as an example. Technical report, National Bureau of Economic Research Working Paper: 8664.

Sant'Anna, P. H. and Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122.

Shepherd, J. M. (2009a). Are appointed judges strategic too? *Duke Law Journal*, 58(7):1589–1626.

Shepherd, J. M. (2009b). The influence of retention politics on judges' voting. *The Journal of Legal Studies*, 38(1):169–206.

Shi, L. (2009). The limit of oversight in policing: Evidence from the 2001 Cincinnati riot. *Journal of Public Economics*, 93(1-2):99–113.

Smyth, R. and Bhattacharya, M. (2003). How fast do old judges slow down?: A life cycle study of aging and productivity in the federal court of australia. *International Review of Law and Economics*, 23(2):141–164.

Stock, J. H. and Wise, D. A. (1990). Pensions, the option value of work, and retirement. *Econometrica*, 58(5):pp. 1151–1180.

Sullivan, D. and von Wachter, T. (2009). Job displacement and mortality: An analysis using administrative data. *Quarterly Journal Of Economics*, 124(3):1265–1306.

Tanaka, H. and Higuchi, M. (1998). Age, exercise performance, and physiological functional capacities. *Advances in exercise and sports physiology*, 4(2):51–56.

Teitelbaum, J. C. (2006). Age and tenure of the justices and productivity of the u.s. supreme court: Are term limits necessary? *Florida State University Law Review*, 34:161–182.

Appendix

A Background and Data

A.1 Aging and Retirement Decisions

Figure A.1: Performance vs. Age for Physical and Cognitive Tasks

(A) 10-K Running Time	(B) Cognitive Performance

Notes. Panel A from Tanaka and Higuchi (1998), showing 10-km race running times for men (white squares) and women (black squares) by age. Panel B from Ballesteros et al. (2009) showing how measures of different factors of intelligence or cognitive performance from psychological tests vary by age. The green lines measure processing speed, the gray lines working memory, blue lines long-term memory, and red lines world knowledge. All are decreasing into old age except knowledge.

Figure A.2: Earnings over the Life Cycle

Real Earnings over the Life Cycle for Workers Who Have At Least Some College Education
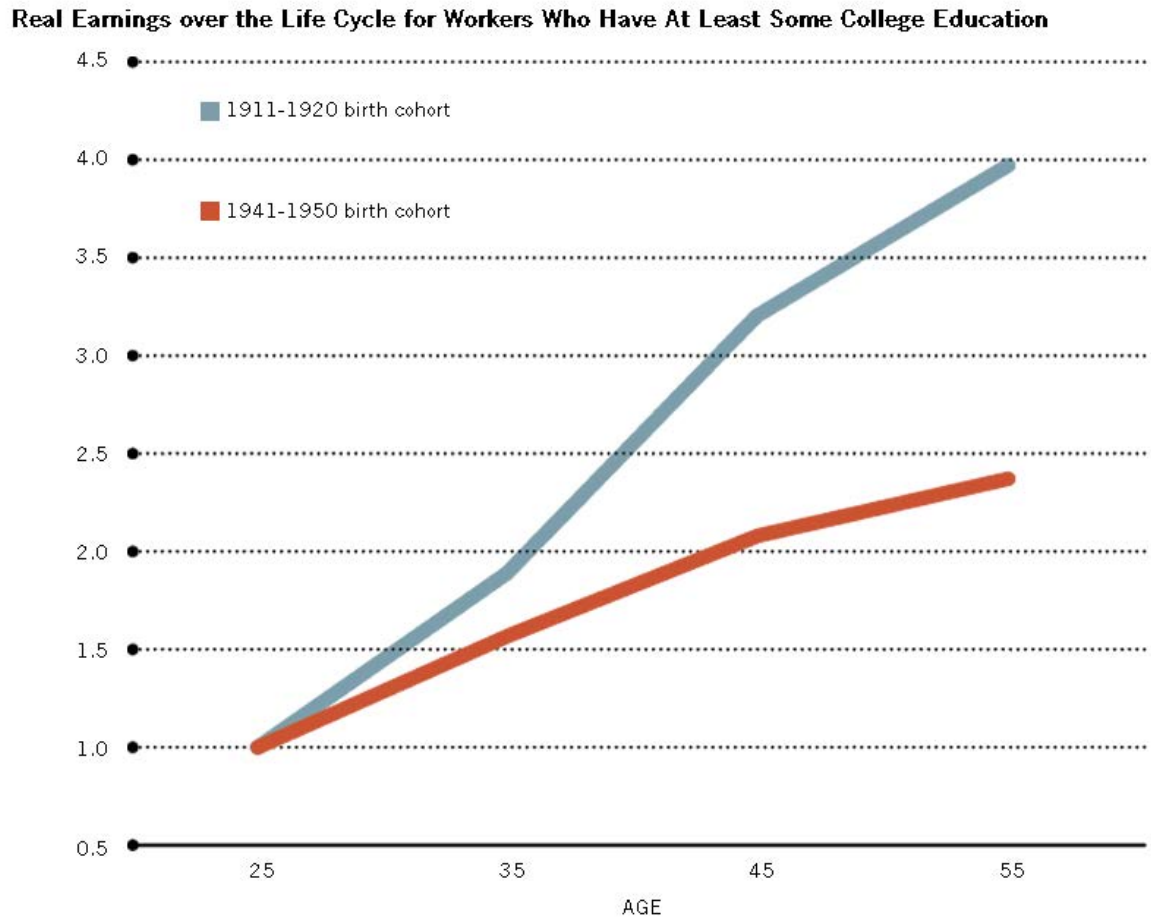
Real earnings over the life cycle for workers with college, 1911-1920 birth cohorts vs 1941-1950 birth cohorts. Source: Kong and Ravikumar (2012).

Figure A.3: Distribution of Judge Birth Decades
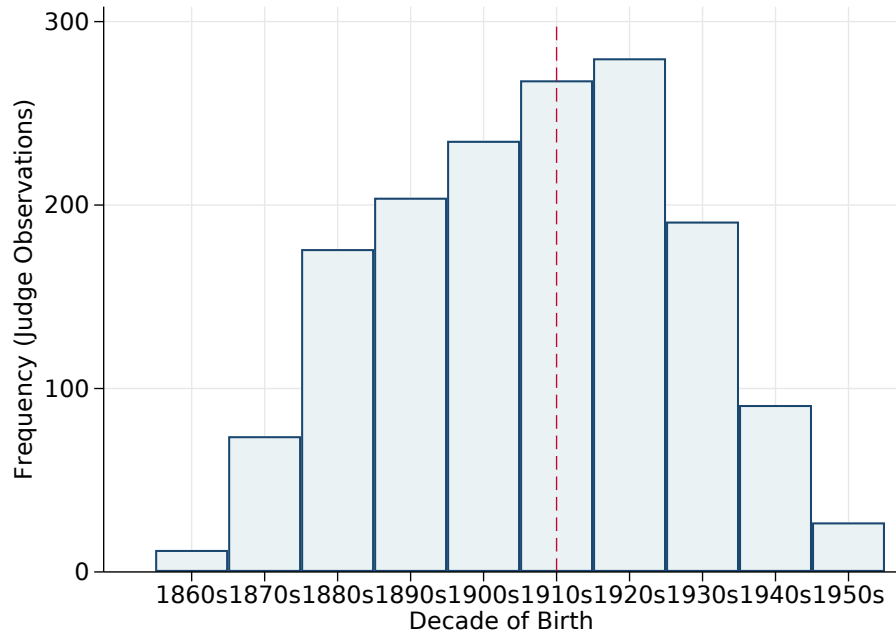
Number of judges in each birth decade cohort. Vertical dashed line at median.

Figure A.4: Average Retirement Age for U.S. Workers and Judges,

(A) Average Retirement Age for U.S. Workers, by Gender, 1962-2013
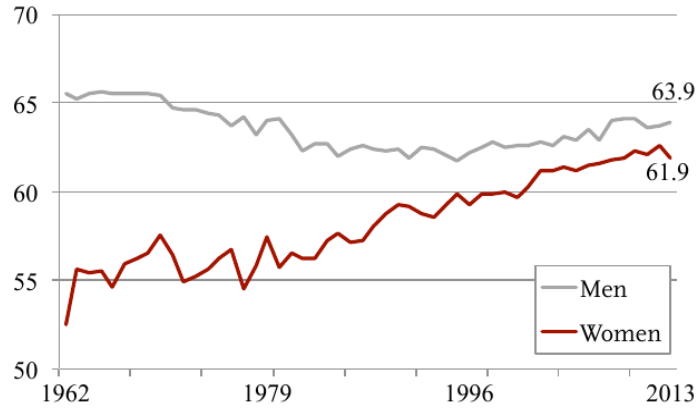
(B) Average Retirement Age for State Supreme Court Judges, 1948-1994
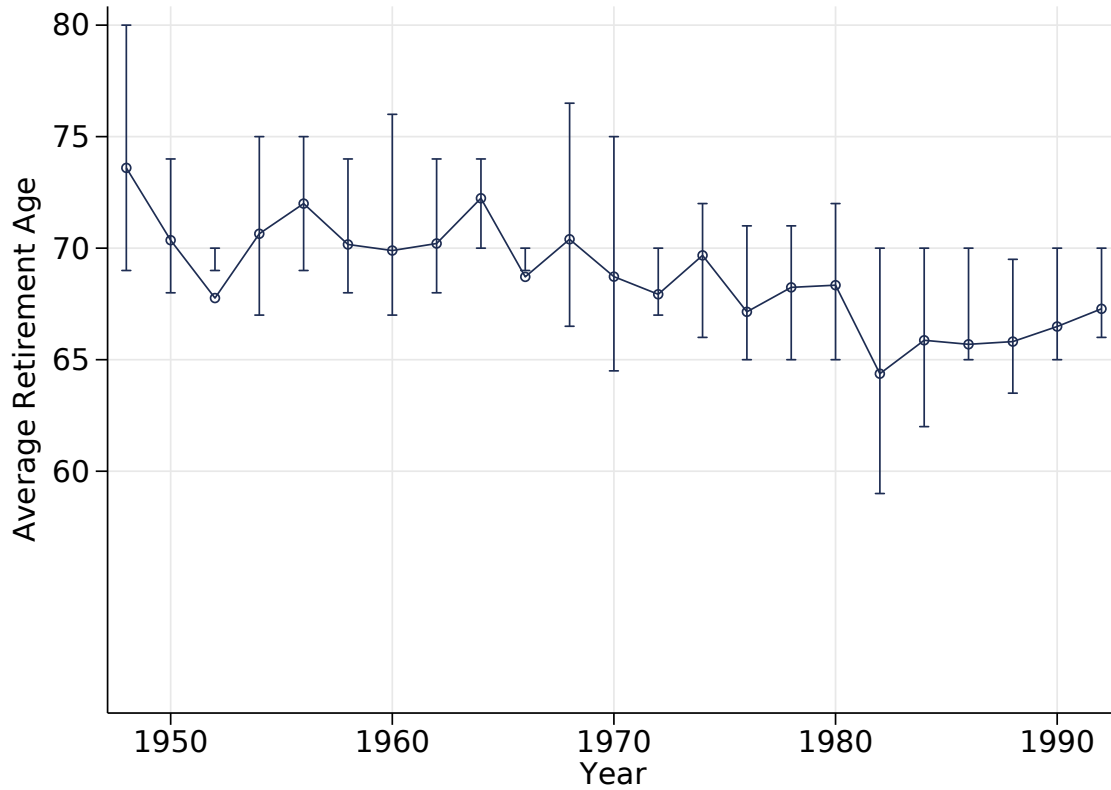
Panel (A): Average retirement age by gender for U.S. workers, computed from CPS by Munnell (2015). Panel (B): Average retirement age of state supreme court judges, by year. Error spikes give 25th and 75th percentiles.

Figure A.5: Post-Judgeship Careers
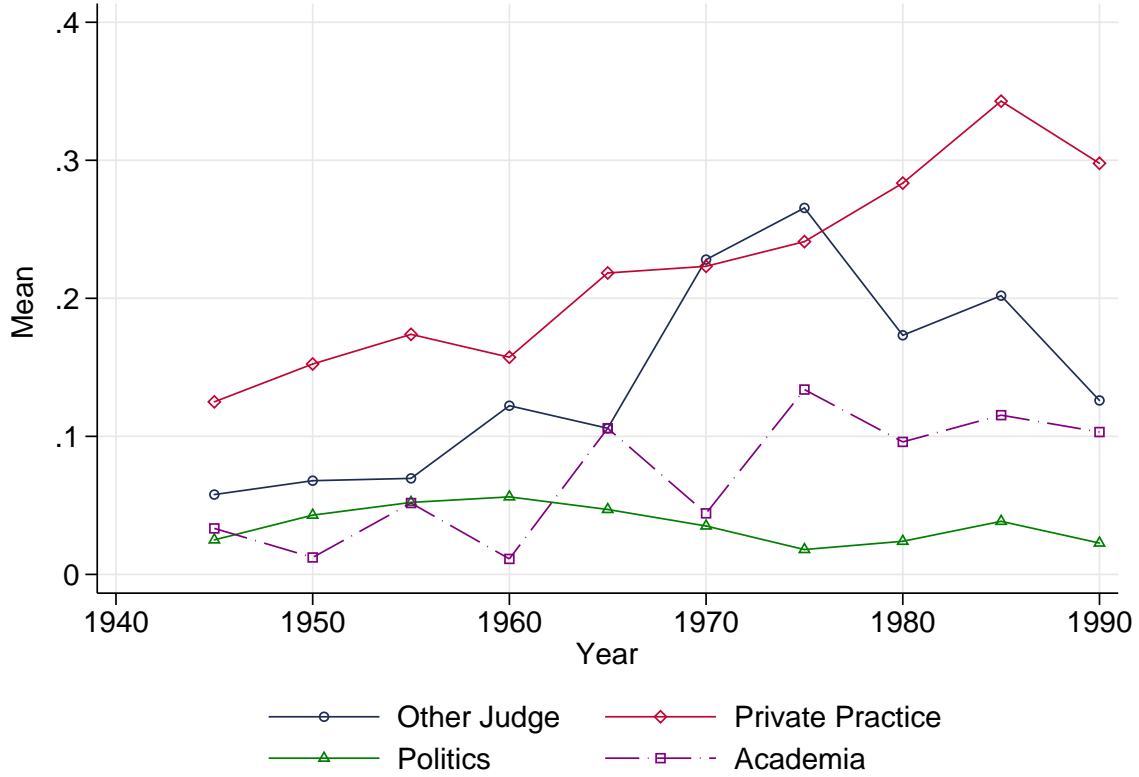
Proportion of judges with documented careers after their state supreme court judgeship, including other judgeship, private practice, politics, and academia. Plotted by five-year bins.

What do judges do after retirement? Figure A.5 shows the trends in these career choices. At the beginning of the sample, few judges took on more work after their judgeship. That has become more common in recent years. If they do take another career, it is usually in private practice as an attorney.

A.2 Mandatory Retirement

Table A.1: Tabulations on Treatment and Control Judges

Reform	# of Reforms	Number of Obs (Judge-Year)			Number of Obs (Court-Year)			Number of Judges		
		Controls	Treated	Total	Controls	Treated	Total	Controls	Treated	Total
Retire at 70	9	11843	1390	13233	2014	202	2216	1291	172	1463
Retire at 72	2	13233	511	13744	2216	62	2278	1412	51	1463
Retire at 75	5	13744	1266	15010	2278	170	2448	1448	143	1591
Any Reform	16	11843	3167	15010	2014	434	2448	1291	366	1657

Notes. Summary tabulations on Retirement Reform Judges. The column gives the number of courts that experience a change in the retirement rule. The second set of columns gives the number of judge-year observations in the control and treatment groups (and total) when a change in retirement rule occurs in that year. The third list of columns gives the number of judges in these respective groups.

Appendix Table A.1 provides tabulations on the relevant treatment variation in the data for mandatory retirement reforms. The first set of columns gives the number of judge-years where at least one treated (selected post-reform) and one control judge (selected pre-reform) is on the court that year. The second set of columns gives the number of court-year observations in the control and treatment groups (and total). The third list of columns gives the number of distinct judges in these respective groups.

Figure A.6: Retirement Rates by Age, by Mandatory Retirement Age
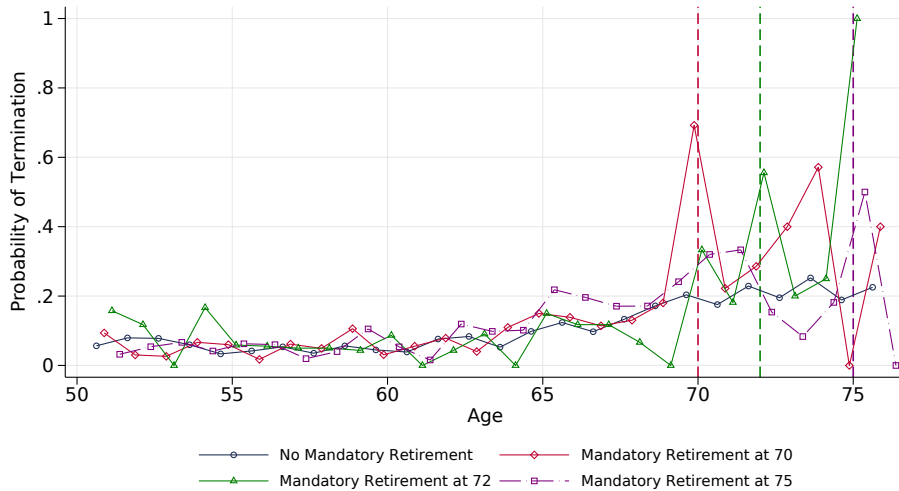
Notes. Probability that a judge retires at a particular age, conditional on working at that age. Plotted separately by mandatory retirement rule.

Appendix Figure A.7 shows further evidence on judges' life cycle, including some statistics on the timing of judge deaths. Panel A shows that the judges have relatively long lifespans, with most living into their eighties. Panel B looks at how judge retirement is related to judge longevity, separately for mandatory retirement (left panel) and voluntary retirement (right panel). The figure shows that with voluntary retirement, judges are much more likely to die within a year of leaving office. This difference supports the idea that mandatory retirement is an impactful policy: judges are more likely to stay in their jobs until death under voluntary retirement. On the other hand, there is still a relatively high chance of death in the first year out of office under mandatory retirement (left panel), which may hint at a causal impact of retirement on mortality (as found in Sullivan and von Wachter, 2009). This is a promising area for future work.[36]

---

[36]Meng et al. (2017) review the mixed evidence about the cognitive impacts of retirement, noting that there is a "major knowledge gap in regards to the impact of retirement on cognitive decline."

Figure A.7: Mandatory Retirement and Deaths on the Job

(A) Judge Age-at-Death Distribution
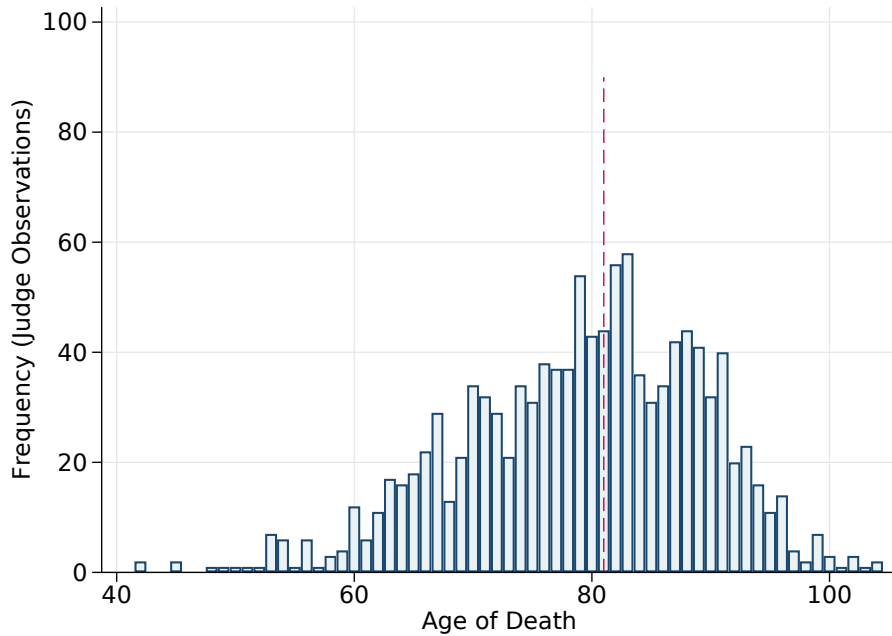
(B) Distribution of Years Between Termination and Death, With/Without Mandatory Retirement
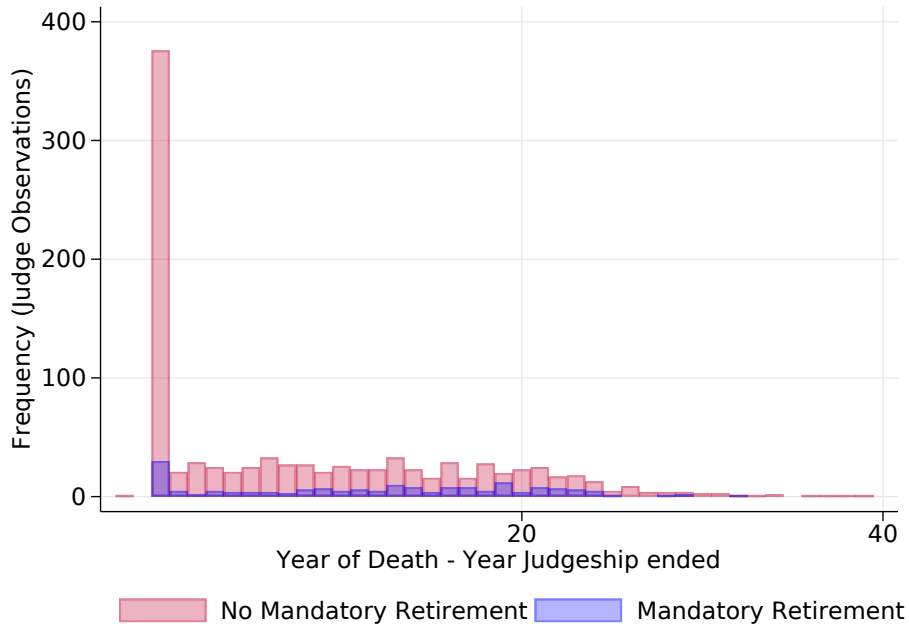
Notes. Distributions of judge age at death (panel A, vertical dashed line at median) and death year minus year judgeship ended (panel B).

Table A.2: Rules on Judge Senior Status, by State

State	Retirement Age	Seniority	Note
Alabama	70	Yes	Supernumerary judge
Alaska	70	No	Only to work on temporary assignments
Arizona	70	No	Only to work on temporary assignments
Arkansas	none	—	No retirement benefits if seek reelection past age 70
California	none	—	
Colorado	72	No	
Connecticut	70	Yes	State referee
Delaware	none	—	
Florida	75	No	Allows temporary assignments
Georgia	none	—	
Hawaii	70	No	
Idaho	none	—	
Illinois	none	—	Retirement Act for age 75, declared unconstitutional in 2009
Indiana	75	No	
Iowa	72	Yes	
Kansas	75	No	
Kentucky	none	—	
Louisiana	70	No	
Maine	none	—	
Maryland	70	No	
Massachusetts	70	No	
Michigan	70	No	
Minnesota	70	No	
Mississippi	none	—	
Missouri	70	Yes	Senior judge
Montana	none	—	
Nebraska	none	—	
Nevada	none	—	
New Hampshire	70	No	
New Jersey	70	No	
New Mexico	none	—	
New York	70	Yes	May serve after 70 until 76
North Carolina	72	No	Only to work on temporary assignments
North Dakota	none	—	
Ohio	70	No	
Oklahoma	none	—	
Oregon	75	No	Legislature may ask retired judges to work on temporary assignments
Pennsylvania	75	Yes	Senior judge
Rhode Island	none	—	
South Carolina	72	—	
South Dakota	70	No	
Tennessee	none	—	
Texas	75	No	Conditions may vary based on Art. 5 of Texas Constitution
Utah	75	No	
Vermont	90	No	
Virginia	73	No	
Washington	75	No	
West Virginia	none	—	
Wisconsin	70	Yes	Can serve a judge on a temporary basis
Wyoming	70	Yes	

A.3 Case Assignment

Table A.3: Summary Statistics on Area of Law and Related Industries

Area of Law	Freq.	Percent	Related Industrial Sector	Freq.	Percent
Criminal Law	191810	21.85	Real Estate	28527	13.64
Civil Procedure	74757	8.52	Law Enforcement	10758	5.14
Evidence	66377	7.56	Automobiles	10206	4.88
Torts	57915	6.6	Insurance	9158	4.38
Damages & Remedies	45073	5.14	Tax	8509	4.07
Contracts	40888	4.66	Construction & Engineering	6332	3.03
Real Property	36408	4.15	Worker's Compensation	5397	2.58
Constitutional Law	34038	3.88	Banking	4917	2.35
Family Law	32191	3.67	Legal & Compliance Services	4682	2.24
Worker's Compensation	22955	2.62	Automobile Insurance	4124	1.97
Insurance Law	19375	2.21	Property Management	4108	1.96
Administrative Law	18264	2.08	Transportation	3890	1.86
Wills, Trusts & Estates	18179	2.07	Child Welfare	3689	1.76
Tax & Accounting	16978	1.93	Employment Services	3679	1.76
Employment Law	14601	1.66	Health & Medical	3478	1.66
Habeas Corpus	13426	1.53	Oil & Gas	3189	1.52
Appellate Procedure	13140	1.5	Railroads	2777	1.33
Professional Responsibility	12052	1.37	Hospitals	2719	1.3
Motor Vehicles & Traffic Law	9644	1.1	Education	2586	1.24
Land Use Planning & Zoning	9122	1.04	Trucking	2097	1
Government	8942	1.02	Bridges & Roads	1751	0.84
Mortgages & Liens	7531	0.86	Agriculture & Farming	1729	0.83
Landlord & Tenant	5499	0.63	Mortgage Lending	1680	0.8
Construction Law	4997	0.57	Manufacturing	1612	0.77
Elections & Politics	4972	0.57	Real Estate Agents & Brokers	1573	0.75
Eminent Domain	4943	0.56	Unions	1485	0.71
Labor Law	4790	0.55	Financial Services	1469	0.7
Government Employees	4773	0.54	Judiciary	1448	0.69
Debtor Creditor	4260	0.49	Politics	1336	0.64
Employee Benefits	4208	0.48	Teachers	1300	0.62
Medical Malpractice	4113	0.47	Medical Procedures	1273	0.61
Personal Property	3994	0.46	Public Works	1223	0.58
Corporate Law	3958	0.45	Life Insurance & Annuities	1155	0.55
Negotiable Instruments	3843	0.44	Apartment Leasing	1127	0.54
Education Law	3803	0.43	Mining & Natural Resources	1115	0.53
Banking & Finance	3380	0.39	Drug Trafficking	1105	0.53
Alcohol & Beverage	3213	0.37	Sewer & Water	990	0.47
Civil Rights	3138	0.36	Electric	985	0.47
Health Law	2950	0.34	Water & Sewer	972	0.46
Transportation Law	2839	0.32	Physicians	966	0.46
Partnerships	2333	0.27	Firearms & Weapons	962	0.46
Natural Resources	2301	0.26	Motorcycles	919	0.44
Legal Malpractice	2285	0.26	Water	904	0.43
Products Liability	2280	0.26	Food & Beverage	888	0.42
Alternative Dispute Resolution	2144	0.24	Commercial Real Estate	883	0.42
Communications & Media	2048	0.23	Property & Casualty Insurance	854	0.41
Environmental Law	1857	0.21	Administration	837	0.4

Table A.4: Case Assignment Rules on State Supreme Courts

Discretionary	Random	Rotating
Arizona	Idaho	Alaska
California	Louisiana	Alabama
Colorado	Mississippi	Arkansas
Connecticut	New Hampshire	Florida
Delaware	New York	Georgia
Hawaii	Ohio	Iowa
Indiana	South Dakota	Illinois
Kansas	Tennessee	Maine
Kentucky	Texas	Minnesota
Massachusetts	Virginia	Missouri
Maryland	Washington	Montana
New Jersey	Wisconsin	North Carolina
Oregon		North Dakota
Pennsylvania		Nebraska
Wyoming		New Mexico
		Nevada
		Oklahoma
		Rhode Island
		South Carolina
		Utah
		Vermont
		West Virginia

List of states by rules for case assignment in state supreme courts. Rules collected by Christensen et al. (2012).

A.4 Judge Performance

Table A.5: Summary Statistics on Outcomes

A. Court-Year Data

	Levels		Logs	
	Mean	S.D.	Mean	S.D.
Positive Cites	1825.6	1534.5	7.185	0.904
All Cites	2182.4	1823.1	7.348	0.922
Out-of-State Cites	425.3	879.9	5.551	0.945
Discussion Cites	654.7	575.9	6.139	0.905
# of Opinions	454.6	616.6	5.664	0.894
Cites per Opinion	5.81	3.56	1.775	0.559
# of Words Written	55791.0	25301.6	10.82	0.468
Addendum Opinions	39.24	44.13	3.154	1.136

B. Judge-Year Data

	Levels		Logs	
	Mean	S.D.	Mean	S.D.
Positive Cites	297.7	273.5	5.360	0.954
All Cites	355.9	322.5	5.521	0.982
Out-State Cites	44.60	81.27	3.297	0.998
Discussion Cites	66.77	53.54	3.912	0.878
# of Opinions	25.73	15.86	3.131	0.565
Cites per Opinion	13.05	12.52	2.419	0.668
# of Words Written	56352.7	32538.4	10.77	0.622
Addendum Opinions	6.400	9.253	1.492	0.995

Notes. Summary statistics (mean and standard deviation) on judicial opinion outcomes, at the court-year level (Panel A) and judge-year level (Panel B).

Figure A.8: Distribution of Citations by Case
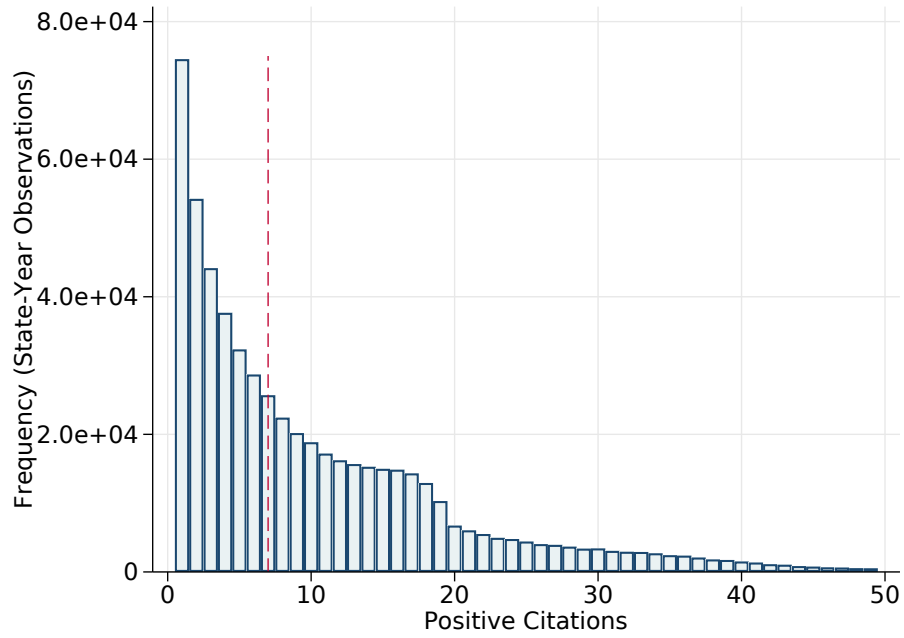
Notes. Histogram of the number of positive cites per case in the dataset. Vertical dashed line at the median.

‘

Figure A.9: Distributions of Outcome Variables - Court-Year

(A) Log Positive Cites

(B) Log Out-of-State Cites

(C) Log # of Opinions

(D) Log Cites per Opinion

Notes. Histograms of court-year performance measures. Vertical dashed line at median.

B   Additional Analysis of Mandatory Retirement Reforms

Table B.1: Balance Tests for Retirement Reform Treatments

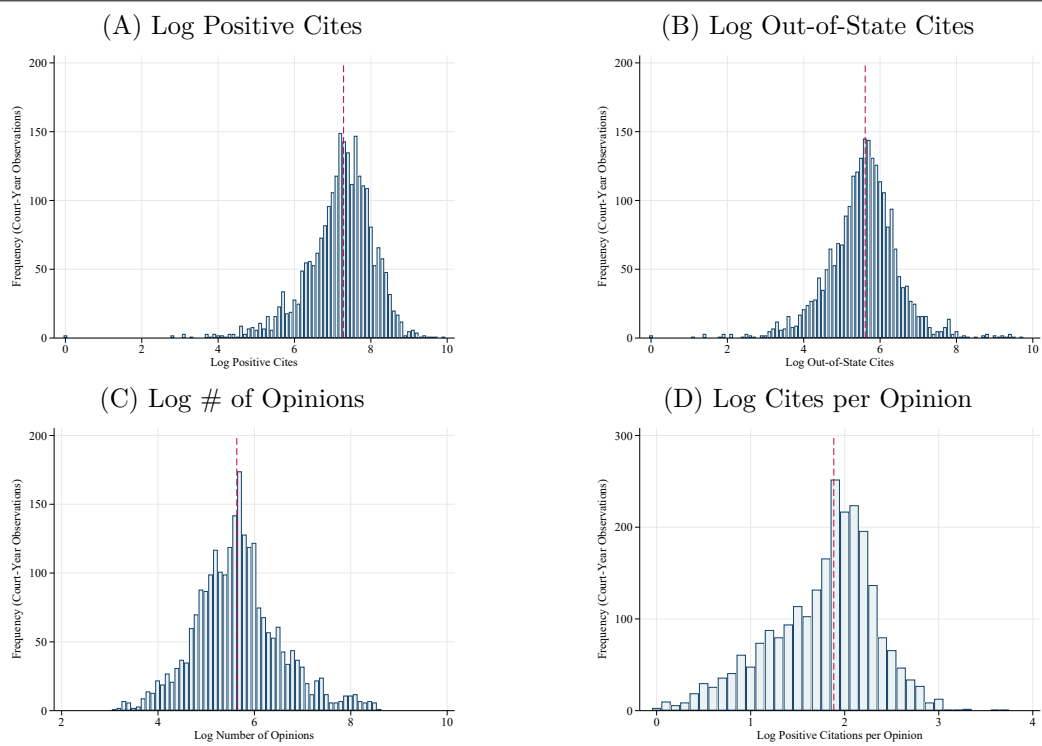| Init X^0	(1)	(2)	(3)	(4)
	Is Reform State		Reform Year	
Judge Age	0.0300*	0.0358*	-0.374	
	(0.0148)	(0.0134)	(1.258)	
Log Population	0.0147		-12.59+	-6.893*
	(0.0744)		(5.764)	(2.798)
Log Income per Capita	0.856*		-17.47	
	(0.399)		(27.03)	
Repub Governor	-0.296*		3.501	
	(0.146)		(11.95)	
Log Positive Cites	-0.0131		0.946	
	(0.0860)		(9.628)	
Case Type PCA 1	0.0288		2.138	
	(0.109)		(4.901)	
Case Type PCA 2	-0.161+		-2.274	
	(0.0883)		(8.196)	
Case Type PCA 3	0.0927		2.751	
	(0.104)		(5.117)	
Case Type PCA 4	0.0903		14.59	13.13*
	(0.108)		(12.31)	(4.832)
Case Type PCA 5	-0.259*	-0.176+	-6.136	
	(0.118)	(0.0924)	(11.96)	
N	50	50	16	16
R-sq	0.302	0.156	0.703	0.514

Notes. Regression coefficients from regressing "is this a treated state" (Cols 1-2), or the year of the reform (Cols 3-4), on initial-period covariates (that is from 1947). Covariates include the average judge age on the court, Log population of the state, Log income per capita, an indicator for having a Republican governor, log positive cites to the court, and the 5 principal components on the areas of law of the cases that year. Population, income, and governor party come from the IPPSR's Correlates of State Policy dataset. Columns 2 and 4 only include the predictors selected by LASSO. Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Table B.2: Effect of Reform on Pre/Post Growth Trend in Positive Citations

	(1)	(2)	(3)	(4)	(5)
	Effect on Log Positive Cites per Court-Year				
Retirement Reform	0.171+	0.178*	0.0619	0.116	0.171+
	(0.0867)	(0.0866)	(0.0839)	(0.0823)	(0.0996)
Pre-Reform × Time	0.00348	0.0154			
	(0.0194)	(0.0225)			
Post-Reform × Time			0.0321*	0.0344*	0.0442+
			(0.0130)	(0.0136)	(0.0226)
Court FE, Year FE	X	X	X	X	X
Court Trends/Windows		X		X	X
Init Court Rules × Year FE					X
Init Case Types × Year FE					X
N	2448	2448	2448	2448	2448
R-sq	0.732	0.828	0.845	0.863	0.872

Notes. DD effect of mandatory retirement reform on log positive citations to a court in eight years after reform, relative to eight years before reform. Observation is a court-year. "Retirement Reform" is a treatment indicator for the eight years after the introduction of mandatory retirement. "Pre-Reform × Time" Is a linear time trend, interacted with an indicator for the 8 years before the reform. "Post-Reform × Time" Is a linear time trend, interacted with an indicator for the 8 years after the reform. Treat Windows means court-specific treatment windows (eight years before and after reform). "Init X" × year FE means initial values are interacted with year. "Init Court Rules" includes a state's 1947 rules for judge selection/retention system, admin office, intermediate appellate court, number of judges, and term length. "Init Case Types" includes a court's 1947 average values for case characteristics (legal area and related industries). "Init Age" includes the initial mean and standard deviation for judge age on the court. Standard errors clustered by state in parentheses. $+$ p<.0.1, * p<0.05, ** p<0.01.

Figure B.1: Event-Study Effect on Performance: Alternative Cite Measures

(A) Effect on Positive Citations (in Levels)

(B) Effect on Log Positive Citations within 10 Years

(C) Effect on Log All Cites (including non-positive)

(D) Effect on Log Discussion Citations
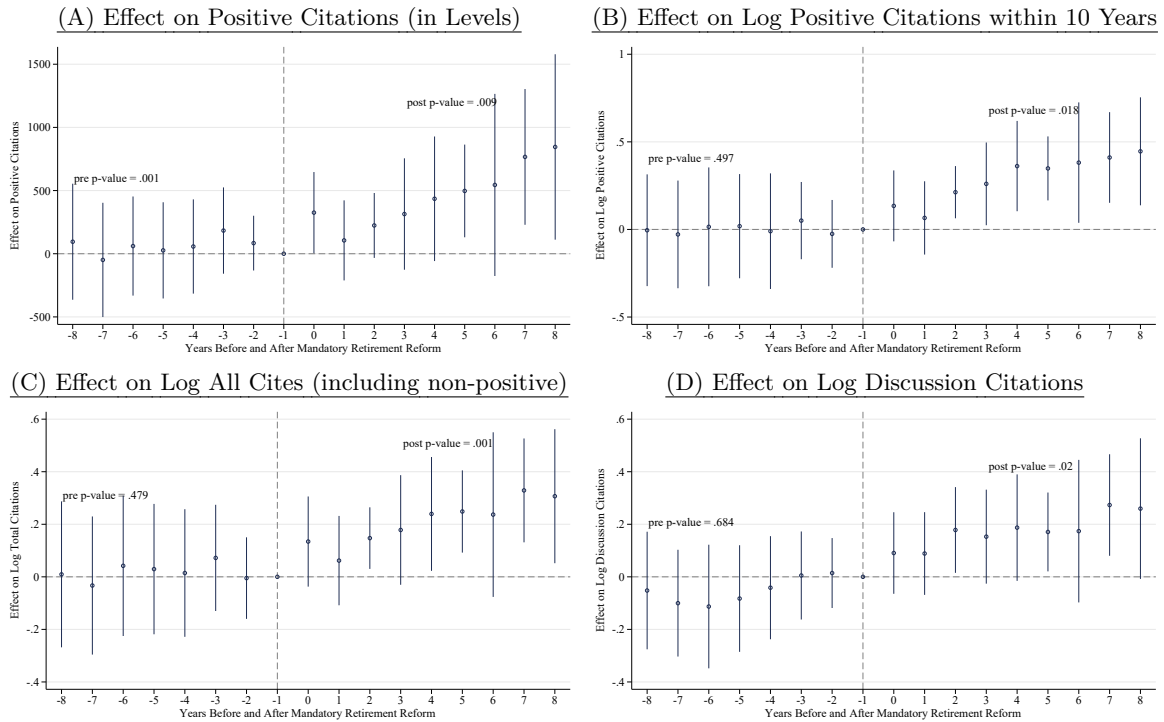
Judge performance before and after reforms implementing retirement ages of 70, 72 or 75. Panel A: outcome is number of positive citations of a judge in a year (in levels, rather than logs). Panel B: outcome is the log positive citations of a judge in a year that were made within eight years of a case. Panel C: outcome is the log total citations of a judge in a year (including non-positive negative cites). Panel D: outcome is the log discussion citations of a judge in a year. Time series is a coefficient plot from the event study regression (2), with coefficients estimated relative to the year before the reform. Regression includes court and year fixed effects. 95% confidence intervals constructed with standard errors clustered by court.

Table B.3: Effect of Reform on Citations: Additional Specifications

	(1)	(2)	(3)	(4)	(5)	(6)
	Cites After 10 years		Has Cite		Drop 5% Outliers	
Ret. Reform	0.282**	0.317**	0.447	0.617+	0.226*	0.281*
	(0.0902)	(0.112)	(0.313)	(0.292)	(0.0911)	(0.103)
Year / Court FE	X	X	X	X	X	X
Trends/Windows		X		X		X
N	2448	2448	2448	2448	2448	2448
R-sq	0.744	0.849	0.783	0.860	0.747	0.844

Notes. Observation is a court-year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. "Cites after 10 Years" means the log of the positive cites to a judge's cases in a year, from cases more than eight years later. "Has Cite" means the proportion of cases with at least one positive citation. "Drop 5% Outliers" is the baseline outcome (log of positive cites to a judge's cases in a year) but for each court-year, the top 5% of cases by cite count are dropped. + p<.0.1, * p<0.05, ** p<0.01.

Table B.4: Effect of Reform on Citations: Different Windows

	(1)	(2)	(3)	(4)	(5)	(6)
	Effect on Log Positive Cites					
Treatment Window	6		14		All	
Retirement Reform	0.140*	0.167**	0.211*	0.255**	0.0547	0.349**
	(0.0593)	(0.0621)	(0.0814)	(0.0889)	(0.114)	(0.116)
Year FE, Court FE	X	X	X	X	X	X
Court Trends/Windows		X		X		X
N	2448	2448	2448	2448	2448	2448
R-sq	0.732	0.825	0.734	0.827	0.731	0.821

Notes. Observation is a court-year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. Court Treat Windows means court-specific treatment windows (specified by the associated number, 6, or 14, or no window for column 6). Standard errors clustered by court in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Table B.5: Effect of Reform on Log Cites: Senior Status Rules

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Effect on Log Citations				
Retirement Reform	0.111	0.265+	0.228**	0.230*	0.195*	0.225*	0.130	0.263*
	(0.101)	(0.136)	(0.0771)	(0.103)	(0.0813)	(0.100)	(0.0947)	(0.115)
× Grandfather Rule	0.158	-0.113						
	(0.135)	(0.167)						
× Finish Term			-0.382**	-0.0761				
			(0.107)	(0.144)				
× Finish Term Half					-0.371**	-0.0125		
					(0.0885)	(0.122)		
× Finish Year							0.189	-0.165
							(0.134)	(0.129)
Year FE, Court FE	X	X	X	X	X	X	X	X
Court Trends/Windows		X		X		X		X
N	1546	1546	1546	1546	1546	1546	1546	1546
R-sq	0.710	0.822	0.712	0.821	0.711	0.821	0.710	0.822

Observation is a court-year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. Coefficients are interacted with respective senior status rules (respectively: the rule not applying to sitting judges, being allowed to finish the term, being allowed to finish terms that are over halfway finished, and being able to finish out the year). Court Treat Windows means court-specific treatment windows (eight years before and after reform). Standard errors clustered by court in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Table B.6: Effect of Reform on Citations: Alternative Clustering

	(1)	(2)	(5)	(6)
Clustering Group	State and Year		None (Robust)	
Retirement Reform	0.189**	0.242**	0.189**	0.242**
	(0.0506)	(0.0416)	(0.0506)	(0.0416)
Court FE, Year FE	X	X	X	X
Court Trends/Windows		X		X
N	2448	2448	2448	2448
R-sq	0.732	0.828	0.732	0.828

Notes. Observation is a court-year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. Court Treat Windows means court-specific treatment windows (eight years before and after reform). Standard errors clustered by court in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Table B.7: Effect of Reform on Log Cites, Alternative Weighting

	(1)	(2)	(3)	(4)
	Effect on Log Positive Cites			
Retirement Reform	0.106	0.217*	0.218**	0.251**
	(0.0978)	(0.100)	(0.0748)	(0.0907)
Weighting	# of Opinions		# of Judges	
Court FE, Year FE	X	X	X	X
Court Trends/Windows		X		X
N	2448	2448	2448	2448
R-sq	0.730	0.826	0.724	0.824

Observation is a court-year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. Court Treat Windows means court-specific treatment windows (eight years before and after reform). Standard errors clustered by court in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Table B.8: Effect of Reform on Log Cites, with Time-Varying Controls

	(1)	(2)	(3)	(4)	(5)	(6)
	Effect on Log Positive Cites					
Retirement Reform	0.091	0.155*	0.237*	0.250**	0.095*	0.148**
	(0.0683)	(0.0748)	(0.0918)	(0.0760)	(0.0423)	(0.0560)
Court FE, Year FE	X	X	X	X	X	X
Court Trends/Windows		X		X		X
Case Controls	X	X				
Rule Controls			X	X		
Lagged Dep. Var.					X	X
N	2448	2448	2446	2446	2391	2391
R-sq	0.819	0.870	0.738	0.831	0.840	0.860

Observation is a court-year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. Court Treat Windows means court-specific treatment windows (eight years before and after reform). Case controls means the first five principal components of the matrix of controls for legal topic and related industries. "Rule controls" means rules for selection and retention of judges and other institutional items. Lagged Dep. Var. means the court-year lag of the dependent variable (log positive cites in the previous year). Standard errors clustered by court in parentheses. $+$ p<.0.1, * p<0.05, ** p<0.01.

Table B.9: Effect of Reform on Log Cites, Additional Institutional Controls

	(1)	(2)	(3)	(4)	(5)	(6)
	Effect on Log Positive Cites per Court-Year					
Retirement Reform	0.181*	0.253**	0.188*	0.253**	0.187*	0.234*
	(0.0791)	(0.0929)	(0.0811)	(0.0938)	(0.0890)	(0.0959)
Court FE, Year FE	X	X	X	X	X	X
Court Trends/Windows		X		X		X
Assign Rule × Year FE	X	X				
Select Rule × Year FE			X	X		
IAC × Year FE					X	X
N	2448	2448	2448	2448	2448	2448
R-sq	0.738	0.833	0.737	0.830	0.750	0.832

Observation is a court-year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. Court Treat Windows means court-specific treatment windows (eight years before and after reform). "Assign × year" means that the regressions include the assignment rule, interacted with year fixed effects, as covariates. Standard errors clustered by court in parentheses. $+$ p<.0.1, * p<0.05, ** p<0.01.

Table B.11: Reform Effect on Number of Opinions and Cites per Case

|	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	# of Opinions			Cites per Case			Out-State Cites / Case		
Retirement Reform	0.0968	0.0266	0.106*	0.0760	0.161**	0.150	0.0429	0.0910+	0.0989
	(0.0700)	(0.0416)	(0.0404)	(0.0683)	(0.0693)	(0.0896)	(0.0512)	(0.0489)	(0.0690)
Year FE, Court FE	X	X	X	X	X	X	X	X	X
Court Trends/Windows		X	X		X	X		X	X
Init Covars × Year FE			X			X			X
N	2448	2448	2448	2448	2448	2448	2448	2448	2448
R-sq	0.787	0.908	0.933	0.580	0.788	0.855	0.807	0.884	0.918

Observation is a court-year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. "# of Opinions" is the number of majority opinions written by a judge in a year. "Cites per Case" is number of citations per published opinion. "Out-of-State Cites / Case" is number of out-of-state citations per published opinion. Court Treat Windows means court-specific treatment windows (eight years before and after reform). Init Covars × Year FE includes initial rule, case type, and age statistics of a court interacted with year. Standard errors clustered by court in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Table B.10: Effect of Reform on Citations: Separately by Maximum Age Imposed

	(1)	(2)	(3)	(4)	(5)	(6)
	Effect on Log Positive Cites					
Maximum Age	70		72		75	
Retirement Reform	0.221+	0.293+	0.258**	0.225**	0.0784	0.158*
	(0.116)	(0.166)	(0.0780)	(0.0556)	(0.125)	(0.0719)
Court FE, Year FE	X	X	X	X	X	X
Court Trends/Windows		X		X		X
N	2448	2448	2448	2448	2448	2448
R-sq	0.732	0.827	0.732	0.819	0.731	0.819

Notes. Observation is a court-year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. Court Treat Windows means court-specific treatment windows (eight years before and after reform). Standard errors clustered by court in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Figure B.2: Event-Study Effect of Reform with Court-Specific Trends
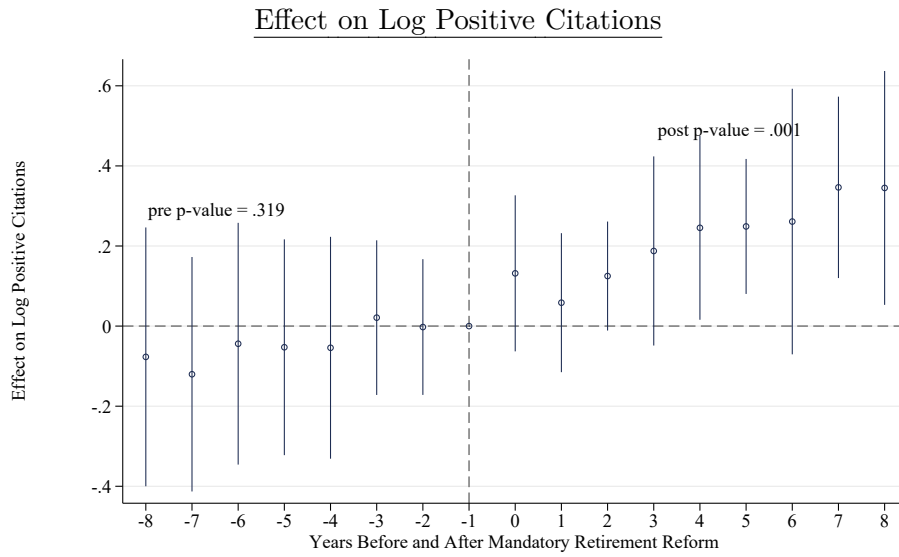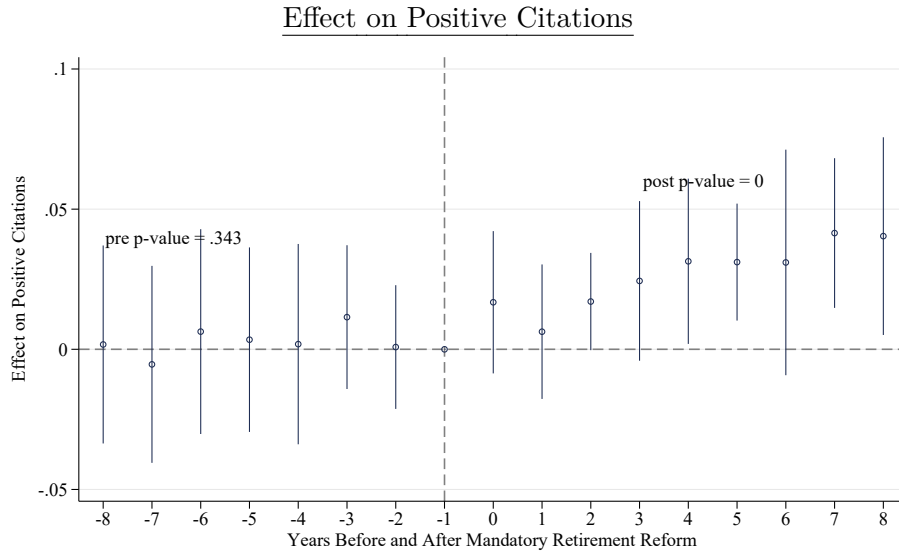
Effect on Log Positive Citations

Court performance before and after reforms implementing retirement ages of 70, 72 or 75. The outcome is the log positive citations of a judge in a year. Time series is a coefficient plot from the event study regression (2), with coefficients estimated relative to the year before the reform. Regression includes court and year fixed effects, but with court-specific trends. 95% confidence intervals constructed with standard errors clustered by court.

Table B.12: Effect of Reform on Inverse Hyperbolic Sine of Citations

|	(1)	(2)	(3)	(4)	(5)
	Inverse Hyperbolic Sine of Positive Cites				
Retirement Reform	0.186*	0.243*	0.260*	0.332**	0.322**
	(0.0764)	(0.0931)	(0.103)	(0.107)	(0.126)
Court FE, Year FE	X	X	X	X	X
Court Trends/Windows		X	X	X	X
Init Court Rules × Year FE			X	X	X
Init Case Types × Year FE				X	X
Init Age × Year FE					X
N	2448	2448	2448	2448	2448
R-sq	0.728	0.824	0.841	0.859	0.869

Notes. DD effect of mandatory retirement reform on inverse hyperbolic sine (asinh) positive citations to a judge's opinions in eight years after reform, relative to eight years before reform. Observation is a court-year. "Ret. Reform" is a treatment indicator for the eight years after the introduction of mandatory retirement. Court Treat Windows means court-specific treatment windows (eight years before and after reform). "Init X" × year FE means initial values are interacted with year. "Init Court Rules" includes a state's 1947 rules for judge selection/retention system, admin office, intermediate appellate court, number of judges, and term length. "Init Case Types" includes a court's 1947 average values for case characteristics (legal area and related industries). "Init Age" includes the initial mean and standard deviation for judge age on the court. Standard errors clustered by court in parentheses. $+$ p$<$.0.1, * p$<$0.05, ** p$<$0.01.

Figure B.3: Event-Study Effect of Reform on Court Performance: Poisson Regression

Effect on Positive Citations

Court performance before and after reforms implementing retirement ages of 70, 72 or 75. Poisson regression with outcome as positive citation counts of a court in a year. Time series is a coefficient plot from the event study regression (2), with coefficients estimated relative to the year before the reform. Regression includes court and year fixed effects. 95% confidence intervals constructed with standard errors clustered by court.
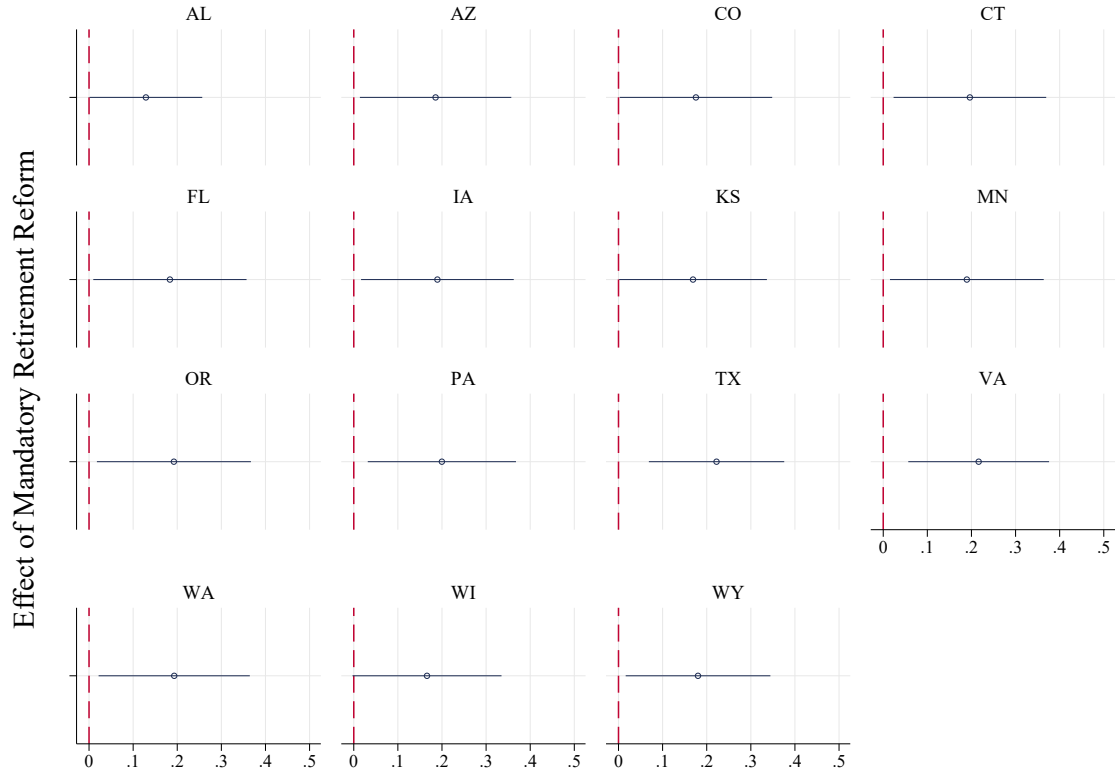
Figure B.4: Event-Study Effect of Reform on Court Performance: Only Reform States

Effect on Log Positive Citations
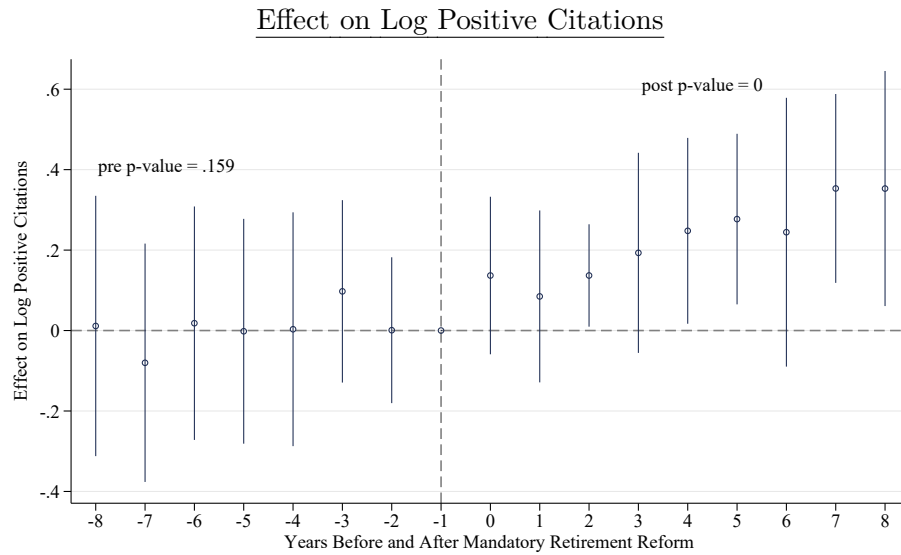
Court performance before and after reforms implementing retirement ages of 70, 72 or 75. Sample limited to reform states. The outcome is the log positive citations of a judge in a year. Time series is a coefficient plot from the event study regression (2), with coefficients estimated relative to the year before the reform. Regression includes court and year fixed effects. 95% confidence intervals constructed with standard errors clustered by court.

Figure B.5: DD Effect of Reform, Dropping each Treated State Individually

Coefficient for the effect of mandatory retirement at ages of 70, 72 or 75 on judge performance. The outcome is the log positive citations of a judge in a year. Each subfigure plots the coefficient from regression 1 excluding one treated state at a time. Includes court and year fixed effects, court-specific windows and trends.

Figure B.6: Event-Study Effect on Performance: Adjustment for Staggered Treatment
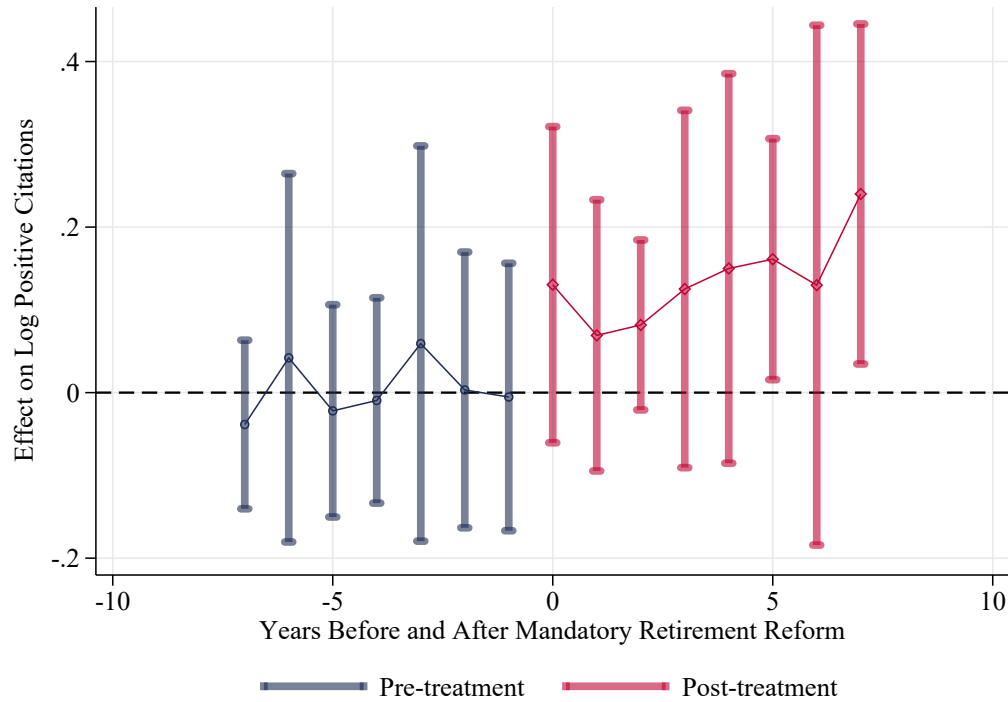
Log positive cites to court before and after reforms implementing retirement ages of 70, 72 or 75. Time series is a coefficient plot from the event study regression (2), with coefficients estimated relative to the year before the reform. Regression includes court and year fixed effects and court-specific event windows. Coefficients and standard errors (clustered by state) adjusted for staggered treatment timing, following the method in Sant'Anna and Zhao (2020), as described in Section 4.1. This is doubly robust diff-in-diff estimator based on inverse probability of tilting and weighted least squares. Produced using the csdid command in stata.

Figure B.7: Event-Study Effect, by Random and Discretionary Case Assignment
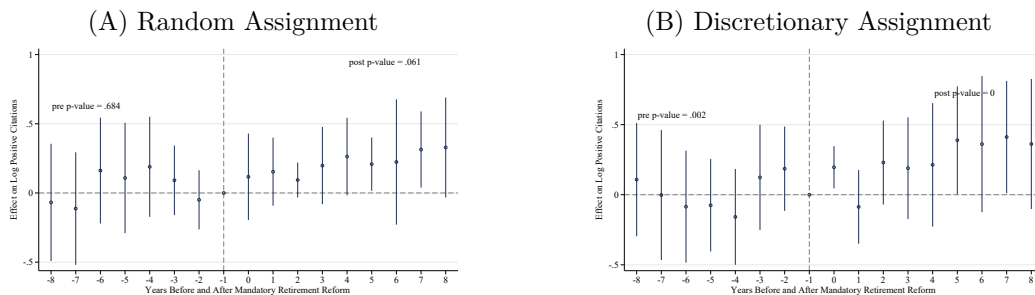
Notes. Court performance before and after reforms implementing retirement ages of 70, 72 or 75. The outcome is the log positive citations to the court in a year. Panel A includes courts with random or rotating assignment of cases. Panel B includes courts with discretionary assignment of cases. Time series is a coefficient plot from the event study regression (2), with coefficients estimated relative to the year before the reform. Regression includes court and year fixed effects. 95% confidence intervals constructed with standard errors clustered by court.

Table B.13: Effect of Mandatory Retirement Reform, Other Behavioral Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Work Output		Caselaw Research		Overruled Rate		Addendum Ops		Dissent Rate	
Ret Reform	-0.0197	0.0086	-0.315	0.0761	-0.0169*	-0.0036	0.312*	0.334*	0.0148*	0.0192*
	(0.0612)	(0.0648)	(0.305)	(0.174)	(0.0084)	(0.0061)	(0.133)	(0.121)	(0.0064)	(0.0069)
Year / Court FE	X	X	X	X	X	X	X	X	X	X
Trends/Windows		X		X		X		X		X
N	2448	2448	2448	2448	2448	2448	2448	2448	2448	2448
R-sq	0.492	0.705	0.603	0.805	0.342	0.477	0.718	0.805	0.481	0.662

Observation is a court-year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. "Work Output" is log number of words written in a year. "Caselaw Research" is number of previous cases cites. "Overruled rate" is being overruled by a higher court. "Addendum Ops" is number of dissenting and concurring opinions (in logs). Court Treat Windows means court-specific treatment windows (eight years before and after reform). Standard errors clustered by state in court. + p<.0.1, * p<0.05, ** p<0.01.

C Additional Material on Mechanisms

C.1 Changes in the Caseload or Case Characteristics

Table C.1: Effect of Retirement Reform, Intermediate Appellate Court Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IAC Cases		IAC / SC Cases		IAC Words		IAC Cites	
Ret Reform	0.434*	-0.158	0.0440	-0.0709	0.0497	-0.0980	0.276*	-0.00169
	(0.208)	(0.147)	(0.0713)	(0.0847)	(0.0882)	(0.0910)	(0.103)	(0.0540)
Year / Court FE	X	X	X	X	X	X	X	X
Trends/Windows		X		X		X		X
N	1445	1445	1445	1445	1445	1445	1445	1445
R-sq	0.817	0.922	0.693	0.821	0.882	0.946	0.810	0.903

Observation is a court-year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. IAC Cases is log number of cases in intermediate appellate courts in the state. IAC / SC cases is the log ratio of IAC cases to state supreme court cases. IAC words is the log words in IAC case opinions. IAC Cites is the log cites to IAC opinions. Court Treat Windows means court-specific treatment windows (eight years before and after reform). Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Figure C.1: Effects of Reform on Caseload and Opinion Authoring Rate

(A) Total Caseload (Log Number of Cases) (B) Opinion Rate (Num Opinions / Num Cases)
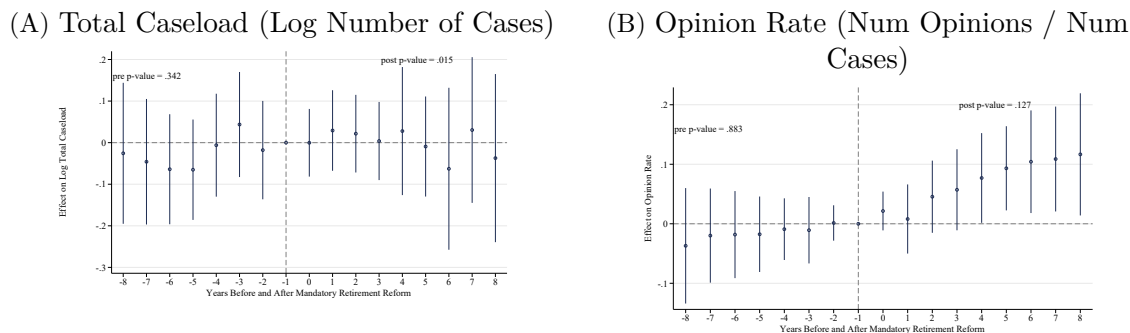
Notes. Caseload and authorship rate before and after reforms implementing retirement ages of 70, 72 or 75. The outcome is the log number of appeals reviewed (Panel A), and the opinion authorship rate (number of opinions, divided by number of cases reviewed, Panel B). Time series is a coefficient plot from the event study regression (2), with coefficients estimated relative to the year before the reform. Regression includes court and year fixed effects. 95% confidence intervals constructed with standard errors clustered by court.

C.2 Life Cycle Effects of Aging

Figure C.2: Dynamic Analysis of Judge Age and Judge Performance
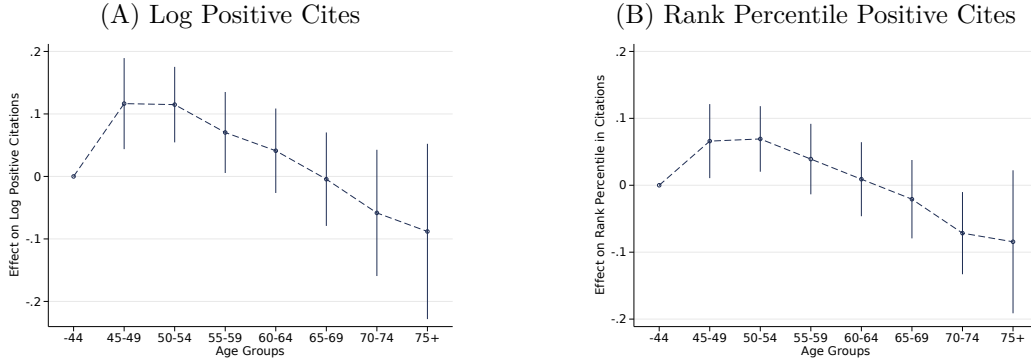
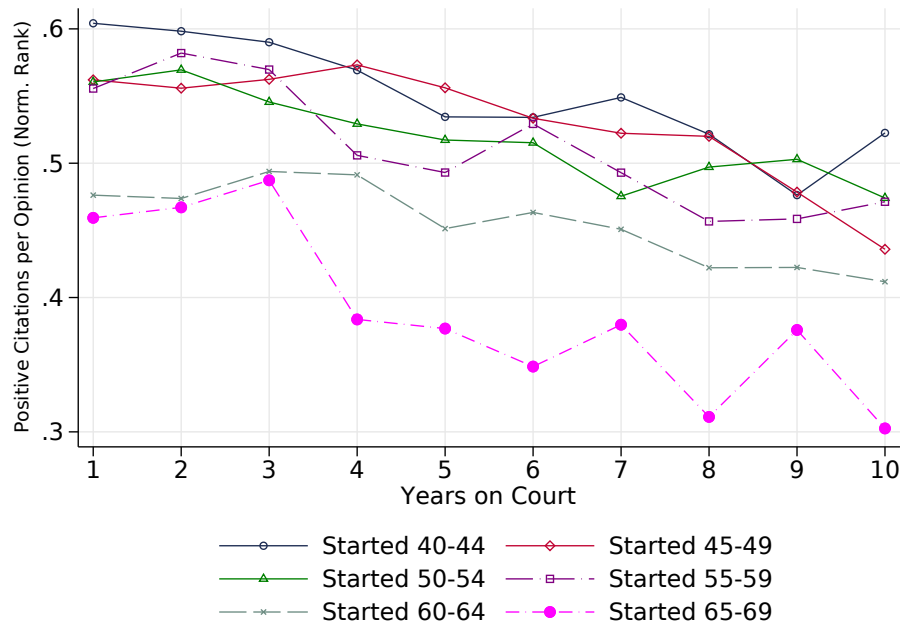| (A) Log Positive Cites	(B) Rank Percentile Positive Cites

Dynamic coefficient plots for estimates of five-year age group differences, relative to the age < 45 group. Observation is a judge working in a year. All graphs contain court-year interacted fixed effects, first year baselines, and cohort fixed effects. Outcomes are in logs or rank percentiles, as indicated. 95% confidence intervals constructed using standard errors clustered by state.

Table C.2: Effect of Mandatory Retirement Reform; Relevance of Judge Experience

|	(1)	(2)	(3)	(4)	(5)	(6)
	Experience		Age		Experience	
Retirement Reform	-1.206+	-1.353+	-1.905**	-1.927**	-0.218	-0.151
	(0.615)	(0.711)	(0.517)	(0.526)	(0.494)	(0.470)
Year FE, Court FE	X	X	X	X	X	X
Court Trends/Windows		X		X		X
Experience Decile FE			X	X		
Age Decile FE					X	X
N	2448	2448	2448	2448	2448	2448
R-sq	0.537	0.642	0.523	0.698	0.672	0.777

Notes. Observation is a court-year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. "Experience" is the years of experience of each judge. Court Treat Windows means court-specific treatment windows (eight years before and after reform). Experience decile FE and age decile FE are fixed effects for the associated deciles computed in the court-year dataset. Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Figure C.3: Performance by Age in First Years of Judgeship

Notes. Time series for average rank percentile (within court year) in positive citations for the first years of a judge career, separately by starting age (indicated in legend).

Figure C.4: Performance by Age in Last Years of Judgeship
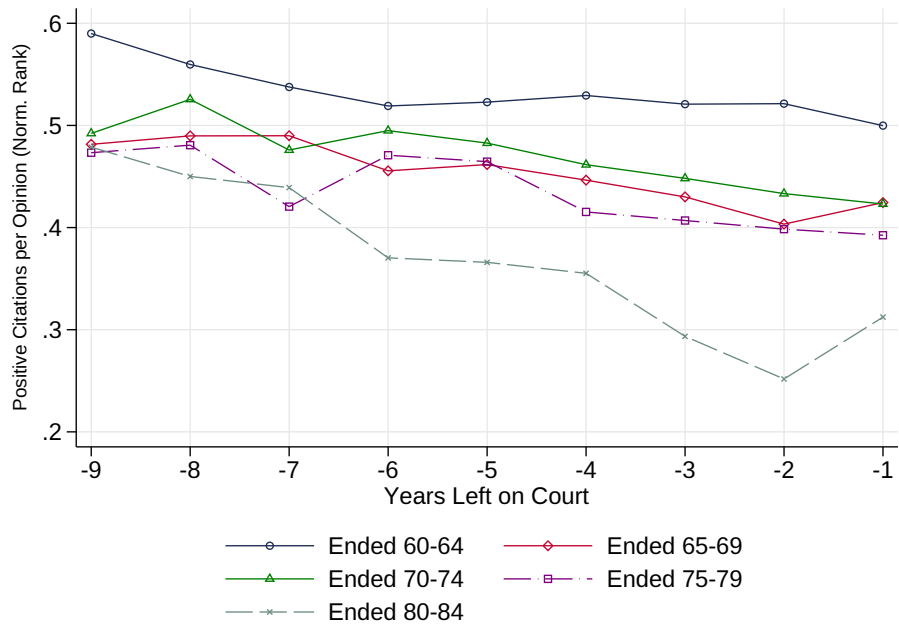
Notes. Time series for average rank percentile (within court year) in positive citations for the last years of a judge career, separately by starting age (indicated in legend).

Table C.3: Life Cycle Effects on Performance: Age vs. Experience

	(1)	(2)	(3)	(4)
		Log Positive Cites		
Judge Age (Years)	-0.00852**	-0.00826**		
	(0.00190)	(0.00150)		
Judge Experience (Years)			-0.00329	-0.00233
			(0.00237)	(0.00216)
Court-Year FE	X	X	X	X
First-Year Baseline		X		X
Experience FE	X	X		
Age FE			X	X
N	14985	14985	14977	14977
R-sq	0.482	0.544	0.469	0.533

Notes. Observation is a judge-year. Judge Age and Judge Experience are years since birth, and years since starting judgeship, respectively. Court-Year FE is interacted court-year fixed effects. First-Year Baseline means a judge's value for the outcome in their first year on the court is included as a control. Experience FE means fixed effects for years of experience, and similarly for Age FE. Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

C.3 Selection on Entry and Exit

Table C.4: Comparing Judges Selected Before/After the Reform

	(1)	(2)	(3)	(4)
	Effect on Log Positive Cites			
Selected Post Reform	0.0509	0.0630	0.0643	0.0923
	(0.0587)	(0.0648)	(0.0782)	(0.0711)
Court × Year FE	X	X	X	X
Age FE	X	X	X	X
Starting Year FE		X	X	X
Inverse Career Weights			X	
Ten Years Before/After				X
N	14969	14968	14955	2030
R-sq	0.674	0.681	0.709	0.588

Observation is a judge working in a year. Inverse career weights means judges are weighted by inverse number of years in the sample. "Ten Years Before/After" means sample is limited to those years before and after the reform. "Selected Post Reform" is an indicator for judges selected after the introduction of mandatory retirement. Standard errors clustered by state in parentheses. $+$ p$<$.0.1, * p$<$0.05, ** p$<$0.01.

Table C.5: Effect on Number of Working Judges and Cites Per Judge

	(1)	(2)	(3)	(4)	(5)	(6)
	Log Number of Judges			Log Cites per Judge		
Retirement Reform	0.0406	0.0493	0.0662*	0.197*	0.198*	0.251**
	(0.0259)	(0.0339)	(0.0325)	(0.0810)	(0.0844)	(0.0923)
Court FE, Year FE	X	X	X	X	X	X
Court Trends/Windows	X	X	X	X	X	X
Init Court Rules × Year FE		X	X		X	X
Init Case Types × Year FE			X			X
N	2448	2448	2448	2448	2448	2448
R-sq	0.814	0.835	0.851	0.802	0.823	0.844

Observation is a court-year. Term are as above. Standard errors clustered by state in parentheses. $+$ p$<$.0.1, * p$<$0.05, ** p$<$0.01.

C.4 Team Effects of Aging

Figure C.5: Event-Study Effect of Reform on Performance, with Judge Fixed Effects
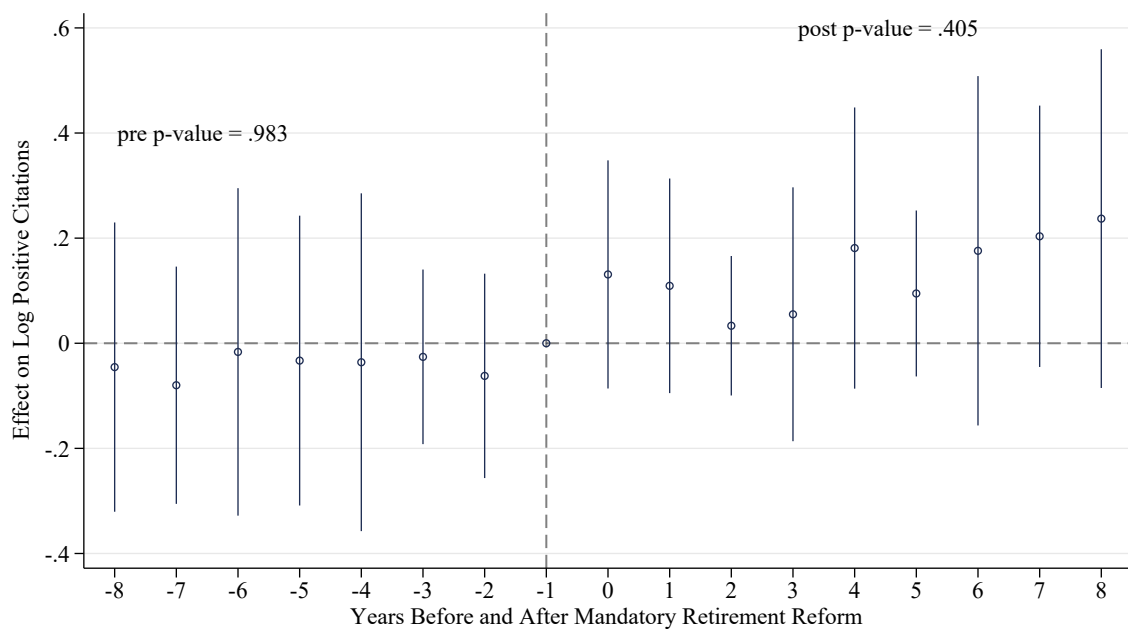
Judge performance before and after reforms implementing retirement ages of 70, 72 or 75. Outcome is log positive citations for a judge in a year. Time series is a coefficient plot from the event study regression (2), with coefficients estimated relative to the year before the reform. Regression includes court and year fixed effects, plus judge fixed effects. 95% confidence intervals constructed with standard errors clustered by court.

Table C.6: Effect of Reform, Other Measures, Judge Fixed Effects

|	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Cites in Levels		Within 10 years		All Cites		Discuss Cites		Out-of-State Cites	
Ret. Reform	22.72	21.10	0.217**	0.185	0.156+	0.138	0.105+	0.111	0.154*	0.180+
	(14.61)	(21.61)	(0.104)	(0.124)	(0.088)	(0.095)	(0.061)	(0.088)	(0.093)	(0.101)
Year / Judge FE	X	X	X	X	X	X	X	X	X	X
Trends/Windows		X		X		X		X		X
N	14905	14905	14905	14905	14905	14905	14905	14905	14905	14905
R-sq	0.585	0.665	0.739	0.809	0.678	0.753	0.667	0.749	0.641	0.718

Notes. Observation is a judge working in a year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. "Cites in Levels" means the outcome is not logged. "Within 10 years" is the log positive cites within eight years of an opinion. "All Cites" is the log number of all citations (positive, negative, and distinguishing) to a judge in a year. "Discuss Cites" is only the positive cites where the latter judge discussed the cited opinion. "Out-of-State Cites" is the count of number of positive citations from courts in other states. "Positive Cites" is the number of positive cites (in levels). Court Treat Windows means court-specific treatment windows (eight years before and after reform). Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Table C.7: Effect of Reform on # of Opinions and Cites Per Opinion, Judge Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	# of Opinions			Cites per Opinion		
Retirement Reform	0.0756+	0.0694	0.0637	0.0544	0.0453	0.0761*
	(0.0369)	(0.0383)	(0.0524)	(0.0442)	(0.0412)	(0.0435)
Judge FE, Year FE	X	X	X	X	X	X
Court Trends/Windows		X	X		X	X
Init Court Rules × Year FE			X			X
Init Case Types × Year FE			X			X
Init Age × Year FE			X			X
N	14905	14905	14905	14905	14905	14905
R-sq	0.648	0.674	0.706	0.815	0.821	0.836

Observation is a judge working in a year. "Retirement Reform" is an indicator for the eight years after the introduction of mandatory retirement. Court Treat Windows means court-specific treatment windows (eight years before and after reform). "Init X" × year FE means initial values are interacted with year. "Init Court Rules" includes a state's 1947 rules for judge selection/retention system, admin office, intermediate appellate court, number of judges, and term length. "Init Case Types" includes a court's 1947 average values for case characteristics (legal area and related industries). "Init Age" includes the initial mean and standard deviation for judge age on the court. Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Figure C.6: Effect of Judge Age on Number of Authored Opinions, and Cites Per Opinion

(A) Log # of Opinions

(B) Log Cites per Opinion
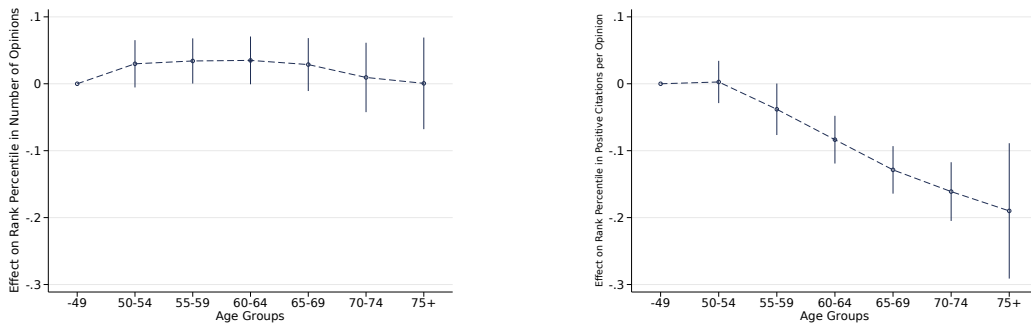
Dynamic coefficient plots for estimates of five-year age group differences, relative to the age < 45 group. Observation is a judge working in a year. All graphs contain court-year interacted fixed effects, first year baselines, and cohort fixed effects. Outcomes are in logs or rank percentiles, as indicated. 95% confidence intervals constructed using standard errors clustered by state.

D The Age Discrimination in Employment Act of 1967 Sec. 621

The Congress hereby finds and declares that

1. in the face of rising productivity and affluence, older workers find themselves disadvantaged in their efforts to retain employment, and especially to regain employment when displaced from jobs;

   (a) the setting of arbitrary age limits regardless of potential for job performance has become a common practice, and certain otherwise desirable practices may work to the disadvantage of older persons;

   (b) the incidence of unemployment, especially long-term unemployment with resultant deterioration of skill, morale, and employer acceptability is, relative to the younger ages, high among older workers; their numbers are great and growing; and their employment problems grave;

   (c) the existence in industries affecting commerce, of arbitrary discrimination in employment because of age, burdens commerce and the free flow of goods in commerce.

   (d) It is therefore the purpose of this chapter to promote employment of older persons based on their ability rather than age; to prohibit arbitrary age discrimination in employment; to help employers and workers find ways of meeting problems arising from the impact of age on employment.