Reviewing Administrative Review

David K. Hausman[†]

In the largest system of federal adjudication—Social Security disability adjudication—outcomes depend more on the randomly assigned judge than on the strength of the case. Does the administrative appeals process use resources effectively to reduce that arbitrariness and limit the discretion of administrative law judges? If not, how and why does it fail? These are empirical questions, and this Article uses a new dataset tracking millions of cases to answer them.

*A system of administrative appeals that efficiently limits the discretion of decisionmakers should display three empirical patterns. First, disappointed claimants should be more likely to appeal the decisions of harsher judges—judges who have lower grant rates than their colleagues in the same hearing office (*claim selection*). Second, when claimants appeal, harsher judges' decisions should be reversed more often than the decisions of their more generous colleagues (*decisionmaking*). Third, judges should try to avoid remands and therefore increase their grant rates after a reversal (*remand aversion*).*

Testing for each of these patterns offers a method of diagnosing problems with systems of administrative review—and helps identify where new resources would be most useful. For example, if litigants rarely appeal decisions of even extreme adjudicators, a quality assurance process might solve the problem by randomly selecting cases for review. If appellate decisionmaking itself is flawed, peer review may be more promising. And if adjudicators are insensitive to remands, training and feedback might be appropriate.

I apply this method to conduct the first quantitative study of administrative appeals in Social Security disability adjudications. I use a previously unreleased database, obtained by Freedom of Information Act request, of nearly four million Social Security Administrative Law Judge

and Appeals Council decisions from 2010 to 2014. Social Security Appeals Council decisions do not show the expected decisionmaking pattern: on average, the Appeals Council fails to distinguish between the decisions of harsher and more generous adjudicators. That finding helps explain why litigants complain that Appeals Council decisions are unpredictable, and the finding leads to a diagnosis: the Appeals Council is both too deferential, because it is unable to conduct holistic factual review, and not deferential enough, because it must reverse decisions for harmless errors.

Introduction

    Why do courts and agencies invest in an appeals process instead of dedicating the same resources to improving initial decisions? There are several possible reasons: that litigants appeal the cases more likely to contain errors, allowing the appeals court to review a smaller number of cases;[1] that the appeals process allows a second decisionmaker to review

---

    1.   *See* Steven Shavell, *The Appeals Process as a Means of Error Correction*, 24 J.L. STUD. 379 (1995).

the work of the first, limiting the discretion of the first decisionmaker and making extreme decisions less likely; and that the appeals process improves the decisions of the initial decisionmaker, who issues rulings in its shadow. But do appeals in fact create these efficiencies?

This Article makes two contributions. First, it proposes a theory of administrative review that offers a new way of understanding the steps at which such review can limit the discretion of line-level decisionmakers. Second, it applies that theory to the Social Security disability appeals process, using an internal administrative dataset to identify the system's strengths and weaknesses.

In the theory, I posit that discretion-limiting systems of review share three empirical features. First, litigants choose whether to appeal (*claim selection*). Litigants appeal more often from harsher decisions—decisions that they believe are more likely to be reversed on appeal. Second, the appeal reaches appellate decisionmakers, who disproportionately reverse harsher decisions (*decisionmaking*). Third, when a remand is handed down, it returns to the original judge, and the theory predicts that such remands lead judges to alter their behavior in the short term, making their decisions more closely match those of their colleagues (*remand aversion*).

To test empirically whether the Social Security Appeals Council limits discretion at these three steps, I take advantage of close-to-random assignment of cases to administrative law judges (ALJs) within the same hearing office. First, I find that litigants do select cases from harsher ALJs for appeal. The claim-selection process therefore works mostly as designed. Second, I find that a rule granting automatic full review of claims challenging dismissals leads to more reversals of more generous ALJs—the opposite of the expected pattern. Third, I find that ALJs barely change their behavior in the aftermath of remands, if at all.

The main results in this Article are causal, exploiting the near-random assignment of cases to ALJs. These results diagnose where the Social Security appeals system is not effective. Interviews with ALJs then point the way to a cure. The appellate decisions of the Social Security Appeals Council themselves fail to promote consistency, not because the Appeals Council is ineffective, but likely because it defers both too little and too much. The Appeals Council often defers to factual determinations that would benefit from review, yet it also often reverses for harmless technical errors.

The theory and results have broad implications. Scholars have long bemoaned dysfunction in mass adjudication. In the Executive Office for

Immigration Review,[2] the Patent and Trademark Office,[3] the National Labor Relations Board,[4] the Securities and Exchange Commission,[5] the Veterans Benefits Administration,[6] and many other state and federal health, safety and environmental agencies, advocates and scholars have found that uncontrolled discretion and inconsistent decisionmaking stand in the way of fair and efficient adjudication.[7]

Yet the systematic study of dysfunction in adjudication has often faced two roadblocks. First, data are scarce. This Article solves that problem with new internal Social Security data, obtained by Freedom of Information Act request, tracking millions of Social Security disability adjudications. Second, even when datasets are available, causal inference may be impossible. This Article solves that problem by exploiting the near-random assignment of cases to administrative law judges. That near-random assignment means that we can attribute differences in grant rates within the same hearing offices to the judges themselves, rather than to differences in their caseloads.

The Article's theory and results also offer a new way to compare the success of appeals processes across administrative domains, and in Part IV of this Article, I explain how this Article's theory can improve our understanding of patent and immigration adjudications. In the patent context, I argue that distinguishing between the selection and decisionmaking steps of review could help advance our understanding. And in the immigration context, I explain how this new theory led me to

---

2.   *See generally* JAYA RAMJI-NOGALES, ANDREW I. SCHOENHOLTZ & PHILIP G. SCHRAG, REFUGEE ROULETTE: DISPARITIES IN ASYLUM ADJUDICATION AND PROPOSALS FOR REFORM (2009); David Hausman, *The Failure of Immigration Appeals*, 164 U. PENN. L. REV. 1177 (2016) (confirming that disparities across immigration judges do not only reflect case assignment and tracking those disparities on appeal); Jaya Ramji-Nogales, Andrew I. Schoenholtz & Philip G. Schrag, *Refugee Roulette: Disparities in Asylum Adjudication*, 60 STAN. L. REV. 295 (2007).

3.   *See, e.g.*, Douglas Lichtman, *Rethinking Prosecution History Estoppel*, 71 U. CHI. L. REV. 151, 170-75 (2004) (discussing the consequences of disparities in the grant rates of patent examiners).

4.   *See, e.g.*, Cole D. Taratoot & Robert M. Howard, *The Labor of Judging: Examining Administrative Law Judge Decisions*, 39 AM. POL. RES. 832, 844-47 (2011) (finding that ALJs exercise considerable discretion, and that the exercise of that discretion reflects their political ideology).

5.   *See* David Zaring, *Enforcement Discretion at the SEC*, 94 TEX. L. REV. 1155, 1185-86 (finding no statistically significant differences in government win rates across ALJs, but noting the extremely high win rate for the government overall).

6.   *See, e.g.*, OFFICE OF THE INSPECTOR GEN., DEP'T OF VETERANS AFFAIRS, REPORT NO. 05-00765-137, REVIEW OF STATE VARIANCES IN VA DISABILITY COMPENSATION PAYMENTS (2005) (documenting cross-state inconsistencies in veterans' disability awards).

7.   For a long and useful list of areas of administrative adjudication in which inconsistency across adjudicators is a central problem, see Daniel E. Ho, *An Experiment of Experimentalism*, 69 STAN. L. REV. 1, 1-5 (2017); and Daniel E. Ho & Sam Sherman, *Managing Street-Level Arbitrariness: The Evidence Base for Public Sector Quality Improvement*, 13 ANN. REV. L. & SOC. SCI. 251, 264-65 (2017).

add to a previous Article on immigration appeals, investigating whether immigration judges change their behavior to avoid remands.

Together, the results and comparisons yield a method of evaluating three prominent proposals for the reform of mass adjudication. First, Jerry Mashaw has advocated the adoption of quality assurance systems to protect claimants' due process rights; such systems, by randomly selecting cases for additional review, mostly alter the first stage of review, taking the case selection process out of claimants' hands. Quality assurance proposals are therefore most obviously needed if the first stage of review—litigant selection—is ineffective. Second, democratic experimentalists have advocated peer review—in which at least a subset of initial decisions happen in teams or receive review by a peer—as a mechanism for enhancing accuracy and consistency in adjudications. These proposals target the decisionmaking process itself (the second stage) and are most needed when that stage is flawed. Third, many scholars have suggested expanding adversarial due process protections in the tradition of *Mathews v. Eldridge*[8] and *Goldberg v. Kelly*.[9] These proposals would make agency adjudications more court-like—for example, by establishing other Article I courts modeled on the Tax Court or improving access to lawyers. Such proposals aim to strengthen adversarial review rather than to augment or replace it; such proposals are most persuasive where they can plausibly assert that the changes would improve the claimant-selection process.

Quality assurance, peer review, and adversarial due process protections all require resources to implement. Although I do not address the budgetary considerations that would necessarily inform the details of implementation, I offer a framework for deciding which of these systems of review is the best use of scarce resources. For disability adjudications, the results here suggest that the selection stage—at which claimants decide whether to appeal—functions reasonably well, although it could be improved by allowing the government to choose cases to contest on appeal. At the decisionmaking stage, by contrast, the Social Security Appeals Council has too limited a mandate to promote consistency in outcomes; I therefore propose expanding that mandate.

I. Constraining Administrative Discretion

A. Why Review

The problem of unconstrained discretion in administrative adjudication has been studied for decades, with evidence from nearly every

---

8.   424 U.S. 319 (1976).
9.   397 U.S. 254 (1970).

large area of adjudication.[10] Such studies have also produced proposals for reform, which, in their most general form, usually call for some type of review—some system whereby the line-level adjudicator's decision receives a second look from another official. Such proposals fall into three broad categories. First, administrative-law scholars—most prominently, Jerry Mashaw—have argued that internal management devices, such as quality assurance systems, are at least as important to accurate and fair adjudication as adversarial due process protections.[11] Second, many scholars have proposed some version of peer review to make decisions more consistent.[12] Finally, some scholars have argued, especially in the immigration context, for strengthening adversarial procedural protections in administrative adjudications, or for moving such adjudications to the Article III courts or to a specialized Article I court.[13]

---

10. Limiting adjudicators' discretion and thereby promoting consistency (or uniformity or predictability) is a near-universal stated goal of administrative adjudication. In the Social Security context, see, for example, Judge Patricia Jonas, Exec. Dir., Office of Appellate Operations, Office of Disability Adjudication and Review, Soc. Sec. Admin., Statement Before the Senate Committee on Homeland Security and Governmental Affairs, Permanent Subcommittee on Investigations (Sept. 13, 2012), https://www.ssa.gov/legislation/testimony_091312.html [https://perma.cc/2WNK-FE8A] (noting that "[o]ver the past five years, the allowance and denial rates have become more consistent throughout the ALJ corps."); see also EXEC. OFFICE FOR IMMIGRATION REVIEW, U.S. DEP'T OF JUSTICE, IMMIGRATION COURT PRACTICE MANUAL (2016), http://www.justice.gov/sites/default/files/pages/attachments/2016/02/04/practice_manual_-_02-08-2016_update.pdf [https://perma.cc/S9EG-4MP9] (noting "the public's desire for greater uniformity in Immigration Court procedures"). Appeals for consistency and uniformity in systems of decisionmaking come not only from policymakers but also from scholars like Jerry Mashaw, who suggests that rational systems of adjudication should be neither completely mechanical nor completely discretionary: "We want discretion to be constrained by programmatically specified values and a perception of true states of the world." JERRY L. MASHAW, BUREAUCRATIC JUSTICE: MANAGING SOCIAL SECURITY DISABILITY CLAIMS 72 (1983). The history of Social Security disability adjudications in particular is a history of repeated and often failing attempts to decrease inconsistency across ALJs. See Margaret H. Taylor, *Refugee Roulette in an Administrative Law Context: The Déjà Vu of Decisional Disparities in Agency Adjudication*, 60 STAN. L. REV. 475, 490-95 (2007).

11. Jerry L. Mashaw, *The Management Side of Due Process: Some Theoretical and Litigation Notes on the Assurance of Accuracy, Fairness, and Timeliness in the Adjudication of Social Welfare Claims*, 59 CORNELL L. REV. 772, 775-76, 822-23 (1974). I know of only one convincing quantitative test of the efficacy of a system of quality assurance. See Daniel E, Ho, Cassandra Handan-Nader, David Ames & David Marcus, *Quality Review of Mass Adjudication: A Randomized Natural Experiment at the Board of Veterans Appeals, 2003-2016*, 35 J.L. ECON. & ORG. 239 (2019); see also David Ames, Cassandra Handan-Nader, Daniel E. Ho & David Marcus, *Due Process and Mass Adjudication: Crisis and Reform*, 72 STAN. L. REV. 1, 48-51 (2020).

12. *See, e.g.*, Ho, *supra* note 7, at 1; Charles F. Sabel & William H. Simon, *Minimalism and Experimentalism in the Administrative State*, 100 GEO. L.J. 53, 79-80 (2011) (noting decentralization and feedback between employees and superiors as hallmarks of an experimentalist administration); see generally Michael C. Dorf & Charles F. Sabel, *A Constitution of Democratic Experimentalism*, 98 COLUM. L. REV. 267 (1998).

13. *See, e.g.*, Stacey Caplow, Peter L. Markowitz, Jojo Annobil, Peter Z. Cobb & Nancy Morawetz, *Accessing Justice: The Availability and Adequacy of Counsel in Removal Proceedings, New York Immigrant Representation Study Report: Part 1*, 33 CARDOZO L. REV. 357, 384 (2011) (comparing relief rates for immigrants with and without lawyers and advocating for funding for lawyers); *Access to Justice for Immigrant Families and Communities: Study of Legal Representation of Detained Immigrants in Northern California*, NORTHERN CAL. COLLABORATIVE FOR

This Article proposes a quantitative method of diagnosing the problems with existing systems of hierarchical, adversarial review. It also articulates a theory that clarifies the relationship between such traditional review and proposed innovations (quality assurance and peer review in particular). The hallmarks of a hierarchical, adversarial system of review are simple: a person who receives an unfavorable decision has the option to have the decision reviewed and potentially reversed by a higher authority. Even many administrative decisionmaking systems that lack the words "appeal" or "review" resemble hierarchical, adversarial review. Consider grievances filed by prisoners against guards[14] or complaints filed against police officers.[15]

Since some parts of a review system may work well while others are dysfunctional, breaking the process into parts is a useful exercise. Hierarchical, adversarial review takes place in three basic steps, outlined below. This Article presents a method of diagnosing problems at each step. It also identifies which elements quality assurance and peer review share, or do not share, with the traditional model of review. Such diagnosis can help policymakers choose among possible proposed reforms.[16]

IMMIGRANT J. 18 (Oct. 2014), http://www.lccr.com/wp-content/uploads/NCCIJ-Access-to-Justice-Report-Oct.-2014.pdf [https://perma.cc/L8GJ-8ZAT] (similar); John D. Montgomery, *Cost of Counsel in Immigration: Economic Analysis of Proposal Providing Public Counsel to Indigent Persons Subject to Immigration Removal Proceedings*, NERA ECON. CONSULTING 3 (May 28, 2014),
http://www.nera.com/content/dam/nera/publications/archive2/NERA_Immigration_Report_5.28.2014.pdf [https://perma.cc/4H72-UPEM] (conducting a cost-benefit analysis of government-provided counsel for immigrants); *see also* Dana Leigh Marks, *An Urgent Priority: Why Congress Should Establish an Article I Immigration Court*, 13 BENDER'S IMMIGRATION BULL. 3 (2008); *Policy Priorities*, FED. BAR ASS'N (Jan. 21, 2021), http://www.fedbar.org/Advocacy/Issues-Agendas.aspx [https://perma.cc/ML5V-BJGA] (advocating the formation of an Article I immigration court).

14. For a list of prisoner-grievance policies, see *Prisoner and Jail Grievance Policies*, U. MICH. L. SCH. (2021), https://www.law.umich.edu/facultyhome/margoschlanger/Pages/PrisonGrievanceProceduresandSamples.aspx [https://perma.cc/L8PY-NTZU].

15. *See, e.g.*, *Complaint Process*, DISTRICT COLUMBIA OFFICE POLICE COMPLAINTS, http://policecomplaints.dc.gov/page/complaint-process-opc [https://perma.cc/6FHM-Q3UL].

16. This Article's goal is descriptive rather than normative, and it leaves to others the defense of the normative proposition that consistent application of rules to similar cases is desirable. But it is important to note that this Article's empirical tests single out a particular type of consistency: consistency *across judges* and *within hearing offices*. The promotion, by administrative review, of that cross-judge, within-hearing-office consistency is not subject to two of the most prominent normative objections to uniformity as a due process value.

The first inapplicable objection is that pursuing uniformity prevents judges from considering relevant differences between cases. For example, pursuing uniformity in grant rates across different regions might mean applying *disuniform* standards to reach the same results despite differences in cases. That objection is not relevant here, since I measure only uniformity across judges within the same court, for whom cases are randomly assigned—who face the same caseload. In other words, this Article is concerned with how often judges who evaluate the same caseload reach opposite conclusions.

The second inapplicable objection is that bright-line rules—the most common method of promoting uniformity—impose high costs. A simple bright-line rule applied to all disability applicants—for example, one holding that only applicants with odd-numbered birth dates receive benefits—would be one way to make outcomes appear uniform across judges with the same

B. A Simple Theory of Review

The standard administrative mechanism for controlling adjudicators' discretion is hierarchical, adversarial administrative review.[17] By hierarchical, adversarial review, I mean a system allowing a disappointed claimant to appeal an adverse decision to an authority that has the power to reverse, remand, or affirm that decision. Such systems are ubiquitous in the administrative state and the courts, and legal scholars have offered theoretical accounts of their purposes and effects. I build on these accounts to offer a simple theory of administrative review as a device for limiting discretion in decisionmaking.

Any system that allows a second decisionmaker to review the work of a first decisionmaker includes at least three stages.[18] First, some set of decisions is selected for review. Second, review takes place. Third, if an error is discovered, some sort of consequence, such as a remand, follows for the original decisionmaker.

In a system of hierarchical, adversarial review, these three stages occur in a distinctive way. First, the losing party chooses whether or not to appeal, and that choice determines which cases receive review (*litigant selection*). Second, the appellate body reviews appealed decisions for legal or factual error, with varying degrees of deference (*decisionmaking*). Third, the appellate body affirms, reverses, or remands the initial decision; upon remand, the case usually returns to the same initial decisionmaker, who is unlikely to greet the remand with pleasure (*remand aversion*).

Thinking of the process in these three simple steps generates expectations about the results. To develop these expectations, I draw on Steven Shavell's influential model of appeal as a means of error

---

caseload. No one would advocate that. Yet the rule/standard tradeoff is not at issue in evaluating the effect of *review* on cross-judge, within-hearing-office consistency.

Reduction of cross-judge, within-hearing-office disparity on review is not evidence of bright-line rule application that raises concerns about the costs of uniformity. To see this, consider an example: if the reviewing body were to reverse all even-numbered-birthday cases, harsher and more lenient judges would be equally likely to be reversed. In other words, application of a bright-line rule by a reviewing court would reduce disparities only with respect to judges' systematically differing application *of that rule*. A reduction in disparities after review is therefore not in itself evidence that the reviewing court is preventing judges from making decisions sensitive to individual facts. More relevant for my purposes, the *lack* of a reduction in cross-judge disparities hardly suggests that the appeals court is applying an individualized fact-sensitive test—though it may imply that the appeals court defers substantially to the trial judge's initial determinations.

17.   Of course, there are other promising proposals for promoting consistency, including quality assurance systems, see *supra* note 3 and accompanying text, and peer review, see Ho & Sherman, *supra* note 7, at 255, 258, 265.

18.   Review by a second decisionmaker is far from the only possible method of limiting adjudicators' discretion, but such review is the topic here. For an extremely useful typology of a broader range of methods of limiting adjudicators' discretion, see Ho & Sherman, *supra* note 7, at 254.

correction,[19] but I tweak the assumptions, leading to different implications.[20] Most important, I assume that litigants have imperfect information about whether the decision in their case was extreme.[21]

At the first stage of review, litigants select cases to appeal. In deciding whether to appeal, I assume that disappointed litigants have some knowledge of whether their cases have merit and, therefore, about whether the denial of benefits was unusual and likely to be reversed on appeal.[22] For claimants with lawyers, this knowledge may come from experience with other cases. Even claimants without lawyers may have some sense of whether a decision is extreme, however, partly because they know whether the decision matches their own view of their disability, and partly because they can observe the demeanor of their ALJ and may even know his or her allowance rate. Litigants, then, are more likely to appeal denials of meritorious claims and, therefore, to appeal from the decisions of harsh judges, who more often deny meritorious claims.[23]

To make this idea concrete, consider Figure 1. In that simplified figure, decisions—both grants and denials—have a distribution of merit, and there are only two judges. Judge A, who is harsh, grants only the top 25% of cases by merit (the cases in the light-grey region, to the right of the dashed line). Judge B, who is generous, grants the 75% of cases before her with the most merit (all decisions to the right of the dotted line). None of the top 25% of cases by merit are appealed, since none of those decisions are denials. For Judge A, appealable denials fall into either the dark-grey or the black sections of the figure—the bottom 75% of decisions by merit. For Judge B, by contrast, the only appealable denials are from the bottom quarter of decisions by merit.[24]

---

19. Shavell, *supra* note 1. Shavell asks the question of why one might use an appeal process rather than investing the same resources in avoiding errors at trial. His theory answers this question with the simple insight that each litigant has knowledge about whether the trial court made an error in his or her case, and that this knowledge makes appeals more efficient because they need only decide cases that litigants have already identified as likely mistaken.

20. Consistency across decisionmakers may partly proxy for accuracy—but only partly—and I do not address other important purposes of administrative review, including its role in policy development and political accountability. For a more comprehensive account of the purposes of administrative review, see generally CHRISTOPHER J. WALKER & MATTHEW LEE WIENER, ADMIN. CONF. OF THE U.S., AGENCY APPELLATE SYSTEMS (2020).

21. See *infra* fig.2 and accompanying text.

22. This resembles Shavell's assumption that litigants know whether an error was made in their case; I assume, similarly, that they know whether a decision was extreme. Note that the empirical tests later in this piece do not test whether judges who are outliers are more likely to be appealed and reversed, but rather whether the allowance rate of a judge is correlated with his or her rate of appeal and reversal. The assumption in this research design is that the incidence of extreme decisions tracks judges' allowance rates.

23. Because there is some cost of appeal, rational litigants do not appeal when they are unlikely to win reversal.

24. I assume that there is some underlying distribution of merit in cases rather than that cases have one correct outcome. If one were to assume that cases had one correct outcome, then judges' differing allowance rates would depend on how many of their denials and grants were mistaken. That model would yield similar predictions, but I prefer to imagine a distribution of merit among cases, with reasonable adjudicators disagreeing about some outcomes.

Figure 1: Denials and Merit
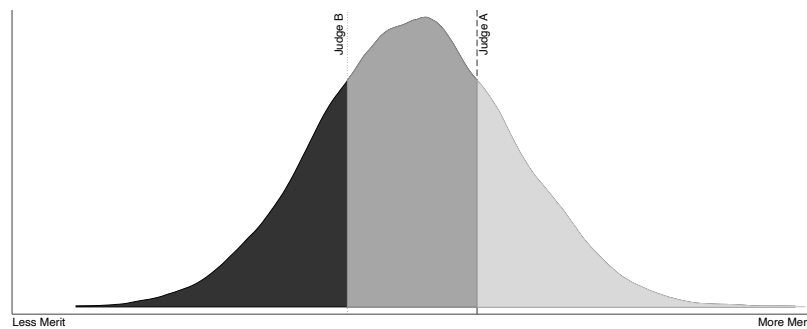
Because I assume that litigants know how meritorious their cases are, and that litigants are more likely to appeal more meritorious cases, litigants are more likely to appeal denials of second and third quartile cases (dark grey in the figure) than denials of first quartile cases (black in the figure). Figure 2 shows the theoretical relationship between merit and the likelihood of appeal; the black dots (cases in the lower 25th percentile of the merit distribution) are less likely to be appealed than grey dots (cases between the 25th and 75th percentiles of the merit distribution). The middle two quartiles include denials only from Judge A. As a result, Judge A not only has more denials total, but a higher *proportion* of his or her denials is meritorious and likely to lead to an appeal. (There are no light-grey dots because there are no denials, and therefore no appeals, from the top 25% of the merit distribution.) The decisions of the harsher judge (Judge A) are more likely to be appealed.

This is *litigant selection*, the first stage of adversarial, hierarchical review. Litigant selection depends on litigants' knowledge and decisions—rather than, for example, random selection—to choose which cases to review. Litigant selection implies that claimants appeal more often after losing before a harsher ALJ.

Figure 2: Merit and Appeals
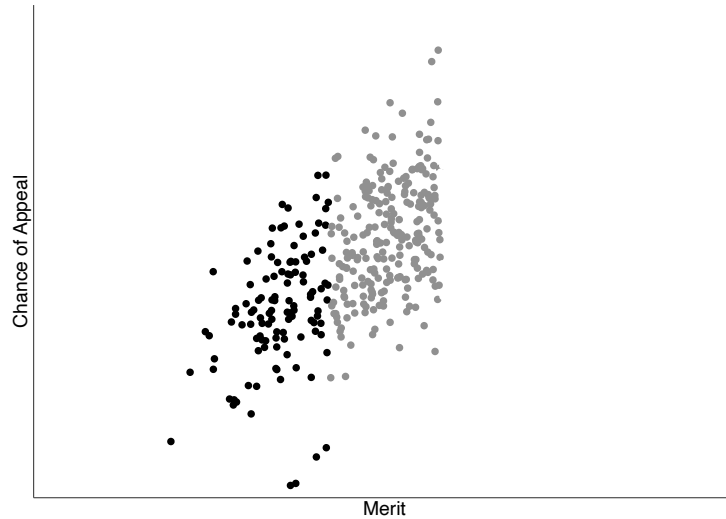
Because litigants are more likely to appeal more meritorious cases, the appellate review body hears cases that have more merit than the average denial. But merit is not the only consideration that determines whether a litigant appeals, and the chance of appeal varies for any given level of merit. As a result, the appeals body must still distinguish among more meritorious and less meritorious cases.

At the second stage of the appeals process, *decisionmaking*, I expect appellate adjudicators to reverse denials in meritorious cases more often than denials in less meritorious ones. Such adjudicators will therefore more often reverse the decisions of harsher judges, who more frequently deny meritorious claims. This follows from the relative weakness of my first assumption: litigants have only *some* knowledge of a decision's merit, not perfect knowledge. Litigants, like adjudicators, make mistakes; they may believe their cases have merit when they do not, and vice versa. I also assume that adjudicators themselves have imperfect information and cannot fully anticipate whether their decisions will be appealed or reversed. Given this mix of decisions, I expect the appeals process to reverse denials in more meritorious cases more often than denials in less meritorious ones. Because, as explained above, the cases appealed from harsh judges should have more merit on average, those appeals should more frequently lead to reversal.

Figure 3 shows this expected pattern graphically. Many of the dots from Figure 2 are no longer present because many cases were not appealed, but among the appealed cases, the grey dots are more likely than

the black dots to result in reversals. Only Judge A has grey dots; Judge A is more likely than Judge B to be reversed on appeal.

Figure 3: Merit and Reversal
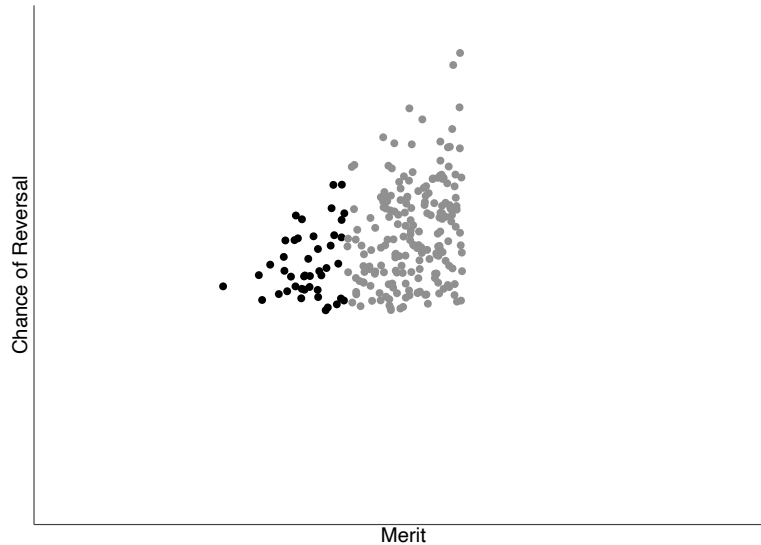
Third, I expect that judges seek to avoid remand. Unlike selection and decisionmaking effects, the effects of line-level judges' efforts to avoid remand are largely invisible, since judges may not react to actual remands but rather anticipate possible remands. One implication of judges' efforts to avoid remands is, however, observable: that judges try harder to avoid reversal *soon* after having experienced a remand. This expectation is grounded in two assumptions. First, I expect that, even if judges largely anticipate remands, they sometimes learn something unexpected, and that remands therefore influence their behavior in the short term, leading them to avoid the error that led to the remand. Second, I expect that judges overestimate the probability of remands when they have recently experienced them. Such overestimation reflects what cognitive psychologists call the availability heuristic.[25] Examples of the availability heuristic are common: for example, "[m]any readers must have experienced the temporary rise in the subjective probability of an accident after seeing a car overturned by the side of the road."[26] This cognitive pattern implies that the effects of remands should be observable in the immediate aftermath of a remand. For Social Security claims, since

25.   Amos Tversky & Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, 5 COGNITIVE PSYCH. 207, 230 (1973).

26.   *Id.*

claimants initiate nearly all appeals, I expect judges to become slightly more generous, on average, after reversals. I call this effect *remand aversion*.[27]

This simple three-stage theory offers a way of thinking about three prominent strains of policy prescriptions for systems of administrative review. It also gives rise to three hypotheses, which I describe and test in Part III below, taking advantage of the near-random assignment of cases to ALJs.

C. Review Proposals

In canvassing three categories of review proposals, I begin with Jerry Mashaw's advocacy of quality assurance systems and other management reforms to protect due process rights in welfare-claims processes. In 1974, Mashaw famously argued that the procedural safeguards of adversarial due process are inadequate for the adjudication of social-welfare claims, partly because welfare claimants may not be able to take advantage of trial-like procedural safeguards.[28] Mashaw therefore suggested a quality assurance system: a management system designed to ensure accuracy in adjudications without the trappings of adversarial litigation.[29]

For illustration, here, drawn from Mashaw's work, is a short description of one such system: the Veterans' Pensions and Compensation disability program.[30] In that program, a system of quality assurance

---

27. Note that the first of the three steps in the theory requires litigants to know—albeit imprecisely—whether a decision reflected an extreme exercise of discretion, and to act on that knowledge, whereas the second and third effects depend on the imperfection of litigants' and judges' knowledge. I believe that both of these assumptions are realistic. Lawyers and claimants have some knowledge of the extremity of the decision in their case, but that knowledge is necessarily limited: some litigants are overconfident, some issues are close, and some lawyers misjudge even issues that are not close. Moreover, lawyers' time and resources are limited, and they—like adjudicators—make errors.

28. Jerry L. Mashaw, *The* Management *Side of Due Process: Some Theoretical and Litigation Notes on the Assurance of Accuracy, Fairness, and Timeliness in the Adjudication of Social Welfare Claims*, 59 CORNELL L. REV. 772, 775 (1974).

29. *Id.* at 791. For another account of a quality assurance system in action, see Deborah A. Chassman & Howard Rolston, *Social Security Disability Hearings: A Case Study in Quality Assurance and Due Process*, 65 CORNELL L. REV. 801 (1981) (describing the implementation of a quality-assurance system for disability adjudications). For a discussion of the problems facing attempts to implement quality-assurance systems in government, see James E. Swiss, *Adapting Total Quality Management (TQM) to Government*, 52 PUB. ADMIN. REV. 356 (1992) (suggesting that total quality management—a system that was widely adopted in postwar manufacturing in Japan, and that more recently became influential in the United States—could be effective in government only with substantial modifications).

30. I summarize Mashaw's account of that system rather than describing it myself—or updating the particulars that may have changed in the four decades since Mashaw's article—since my goal here is conceptual rather than empirical. On the issue of the inadequacy of adversarial procedures, see Jerry L. Mashaw, *Conflict and Compromise Among Models of Administrative Justice*, 1981 DUKE. L.J. 181, 211-12 (1981) (suggesting that courts should consider nonadversarial procedures for protecting due process and decrying the false choice between "the legal chauvinism

supplemented an administrative appeals process. Each day, reviewers in regional offices considered a random sample of initial disability decisions, looking for procedural and substantive errors.[31] They logged these errors in monthly reports sent to a national Office of Appraisal.[32] The national office then conducted its own review of a random sample of cases and noted where its review did not match the findings of the regional office.[33] Finally, individual determinations of error were returned to the initial adjudicator, who was asked to agree or disagree with the appraisal of his or her error.[34]

This quality assurance system was both similar to and different from a classic system of hierarchical, adversarial review. Instead of relying on disappointed claimants to generate appeals, the quality assurance system randomly selected cases for additional review. The second stage of the review process was similar to the appellate decisionmaking stage in hierarchical review: although the decisionmaker did not receive briefs from the parties, he or she did, like an Administrative Appeals Judge, evaluate the decision below for procedural and substantive error. Finally, the last stage, in which decisions were returned to the initial decisionmaker, resembled the remand stage in hierarchical, adversarial review, except that the decisionmaker was not asked to redo the case, but rather to register agreement or disagreement with the finding of error.

In sum, the largest difference between quality assurance and adversarial, hierarchical review is in the selection of cases—the first step of the review process. Favoring a quality assurance system over more traditional review therefore usually depends on the view that the adversary selection process for appeals is inadequate. Mashaw recognizes this, and he notes that appeals decisions themselves could produce information about patterns in errors in decisionmaking, but he doubts that they do, because appeals adjudicators hear only individual cases and are unable to perceive larger patterns.[35] His view is plausible; it relies on the assumption that litigants often do not select mistaken or extreme decisions for appellate review. If litigants do select such decisions consistently, then patterns in those cases will reach appellate decisionmakers, and those decisionmakers should be able to correct them.

I share Mashaw's view that selection problems trouble many systems of administrative review. Of course, even if quality assurance programs are able to address selection problems through the random selection of cases,

---

of a ubiquitous demand for trial-type procedure and an unexamined deference to administrative or legislative choice of some alternative mode of decisionmaking").

31.   Mashaw, *supra* note 28, at 794.
32.   *Id.*
33.   *Id.*
34.   *Id.* at 795.
35.   *Id.* at 785.

they may not achieve their aims. In the most credible study to date, Daniel Ho and his coauthors find that a quality review program at the Board of Veterans Appeals, which randomly selected cases for review and potential feedback from a quality review team, did little to reduce numbers of appeals or reversals.[36]

The effectiveness of claimant-driven selection is at least partly measurable. Below, I offer evidence that selection in Social Security appeals, although far from perfect, is functional at the ALJ level. For example, it functions much better than the analogous selection process in immigration appeals. Of course, quality assurance systems include features beyond random selection. For example, the Veterans' Pensions and Compensation quality assurance system not only altered the selection of cases for appeal but also led to the creation of statistical reports on outcomes that were shared with regional offices. Absent a selection problem, however, such reports could also be compiled from decisions of a standard adversarial appellate body. Finally, perhaps a quality assurance system would apply a different standard of review, focusing less on procedure and more on substance. The results below suggest that such a change could be useful for disability appeals, but it need not be paired with random selection of cases for review.

Peer-review proposals, unlike quality assurance proposals, squarely attempt to improve the decisionmaking process itself—the second stage of review.[37] In fact, one way to understand peer review is as a form of quality assurance that more directly targets inconsistency in decisionmaking.[38] Rather than face review by an appellate body, adjudicators are paired (or placed in larger groups) and reach a decision together or provide each other feedback on decisions. Peer review may promote consistent decisionmaking through "the informal pressures of pride and shame," and through a requirement that decisions "be explained and subjected to examination."[39]

As an example, consider the peer-review procedures for food safety inspectors recently instituted by the Public Health Department of Seattle and King County. These procedures offer an especially good example because of their randomized implementation; the resulting study has yielded the most persuasive empirical evidence to date that peer review makes decisionmaking more consistent.[40] In that study, some of the

---

36.   Ames et al., *supra* note 11; Ho et al., *supra* note 11, at 252-58.

37.   For a discussion of experimentalist administration, see Sabel & Simon, *supra* note 13, at 78-93. One of Sabel and Simon's examples combines elements of quality assurance and peer review: the Quality Service Review program in child-welfare programs in Utah. That program used stratified random sampling to identify cases for review by a two-person team. *Id.* at 91-93.

38.   *See, e.g.*, Ho, *supra* note 7, at 79 (contrasting peer review with "more conventional forms of quality assurance").

39.   Sabel & Simon, *supra* note 12, at 91.

40.   Ho, *supra* note 7.

county's health inspectors were randomly assigned to a program in which they conducted restaurant inspections together with one other inspector, rather than alone. The peer review included not only these paired inspections but also weekly training sessions at which difficult issues were discussed. These sessions produced guidance memoranda on difficult issues, which aided further training. These initiatives significantly decreased inconsistency across inspectors, not only for inspections conducted jointly, but also for individual inspections conducted by inspectors who had participated in the peer review.[41]

Peer review makes review nonhierarchical, substituting the review of a peer for the review of a higher authority. That substitution most directly alters the second and third stages of review: it means that review and initial decisionmaking take place simultaneously, and that feedback (the analogue of remand in an adversary, hierarchical system) also occurs simultaneously. Peer review therefore seems best suited as an intervention for adversarial systems of review in which the second and third stages do not function well. (In principle, peer review could leave the selection process unaltered, with review requested by claimants themselves. But in practice, peer-review proposals, like quality assurance proposals, tend not to make review depend on claimants' decision to appeal.)

Table 1 compares the traditional adversarial model of review with quality assurance and peer-review models. Quality assurance and peer review both eliminate claimants' role in selection; peer review also dispenses with the hierarchical nature of the review itself.[42] Finally, all three types of review offer decisionmakers feedback in cases of error, but the form of that feedback varies.

These categories, of course, bleed into each other. For example, both quality assurance processes and peer-review proposals may include a debriefing component, with the reviewer discussing any errors directly with the original decisionmaker. Both also favor selection processes that do not depend on action by disappointed claimants. And one can imagine useful initiatives that combine features of more than one category. For example, Jonah Gelbach and David Marcus propose that federal courts should engage in more "problem-oriented" decisionmaking when reviewing Social Security and immigration decisions—that is, that courts should pay attention to patterns across multiple cases and craft their opinions

---

41.   For a detailed account of the procedures, see *id.* at 31-38.

42.   Many standard systems of review incorporate an element of peer review: panel decisionmaking. When the Social Security Appeals Council decides to grant full review of a case—which almost always means remand—the signature of a second Administrative Appeals Judge is required on the opinion. And before streamlining at the Board of Immigration Appeals in the early 2000s, many decisions were reached by panels of three Board members. Of course, the federal Courts of Appeals also decide cases in panels of three.

accordingly.[43] That attention to patterns would integrate some of the benefits of the pattern-driven selection that is more common in quality assurance proposals.

Table 1: Review Proposals Compared

	Adversarial, Hierarchical Review	*Quality Assurance*	*Peer Review*
Selection	Claimants appeal	Random or pattern-based	Random or pattern-based
Decision	Formal legal decision by higher authority	Legal decision by higher authority	Decision by agreement or with feedback among two or more peers
Follow-up	Remand	Feedback to initial decisionmaker	Discussion between peer decisionmakers

Finally, a third group of reformers does not suggest an overhaul to systems of adversarial, hierarchical review but rather proposes making such systems more closely resemble appeals in Article III courts. Scholars in this group often suggest making adjudicators more independent or improving legal representation for litigants. In Social Security disability adjudications, ALJs themselves have fiercely defended their independence, seeing it as a critical element of fair adjudication.[44] Other scholars have called for more adversarial process in the form of a lawyer representing the government in ALJ hearings.[45] And in the immigration courts, for example, where the government always has representation but

---

43.   Jonah B. Gelbach & David Marcus, *Rethinking Judicial Review of High Volume Agency Adjudication*, 96 TEX. L. REV. 1097, 1134-48 (2018).

44.   *See, e.g.*, Ass'n of Admin. Law Judges, Inc. v. Heckler, 594 F. Supp. 1132, 1133 (D.D.C. 1984), *amended sub nom.* Ass'n of Admin. Law Judges, Inc. v. Heckler, No. 83-0124, 1985 WL 71829 (D.D.C. July 2, 1985) (involving a challenge to procedures limiting the independence of ALJs).

45.   *See, e.g.*, David H. Autor & Mark G. Duggan, *The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding*, 20 J. ECON. PERSP. 71, 93 (2006) (recommending attorney representation for SSA at ALJ hearings and noting that the Social Security Advisory Board has made the same recommendation).

immigrants often do not, many have called for government-provided legal representation.

Thinking of review in three stages, and testing whether each stage is functional, offers a place to start for policymakers contemplating reforms. Flaws in the selection process, for example, invite quality assurance solutions, whereas flaws in appellate decisionmaking invite peer review. Table 2 shows how the diagnosed problems line up with the possible solutions.

Table 2: Review Diagnosis and Possible Policy Responses

	Adversarial Solution	*Management Solution*
Problems with Litigant Selection	Government-funded representation	Quality assurance— random selection of cases for review
Problems with Decisionmaking	Altered standard of review	Peer review
Problems with Remand Avoidance	Better functioning of first two stages to increase number of remands	Training and feedback for ALJs

This Article uses data from disability adjudications to offer an example of such diagnosis.

II. Social Security Adjudications

A. *Why Study Disparities in Social Security Adjudications?*

Social Security Administrative Law Judges decided an average of over 750,000 cases per year from 2010 to 2014—about twice as many cases as the federal district courts.[46] As the largest single system of adjudication in the western world,[47] but with a relatively homogeneous caseload, the

---

46. *See Federal Judicial Caseload Statistics 2015*, U.S. COURTS, http://www.uscourts.gov/statistics-reports/federal-judicial-caseload-statistics-2015 [https://perma.cc/D5HC-93EE] (noting 361,689 combined civil and criminal filings in the federal district courts in 2015).

47. Barnhart v. Thomas, 540 U.S. 20, 28-29 (2003) ("[T]he Social Security hearing system is probably the largest adjudicative agency in the western world.") (citation and internal quotation marks omitted).

disability adjudication system offers an excellent test case for theories of appeal and administrative review. This study uses newly available individual-level case-tracking data to examine whether and how the Social Security Appeals Council reduces disparities in ALJs' decisions to grant benefits.

The disability process employs over 8,000 examiners and over 1,500 ALJs, and the stakes of disability adjudications are high. District Judge Jack B. Weinstein has described disability appeals as in many respects "more wrenching than sentencing determinations . . . . [G]rant or denial . . . may mean permitting [claimants] to live in dignity or consigning them to poverty and charity."[48] Moreover, the accuracy of the Social Security adjudication system has important policy implications. The SSDI caseload has nearly tripled since 1989.[49] Inaccurate denials harm claimants; inaccurate allowances strain public resources and discourage claimants from working.

Scholars have long wrestled with large disparities across ALJs and have proposed a wide variety of policy changes that might enhance consistency. This Article advances that literature by testing whether Appeals Council review in fact already promotes consistency across ALJs. I apply the three-step theory of review to examine the mechanisms through which Appeals Council review might function: litigants' decisions about whether to appeal, Appeals Council decisions about whether to affirm the ALJ's decision, and ALJs' reactions to remands.

The Social Security Administration has repeatedly attempted to reduce the size of disparities across ALJs and especially to rein in ALJs with high allowance rates. Those attempts have met resistance from ALJs determined to protect their decisional independence. In *Nash v. Califano*,[50] a group of ALJs challenged a series of actions taken by SSA to supervise them; those actions included monthly quotas for decisions and explicit consideration of ALJs' grant rates; deviations from the average were discouraged.[51] The court held that the ALJs had standing to pursue their claim, and it certified a class.[52] Soon after, Congress passed the so-called Belmon Amendment, which required the Appeals Council to resume review, on its own motion, of ALJs' decisions to grant benefits;[53] SSA chose to target judges with high allowance rates. In response, the

---

48. Jack B. Weinstein, *Equality and the Law: Social Security Disability Cases in the Federal Courts*, 35 SYRACUSE L. REV. 897, 898 (1984).

49. Nicole Maestas, Kathleen J Mullen & Alexander Strand, *Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt*, 103 AM. ECON. REV. 1797, 1798 fig.1 (2013).

50. 613 F.2d 10 (2d Cir. 1980).

51. *Id.* at 14.

52. Nash v. Califano, 31 Fed. R. Serv. 2d 1119 (W.D.N.Y. 1981).

53. Pub. L. 96-265, § 304, 94 Stat. 441, 453-57 (1980).

Association of Administrative Law Judges brought suit against the Secretary of Health and Human Services, arguing that such targeting violated the Administrative Procedure Act.[54] Once again, however, the court did not resolve the underlying dispute. Although it concluded that SSA's implementation of the Belmon Amendment had violated the APA, it nonetheless dismissed the case and declined to grant prospective relief because SSA had, during litigation, ceased targeting individual ALJs with high allowance rates.[55]

B. The Disability Claims Process

Examiners in state Disability Determination Services offices are the first officials to review applications for disability, mostly using written medical evidence. If the examiner rejects an application, the claimant can request "reconsideration"—a written appeal, once again evaluated by a state DDS office. If the application is rejected at the reconsideration stage, the claimant can appeal again, this time in an oral hearing before a federal ALJ. Next, if the Appeals Council denies review or affirms the decision below, the claimant may appeal once again in federal district court and appeal the decision of that court to a Court of Appeals, and even on to the Supreme Court.

At every stage of review—initial review, reconsideration, ALJ hearing, and Appeals Council review—the determination of disability takes place in five steps.[56] First, SSA asks whether the claimant is currently employed—in SSA terminology, whether the claimant is engaged in substantial gainful activity (sometimes abbreviated SGA). If so, then the claimant is not disabled, and the inquiry ends. Second, SSA determines whether the claimant has a sufficiently severe and lasting medical impairment, or set of impairments, to qualify as disabled. If not, the case ends there. Third, SSA considers whether one or more of these impairments fits within a predetermined category of disabling impairments—a so-called "listing" found in an Appendix to the regulations governing disability determinations.[57] If the impairments fit a listing, the claimant is considered disabled, and the inquiry ends. If the impairments do not fit a listing, SSA considers the extent to which these impairments may nonetheless prevent work—in SSA jargon, the claimant's "residual functional capacity."[58] At the fourth stage, SSA evaluates whether the

---

54. Ass'n of Admin. Law Judges, Inc. v. Heckler, 594 F. Supp. 1132, 1133 (D.D.C. 1984), *amended sub nom.* Ass'n of Admin. Law Judges, Inc. v. Heckler, No. 83-0124, 1985 WL 71829 (D.D.C. July 2, 1985); *see* 5 U.S.C. § 556 (2018) (specifying duties of employees presiding at hearings); 5 U.S.C. § 4301(2)(E) (2018) (prohibiting performance ratings for ALJs).

55. *Heckler*, 594 F. Supp. at 1143.

56. Regulations set out the same five-step evaluation process for SSDI and SSI. *See* 20 C.F.R. § 404.1520(a) (2020) (SSDI); *id.* § 416.920 (2020).

57. Listing of Impairments, 20 C.F.R. pt. 404, subpt. P, app. 1 (2020).

58. *See* 20 C.F.R. § 404.1520(a)(4) (2020).

claimant's residual functional capacity would allow her to do the type of work that she has done in the past. If so, she is not disabled. If not, then the final stage of the inquiry concerns whether the claimant could adjust and perform some other type of work in the economy. At this last stage, SSA considers not only evidence of the claimant's impairments but also other characteristics, such as age and education.

This five-step test applies regardless of whether a claimant is seeking Supplemental Security Income (SSI) or Social Security Disability Insurance (SSDI), but for each program, the ALJ must make an additional determination before the final award. To grant Supplemental Security Income, the ALJ must determine that the claimant is indigent, and to grant Social Security Disability Income, the ALJ must determine that the claimant is insured—in other words, that the claimant has paid Social Security taxes for long enough to receive benefits.[59]

This Article's analysis begins at the level of ALJ review, which is already the third stage of review.[60] By beginning the analysis at the ALJ review stage, I consider only about one third of all disability claims that are denied at the initial level; two thirds never proceed past reconsideration.[61] Yet ALJ review is nonetheless a meaningful place to start. ALJ hearings are formal adjudications under the Administrative Procedure Act, and they include a range of procedural protections: claimants have a right to counsel (at their own expense), and they may present witnesses and appear in person or by video conference.[62]

Within sixty days of receiving a full or partial denial of benefits from an ALJ, a claimant may request review, in writing, by the Appeals Council.[63] The request for review is accomplished by filling out a simple one-page form (HA-520).[64] The claimant need not make legal arguments or provide briefing, and there is no filing fee.

---

59. HAROLD J. KRENT & SCOTT MORRIS, ADMINISTRATIVE CONFERENCE OF THE U.S., ACHIEVING GREATER CONSISTENCY IN SOCIAL SECURITY DISABILITY ADJUDICATION: AN EMPIRICAL STUDY AND SUGGESTED REFORMS 5-6 (2013).

60. In ten so-called "prototype" states, the reconsideration stage of review has been eliminated, so ALJ review makes up the second stage of the process. *See DI 12015.100 Disability Redesign Prototype Model*, SOC. SECURITY ADMIN. (Jan. 3, 2020), https://secure.ssa.gov/poms.nsf/lnx/0412015100 [https://perma.cc/MXT9-ZRXK].

61. In 2010, there were a total of 2,838,485 total disability-adjudication outcomes at all levels of the system; during the period of this study (2010-2014), there were an average of more than 800,000 ALJ-level outcomes per year. *See* Office of Ret. & Disability Policy, *Annual Statistical Report on the Social Security Disability Program, 2011*, SOC. SECURITY ADMIN., tbl.59, https://www.ssa.gov/policy/docs/statcomps/di_asr/2011/sect04.html#table59 [https://perma.cc/YTG2-EAP8].

62. 20 C.F.R. § 404.950 (2020).

63. *Information About Requesting Review of and Administrative Law Judge's Hearing Decision*, SOC. SECURITY ADMIN., https://www.ssa.gov/appeals/appeals_process.html [https://perma.cc/6SUA-P9S4].

64. *Request for Review of Hearing Decision/Order*, SOC. SECURITY ADMIN., https://www.ssa.gov/forms/ha-520.pdf [https://perma.cc/TW4K-4S5Y].

Review within the Appeals Council takes place in two stages. First, an analyst—who may or may not be an attorney—reviews the appeal and makes a recommendation about whether it should receive "review"—a term of art within the Appeals Council that means review by two Administrative Appeals Judges.[65] If the analyst recommends denial of review, the appeal goes to one of fifty-six Appeals Officers, who are attorneys, but who lack the authority to grant review. Instead, they issue only denials, though they may forward cases on to an Administrative Appeals Judge if they disagree with the analyst and believe that review should be granted. When either the analyst or the Appeals Officer believes that an appeal should receive review, that appeal makes its way to one of seventy Administrative Appeals Judges, who are authorized to grant review.[66] That Administrative Appeals Judge may still decide to deny review, but more often he or she brings in a second Administrative Appeals Judge.[67] If that second judge agrees that it should be reviewed, the appeal proceeds to consideration on the merits—what the Appeals Council calls review. Between FY 2010 and FY 2014, the rate at which the Council granted review ranged from 25% to 16%.[68]

Once review is granted, the Appeals Council (through two Administrative Appeals Judges) may reach one of several decisions. First, it may fully or partially affirm or reverse the ALJ's decision.[69] Second, it may dismiss a claim for benefits on any ground available to the ALJ.[70] Third, it may vacate the ALJ's decision and remand the case to the ALJ. In practice, however, a decision to have a second AAJ conduct review is a decision to reverse: 97% of review decisions are remands or favorable dismissals. (Figure 4 shows this process.)

The Appeals Council may also decide to initiate review on its own motion, allowing it to review a small number of allowances in addition to the many denials that are appealed by claimants.[71] Such review is known as pre-effectuation review, since it precedes the effectuation of benefits that follows a grant by an ALJ. That review must be initiated within sixty days of the final ALJ decision.[72] SSA's regulations foresee review of cases

---

65.   Telephone Interview with Judge (Apr. 28, 2016).

66.   KRENT & MORRIS, *supra* note 59, at 53.

67.   *I-3-8-11. Administrative Appeals Judge's Actions*, SOC. SECURITY ADMIN. (June 5, 2015), https://www.ssa.gov/OP_Home/hallex/I-03/I-3-8-11.html [https://perma.cc/69F7-8RWW].

68.   *AC Grant Review Actions as a Percentage of All AC Dispositions*, SOC. SECURITY ADMIN., https://www.ssa.gov/appeals/DataSets/AC02_AC_GrantReview_All_Dispositions.html [https://perma.cc/3323-6X4F]. Note that the data furnished to me by FOIA request did not include a variable indicating whether review was granted; instead, I am able to identify cases that involve more than one AAJ.

69.   *GN 03104.350 Appeals Council (AC) Jurisdiction*, SOC. SECURITY ADMIN., at A.2, https://secure.ssa.gov/poms.nsf/lnx/0203104350 [https://perma.cc/4AJR-YSCN].

70.   *Id.*

71.   20 C.F.R. § 404.969 (2020).

72.   *Id.*

identified through random and selective sampling, as well as of cases identified as wrongly decided by another branch of the agency.[73] Such review is relatively rare; in 2011, it accounted for less than 4% of the Council's caseload.[74] Unfortunately, such cases are imperfectly identified in the database released to me. Although I was able to identify cases in which the Appeals Council conducted review despite a grant of benefits, and I removed those from the analysis below, I could not identify own motion review cases where the claimant lost or partly lost. Those cases therefore remain in the data, but given their small number, they are extremely unlikely to have affected the results here.

---

73.   *Id.* ("We may refer a case requiring effectuation to the Appeals Council if, in the view of the effectuating component, the decision cannot be effectuated because it contains a clerical error affecting the outcome of the claim; the decision is clearly inconsistent with the Social Security Act, the regulations, or a published ruling; or the decision is unclear regarding a matter that affects the claim's outcome.").

74.   KRENT & MORRIS, *supra* note 59, at 52.

Figure 4: Within-Agency Disability Appeals Process

Initial Application
&
Reconsideration
(Not in Dataset)

↓

ALJ Hearing

↓

Request for Review
by Claimant

↓

Analyst Review

Appeals
Officer   Administrative
                           Appeals
                           Judge

Deny   Review
Review   (with Second
                           AAJ)

                    Affirm   Remand
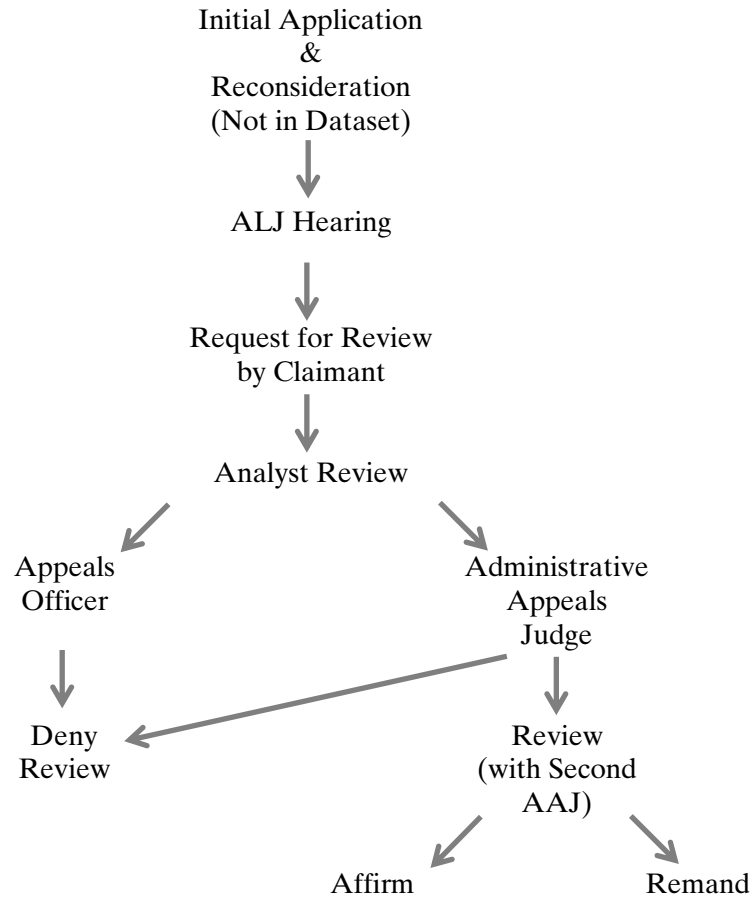
C. Scholarship on Social Security Adjudications

Disparities in ALJ decisionmaking have prompted handwringing from both lawmakers and scholars. In 2011, Sam Johnson, the Chairman of the Subcommittee on Social Security (Committee on Ways and Means) asked the Inspector General to provide information on ALJ workloads, policies, and procedures. Part of the Inspector General's response was a report of 24 ALJs who were outliers in terms of their grant rate or their workload.[75] The report noted large disparities—for example, allowance rates varying from 9% to 95% within a single office[76]—but also found that

---

75.   OFFICE OF THE INSPECTOR GEN., SOC. SECURITY ADMIN., A-12-11-01138, OVERSIGHT OF ADMINISTRATIVE LAW JUDGE WORKLOAD TRENDS 8 (2012).
76.   *Id.*

many of the 24 outliers in its sample appeared less extreme when the nature of their caseload was taken into account.

The size of ALJs' caseloads does not appear to drive the disparities. Harold J. Krent and Scott Morris used disposition data from SSA to examine correlates of ALJ allowance rates,[77] and they found a modest positive association between allowance rates and cases processed. They suggested, however, that time pressure did not play a large role in encouraging allowances. In particular, they examined decisions during the month of September, when the fiscal year is ending and the deadline for ALJs' annual case target looms, and they found that, despite completing more cases in that month, ALJs did not increase their allowance rate.[78]

Economists, meanwhile, have seen the variance in adjudicators' dispositions as an opportunity for causal inference. An important recent article by Nicole Maestas and coauthors combines administrative data from the earlier stages of the disability adjudication process with earnings data to study the effect of benefits receipt on employment and income.[79] The authors take advantage of (nearly) random assignment of disability applications to state-level examiners within the same Disability Determination Services office to offer a convincing estimate of this causal effect. This effect is large: the authors estimate that, for applicants with marginal claims to benefits—that is, those whose claims some examiners would grant and others deny—granting benefits reduces the chances of employment by 28% in the two years after receipt and 16% four years after.[80] These estimates highlight the importance of benefits decisions: grants and denials carry serious consequences, not only for claimants and government expenditures but also for the labor supply.

I focus on a related but distinct set of costs: the costs of inconsistent and unpredictable benefits decisions. Maestas et al. measure part of one of these costs: they find that the longer the disability determination process takes, the lower the chance that a claimant will later find employment if denied benefits.[81] If the Appeals Council more effectively reduced disparities across ALJs, this cost might initially increase, since more claimants would have a chance of obtaining reversals of benefits denials. Over time, however—especially with expanded review of benefits grants, rather than just of denials—a more effective review process would reduce

---

77.   KRENT & MORRIS, *supra* note 59, at 10-11.

78.   *Id.* at 26-28. Although the focus of this Article is on ALJ decisionmaking, federal district judges display inconsistency in deciding such cases as well, with large disparities in remand rates both within and across district courts. Harold J. Krent & Scott Morris, *Inconsistency and Angst in District Court Resolution of Social Security Disability Appeals*, 67 HASTINGS L.J. 367, 404 n.164 (2016).

79.   Maestas et al., *supra* note 49.

80.   *Id.* at 1818-19.

81.   *Id.* at 1822 (estimating a depreciation in the chance of employment of approximately 0.4% per month over four months).

the costs of uncertainty by decreasing disparities across adjudicators and ultimately lowering the number of initial disability claims.

Finally, scholars and officials have collaborated to describe possible reforms to the system. In 1990, Charles H. Koch, Jr. and David A. Koplow released an Administrative Conference study on the role and effectiveness of the Social Security Appeals Council.[82] Koch and Koplow pointed out that objective measurement of the accuracy of disability decisions is likely impossible, for several reasons. The appellate outcome is far from the last word on accuracy: review is asymmetric, with only denials of benefits reaching appeal, and cases change throughout the process, since new evidence may be introduced on appeal.[83] Koch and Koplow concluded that the Appeals Council was overwhelmed by its caseload and that any significant reforms would need to start with mechanisms for reducing that caseload. They found that a typical day for an Appeals Council member began with the receipt of two or three dozen new cases, meaning that simple denials might receive only a few minutes of attention each.[84] Perhaps more disturbing still, Koch and Koplow concluded that "ALJs generally do not see the Appeals Council as a legitimate appellate body."[85] ALJs' common criticisms included the view that Appeals Council adjudicators lacked training and experience and that the Council was subject to political manipulation.[86]

More optimistically, in a 2015 article, Gerald K. Ray and Jeffrey S. Lubbers describe a collaboration between the Administration Conference and the Appeals Council to learn from the adjudication data that SSA collects.[87] They emphasize the role of electronic processes in improving the quality of decisions, of decision trees in identifying noncompliance with agency policy, and of data analysis in identifying anomalies. These steps arose partly from Administrative Conference studies urging further use of the adjudication data at SSA's disposal.[88] The changes also broaden quality assurance measures that have long been in place. In 1981, Deborah A.

---

82. Charles H. Koch, Jr. & David A. Koplow, *The Fourth Bite at the Apple: A Study of the Operation and Utility of the Social Security Administration's Appeals Council*, 17 FLA. ST. U. L. REV. 199 (1990).

83. *Id.* at 271-75.

84. *Id.* at 239-40.

85. *Id.* at 294.

86. *Id.* at 293-94.

87. Gerald K. Ray & Jeffrey S. Lubbers, *A Government Success Story: How Data Analysis by the Social Security Appeals Council (with a Push from the Administrative Conference of the United States) Is Transforming Social Security Disability Adjudication*, 83 GEO. WASH. L. REV. 1575 (2015).

88. *See* Krent & Morris, *supra* note 78, at 370-71; *see also* ADMIN. CONFERENCE OF THE U.S., SSA DISABILITY BENEFITS PROGRAMS: THE DUTY OF CANDOR AND SUBMISSION OF ALL EVIDENCE (2012) (considering whether SSA should require claimants to disclose evidence unfavorable to their case); STEPHANIE J. TATHAM & MATHEW LEE WIENER, ADMIN. CONFERENCE OF THE U.S., EVALUATING SUBJECTIVE SYMPTOMS IN DISABILITY CLAIMS 54-55 (2012).

Chassman and Howard Rolston took up Mashaw's call for further use of quality assurance, describing the implementation of such a system for disability adjudications.[89] They concluded that the program had only had a limited impact so far, but that it had the potential to lead to significant improvements through the use of discriminant analysis to identify errors. The authors also observed quality assurance results that matched the selection results in this Article: errors were much more common in appealed denial decisions than in those that remained unappealed.[90]

Finally, a 2016 study by Jonah Gelbach and David Marcus offers a thorough and useful account of the role of judicial review in Social Security disability determinations.[91] Gelbach and Marcus largely focus on the role of federal district courts, but in the process they also offer one of the best accounts of the actual work processes of ALJs and the Appeals Council.[92] Ultimately, they provide a rich account of the reasons for frequent remands from federal district court (as well as for variation in remand rates across districts); they also offer a series of detailed policy recommendations. Most relevant here is their recommendation that the agency focus more on experimentalist attempts to facilitate mutual learning and perhaps less on supervision for technical policy compliance.[93]

These proposals are only the tip of the iceberg. Other scholars have canvassed policy alternatives ranging from a specialized Article I court for review of disability claims to the replacement of the all-or-nothing disability judgment with a graduated disability scale, similar to the scale used for veterans' benefits.[94] What scholars have not yet produced is an empirical appraisal of each stage of the Appeals Council process. Such an appraisal helps in choosing among the many proposed reforms.

III. Reviewing Appeals Council Review

A. *Data and Descriptive Statistics*

The data come from a FOIA request to the Social Security Administration, submitted in March 2014 and fulfilled in September 2015. The dataset includes all cases appealed to ALJs, including cases arising through continuing review of initial grants of benefits, that were closed during fiscal years 2010-2014. (The fiscal year runs from October 1-

---

89.   Deborah A. Chassman & Howard Rolston, *Social Security Disability Hearings: A Case Study in Quality Assurance and Due Process*, 65 CORNELL L. REV. 801, 801-03 (1981).

90.   *Id.* at 819.

91.   JONAH B. GELBACH & DAVID MARCUS, ADMIN. CONFERENCE OF THE U.S., A STUDY OF SOCIAL SECURITY DISABILITY LITIGATION IN THE FEDERAL COURTS (2016).

92.   *Id.* at 17-30.

93.   *Id.* at 168-71.

94.   *See, e.g.*, Paul Verkuil & Jeffrey Lubbers, *Alternative Approaches to Judicial Review of Social Security Disability Cases*, 55 ADMIN. L. REV. 733, 736 (2003) (describing Article I proposals); Autor & Duggan, *supra* note 45, at 94.

September 30.) Because appeals can take more than a year to be decided, I use only ALJ decisions reached between 2010 and 2012. In order to take advantage of the quasi-experiment provided by random assignment of cases to judges, and to match cases across levels of adjudication, I use a subset of the dataset for the estimates described below. Appendix A provides the details of these sample restrictions, as well as a description of the ways in which the assumption of random assignment of cases to judges does not hold in many hearing offices and time periods. In the subset of hearing offices and time periods used in the analysis below, observable patterns in the data are consistent with random assignment.

The data include the dates on which hearings and appeals were requested and decided, as well as the dates of the final dispositions by ALJs and the Appeals Council. Each case includes the sex and birth date of the claimant, as well as whether he or she was represented by a lawyer (although this field is inconsistently left blank). Crucially, the data include unique identifiers for the ALJ who decided the case as well as the hearing office at which it was decided; an identifier for the Appeals Officer or Administrative Appeals Judge who decided the case is also included. Table 3 shows descriptive statistics, displaying the age and sex of successful and unsuccessful claimants, as well as whether they are represented by an attorney. In both datasets, SSI and SSDI claims are treated separately, but I combine them for the bulk of the analysis because both claims require a determination of disability.[95] I categorize an outcome before an ALJ as a win if, for either an SSDI or SSI claim (or both), the ALJ makes a decision favorable to the claimant. The results in Tables 3 & 4 come from the restricted sample that I use throughout the Article, but the numbers in the full sample are similar.

Table 3 shows the winnowing of cases during the appeals process, displaying the percentage of claimants who appeal (among those who lost at least one of their claims before the ALJ), the percentage who reach an Administrative Appeals Judge, the percentage who reach a panel of two judges, and the percentage who obtain reversal or remand. Figure 4 accompanies Table 4, showing the within-agency appeals process in the form of a flow chart.

---

95.   Combining the two types of claims makes sense because I am interested in whether a claimant obtained benefits; I only count a decision as negative if an individual obtained neither type of benefits.

Table 3: Outcomes by Age, Sex, and Representation

	Age (Years)	Percent Male	Percent Represented
Unsuccessful Claimants	39	50%	57%
Successful Claimants	45	51%	80%
Total	42	50%	69%

Table 4: Winnowing of Cases on Appeal

Stage of Proceeding	Number of Cases	Percent of Previous Stage
Initial Hearing	663,901	100%
Appealable Loss (SSDI or SSI)	325,915	49%
Appeal	143,242	44%
Review by at Least One AAJ	76,910	54%
Review by Two AAJs	26,828	35%
Remand or Favorable Dismissal	26,108	97%

Nearly half of all disappointed claimants file an appeal—a large proportion, particularly given the relative rarity of remand. This may reflect the lack of a filing fee and little need for briefing. Once an appeal is filed, actual review—in which two Administrative Appeals Judges evaluate the case on its merits—is rare. About half of all appeals reach an Administrative Appeals Judge, and review occurs in fewer than 20%. The rest are directed by an analyst to an Appeals Officer for denial. If a case receives review by two AAJs, finally, it is nearly certain to result in a remand or a favorable dismissal; the effective decision to remand occurs with the decision to review.

B. Litigant Selection

The theory of review predicts that litigants are more likely to appeal the decisions of judges who are systematically harsher. In the context of disability adjudications, I expect claimants to appeal more often when they are denied benefits by a harsher judge. I expect this partly because the

Social Security Administration makes public the number of decisions each ALJ has reached each year, as well as how many were awards and how many were denials.[96] Websites make the data easily available to claimants and their lawyers.[97] Because claimants know both the overall grant rate of their ALJ, and whether they consider the denial of their benefits unfair, I expect that:

> H_1: Assignment to an ALJ with a lower allowance rate (compared to the rates of others in his or her hearing office) increases the probability that a claimant will appeal a denial of benefits.[98]

Figure 5 shows a visual test of this hypothesis. As predicted, claimants are more likely to appeal the decisions of harsher judges.[99] Each dot represents one judge; the vertical axis shows claimants' rate of appeal (by judge), and the horizontal access shows the judge's allowance rate relative to the rates of the other judges in his or her hearing office, with the average set to zero.[100] Looking from top left to bottom right, as the ALJ's allowance rate increases, the likelihood that his or her decisions will be appealed decreases.

The same pattern holds in linear regressions (Table 5) that include office-year fixed effects (which controls for the idiosyncratic characteristics of individual hearing offices and individual years).[101] The regressions also control for claimants' age and sex, as well as whether each claimant had a lawyer.[102] In any case, I show estimates with and without controls, so the reader can choose among specifications. The estimates suggest a 31-38% percent difference in rates of appeal between judges with 0% and 100% relative allowance rates. This translates into a 13-15 percentage point decrease in the chance of appeal from a denial when moving from a judge

---

96. *ALJ Disposition Data FY 2021*, Soc. Security Admin., https://www.ssa.gov/appeals/DataSets/03_ALJ_Disposition_Data.html [https://perma.cc/D3HL-J7KQ].

97. *See, e.g.*, *Disability Judges: Administrative Law Judge Case Statistics*, Disability Judges (2021), https://www.disabilityjudges.com [https://perma.cc/D89C-7RR5] (allowing allowance-rate searches by judge name and state).

98. Although assignment of cases to judges in the sample appears close to random, that randomness does not mean that judge allowance rates themselves are randomly assigned. Instead, random assignment to a judge means random assignment to a cluster of characteristics, of which the allowance rate is one measure.

99. Note that the appeal rate includes only cases in which the claimant was denied benefits (either SSDI or SSI), so this correlation is not simply mechanical.

100. I consider disparities only within (rather than across) hearing offices, and I test for random assignment of cases to judges. I find strong evidence that cases are not, in fact, randomly assigned in many hearing offices. In order to draw valid inferences about the size of disparities, I concentrate on a subset of hearing offices and time periods in which assignment appears to have been random. The Appendix offers details.

101. I use linear probability models throughout.

102. Given the assumption of random assignment within courts, controlling for these claimant characteristics should not be necessary to correct bias, but the controls may increase the precision of the estimates.

20% less generous than his or her office average to one 20% more generous. Since the rate of appeal is just under 45% during this period, that is approximately a 30% relative increase in the chance of appeal—a large increase. Table 5 shows these regression estimates.

Figure 5: Claimant Selection
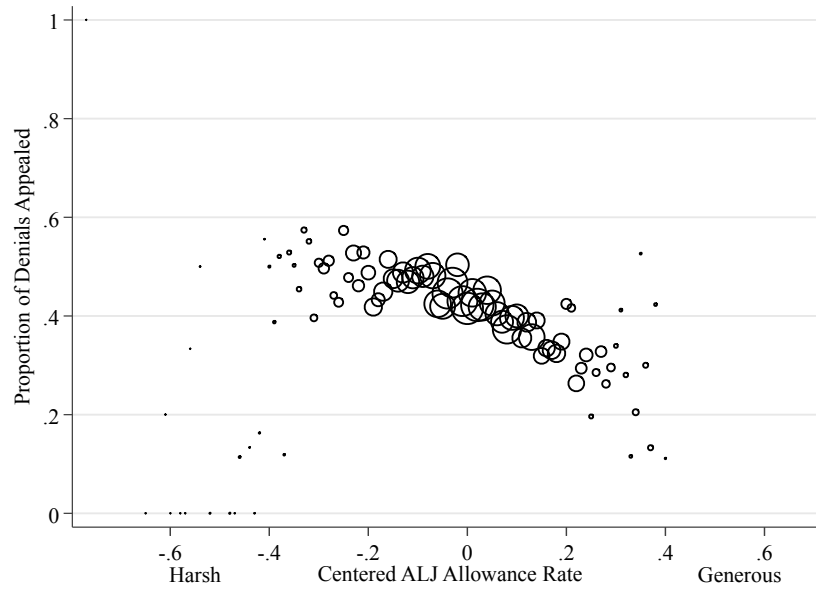
Note: Each circle represents 1/100th of the range of ALJ allowance rates (centered around each hearing office mean); each circle contains the cases of one or more ALJs within that range. Larger circles indicate more cases.

Table 5: Claimant Selection

Dependent variable is 1 if claimant appeals

	(1)	(2)	(3)
Demeaned ALJ Allowance Rate	-0.38*** (0.015)	-0.31*** (0.013)	-0.25*** (0.016)
Sex (Male=1)		0.025*** (0.0018)	0.025*** (0.0018)
Representation		0.26*** (0.0032)	0.25*** (0.0032)
Age		0.0013*** (0.000085)	0.0013*** (0.000085)
Representation*ALJ Allowance Rate			-0.12*** (0.025)
N	325915	325734	325734

Standard errors in parentheses
Models include year and office fixed effects
Standard errors clustered on ALJ
$^*\ p < 0.05,\ ^{**}\ p < 0.01,\ ^{***}\ p < 0.001$

The first stage of the review process works as expected. Litigants select more extreme opinions for appeal, sparing the Appeals Council the task of sifting through all ALJ decisions. This result is credible: because the assignment of the ALJ is nearly random, we can interpret these results as showing the causal effect of that assignment.

There is, however, one wrinkle. The effect of ALJs' allowance rates on claimants' choices about whether to appeal depends partly on whether those claimants are represented. That pattern is shown by the last row in Table 5—the interaction term (Representation*ALJ Allowance Rate). That term represents the association between appeals and the *combination* of representation and the judge's grant rate. The negative coefficient means that claimants without a lawyer (or other representative) are less influenced by the allowance rate of their ALJ than those who are represented.[103] Note, however, that the presence of an attorney is *not*

---

103. The dataset does not indicate whether the representative is an attorney; nonattorneys may represent claimants as well. Note that the coefficient on representation alone

randomly assigned; for example, claimants with better cases may have more success in finding a lawyer to take their case on contingency. The presence of a lawyer therefore might *or might not* suggest that the lawyer has a causal effect. It is also possible that some characteristic of the claimant, leading to representation, has an effect on whether the claimant appeals, and that the regressions are picking up that effect. In sum, it is clear that being assigned to a harsher judge makes claimants more likely to appeal. Having a lawyer might also make them more likely to appeal—but the evidence for that is not causal, since attorneys are not randomly assigned.

Claimant selection of cases works: claimants systematically choose more often to appeal the decisions of harsher judges. This does not mean, however, that the selection process is perfect. It may well be, for example, that unrepresented claimants too often fail to appeal, and that random sampling of denials in their cases would enhance consistency. Moreover, the observed effect, though large, could be larger still; in an ideal world, harsher judges might be even more likely to see their decisions appealed. And we know for certain that there is no selection process for generous ALJs' decisions, since the government cannot appeal. Yet this finding should, at a minimum, lead policymakers to hesitate before implementing a reform—such as a quality assurance system that randomly selects cases for review—that eliminates the role of claimants in selecting cases for appeal.

C. Decisionmaking

After claimants have chosen cases for appeal, the Appeals Council must evaluate those appeals for (1) an abuse of discretion by the ALJ, (2) legal error, (3) findings not supported by substantial evidence, and (4) "a broad policy or procedural issue that may affect the general public interest."[104] The Appeals Council may also consider new evidence on appeal, but only if it relates to the claimant's condition before the ALJ hearing.[105] Assuming that harsher ALJs are more likely to make legal errors and factual findings that harm the claimant's chances, and that harsher ALJs are more likely to abuse their discretion in that direction, those ALJs should more often be reversed when claimants appeal—as long as, as I have suggested earlier, claimants do not have *perfect* information about whether a decision was extreme. I test the following hypothesis:

---

simply shows that people with lawyers are more likely to appeal—an unsurprising result that could reflect the effect of either the presence of a lawyer or some other systematic difference between represented and unrepresented cases.

104. 20 C.F.R. § 404.970 (2020).

105. *Id.*

H_2: Appeals from ALJs with lower allowance rates are more likely to result in reversals than appeals from those ALJs' more generous colleagues.

Testing this hypothesis yields the opposite of the expected result: the *higher* the ALJ's allowance rate, the higher his or her reversal rate. Figure 6 shows this visually. This means that more generous ALJs—those who more rarely deny claims—are nonetheless more likely to be reversed on the rare occasions when they do deny such claims. Regressions with office-year fixed effects and other controls confirm this strange pattern. The results are in Table 6 below. This second main result relies on the near-random assignment of judges to cases, and it therefore has a causal interpretation: assignment to a harsher judge does not, on average, cause claimants to appeal more often.

Notably, however, the pattern in Figure 6 looks slightly nonlinear. In fact, Figure 6 might at first seem to show that the decisionmaking process is working, since more extreme judges are more likely to be reversed, regardless of whether they are generous or harsh. But recall that the reversals here are only of *denials*. The appeals process is therefore making both very harsh and very generous judges more generous.

Judges who are extremely harsh relative to their peers (more than 20 percentage points below their hearing office average) may be slightly more likely to be reversed, as the theory predicts. But so are judges who are more generous than their peers, and the effect is stronger for those generous judges. That pattern might make sense if the government were also appealing, but it is harder to make sense of given that the only appeals in the data here are those in which the claimant lost before the ALJ. I call this pattern only slightly nonlinear, and do not draw strong conclusions from it, because (in sharp contrast to Figure 5) the trend looks roughly flat through most of the data.

Figure 6: Reversals at the Appeals Council
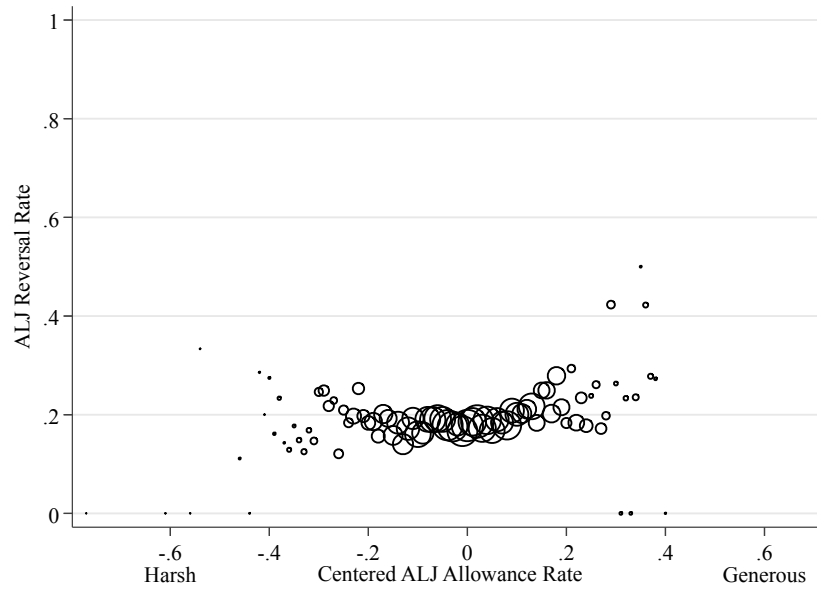
Note: Each circle represents 1/100th of the range of ALJ allowance rates (centered around each hearing office mean); each circle contains the cases of one or more ALJs within that range. Larger circles indicate more cases.

Table 6: Appellate Decisionmaking

Dependent variable is 1 if claimant obtains reversal or remand

	(1)	(2)	(3)
Demeaned ALJ Allowance Rate	0.047[*] (0.018)	0.057[**] (0.018)	0.10[***] (0.024)
Sex (Male=1)		-0.00067 (0.0020)	-0.00067 (0.0020)
Representation		-0.025[***] (0.0026)	-0.027[***] (0.0028)
Age		0.0019[***] (0.000081)	0.0019[***] (0.000081)
Representation*ALJ Allowance Rate			-0.062[**] (0.023)
N	142904	142770	142770

Standard errors in parentheses
Models include year and office fixed effects
Standard errors clustered on ALJ
[*] $p < 0.05$, [**] $p < 0.01$, [***] $p < 0.001$

What might explain this pattern? One possibility is that judges at both ends of the distribution are more careless, and that the appeals process is detecting those errors. But it remains troubling that claimants denied benefits by an ALJ 40% more generous than the hearing office mean are actually *more* likely to succeed on appeal than claimants denied benefits by a median ALJ. One would have expected the opposite: that only truly meritless cases would be denied by extremely generous ALJs, and that those denials would be very unlikely to be overturned on appeal.

The most plausible explanation for this strange pattern is that different types of cases are appealed from harsher and more generous judges. Denials of benefits by generous ALJs are much more likely to be dismissals (usually for failure to appear). This is unsurprising, though not immediately obvious. Because generous ALJs are more likely to exercise their discretion in favor of claimants, a smaller proportion of their denials should reflect exercises of discretion against claimants, and a correspondingly larger proportion should be relatively nondiscretionary denials, such as dismissals. Dismissals, unlike other denials, automatically

reach an Administrative Appeals Judge.[106] Dismissals may therefore be driving the unexpected pattern in the data.

Controlling for ALJs' dismissal rate offers evidence for this view. Table 7 shows the same basic regressions as in Table 6, but the second and third models now control for the rate at which ALJs dismiss cases. Adding that term to the regression eliminates the counterintuitive effect of ALJ allowance rates on reversal rates: the appeals process targets dismissals for particularly stringent review, and that targeting happens to reverse generous judges' denials. Here, I must acknowledge that dismissals are not randomly assigned, and that this result is not a simple causal one: it is possible that those judges with higher dismissal rates are different in other respects that drive these results. Moreover, the association with dismissals is smaller when I include other variables in the model (though it persists when I disaggregate the stages of review below). Dismissals may therefore be systematically associated with representation, age and gender. Still, given the qualitative evidence, I think it is likely that dismissals partly drive the pattern shown in Figure 6.

---

106.   Interview with Appeals Council Official, in Falls Church, Va. (May 20, 2016).

Table 7: Dismissal Rates and Reversal Rates

Dependent variable is 1 if claimant obtains reversal or remand

	(1)	(2)	(3)
Demeaned ALJ Allowance Rate	0.047[*] (0.018)	0.0020 (0.022)	0.033 (0.027)
Demeaned ALJ Dismissal Rate		0.13[***] (0.028)	0.053 (0.030)
Sex (Male=1)			0.0031 (0.0020)
Representation			0.012[***] (0.0027)
Age			0.0022[***] (0.000081)
Representation*ALJ Allowance Rate			-0.019 (0.024)
N	142904	142904	134096

Standard errors in parentheses
Models include year and office fixed effects
Standard errors clustered on ALJ
[*] $p < 0.05$, [**] $p < 0.01$, [***] $p < 0.001$

    Although controlling for dismissal rates accounts for the perverse pattern of more reversals for more generous judges, it does not generate the expected, opposite pattern. Why not? Here, I offer a guess based on interviews and aggregate statistics. ALJs said that the Appeals Council rarely questions ALJs' overall findings and conclusions; instead, it focuses on the adequacy of the ALJ's articulated rationale at each step of the opinion. This approach partly reflects the deference that the Appeals Council owes ALJs' findings: the Council affirms those findings if they are supported by substantial evidence, but it reverses any legal error, since it does not defer to ALJs' legal determinations.

    Perhaps, then, dismissals are one portion of a more general story. If the Appeals Council focuses particularly on clear errors in the ALJ's reasoning, it may be less likely to reverse on relatively discretionary grounds. Unfortunately, the individual-level data released to me contain little information about the reasons for either ALJ or Appeals Council

decisions. The Appeals Council does, however, publish aggregate statistics on the most common reasons for remand.

These data support the view that the Appeals Council focuses on errors in ALJs' articulation of reasons. Table 8, reproduced from the SSA website,[107] shows the most common remand reasons in 2014. (These reasons have changed little from year to year.)

Table 8: Remand Reasons

Percent of Cited Reasons	Remand Reason
8.7	Inadequate Evaluation: Mental Limitations
6.2	New Evidence on Appeal
3.9	Failure to Discuss Doctor's Opinion
3.8	Inadequate Consideration: Mental Disorder
3.8	Inadequate Evaluation: Limits on Exertion
2.7	Inadequate Evaluation: Limits on Manipulation
	Inadequate Rationale: Weight Given
2.6	Consultative Examination
2.4	Failure to Discuss Source
	Inadequate Articulation: Reasons to Reject
2.4	Doctor's Opinion
2.2	Other

Eight of the nine reasons (excluding Other) concern the adequacy of the ALJ's articulation of reasons for the decision. The ninth—new evidence presented for review—reflects a factor outside the control of the ALJ. This list reflects the Appeals Council's focus on the adequacy of ALJs' reasoning; unlike a federal court of appeals, the Appeals Council cannot affirm on any ground that the record supports. For example, in some cases, a nonexamining source's opinion might be useless (or at least relatively unhelpful), but a remand could result if that opinion is not identified or discussed. The same point applies to the other remand reasons.

Perhaps the counterintuitive pattern in reversals also reflects decisionmaking procedures within the Appeals Council. The individual-level data allow a more detailed look at the stages of that process. Recall that the first Appeals Council employee to look at a request for review is

---

107. *Top 10 Remand Reasons Cited by the AC on Remands of RRs or Own Motion Reviews*, SOC. SECURITY ADMIN., https://www.ssa.gov/appeals/DataSets/AC07_Top_10_RR_and_OM.html [https://perma.cc/7ZWH-SPYX].

an analyst, who may or may not be an attorney. The analyst uses a seven-page electronic worksheet to assess whether the ALJ complied with SSA policies at each step of the decision.[108] The details of the worksheet are confidential, but Appeals Council officials describe it as based on common problems identified in ALJ opinions—the types of problems listed in Table 8 above. After filling out the worksheet, the analyst makes a recommendation; if the analyst recommends review, the case goes on to an Administrative Appeals Judge, and if the analyst recommends denial of review, the case usually, but far from always, goes to an Appeals Officer. The database records whether a request for review was denied by an Appeals Officer or an Administrative Appeals Judge.[109] It is at this stage of Appeals Council review that the surprising pattern in decisionmaking is strongest: judges with high allowance rates are much more likely to have requests for review decided by an AAJ. Again, this pattern mostly disappears with a control for judges' dismissal rates. Tables 9 and 10 show the effect of judges' allowance rates at each stage of the Appeals Council decisionmaking process. Table 9 shows these effects without controlling for judges' dismissal rates; Table 10 includes that control variable.

Here, I should pause to emphasize that this set of regressions showing the relationship between dismissal rates, representation, and remands does not have a straightforward causal interpretation. Neither dismissal nor representation is randomly assigned. Unlike Tables 5 and 6, then, in which the coefficients have a straightforward quantifiable interpretation, Tables 7-10 offer only clues about the reasons for the puzzling decisionmaking pattern—quantitative clues that require qualitative context.

---

108. Interview with Appeals Council Official, in Falls Church, Va. (May 20, 2016).
109. I distinguish between the two types of adjudicators using the fact that Appeals Officers can issue only denials of review. I therefore categorize adjudicators with zero percent review rates as Appeals Officers. Since Appeals Officers can and sometimes do become Administrative Appeals Judges, I perform this categorization by month.

Table 9: Effect of ALJ Allowance Rates by Stage of Appeals Council Decisionmaking

	(1) Appeal Reaches an AAJ	(2) Appeal Receives Review (Given that It Reaches an AAJ)	(3) Reversal or Remand (Given Review)
Demeaned ALJ Allowance Rate	0.12*** (0.027)	0.068* (0.033)	0.029* (0.014)
Sex (Male=1)	-0.0015 (0.0025)	0.0025 (0.0033)	-0.0030 (0.0021)
Representation	-0.075*** (0.0036)	0.0097* (0.0042)	-0.0091*** (0.0022)
Age	0.0015*** (0.00010)	0.0027*** (0.00013)	0.000011 (0.000078)
Representation*ALJ Allowance Rate	-0.054 (0.028)	-0.028 (0.034)	-0.051** (0.017)
N	142601	76842	26794

Standard errors in parentheses
Models include year and office fixed effects
Standard errors clustered on ALJ
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Effect of ALJ Allowance Rates by Stage of Appeals Council Decisionmaking (Accounting for Dismissals)

	(1) Appeal Reaches an AAJ	(2) Appeal Receives Review (Given that It Reaches an AAJ)	(3) Reversal or Remand (Given Review)
Demeaned ALJ Allowance Rate	0.057* (0.028)	0.034 (0.037)	0.022 (0.015)
Demeaned Dismissal Rate (Among Denials)	0.16*** (0.024)	0.087* (0.042)	0.018 (0.014)
Sex (Male=1)	-0.0014 (0.0025)	0.0026 (0.0033)	-0.0030 (0.0021)
Representation	-0.073*** (0.0035)	0.011* (0.0042)	-0.0089*** (0.0023)
Age	0.0015*** (0.00010)	0.0027*** (0.00013)	0.000012 (0.000078)
Representation*ALJ Allowance Rate	-0.048 (0.028)	-0.024 (0.034)	-0.050** (0.017)
N	142601	76842	26794

Standard errors in parentheses
Models include year and office fixed effects
Standard errors clustered on ALJ
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Interviews with former AAJs, ALJs, and an Appeals Council analyst suggest a possible explanation for the particularly large effect of judge allowance rates at the initial stage of sorting requests for review between Appeals Officers and Administrative Appeals Judges. When analysts are unsure of the right outcome,[110] they are more likely to bring the case to an AAJ, who has the power to issue either decision. Analysts are perhaps more likely to be unsure about their recommendation when they find some error in the ALJ's articulation of reasons for the decision, even if that error is minor. ALJs with high allowance rates may have higher rates of policy

---

110.   Telephone Interview with Former Appeals Council Analyst (May 19, 2016).

noncompliance in the articulation of the rationale for denials. Denials by more generous judges may also more commonly be easy cases, in which the ALJ need not strain to reach a denial decision. Paradoxically, the strength of the evidence underlying a denial might lead the ALJ not to articulate a rationale as carefully, making the Appeals Council analyst unsure whether to refer a case to an Appeals Officer or an Administrative Appeals Judge.[111]

This hypothesis is consistent with the views of ALJs, who complain about the technical nature of the remands that they receive. One ALJ, for example, when asked to describe examples of remands, described two highly technical errors.[112] First, he complained that the Appeals Council is too demanding of ALJs in presenting the reasons for dismissals for failure to appear at a hearing. "Dismissals consume an inordinate amount of time," he said. "No other court requires a judge to jump through all the hoops the Appeals Council does to reach a claimant. We almost have to prove that the claimant got the notice." Second, he described being reversed because he had failed to enter a piece of evidence into the record—and then realizing that his failure reflected the fact that the evidence had been submitted the day after he had issued his decision. (Recall that the Appeals Council will accept new evidence as long as it is related to the period before the ALJ hearing.)

Other ALJs offered similar accounts. One retired judge, who served both on the Appeals Council and as an ALJ, said he thought that the Appeals Council rarely reversed on the ground that an ALJ abused his or her discretion or reached a factual finding unsupported by substantial evidence. "I don't think they are substituting their judgment," he said.[113] Much more common were reversals on the basis of new evidence or for legal error.

To sum up, I have offered causal evidence that assignment to a harsher judge does not make claimants more likely to obtain a remand after appealing. The decisionmaking stage of Appeals Council review does not promote uniformity in ALJ decisionmaking. My hypotheses about the mechanisms driving this result—the checklist-like focus on procedure at the Appeals Council, combined with deferential review of factual conclusions and the lack of a harmless error doctrine—depend more on the qualitative evidence and are therefore more speculative.

Of course, the lack of a consistency-enhancing effect of the decisionmaking process is offset by the effect of the selection process. Taken together, the selection and decisionmaking phases of Appeals

---

111. This possibility suggests a potential tradeoff between error correction and consistency promotion: it may be that the most generous ALJs do make more errors. But the result—that they receive remands at a higher rate, pushing their allowance rates higher still—seems unintended.

112. Telephone Interview with Administrative Law Judge (Apr. 27, 2016).

113. Telephone Interview with Former Administrative Law Judge (May 11, 2016).

Council review do exert a small consistency-promoting (discretion-limiting) effect. Table 11 shows estimates of the overall effect of the process—that is, the effect of ALJ allowance rates on whether denials are *both* appealed and reversed. These estimates therefore combine selection and decisionmaking effects. Since the consistency-enhancing effect of the selection process is larger than the consistency-defeating effect of the decisionmaking process, the effect of the process overall is to promote consistency, but to a very modest degree. The results suggest about a 1-2 percentage-point decrease in the chance of appeal and remand when moving from a judge 20% less generous than his or her office average to a judge 20% more generous. In other words, the appeals process does limit discretion but not enough to make a large dent in the cross-ALJ disparities.

Table 11: Overall Effect of ALJ Allowance Rates on Appeals Council Reversals

	(1)	(2)	(3)
Demeaned ALJ Allowance Rate	-0.056*** (0.0084)	-0.042*** (0.0083)	-0.026*** (0.0077)
Sex (Male=1)		0.0048*** (0.00095)	0.0049*** (0.00095)
Representation		0.038*** (0.0014)	0.038*** (0.0014)
Age		0.00098*** (0.000038)	0.00098*** (0.000038)
Representation*ALJ Allowance Rate			-0.026* (0.011)
N	325915	325734	325734

Standard errors in parentheses
Models include year and office fixed effects
Standard errors clustered on ALJ
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D. Remand Avoidance

The selection phase of Social Security Appeals Council review promotes consistency and thereby limits the discretion of ALJs. The decisionmaking phase does the opposite. Together, they have a slight

consistency-enhancing effect. In the third phase, how do ALJs themselves react when they receive remands and must conduct new hearings? Recall the third implication of the theory: line-level adjudicators react to remands by avoiding the type of decision that led to remand, and they react most often in the immediate aftermath of a remand. Since remands are relatively rare for individual ALJs—most months pass without a remand for any given ALJ—and since remands arrive long after an initial decision, I am able to test whether hearing a case after remand temporarily influences a judge's decisions. I expect that:

> H_3: Hearing more cases after remand in a given time period causes an ALJ to issue fewer denials in other, unrelated cases in that time period. This effect is larger soon after post-remand hearings.

Before describing the test of this hypothesis, let me pause to address a counterargument: suppose that individual remands contain very little information about the chance of reversal in other, similar cases. If so, ALJs would correctly ignore them. Remand decisions might contain little information if they were made randomly, but the results above do not have that implication. Instead, I found a small but not trivial overall consistency-promoting effect from the selection and decisionmaking processes, taken together. This finding suggests that ALJs will avoid making similar errors in the period immediately after receiving a remand and that their allowance rates will rise as a result.

To test this third hypothesis, I use panel regressions that measure whether judges issue a higher proportion of allowances, on average, in periods in which they decide additional cases after remand. I run these regressions for four different time period lengths: one week, two weeks, one month, and two months. The regressions test for an effect within each time period for each judge. Since hearings after remand arrive on an ALJ's schedule a long (and varying) time after the initial decision, there is no reason to believe that patterns in initial decisions drive the effect. Still, to interpret these regressions as measuring the causal effect of receiving a remand, we must assume that the grant rates of ALJs who received remands would have followed the same trend, had they not received remands, as those who in fact received no remands.

The results, presented in Table 12, weakly suggest that this remand avoidance effect may exist, but that it is small. An additional decision after remand in a given week makes an ALJ a little less than 0.4% more likely to issue allowances in that week. The effect is statistically significant but very small; remands are not a large factor in ALJs' decisionmaking, even in the week in which they must make a decision in light of the remand. As expected, the effect diminishes as the time period lengthens: remands influence ALJs more soon after they must hold new hearings based on

those remands.[114] Table 13 shows the same results, this time controlling for the age and sex of the claimant and the volume of cases before the ALJ in each time period. The results are similar.

I urge extreme caution in interpreting these results, however. In the Appendix, although I find similar results in one other specification, I also find that remands appear to have an inconsistent, opposite effect as of the date the appeals are *decided* (as opposed to the date when the ALJ hearing in the remanded case is held). I would not overinterpret this contrary evidence, either; ALJs told me that they became aware of remands only when those remands came before them. But the sensitivity of the analysis to the use of different dates should reduce confidence in the results. There is also a practical reason to doubt that remands have a large effect on ALJ behavior. Although ALJs determine the outcomes of cases, they do not write the first drafts of the decisions themselves. That task is completed by decision writers.[115] Remands may therefore hold more obvious lessons for decision writers than for ALJs.[116]

If ALJs' remand avoidance does in fact have a small impact on outcomes, it is worth noticing that the effect might be larger if remands were more frequent. For example, if some ALJs were to receive ten to twenty remands in a month, these results might imply a one-to-two percentage increase in allowance rate for those ALJs. Although the effect of remands might attenuate as more are received, the opposite is also possible: more remands might have increasing effect as they have a discernible impact on a judge's workload.

In sum, there is weak, limited evidence of a small effect of remands on ALJ behavior. Soon after receiving a remand, ALJs change their behavior in a way that may make further remands less likely. If this effect is real, decreased deference from the Appeals Council—and the resulting increase in remands—could create a virtuous cycle, causing ALJs to alter their decisions and cross-ALJ disparities to decrease.

---

114.   This larger effect in smaller time periods may reflect both proximity in time to the remand and the fact that, in shorter periods, the same number of remands makes up a larger proportion of the relevant cases that the ALJ is deciding.

115.   For a discussion of the role of decision writers, see generally OFFICE OF THE INSPECTOR GEN., SOC. SECURITY ADMIN,, A-12-18-50579, DECISION-WRITING BACKLOG IN THE OFFICE OF HEARING OPERATIONS (2019).

116.   I am grateful for Harold Krent for this point.

Table 12: ALJs React to Remands

Dependent variable is 1 if claimant wins

	(1) One Week	(2) Two Weeks	(3) One Month	(4) Two Months
Number of Remands	0.0037*** (0.00056)	0.0020*** (0.00046)	0.0012*** (0.00035)	0.00061* (0.00028)
N	2576337	2576340	2576341	2576341

Standard errors in parentheses (clustered on office-ALJ)
Models include time period and office-ALJ fixed effects
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: ALJs React to Remands (with Control Variables)

Dependent variable is 1 if claimant wins

	(1) One Week	(2) Two Weeks	(3) One Month	(4) Two Months
Number of Remands	0.0037*** (0.00055)	0.0023*** (0.00044)	0.0017*** (0.00034)	0.00099*** (0.00026)
Sex (Male=1)	-0.00064 (0.0012)	-0.00064 (0.0012)	-0.00065 (0.0012)	-0.00065 (0.0012)
Age	0.000023*** (0.00000024)	0.000023*** (0.00000024)	0.000023*** (0.00000024)	0.000023*** (0.00000024)
Total Cases	0.00021* (0.00011)	0.000081 (0.000076)	-0.000046 (0.000052)	-0.000074* (0.000034)
N	2572567	2572570	2572571	2572571

Standard errors in parentheses
Models include month and office-ALJ fixed effects
Standard errors clustered on office-ALJ
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

IV. Comparisons and Implications

This Article has proposed a method of diagnosing problems in systems of review. In the Social Security context, that method offers reasons for reformers to focus their attention on decisionmaking processes and criteria at the Appeals Council. The same method can yield insights

not only for other agencies, but also for other areas in which the government must supervise decisionmakers. This Part first surveys two areas where the empirical results are already in: patent and immigration adjudication. In both areas, the method reveals the systems' relative strengths and weaknesses. This Part concludes by comparing all three areas of adjudication (patent, immigration, and disability) and discussing lessons for the disability adjudications.

Why compare Social Security adjudications with patent and immigration adjudications? The systems share not only institutional features but also a policy challenge: large disparities across adjudicators in grant rates.[117] In Figure 7, I compare the distribution of ALJs' allowance rates with the equivalent distributions for immigration judges, patent examiners, and federal district judges making decisions about whether to incarcerate a criminal defendant. The disability, immigration, and incarceration rates are centered around zero (that is, demeaned): I subtract the hearing office (or immigration court) average from each judge's allowance rate. This means that the rates reflect within-office comparisons only—except in the patent context. To create this comparison, I rely on Frakes and Wasserman's work on patent adjudications, as well as Crystal Yang's work on sentencing disparities, and on my past work on immigration adjudications.[118]

The comparison supports widespread concerns about disparities in disability, immigration, and patent adjudications: disparities in federal judges' incarceration decisions are far smaller than those in any of the three agencies. This is not to say that federal judges might not generate similarly disparate outcomes in other contexts. In particular, Krent and Morris find large disparities in Social Security outcomes both within and across federal district courts,[119] but I lack the data to place those disparities on the same scale here. In addition, disparities across patent examiners appear larger than those across immigration judges and Social Security ALJs, but the immigration and disability disparities shown below are disparities *within* hearing offices and immigration courts, whereas the patent disparities do not include that adjustment.[120]

---

117.   *See, e.g.*, Michael D. Frakes & Melissa F. Wasserman, *Patent Trial and Appeal Board's Consistency-Enhancing Function*, 104 Iowa L. Rev. 2417, 2427-29 (2019); Jaya Ramji-Nogales et al., *supra* note 2, at 325-39.

118.   Crystal S. Yang, *Have Inter-Judge Sentencing Disparities Increased in an Advisory Guidelines Regime? Evidence from* Booker, 89 N.Y.U. L. Rev. 1268, 1300-26 (2014).

119.   Krent & Morris, *supra* note 77, at 404 n.164.

120.   Frakes and Wasserman make a similar adjustment with a variety of control variables, and that adjustment does decrease the dispersion. *See* Frakes & Wasserman, *supra* note 117, at 2429 fig.2. To make this comparison, I use their reported standard deviation in examiner grant rates. *Id.* at 2427.
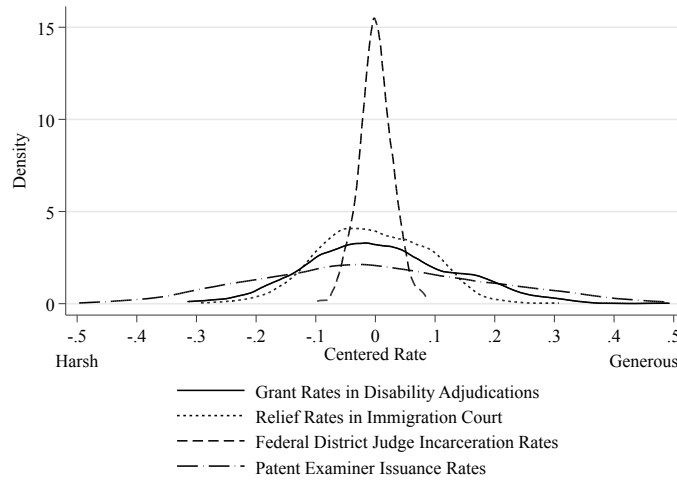
Figure 7: Disparities Compared

Figure 7 shows a remarkable commonality across these three agencies. In all three, the adjudicator's identity is often decisive for the outcome of the case. And for each agency, we now have some evidence on how and whether review processes increase consistency. The theory proposed in this Article makes meaningful comparisons possible and leads to suggestions for further research.

A. Patent Adjudication

Patent adjudications offer a natural comparison for Social Security disability adjudications, with a similar scale (over 600,000 decisions in 2017 alone[121]) and a similar agency appeals process. And Melissa Wasserman and Michael Frakes, two leading empirical scholars of patent adjudications, have recently assessed the consistency-enhancing effect of a new administrative appeals process using methods similar to those proposed here. Their important finding—that an expanded patent appeals process increases consistency across patent examiners—illustrates the usefulness of the method proposed here. This method also leads to a suggested extension of their work.

Frakes and Wasserman study the effect of the Patent Trial and Appeal Board ("PTAB") on the decisionmaking of examiners—line-level decisionmakers.[122] Frakes and Wasserman find that the patent rejections of restrictive examiners are more likely to be reversed on appeal than those

---

121. Frakes & Wasserman, *supra* note 117, at 2418.
122. *Id.* at 2429.

of lenient examiners and that the patent issuances of lenient examiners are more likely to be targeted for PTAB review than the issuances of restrictive examiners.[123] In addition, they find that examiners alter their grant rates in response to reversals by the PTAB—but also that reversals of patent rejections deter examiners more than reversals of patent issuances.[124] They conclude, inter alia, that the PTAB could decrease mistaken patent issuance by more often targeting lenient patent examiners for review.

This Article's theory suggests a possible extension of Frakes and Wasserman's work. They explain that the PTAB does not choose on its own which cases to target: in the case of patent rejections, disappointed applicants choose whether to appeal, and in the case of patent issuances, third party petitioners may begin proceedings before PTAB.[125] Yet Frakes and Wasserman choose, for brevity, to collapse their study of the selection and decisionmaking stages into a single stage.[126] This Article's theory and empirical results suggest that by distinguishing between patent appeals' selection and decisionmaking aspects, scholars could build on their work to pinpoint which parts of the patent appeals process are most and least successful. In particular, distinguishing between the first two stages of the process could either reinforce or alter their recommendation that PTAB target lenient examiners for review. If third-party petitioners are already targeting lenient examiners heavily, but PTAB rarely reverses their decisions, then the problem would lie with PTAB's decisionmaking rather than PTAB's targeting. If, on the other hand, third-party petitioners are less likely to appeal the decisions of lenient examiners than applicants are to appeal the decisions of restrictive examiners, then the problem would indeed lie in the failure to select enough cases of lenient examiners for review.[127]

In sum, the patent context illustrates the wide use of this Article's theory: breaking the appeals process down into its three steps can help agencies determine where problems are hidden.

B. Immigration Court

Immigration proceedings also offer a natural comparison case for Social Security disability adjudications. Unlike disability hearings before

---

123.   *Id.* at 2420, 2433

124.   *Id.* at 2437-43, 2445.

125.   *Id.* at 2433.

126.   *Id.*

127.   Frakes and Wasserman actually perform different analyses for appeals of rejections and for appeals of issuances. For appeals of rejections, they examine whether the decisions of more restrictive examiners were *both* appealed and reversed. For challenges to issuances, they examine only whether the issuance received review. Although both of these analyses are valuable, neither distinguishes between the selection and decisionmaking stages of review.

ALJs, immigration court decisions are not formal adjudications under the Administrative Procedure Act,[128] but both take place in court-like settings, with a judge presiding in a robe. In both systems, the claimant/immigrant has the right to a lawyer at his or her expense,[129] though in disability adjudications, the possibility of benefits payments allows lawyers to work on contingency, making them available even to indigent claimants.[130] And both systems include hierarchical, adversary systems of review. In fact, by curious coincidence, the Board of Immigration Appeals—the administrative appeals body for the immigration courts—and the Social Security Appeals Council both occupy the same nondescript office tower in Falls Church, Virginia.

In previous work on immigration appeals, I took advantage of similar near-random assignment of cases to judges to study the effect of the appeals process. I found that, when the government appeals, the process works as expected at both the first and second stages of review, but when immigrants appeal, the process fails to promote uniformity at either stage.[131] By considering each stage of review, the analysis pointed to the source of the problem: immigration judges not only decide whether to order immigrants removed but also affect whether immigrants appeal. They do so partly by controlling the length of immigration proceedings, which (in nondetained cases) routinely last years. The longer an immigrant has before his or her final hearing, the more likely he or she is to have a lawyer, and therefore to appeal an adverse decision. Harsher immigration judges order immigrants deported sooner in their proceedings, making them less likely to appeal. This dynamic prevents the appeal selection process from functioning correctly: since many meritorious cases never reach the Board of Immigration Appeals, Board review does not promote consistency at the second stage, either.

Finally, developing this three-step theory of review led me to realize that my previous work on immigration appeals was incomplete: it only assessed the first two stages of review. As a result, I returned to the data to assess the third stage of review.[132] I found that the third and last stage of

---

128. *See generally* 5 U.S.C. § 554 (2018).

129. *See* 42 U.S.C. 406(a) (2018) (allowing SSA to set procedures for claimant representation); 8 C.F.R. § 1240.10(a)(1) (2015) (requiring immigration judges to inform immigrants of this right).

130. *See* 42 U.S.C. 406(b) (2018); *SSA's Fee Authorization Processes*, SOC. SECURITY ADMIN., https://www.ssa.gov/representation/overview.htm#&sb=0&a0=0 [https://perma.cc/X374-WLLJ].

131. Hausman, *supra* note 2. Note that this analysis was limited to nondetained immigration court cases—those in which noncitizens were not imprisoned while fighting deportation.

132. See David K. Hausman, *Do Remands From the Board of Immigration Appeals Change Immigration Judges' Behavior? Addendum to* The Failure of Immigration Appeals 1, https://5a454719-8c50-4c42-b5c3-

the immigration appeals process may work partly as expected: immigration judges try to avoid remands, but only very slightly. When immigration judges decide cases following remands in which the noncitizen won, they temporarily become more likely to grant relief. This effect is very small, even though it is three times larger for immigration judges than for Social Security ALJs: an additional decision after a noncitizen-initiated remand in a given week makes an immigration judge 1.4% more likely to allow noncitizens in that week to remain in the country. Conversely, when immigration judges decide cases after a remand following a government appeal, immigration judges become temporarily more likely to order a noncitizen removed from the country.[133]

Extending the previous analysis to the third step also leads to a refinement of the previous policy implications. Given that remands do have some effect on immigration judge behavior, there is all the more reason to think that new review mechanisms—such as random selection of pro se immigration judge decisions for review by the Board of Immigration Appeals—would reduce disparities. But the limited size of the remand aversion effect also suggests that mechanisms for increasing the salience and cost of remands to immigration judges could be effective.

C. Policy Implications

Comparing disability, patent, and immigration adjudications at each step of the review process helps in evaluating policy responses. Table 14 makes these comparisons and includes likely policy solutions. (Two of the columns reproduce Table 2 *supra*, and three of the columns are specific to the patent, immigration, Social Security contexts.)

In patent adjudications, Frakes and Wasserman found that the selection and decisionmaking processes *together* have a consistency-enhancing effect, but the existing evidence does not allow us to distinguish which is more important. As a result, it is not clear whether, to improve the process, the agency should work more on selection or on the decisionmaking of the PTAB itself. Of the three systems, the PTAB has the largest remand effect: the first rejection reversal issued to patent examiners increases their grant rate by around five percentage points.[134] But the effect is much smaller (under two percentage points) for reversals of patent issuances.[135]

In immigration appeals, by contrast, the selection process is broken, and policy responses should center on that step. The selection step *does*

669c066b24e7.filesusr.com/ugd/a3c925_984cbad1ba164f2db58f28fa67cd71ad.pdf [https://perma.cc/9GDR-E9DJ].

133.   *Id.* The effect of government appeals, although similar in size to that of immigrant appeals, is not statistically distinguishable from zero.

134.   Frakes and Wasserman, *supra* note 117, at 2442.

135.   *Id.* at 2443.

work for immigrants who have lawyers, suggesting that universal government-funded representation could begin to solve the problem without disturbing the adversarial nature of the process. A quality assurance system might address these problems: with random sampling of cases for review, meritorious cases that are assigned to harsh, quick judges (and therefore are rarely appealed) would receive review.

Table 14: Immigration and Disability Review Diagnosis and Possible Policy Responses

	Patent Adjudication	Immigration Courts	Disability Adjudication	Adversarial Solution	Management Solution
Litigant Selection	Functional?	Problematic for immigrants, works for government	Functional	Government-funded representation	Quality assurance—random selection of cases for review
Decision-making	Functional?	Problematic for immigrants, works for government	Problematic	Altered standard of review	Peer review?
Remand Avoidance	Functional, especially for rejection reversals	Functional (mostly)	Likely problematic	Better functioning of first two stages	Training and feedback for ALJs

In Social Security disability adjudications, finally, the healthiest stage of the Appeals Council process is the first one, at which claimants choose whether to appeal. Many claimants appear to know whether the decisions in their cases were worth appealing. I therefore urge caution before replacing the appeals process with a quality assurance system; a major feature of such system would be random sampling of cases for review, eliminating claimants' role in selection. Perhaps such sampling would still be useful, but these findings suggest that random selection should supplement, rather than replace, the current system of claimant selection. Of course, since the government is unrepresented and cannot request Appeals Council review, these findings do *not* suggest that the Appeals Council should cease random sampling of *allowances*. The findings do, however, imply that providing the government with a lawyer in ALJ hearings would have an advantage over random sampling: that lawyer could select which allowances to appeal, freeing the Appeals Council from relying on random sampling.

The decisionmaking process, by contrast, does not promote consistency in outcomes. The Appeals Council's failure to promote uniformity, as well as its accompanying focus on errors in ALJs' rationales for their decisions, does not suggest that the Appeals Council is unsuccessful in its mission, but rather that its mission—to standardize the procedures by which the ALJ conducts a hearing and articulates a rationale for decision—should be expanded.

The Appeals Council may defer both too much and too little to the decisions of ALJs. First, on the overall assessment of the record, the level of deference to the ALJ's conclusions may be too high: with such wide variance across ALJs in allowance rates, ALJs' advantage over Administrative Appeals Judges in finding facts is open to doubt. Indeed, the need for agreement between two adjudicators on findings of fact may be one reason that peer review holds such promise. The Appeals Council could take a step in this direction with a weaker form of deference to the ALJ's conclusions. This could be especially useful in the first stage of review, when analysts screen cases; with weaker deference at that stage, more frequent review by an Administrative Appeals Judge could reveal many more cases in which ALJs' findings were not supported by substantial evidence.

Yet the Appeals Council's failure to reverse the decisions of harsher judges more often might reflect not only too much deference to harsh judges' decisions but also too little deference to generous judges' decisions. Imagine a case in which the claimant clearly is not entitled to benefits, and the ALJ therefore gives the evidence a more cursory treatment than in a case that is closer to the line. Recall, for example, one frequent reason for remand—that the ALJ failed to identify or discuss the opinion of a nontreating physician. Such a failure may be more likely where the reasons for denial were clear, and that opinion was unlikely to affect the outcome of the case. Clear cases make up a higher proportion of generous ALJs' denials, since they more often grant benefits. However, such cases may be as likely as difficult cases to lead to remands by the Appeals Council, since analysts screen for errors using a checklist and make no formal provision for harmless errors.[136]

This hypothesis fits other complaints about the process. Judge David Hatfield has suggested in congressional testimony that the standard of review is not deferential enough and that the Appeals Council defers less to findings of fact than the substantial evidence standard would suggest it

---

136.   *Cf.* Eric Schnaufer, *Proving Harm at the Appeals Council and on Judicial Review*, SCHNAUFER.COM 3 (May 5, 2012), http://www.schnaufer.com/Proving-Harm.pdf [https://perma.cc/BZ4Q-JCC2] (arguing that the Appeals Council and the courts apply an implicit harmless-error test).

should. [137] Yet the requirements to which he most objects are "articulation" requirements: requirements that the ALJ discuss and consider each piece of evidence in order to reach a legally sufficient decision.[138] In other words, the Appeals Council does not subject the ALJ's substantive assessment of the evidence to de novo scrutiny but rather applies strict articulation requirements. As Hatfield suggests, this problem might be solved with a harmless error standard, or with the understanding, well accepted in the federal courts of appeal, that an appellate court may affirm on any ground that the record supports—not only on grounds that the district court articulated.[139]

An obvious objection to this proposal is that it asks the Appeals Council to overstep the bounds of its current role: if the factfinding responsibility rests with the ALJ, the Appeals Council should not itself weigh the evidence. This fits a limited understanding of the Appeals Council's role, in which it protects the integrity of the hearing process without subjecting ALJs' conclusions to searching review. This is a role that the Appeals Council appears to fulfill successfully, particularly with its data-driven advances of the last few years.[140]

Underlying this view, however, is an assumption that the ALJ is in a significantly better position than the Appeals Council to weigh the evidence. That is at least questionable—medical records, which are available on appeal, are the central source of evidence in most cases—and must be weighed against the consistency-enhancing effect of more holistic review. Moreover, asking the Appeals Council to consider whether it can affirm on the basis of other evidence in the record is asking the Council to grant ALJs *more* deference, not less. Granted, this suggestion goes hand in hand with the view that on larger evidentiary questions, the Council should defer less. Overall, however, I argue not for more or less deference, but rather for an approach to appellate adjudication that *does* substitute the judgment of the appeals judge for the ALJ—albeit with deference.

The findings here offer additional reason to support some existing proposals for reform. For example, Michael Asimov and Jeffrey S. Wolfe have advocated the establishment of a Social Security Tribunal, modeled on a similar claims tribunal in the UK, that would increase the number of adjudicators who perform hearings, partly by allowing non-ALJs to

---

137. *Hearing on the Challenges of Achieving Fair and Consistent Disability Decisions: Hearing Before the Subcomm. on Soc. Security of the H. Comm. On Ways and Means*, 113th Cong. 4-5 (2013) (statement of David G. Hatfield, A.L.J.).

138. *Id.* at 62.

139. *See, e.g.*, Frederick v. Marquette Nat. Bank, 911 F.2d 1, 2 (7th Cir. 1990) ("[W]e can affirm on any ground fairly supported by the record.").

140. *See generally* Ray & Lubbers, *supra* note 87.

conduct those hearings.[141] Adjudication would take place in a tribunal separate from SSA,[142] and although the agency would not be represented in hearings, it would have the power to appeal allowances, ending the asymmetry in appeals. If the government's choices in appeals were similar to those of claimants, such a reform could harness the selection process to promote consistency. The proposal would also expand the role of the Appeals Council (or rather its new equivalent), allowing it to take on more of the characteristics of an appellate court. Perhaps such a larger role would free the Council to look beyond the dismissals and technical errors that form its current focus. In sum, these suggested reforms are consistent with the findings of this Article: the new tribunal would expand the selection process for appeals, which seems to function reasonably well, and would offer the new Appeals Council greater freedom in its review decisions. The current Appeals Council's inability to promote uniformity in its decisionmaking suggests that a new Appeals Council should have the freedom to approach the record in its cases holistically: it should simultaneously defer more to small evidentiary decisions and less to the ALJ's assessment of the record as a whole.

Such expanded Appeals Council review is only one possible response to the findings here. Another response might be to *reduce* the role of the Appeals Council. In its most extreme form, this could mean the elimination of the Appeals Council: claimant selection could continue to function by means of direct appeals to the federal district courts. Indeed, the filing fee and difficulty of litigating a case in federal court might heighten the sorting effect of the selection phase. And perhaps the district courts themselves would perform a holistic assessment of the record. The Appeals Council could persist as a policy body, issuing opinions on common legal issues and perhaps performing training tasks. Variations on this proposal have been prominent recently: Krent and Morris, in their 2013 ACUS report, proposed reducing the Appeals Council's role to that of an auditor, reviewing a random sample of judges' allowances and denials.[143] Perhaps limiting the number of cases reviewed by the Appeals Council in this way could improve the quality of its review. Yet Gelbach and Marcus's findings concerning inconsistency and inefficiency in federal-court litigation counsel caution.[144]

---

141. Michael Asimov & Jeffrey S. Wolfe, *Thinking Outside the APA Box: A New Social Security Tribunal*, 38 ADMIN. & REG. L. NEWS 3 (2013).

142. Asimov and Wolfe also envision these judges occasionally sitting in panels—a form of peer review.

143. KRENT & MORRIS, *supra* note 77, at 72-78.

144. GELBACH & MARCUS, *supra* note 91, at 146-67 (suggesting reforms to increase consistency and efficiency in federal-court decisionmaking). The results here are also consistent with Gelbach and Marcus's view that remands generally fail to offer systematic, issue-based feedback for ALJs. Given the small and inconsistent effect of remands on ALJs' decisionmaking

Finally, tweaks to the appeals selection mechanism could reduce the Council's workload: the Council could, like the district courts, put in place a filing fee for appeals and require more extensive briefing. At present, there is no fee for requests for review.[145] Higher barriers to appeal might discourage appeals in cases without merit, but they also might discourage appeals in meritorious cases—a problem that a generous fee waiver provision could address. Eliminating claimant-initiated appeals to the Appeals Council would in any case have a similar, but likely more severe, effect: the current filing fee for a civil action in federal district court is $400, and litigation expenses in federal court are likely to be higher than before the Appeals Council.

These proposals only scratch the surface of the many possible policy responses; my purpose here is not to advocate specific reforms, but rather to indicate the directions in which the empirical analysis could point policymakers.

Conclusion

Effective discretion-limiting systems of review share three distinctive features. First, litigants more often appeal from the decisions of adjudicators who are systematically hostile to their claims. Second, those adjudicators' decisions are more likely to be reversed. Third, adjudicators respond to remands by at least temporarily avoiding decisions that might lead to similar appeals.

All three of these patterns are observable. This Article tests for their presence in Social Security Appeals Council review. Only the first pattern is clearly present: claimants appeal more often from denials by harsher Administrative Law Judges. In the Appeals Council decisionmaking process itself, harsher judges are *less* often reversed than their generous colleagues. This anomalous pattern likely follows from the types of denials issued by generous judges, as well as the common issues that Appeals Council review targets. First, dismissal decisions (mostly for failure to appear) make up a larger proportion of generous judges' denials, and they automatically receive more extensive review. Second, Appeals Council analysts focus on errors in ALJs' articulation of reasons for their decisions rather than on the overall degree of support in the record for an ALJ's decision. In light of these results, I suggest that the Social Security Administration allow the Appeals Council to perform more searching review.

---

and the relatively small number of remands, they appear unlikely to play this role absent a policy initiative to generate feedback for ALJs.

145. *See Information About Requesting Review of an Administrative Law Judge's Hearing Decision*, Soc. Security Admin., https://www.ssa.gov/appeals/appeals_process.html [https://perma.cc/5KPC-HBBX].

The method proposed here can also shed light on other areas of adjudication. Applying the same method in the immigration courts, the glaring problem is at the selection stage; immigrants without lawyers are unable to appeal, even when the decision exiling them comes from an especially harsh immigration judge. And in the patent context, work by other scholars reveals an asymmetry between review of rejections and issuances: examiners are more likely to change their behavior in response to reversals of rejections than to reversals of issuances.

Empirical study of each stage of review should be a first step in policy change, not only for disability adjudications but for many other types of adjudications that share its structural features, from parking citation contestations to patent appeals. As data on such adjudications become more and more widely available, this empirical method for studying systems of review can and should be put to wide use.

Appendix

A. Case Assignment

Assignment of cases to ALJs is formally random, but there are many reasons to expect violations of strict random assignment. Throughout the period covered by the data, SSA has had in place a rotation policy for case assignment: the cases follow a routine rotation through ALJs, with the oldest case being assigned first. The rotation is only disturbed when some special situation requires it, and beginning in June 2011, ALJs themselves no longer have any role in that reassignment.[146] That change was accompanied by a change in the CPMS electronic database preventing ALJs from assigning themselves cases. Hearing office management teams are responsible for case assignment but can delegate the task to lead case technicians and master docket clerks.[147]

Violations of random assignment remained possible after that change. For example, a Congressional Response Report on outlier judges found that some ALJs within the same hearing office were assigned more "on the record" decisions than others—decisions without an in-person hearing.[148] Such decisions occur either when the ALJ determines that no hearing is necessary or the claimant waives the right to a hearing. One ALJ in the Congressional Response Report study reported that some ALJs in that hearing office preferred not to take cases to be decided without a hearing.[149] Such cases overwhelmingly end with a favorable decision; when

---

146. OFFICE OF THE INSPECTOR GEN., SOC. SECURITY ADMIN., A-12-11-01138, OVERSIGHT OF ADMINISTRATIVE LAW JUDGE WORKLOAD TRENDS 9-10 (2012).

147. *Id.* at 10.

148. *Id.* at 11.

149. *Id.* at 11-12.

a favorable decision is in doubt, a hearing is usually held. The nonrandom assignment of these cases is therefore likely to increase disparities spuriously.

Another review of Social Security case-assignment procedures found that over half of cases had not been assigned using the first-in-first-out procedure.[150] More than half of these exceptions occurred in on-the-record decisions, some issued by Senior Attorney Adjudicators and others by ALJs. Still other exceptions reflected dismissals, remands, and critical cases (such as cases of terminal illness).

The Congressional Response Report on outlier ALJs also documented more disturbing violations of random assignment of cases to judges. For example, it found one ALJ for whom fifty-nine percent of cases involved a single claimant representative.[151] The ALJ in question said that he heard so many cases from that representative because he perceived those requests as "priority" matters and because he did not travel to remote hearing sites.[152]

Assignment may also be nonrandom when cases are reassigned from one ALJ to another. According to the agency's guidelines, cases may be transferred after an ALJ is absent for twenty days, but the Congressional Response Report found that transfers often occur sooner than that.

In sum, there is ample to reason to suspect that cases are often not randomly assigned to ALJs. In order to test the assumption of random assignment, I perform balance tests.

B. Balance and Sample Restrictions

I examine balance across three characteristics of claimants that should be unaffected by their ALJ: their sex, their age (using a dichotomous variable for below or above the median age), and the type of claim they have brought (using a dichotomous variable that is 1 when the claimant has concurrent SSDI and SSI claims and 0 otherwise). To test balance across these characteristics, I divide up the data by office and by three-month period (i.e. four periods per year), further restrict the sample by including only judges who heard at least forty-five cases during each period (excluding the ten percent of judge-periods with the fewest observations), and exclude periods with fewer than two judges. I then perform χ^2 tests for independence across judges and exclude court-three-month periods in which at least one of the three tests (for sex, median age, and claim type)

150. OFFICE OF THE INSPECTOR GEN., SOC. SECURITY ADMIN., A-12-10-20169, OFFICE OF DISABILITY ADJUDICATION AND REVIEW'S SCHEDULING PROCEDURES FOR HEARINGS 12 (2011).

151. OFFICE OF THE INSPECTOR GEN., SOC. SECURITY ADMIN., *supra* note 144, at 12.

152. *Id.*

yields a p-value below 0.1. This results in the exclusion of a little more than two thirds of all observations (as shown in Table 15 below).

Further sample restrictions follow from peculiarities of the data. SSA provided a dataset containing records of 800,099 outcomes before the Appeals Council between FY2010 and FY2014. An electronic-folder number[153] that is present in both the ALJ and Appeals Council datasets allows cases to be matched across the two levels of adjudication. Of the 800,099 observations, 83,860, or about 10%, are missing this unique identifier; this means that they were paper-folder cases.[154] These are unusable for this Article: they cannot be matched with original ALJ dispositions. Next, I exclude observations with duplicate request and decision dates; nearly all of these observations also share the same outcome, and they therefore appear to reflect unneeded duplicates in the system.[155] Finally, I exclude all appeals after the first. This leaves 681,903 observations, one for each unique case. Merging with the ALJ data reduces the number of observations by another 69,826 cases, for a simple reason: many of the appeals that were completed between FY 2010 and FY 2014 had outcomes at the ALJ level from before FY 2010, and those cases are not present in the dataset provided by SSA.

The ALJ dataset presents similar issues. It includes 3,828,067 observations and 3,486,630 unique cases; when a case comes back to an ALJ on remand, it appears in the dataset twice. Unfortunately, 202,797 observations are missing unique identifiers that would allow them to be matched with appeals or other ALJ-level observations.[156] Of the remaining 3,625,270 observations, 50,190 have duplicate filing and disposition dates; the large majority of these share outcomes but not claim types. In order to avoid double-counting, I drop the 25,463 duplicates. Next, in order to concentrate on the outcome before the first appeal, I look at the last final disposition date before appeal.[157] This eliminates hearings on remand from the sample, as well as hearings before the final hearing that precedes appeal, lowering the total number of observations to 3,465,274. Finally, since my interest is in judges' behavior, I remove observations from judges within a single office who heard fewer than 310 cases over the five years under study, removing 5% of the observations.[158]

---

153.   This identifier is called the "efldr_num," which the documentation provided by SSA calls the "system-generated unique identifier of the Electronic Folder."

154.   Telephone Interview with Judge (Apr. 28, 2016). Such cases have become less common in the data over time.

155.   Of 3,388 cases with duplicate request and decision dates, 3,187 also have identical outcomes, so their exclusion introduces minimal bias.

156.   Unfortunately, these observations with missing identifiers are not duplicates of other proceedings that have such identifiers. Outcomes in cases with and without missing identifiers are quite similar, however, suggesting that bias from the missing data should be limited.

157.   In the very rare cases in which there are duplicate final disposition dates but different filing dates, I look at the later-filed case.

158.   None of the remaining offices had fewer than two judges, making cross-judge, within-office comparisons possible.

Table 15: Sample Restrictions

Appeals Council Data

Full dataset	800,099
Remove observations without unique IDs	716,239
Remove duplicate request and appeal dates	712,851
Keep the first appeal in each case	681,903

ALJ Data

Full dataset	3,828,067
Remove observations without unique IDs	3,625,270
Remove cases with duplicate request and appeal dates	3,599,807
Keep the final hearing before appeal, eliminating earlier hearings before and hearings on remand after appeal	3,465,274
Remove judge/office pairs with fewer than 310 hearings	3,292,134
Remove judge-court-three-month-period pairs with fewer than 49 cases	2,995,406
Remove periods with just one judge	2,979,667
Remove periods with evidence of nonrandom assignment	1,085,699
Remove cases with ALJ dispositions after 2012	663,901

C. Selection and Decisionmaking Regressions

To estimate the effect of ALJs' allowance rates on (1) whether claimants appeal, (2) whether their appeals reach an Administrative Appeals Judge, (3) whether their appeals receive review by two Administrative Appeals Judges and (4) and whether their appeals are remanded, I use a linear probability model in which only the dependent variable and the sample vary. That model has the form:

$$Y_{ijct} = \beta_0 + \beta_1 X_{ijct} + \beta_2 J_{jc} + \beta_3 R_{ijct} + \beta_4 (R_{ijct} * J_{jc}) + \beta_5 G_{jc} + \Gamma_{ct} + e_{ijct}$$

where Y_{ijct} is the dichotomous dependent variable (appeal, AAJ decision, review, or remand) for case i before ALJ j in hearing office c in year t, X_{ict} is a vector of control variables that includes sex and date of birth, J is each ALJ's demeaned allowance rate within a given hearing office, R is an indicator variable for whether a claimant has a legal representative, and G is each ALJ's demeaned dismissal rate (among denials—note that this variable is only included in the dismissals-related results above).

D. Remand Regressions

To test whether remands affect short-term ALJ behavior, I use a panel regression model of the form:

$$Y_{ijt} = \beta_0 + \beta_1 X_{ijt} + \beta_2 J_{jt} + \beta_3 D_{jt} + \Gamma_{jt} + e_{ijt}$$

where Y_{ijt} is a binary outcome for disability case i before ALJ j in time period t, where J is the number of remands for judge j in time period t, D is the total number of cases completed by judge j in time period t, X is a vector of control variables (age and sex), and Γ represents fixed effects for judges and time periods. A weakness of the analysis, especially compared to the analogous findings in immigration court, is the inability to consider remands after own motion review by the Appeals Council. Unfortunately, I am unable to identify such remands reliably in the data.[159] That means that both types of remands are included in the monthly total for judges. Remands after requests for review overwhelmingly outnumber own motion remands, so the bias created by this inclusion should be small. Nonetheless, I suggest more caution in interpreting these findings than in interpreting the findings at the first two stages of appeal.

Since the panel regressions do not rely on the assumption of random case assignment to judges, I use the full dataset, removing judges with

---

159.   *See supra* text accompanying note 74

relatively few cases and remands, and excluding cases decided without a hearing. Table 16 shows sample restrictions.

For the ALJ remand analysis, I also employ two other models. In the first of those specifications, I run the same regression as above, but I make J an indicator variable for whether an ALJ held a hearing on remand in the last week, the last two weeks, the last month, or the last two months before each hearing. For these regressions, time period t corresponds to months. Results are in Table 17 below; they show the same pattern as the regressions presented in the text above.

In the second of those specifications, I make J an indicator variable for whether the Appeals Council *issued* a remand for a case from that ALJ in last week, the last two weeks, the last month, or the last two months before each hearing. Again, time period t corresponds to months.

Comparing these two sets of results, the remand hearing date results show not only the expected direction, but also the expected pattern, with the effect decreasing as the hearing becomes more distant. By contrast, the remand-date results become stronger as the time period grows longer—a strange pattern. Overall, since ALJs are more likely to become aware of remands when they hold hearings on them, I am inclined to credit the results in the main text and Table 17 over those in Table 18, but my confidence in both is low.

Table 16: Sample Restrictions for ALJ Remand Analysis

Full dataset	3,828,067
Remove observations without unique IDs	3,625,270
Remove cases with duplicate request and appeal dates	3,599,807
Remove judge/office pairs with fewer than 310 hearings	3,426,989
Drop duplicate records after remand	3,426,935
Remove judges with fewer than 14 remands	3,256303
Remove cases decided without hearing, since the timing of the decision in those cases is more difficult to pinpoint	2,656,559
Exclude remands themselves	2,576,342

Table 17: Chance of Benefits Grant by Time Since Last Remand Hearing

Dependent variable is 1 if claimant wins

	(1) Remand Hearing in Last Week	(2) Remand Hearing in Last Two Weeks	(3) Remand Hearing in Last Month	(4) Remand Hearing in Last Two Months
At Least One Remand	0.0057*** (0.00092)	0.0041*** (0.00086)	0.0025** (0.00091)	0.00088 (0.0012)
Sex (Male=1)	-0.00065 (0.0012)	-0.00065 (0.0012)	-0.00065 (0.0012)	-0.00065 (0.0012)
Age	0.000023*** (0.00000024)	0.000023*** (0.00000024)	0.000023*** (0.00000024)	0.000023*** (0.00000024)
N	2,572,571	2,572,571	2,572,571	2,572,571

Standard errors in parentheses
Models include month and office-ALJ fixed effects
Standard errors clustered on office-ALJ
$^{*} p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$

Table 18: Chance of Benefits Grant by Time Since Last Appeals Council Remand Decision

Dependent variable is 1 if claimant wins

	(1) Remand in Last Week	(2) Remand in Last Two Weeks	(3) Remand in Last Month	(4) Remand in Last Two Months
At Least One Remand	-0.00069 (0.00091)	-0.0028** (0.00088)	-0.0036*** (0.00095)	-0.0053*** (0.0012)
Sex (Male=1)	-0.00065 (0.0012)	-0.00065 (0.0012)	-0.00065 (0.0012)	-0.00065 (0.0012)
Age	0.000023*** (0.00000024)	0.000023*** (0.00000024)	0.000023*** (0.00000024)	0.000023*** (0.00000024)
N	2,572,571	2,572,571	2,572,571	2,572,571

Standard errors in parentheses
Models include month and office-ALJ fixed effects
Standard errors clustered on office-ALJ
$^{*}\ p < 0.05, ^{**}\ p < 0.01, ^{***}\ p < 0.001$