1-1-2016

Inconsistency and Angst in District Court Resolution of Social Security Disability Appeals

Harold J. Krent
IIT Chicago-Kent College of Law, hkrent@kentlaw.iit.edu

Scott Morris
Illinois Institute of Technology

Inconsistency and Angst in District Court Resolution of Social Security Disability Appeals

HAROLD J. KRENT* AND SCOTT MORRIS**

This study of federal court decisionmaking asks whether characteristics of a jurist including age, race, gender, and work experience, can affect results in the context of the nation's most frequently litigated administrative law dispute—social security disability claims. SSDI cases by and large are similar, turning most frequently on claims of mental illness and muscular skeletal pain. Thus, there is ample room for discretion among ALJs and federal judges in determining whether an applicant is entitled to benefits.

The results are remarkable both in what they showed and did not show. First, decisionmaking patterns among district court judges and magistrates both reveal the same kind of inconsistencies that plague ALJ adjudication more generally. The results of an SSDI appeal might turn more on the hap of which judge or magistrate is slated to review the appeal than on the merits of the case.

Second, if the cases are similar, the question arises as to what explains the difference in outcomes. Again, the results are striking in that no correlation can be drawn between results and the race, gender, seniority, and job experience of the jurist. Nor can they be explained by geography or the percentage of disabled within the region.

Third, although sociological attributes did not explain much of the variation in resolution of the cases, we noted a substantial correlation between remand rates and the circuit in which the judges and magistrates sat. Remand rates from both judges and magistrates in the Tenth, Seventh, and Ninth Circuits, for instance, were almost double those from judges and magistrates in the First and Fourth Circuits. The statistics strongly suggest that the "culture' within a particular judicial circuit makes a substantial difference in such decisionmaking.

---

* Dean and Professor of Law, Chicago-Kent College of Law, Illinois Institute of Technology.

** Professor, Department of Psychology, Illinois Institute of Technology. We would like to thank Hal Bruff, Richard Posner, Gerald Ray, and Greg Sisk for comments on earlier drafts. We would also like to thank a bevy of Chicago-Kent students for help in coding cases, and particularly Peter Cassata for his stewardship of the coding project.

TABLE OF CONTENTS

INTRODUCTION 368
I. THE STRUCTURE OF SSI AND SSDI ADJUDICATION 373
II. THE INCONSISTENCY PROBLEM AT THE ALJ LEVEL 377
A. STATISTICAL ANALYSIS OF ALJ VARIANCES 379
1. Distribution of Favorable Rates 379
B. EXPLANATION OF VARIABILITY IN ALLOWANCE RATE 382
III. DISPOSITIONS AT THE FEDERAL COURT LEVEL 384
A. THE DATA 385
1. District Court Judges 385
2. Magistrate Judges 386
3. District Court Characteristics 387
4. Cases 387
5. Database Structure 388
6. Database Compilation 388
B. VARIABILITY IN REMAND RATES 389
C. SOURCES OF VARIANCE—DISTRICT COURT JUDGES AND MAGISTRATES 392
D. VARIATION IN JUDICIAL CIRCUITS 395
CONCLUSION 404

INTRODUCTION

This empirical study examines appeals in the federal district court from the Social Security Agency's denial of benefits. The results reveal striking inconsistency across the country, a finding seemingly at odds with the goal of judicial impartiality enshrined in our very social fabric. To the extent we believe in a rule of law and not a rule of man, judicial decisions should be resolved the same no matter the experience or political leanings of the jurist.[1] Our legal system furnishes the tools and precedents to generate "right" answers. Supreme Court appointees themselves have used the metaphor of the baseball umpire to describe judging: just calling strikes or balls in an objective fashion.[2] From time to time, we second-guess umpires' calls, but rarely question whether they favor one side over the other.

1. As Judge Harry Edwards stated, "members of the federal judiciary strive, most often successfully, to decide cases in accord with the law rather than with their own ideological or partisan preferences." Harry T. Edwards, *The Judicial Function and the Elusive Goal of Principled Decisionmaking*, 1991 WIS. L. REV. 837, 838; Patricia M. Wald, *Regulation at Risk: Are Courts Part of the Solution or Most of the Problem?*, 67 S. CAL. L. REV. 621, 645 (1994).

2. For one of many assessments of Chief Justice John Roberts' famous analogy, see Bruce Weber, *Umpires v. Judges*, N.Y. TIMES, July 11, 2009, at WK4.

The public, however, has never believed fully in the judicial impartiality rubric, and political charges have been leveled against judges even before *Marbury v. Madison*.[3] The realists a few generations ago famously wrote that jurists must be understood within the economic and social milieu of their upbringing and times.

Political scientists and, more recently, law professors have begun to code judicial opinions on an ideological scale to determine if judges appointed by conservative presidents tend to decide cases conservatively, and if judges appointed by more liberal presidents tend to decide cases more liberally. Not surprisingly, studies of politically charged cases have demonstrated such an intuitive link;[4] jurists cannot escape their ideology. Such empirical evidence does not shatter any goal of judicial impartiality, but does suggest that the political ideology of the jurist, at least as measured by the politics of the appointing entity,[5] can make a difference in judging.[6] There is so much room for discretion in interpreting constitutional or statutory language, as well as in assessing the importance of precedents, that many outcomes in high salience cases likely turn on the ideology of the judge.[7] Indeed, courts limit judge shopping to preserve an external appearance of fairness.[8]

Other scholars have bemoaned the "thinness" of the ideological coding, which typically focuses on the ideology of the President as the appointing entity.[9] It seems that jurists' views cannot always be captured by reference to such clunky measures, and those views might evolve over time.

Moreover, some academics suggest that culture as opposed to ideology wields a larger role in predicting judicial behavior.[10] If we understand the familial, religious, and educational background of a jurist, we may gain a

3. 4 U.S. 137 (1803). *See, e.g.*, Alex Kozinski, *The* Real *Issues of Judicial Ethics*, 32 HOFSTRA L. REV. 1095, 1096 (2004) (discussing Chief Justice Marshall's conflicts of interest in *Marbury*).

4. *See, e.g.*, C. K. ROWLAND & ROBERT A. CARP, POLITICS AND JUDGMENT IN FEDERAL DISTRICT COURTS 24–57 (1996); Frank B. Cross & Blake J. Nelson, *Strategic Institutional Effects on Supreme Court Decisionmaking*, 95 NW. U. L. REV. 1437, 1445–52 (2001); Michael D. Gilbert, *Does Law Matter? Theory and Evidence from Single-Subject Adjudication*, 40 J. LEGAL STUD. 333 (2011).

5. *See, e.g.*, Richard L. Revesz, *Environmental Regulation, Ideology, and the D.C. Circuit*, 83 VA. L. REV. 1717 (1997).

6. Ideology can also be measured in part by deviation from expected results as generated by surveys. Gilbert, *supra* note 4, at 349–50.

7. *Id.* at 350–51.

8. Federal courts, for instance, rarely inform litigants at the time of filing which judges are assigned to the case.

9. *See, e.g.*, Lee Epstein et al., *How Business Fares in the Supreme Court*, 97 MINN. L. REV. 1431, 1462–68 (2013).

10. *See, e.g.*, Dan M. Kahan & Donald Braman, *Cultural Cognition and Public Policy*, 24 YALE L. & POL'Y REV. 149, 150 (2006); Paul M. Secunda, *Cultural Cognition at Work*, 38 FLA. ST. U. L. REV. 107, 117 (2010).

better understanding of her judging.[11] Culture might not be as easily verified as the litmus test of the politics of the appointing entity, but a richer view might emerge through such efforts.

Indeed, political scientists,[12] and to a lesser extent legal academics,[13] have attempted to correlate attributes of judges with judicial decisions. For instance, academics have mapped votes on the constitutionality of the Sentencing Guidelines[14] and on Free Exercise Clause claims[15] against the race and religion of particular jurists. Some correlations were found, furthering the observation that the assignment of particular judges to hear particular claims can and has altered the outcome of cases.

This study sheds new insight on the link between judicial outcomes, judicial ideology, and demographic factors. We ask whether characteristics of a jurist other than ideology, including age, race, gender, and work experience, can affect results in the context of the nation's most frequently litigated administrative law dispute—Social Security disability claims.

Over 10,000 Social Security disability cases are resolved each year in the federal court system.[16] As a consequence, almost every district court judge and magistrate decides a significant number of such appeals from claimants each year. Social Security disability disputes, therefore, generate enough cases to permit meaningful comparisons among jurists.

What is remarkable about Social Security cases is that, by and large, they are similar.[17] All appeals arise in cases where the state Disability Determination Service ("DDS") has denied an initial claim for disability and a Social Security Administration ("SSA") Administrative Law Judge ("ALJ"), after a *de novo* hearing, has upheld the denial.[18] (The SSA

11. Judges' religion has been found highly correlated with views of the First Amendment's religious clauses. Gregory C. Sisk & Michael Heise, *Judges and Ideology: Public and Academic Debates About Statistical Measures*, 99 NW. U. L. REV. 743, 764–65 (2005).

12. *See, e.g.*, Jilda M. Aliotta, *Combining Judges' Attributes and Case Characteristics: An Alternative Approach to Explaining Supreme Court Decisionmaking*, 71 JUDICATURE 277 (1988); Frank B. Cross & Emerson H. Tiller, *Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals*, 107 YALE L.J. 2155 (1998).

13. *See, e.g.*, Michael Heise & Gregory C. Sisk, *Religion, Schools, and Judicial Decision Making: An Empirical Perspective*, 79 U. CHI. L. REV. 185 (2012).

14. Gregory C. Sisk et al., *Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning*, 73 N.Y.U. L. REV. 1377 (1998).

15. *See* Sisk & Heise, *supra* note 11.

16. *Federal Court Review Process*, SOC. SEC., https://www.socialsecurity.gov/appeals/court_process.html#&ao=1 (last visited Feb. 8, 2016). From October 2013 to September 2014, the SSA received over 18,000 new court cases. *Id.*

17. HAL DAUB ET AL., SOC. SEC. ADVISORY BD., IMPROVING THE SOCIAL SECURITY ADMINISTRATION'S HEARING PROCESS 6 (2006).

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

cannot directly appeal an ALJ's grant of benefits.[19]) Although the disabilities vary with geography and demographics,[20] twenty percent of all beneficiaries suffer from some type of mental disorder.[21] There are few clear markers whether someone with a mood disorder is capable of gainful employment.[22] Also, an even greater percentage of cases turn on muscular skeletal pain,[23] which also is notoriously difficult to ascertain on any objective basis.[24] Lab results are rarely dispositive—rather, most of the cases turn on the claimant's testimony and letters from physicians.[25] Additionally, the percentage of recipients falling into those two categories has been growing.[26] In other words, there is ample room for discretion among ALJs and federal judges in determining whether an applicant is entitled to benefits when the claim turns on the presence of debilitating mental illness or on the amount of pain experienced. Pain cannot be assessed on an objective scale, and tolerance for pain differs markedly.[27] Although the cases assigned to a particular jurist vary in terms of subject matter, over time the pool of cases that each jurist sees reflects the same type of claims.[28] Thus, this Article constitutes the first study to assess and compare the decisionmaking of virtually all district court judges resolving nearly identical challenges.[29]

---

19. SSA, however, has the authority to review an ALJ's decision before it becomes finalized through a mechanism termed "own motion review." *See infra* note 72 and accompanying text.

20. *See infra* text accompanying notes 82–84.

21. SOC. SEC. ADMIN., ANNUAL STATISTICAL REPORT ON THE SOCIAL SECURITY DISABILITY INSURANCE PROGRAM 105–10 tbl.40 (2012). That percentage evidently has risen recently. Javier Meseguer, *Outcome Variation in the Social Security Disability Insurance Program: The Role of Primary Diagnosis*, 73 SOC. SEC. BULL. 39, 51 (2013) (reporting a historical figure of seventeen percent).

22. Richard J. Pierce, Jr., *What Should We Do About Social Security Disability Appeals?*, REG., Fall 2011, at 35.

23. *Id.*

24. Meseguer, *supra* note 21, at 41; *see also* Pierce, *supra* note 22, at 35 ("[T]here are no objective diagnostic criteria that can be used to measure the degree of an applicant's anxiety, depression, or pain.").

25. *See, e.g.*, Jon C. Dubin, *Poverty, Pain, and Precedent: The Fifth Circuit's Social Security Jurisprudence*, 25 ST. MARY'S L.J. 81, 123 (1993).

26. Mark Duggan & Scott A. Imberman, *Why Are the Disability Rolls Skyrocketing?: The Contribution of Population Characteristics, Economic Conditions, and Program Generosity*, *in* HEALTH AT OLDER AGES: THE CAUSES AND CONSEQUENCES OF DECLINING DISABILITY AMONG THE ELDERLY 337, 363–67 (David M. Cutler & David A. Wise eds., 2009).

27. Pierce, *supra* note 22, at 35.

28. Appeals from district court judges in particular are few and far between, so judges rarely have to be concerned with the views of superintending appellate judges. Others have charted the impact of potential appellate review. *See, e.g.*, Revesz, *supra* note 5, at 1729. Moreover, no "panel" effect is presented in our study because we focused only on trial judges. In any event, due to the low salience of SSDI cases, we doubt that there would be much of a panel effect. *Id.* at 1732.

29. For an empirical study of asylum decisions by administrative judges and appellate courts, see Jaya Ramji-Nogales et al., *Refugee Roulette: Disparities in Asylum Adjudication*, 60 STAN. L. REV. 295 (2007).

Moreover, to our knowledge, no prior study has assessed the decisional pattern among the approximately 500 magistrate judges who assist district court judges. Magistrate judges are appointed by the federal courts they serve, so no ideology assessment is feasible given that one cannot capture ideology through the political leanings of the appointing authority, which in this case is a group of district court judges. Nonetheless, we were able to code other characteristics of most magistrates, including prior experience, age, and seniority in an effort to gauge the correlations between those factors and decisions on the appeals from denials of the Social Security disability claims. In turn, the results for the magistrates serve as a check on the results reached for the district court judges themselves.

The results are remarkable both in what they did and did not show. We reach three principal conclusions. First, decisionmaking patterns among district court judges and magistrate judges reveal the same kind of inconsistencies that plague ALJ adjudication more generally. Inspector General reports and testimony before Congress have focused on the variance in ALJ grant rates as reasons to reform the SSA adjudication system. Professor Richard Pierce has stressed that "[s]tudies of ALJ disability decisionmaking have documented massive unexplained differences in the rate at which ALJs grant or deny benefits."[30] Indeed, he has seized on the unexplained difference in grant rates among ALJs as a reason to gut the entire ALJ corps.[31] But, we determined that the differential in grant rates at the ALJ level also exists at the judicial level—some judges and magistrates affirm 100% of all SSDI and SSI appeals, while some affirm none.[32] Thus, while ALJs have been attacked repeatedly for such inconsistencies, district court decisionmaking fares little better. The results of an SSDI appeal may turn more on the hap of which district judge or magistrate judge is slated to review the appeal than on the merits of the case. Of course, simply noting the disparity in judicial decisionmaking does not further our understanding of *why* such a disparity exists.

Second, if the cases are by and large similar, the question arises as to what explains the difference in outcomes. Again, the results are striking in that the different outcomes cannot be explained by the sociological factors that others have investigated. No correlation can be drawn between results and the race, gender, seniority, and prior job experience of the jurist. Nor can the results be explained by geography or the percentage of disabled within the region of the appeal. And, there was only a modest correlation between judicial ideology as measured by the

30. Pierce, *supra* note 22, at 35.

31. *Id.* at 39; *see also infra* Figure 1.

32. *See infra* Figures 4 & 5 and text accompanying notes 82–84.

politics of the appointing judge and the judicial decision—conservative judges tended to dismiss more appeals. In other words, sociological factors evidently played far less of a role in deciding SSDI and SSI cases than one would have expected.

Third, although the sociological attributes did not explain much of the variation in resolution of the cases, we noted a substantial correlation between remand rates and the circuit in which the judges and magistrates sat. Remand rates from both judges and magistrates in the Seventh, Ninth, and Tenth Circuits, for instance, were almost double those from judges and magistrates in the First and Fourth Circuits.[33] To be sure, we cannot discount the possibility that personal beliefs—not captured in the attributes we tested for—play a significant role in district court and magistrate decisionmaking. Nonetheless, the statistics strongly suggest that the "culture" within a particular judicial circuit makes a substantial difference in such decisionmaking.

I. The Structure of SSI and SSDI Adjudication

The Supplemental Security Income ("SSI") Program prevents the aged, blind, and disabled from falling below the poverty line.[34] To be eligible for SSI, applicants must be both indigent and disabled.[35] Social Security Disability Insurance benefits ("SSDI") in contrast are not based on the financial wherewithal of the claimant, but rather are predicated on a determination that claimants are both insured and disabled.[36] The programs share the same definition of "disability": the inability "to engage in any substantial gainful activity by reason of any medically determinable physical or mental impairment which can be expected to result in death or which has lasted or can be expected to last for a continuous period of not less than [twelve] months."[37] SSA must engage in the following five-step process to determine whether an individual is disabled[38]:

33. *See infra* Figures 6 & 7.

34. 42 U.S.C. § 1381 (2015); 20 C.F.R. § 416.110 (2015). This background section is adapted from Krent & Morris, *supra* note 18.

35. Nobles v. Comm'r of Soc. Sec. Admin., 2002 WL 553735, at *1 (E.D. Tex. Apr. 10, 2002) (citing H.R. Rep. No. 92-231 (1972)); *see also* 42 U.S.C §§1382(a), 1382c(a)(3)(A)–(C) (2015).

36. 42 U.S.C. § 423 (2015). Social Security deducts an amount from a worker's paycheck if the worker earns sufficient wages. Every quarter of the year in which the worker has Federal Insurance Contributions Act ("FICA") taxes is considered a quarter of coverage. Most claimants must work forty quarters of coverage to be considered "insured." *See* 20 C.F.R. §§ 404.130, 404.140 (2015).

37. 42 U.S.C. §§ 423(d)(1)(A), 1382c(a)(3)(A) (2015).

38. 20 C.F.R. § 416.920(a)(4) (2015).

> (1) If SSA finds the individual to be engaged in substantial gainful activity, the Commissioner will find the individual *not disabled*.[39]
>
> (2) If the individual does not have a severe medically determinable physical or mental impairment or a combination of such impairments that is severe and meets the duration requirement, SSA will find the individual *not disabled*.[40]
>
> (3) If SSA determines that an impairment meets or equals one of the listings in Appendix 1 to Subpart P of Part 404 of 20 C.F.R. section 416.920, SSA will find the individual *disabled*.[41]
>
> (4) If SSA determines, based on the individual's residual functional capacity and past relevant work, that the individual can do her past work, the individual will be found *not disabled*.[42]
>
> (5) SSA will find the individual disabled if the individual cannot make an adjustment to perform other work in the economy. If the individual can make such an adjustment, SSA will find the individual *not disabled*.[43]

Social Security disability claimants must complete an application with the local Social Security office.[44] The local office determines if the applicant is indigent when applying for SSI or insured when applying for SSDI.[45] If the claimant does not qualify, a notice of denial is mailed to the claimant; if the claimant is qualified, the file is sent to a state government agency operating as a Disability Determination Service ("DDS") under contract with SSA.[46] DDS may then gather medical documents or order an examination by a contracting physician, termed a consultative exam, to make a decision regarding the claimant's disability status.[47] DDS approval rates in the states vary considerably.[48]

---

39. *Id.* § 416.920(b).

40. *Id.* § 416.920(c).

41. *Id.* § 416.920(d). The Appendix 1 "listings" include the following categories of impairments: growth, musculoskeletal system, special senses and speech, respiratory system, cardiovascular system, digestive system, genitourinary, hematological, skin, endocrine system, congenital disorders that affect multiple body systems, neurological, mental, malignant neoplastic diseases, and immune system. 20 C.F.R. pt. 404, subpt. P, App. 1 (2015).

42. 20 C.F.R. § 416.920(e) (2015).

43. *Id.* § 416.920(f).

44. James A. Maccaro, *The Treating Physician Rule and the Adjudication of Claims for Social Security Disability Benefits*, 41 Soc. Sec. Rep. Serv. 833, 833 (1993).

45. 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, 219 (1990).

46. *Id.*

47. *Id.*

48. *See* Jon C. Dubin & Robert E. Rains, Am. Const. Soc. Law & Policy, Scapegoating Social Security Disability Claimants (and the Judges Who Evaluate Them) 4 (2012); *see also* Soc. Sec. Advisory Bd., Aspects of Disability Decision Making: Data and Materials 43–44 (2012).

In all but ten states, a dissatisfied claimant may ask for "reconsideration."[49] Reconsideration involves the same procedures as an initial determination, but a different team in the same office makes the decision.[50] The ten states that have abandoned the reconsideration level of review are termed "prototype" states.[51] As a whole, states approve about forty percent of disability claims.[52]

A claimant may appeal a decision within sixty days,[53] and about one-third of those whose claims were denied do in fact appeal. An ALJ presides over the appeal, conducting an in-person hearing with the claimant for the first time in the process.[54] No deference is afforded to the DDS determination, and the ALJ may consider additional medical examinations and vocational or medical expert testimony, as well as personally question the claimant or other witnesses.[55] Although ALJs preside at the hearings, they do not typically write the opinions. Rather, they provide instructions to decision writers, who need not be attorneys, to write the opinions after they decide whether to grant or deny benefits.[56] The decision writers do not attend the hearings. SSA management, as opposed to ALJs, supervises the decision writers and direct decision writers to rotate among the ALJs.[57] Senior attorney advisors, drawn from the ranks of decision writers, can review claims and order payment of benefits on their own.[58] If the senior attorney advisors conclude that the case for payment is not clear, they transmit the files to the ALJs to schedule hearings.

In contrast to most administrative adjudications, the agency is not represented at the hearing, while attorneys represent almost eighty percent of claimants.[59] The percentage of cases in which claimants are

49. 20 C.F.R. § 404.907 (2015). Ten states do not permit reconsideration. *See* SOC. SEC. ADMIN., DI 12015.100 DISABILITY REDESIGN PROTOTYPE MODEL (2012).

50. 20 C.F.R. § 404.915 (2015).

51. *See supra* note 49.

52. For more specific figures and differences in allowance rates among the states, see SOC. SEC. ADVISORY BD., *supra* note 48, at 44 tbl.39.

53. 20 C.F.R. § 404.909 (2015); *id.* § 416.1409.

54. 20 C.F.R. § 404.929 (2015); 42 U.S.C. §§ 405, 1383 (2015).

55. *See* 20 C.F.R. §§ 404.929, 416.1465 (2015).

56. *Clearing the Backlog: Hearing on Office Performance Before the Subcomm. on Soc. Sec. of the H. Comm. on Ways & Means*, 110th Cong. 2 (2008) (statement of Hon. Ronald G. Bernoski, President, Ass'n of Admin. Law Judges).

57. *See* Jeffrey S. Wolfe, *Civil Justice Reform in Social Security Adjudications*, 64 ADMIN. L. REV. 379, 408–09 (2012).

58. *Securing the Future of the Social Security Disability Insurance Program: Hearing Before the Subcomm. on Soc. Sec. of the H. Comm. on Ways & Means*, 112th Cong. 3 (2012) [hereinafter *Hearing*] (statement of Hon. Michael J. Astrue, Comm'r, Soc. Sec. Admin.).

59. Among dispositions issued in Fiscal Years ("FY") 2009 to 2011, we found a claimant representative to be present in seventy-seven percent of hearings. This percentage was the same in each of the three years examined. We determined a correlation between representation and successful appeals. For an average ALJ, the expected allowance rate was sixty-four percent when a

represented has soared in the past thirty years.[60] The adjudication system has been beset with delays, with sizeable backlogs preventing needy individuals from receiving their due.[61] Although substantial progress has been made to minimize delays in claim adjudication, the 350-day median wait for a hearing before an ALJ is still unconscionable from the perspective of claimants who frequently are in dire need.[62] ALJs have the duty to develop the record where needed, irrespective of whether the claimant is represented by counsel.[63] ALJs resolve roughly 800,000 cases a year.[64] ALJs in the past several years have determined that disability is warranted in roughly sixty percent of the cases decided, although the percentage has declined more recently.[65] The allowance rate was sixty-one percent in Fiscal Year ("FY") 2009 and FY 2010, and then dropped to fifty-six percent in FY 2011.[66] The sharply divergent allowance rates among ALJs[67] strongly suggest that the claims are being resolved in an inconsistent manner. ALJs labor under an SSA target of adjudicating 500 to 700 cases a year.[68]

---

representative was present, but only forty-seven percent in hearings without a representative. However, it is unclear whether the positive correlation exists because of the effectiveness of the representatives or due to the fact that representatives agree to help claimants only in stronger cases.

60. The percentage of represented cases in 1971, for instance, was twenty. *See* Wolfe, *supra* note 57, at 406; *see also* SOC. SEC. ADVISORY BD., *supra* note 48, at 60.

61. Reducing the backlog has been an SSA priority. *See* SOC. SEC. ADVISORY BD., *supra* note 48, at 2–3; *Hearing*, *supra* note 58.

62. SOC. SEC. ADVISORY BD., *supra* note 47, at 66 fig.61. The delay in Veterans disability cases can be just as long. *See, e.g.*, Michael Martinez, *Obama Says Backlog Reduced in Veterans' Disability Claims*, CNN (Aug. 10, 2013, 4:09 PM), http://www.cnn.com/2013/08/10/us/obama-veterans/index.html.

63. *See, e.g.*, Hildebrand v. Barnhart, 302 F.3d 836, 838 (8th Cir. 2002).

64. SOC. SEC. ADMIN., ANNUAL PERFORMANCE PLAN FOR FISCAL YEAR 2013 13 (2012).

65. There is no objective way to measure whether ALJs or state DDS systems measure "disability" more accurately. The ALJ proceeding is more in line with due process given that testimony is allowed. For one intriguing study concluding that ALJs are more likely to get it "right," see HUGO BENÍTEZ-SILVA ET AL., NAT'L BUREAU OF ECON. RESEARCH, HOW LARGE ARE THE CLASSIFICATION ERRORS IN THE SOCIAL SECURITY DISABILITY AWARD PROCESS? (2004).

66. For decisions issued by ALJs, the allowance rates were fifty-nine percent in FY 2009, fifty-eight percent in FY 2010, and fifty-three percent in FY 2011. However, these figures do include decisions by senior attorneys, which are always favorable. Including the senior attorney decisions, the allowance rate was sixty-one percent in FY 2009 and FY 2010, and dropped to fifty-six percent in FY 2011. KRENT & MORRIS, *supra* note 18, at 22, 22 n.127.

67. Using data provided by SSA on adjudication outcomes from FY 2009 to FY 2011, we earlier found an average allowance rate of fifty-six percent and a standard deviation of fifteen percent. The yearly allowance rates ranged from four percent to ninety-eight percent, with ninety-five percent of the rates falling between twenty-six percent and eighty-five percent. *See ALJ Disposition Data*, SOC. SEC. ADMIN., ssa.gov/appeals/DataSets/03_ALJ_Disposition_Data.html (last visited Feb. 8, 2016) (listing disposition rates for all SSA ALJs).

68. SSA's Chief ALJ articulated the goal in a 2007 Memorandum. OFFICE OF THE INSPECTOR GEN., SOC. SEC. ADMIN., A-12-11-01138, OVERSIGHT OF ADMINISTRATIVE LAW JUDGE WORKLOAD TRENDS 3 n.4 (2012); *see also* DARRELL ISSA, HOUSE COMM. ON OVERSIGHT & GOV'T REFORM, MISPLACED PRIORITIES: HOW THE SOCIAL SECURITY ADMINISTRATION SACRIFICED QUALITY FOR QUANTITY IN THE DISABILITY DETERMINATION PROCESS 34–35 (2014).

If the claimant is dissatisfied with the ALJ's decision, the claimant has sixty days to appeal the adverse decision to SSA's Appeals Council,[69] although the Council is not required by law to review each appeal on the merits.[70] The Appeals Council will review a case if: (1) the ALJ committed an abuse of discretion; (2) there is an error of law; (3) the ALJ's decision was not supported by substantial evidence; or (4) there is a broad policy issue that might affect the public interest.[71] The Appeals Council may affirm, modify, reverse, or remand the ALJ's decision.[72] The Appeals Council also may exercise the authority to review cases on its own motion prior to effectuation. An Appeals Council determination represents the SSA's final decision. A claimant may then appeal the agency's decision to a federal court within sixty days.[73] When cases are appealed to federal court, the reviewing court must uphold the agency's findings so long as the findings are supported by "substantial evidence."[74] The court may affirm, modify, or remand the decision.[75] Courts generally only see appeals from denials of claims, so in that sense they do not consider a representative sampling of ALJ decisions. They reverse outright in slightly more than five percent of the cases.[76] They remand at a rate of close to fifty percent, and most of the remanded cases result in an eventual grant of benefits.[77] In FY 2011, the percentage remanded dipped to forty-two percent.[78]

II. The Inconsistency Problem at the ALJ Level

The problem of inconsistency plagues ALJ adjudication of Social Security disability claims. SSA publishes the claim allowance rate of each ALJ on its website,[79] and the dramatic variance erodes faith in the integrity of the adjudication system. Indeed, the disparities are so great that claimant representatives in 2013 vigorously and successfully lobbied

69. 20 C.F.R. § 404.968 (2015).

70. *Hearing*, *supra* note 58.

71. 20 C.F.R. §§ 404.970, 416.1470 (2015).

72. 20 C.F.R. § 416.1479 (2015).

73. *Id.* § 404.981; *id.* § 416.1481.

74. 42 U.S.C. § 405 (2010); *see also* 5 U.S.C. § 7123(c) (2011) (providing for substantial evidence review under the Administrative Procedure Act).

75. 42 U.S.C. § 405(g) (2010).

76. In FY 2011, the reversal rate dipped to three percent. *Hearing*, *supra* note 58, at 12.

77. With respect to district court remands, roughly two-thirds of the remanded claims are paid. U.S. Gov't Accountability Office, GAO-07-331, SSA Has Taken Steps to Address Conflicting Court Decisions, but Needs to Manage Data Better on the Increasing Numbers of Court Remands 16 (2007); *see also* Martinez v. Astrue, 630 F.3d 693, 695–96 (7th Cir. 2011) (summarizing available data); Paul R. Verkuil & Jeffrey S. Lubbers, *Alternative Approaches to Review of Social Security Disability Cases*, 55 Admin. L. Rev. 731 (2003).

78. Krent & Morris, *supra* note 18, at 9, 9 n.56 (referring to follow-up email from Rainbow Forbes, Appeals Officer in the Appeals Council of Sept. 28, 2012).

79. *ALJ Disposition Data*, *supra* note 67.

Congress and the Secretary of SSA to block modest reforms that would have minimized claimants' ability to forum shop so as to select an ALJ with a favorable claim allowance percentage.[80] In this Part, we describe the extent of the disparities in ALJ adjudication to set a baseline before turning our eye to adjudication among district court judges and magistrates.[81]

Inconsistencies among ALJs are alarming. For instance, in FY 2011, ten percent of ALJs had favorable rates in excess of seventy-two percent, while ten percent had favorable rates less than thirty-four percent. Twenty-five ALJs found for claimants at least ninety percent of the time, and about the same number found against claimants eighty percent of the time. Given the similarity in the pools of claims nationwide, such disparities should not arise.

Claims assigned to particular ALJs differ, even when assigned in a random manner, and some types of claims predominate more in particular geographical jurisdictions. For instance, there is a three times higher percentage of disability claims based on the nervous system in Colorado as in Vermont, and four times greater percentage based on injuries in West Virginia than in South Dakota.[82] One study concluded that variation in disability across states can be explained by the education level of the population in the state, the prevalent types of industries, and the percentage of immigrants.[83] Nonetheless, substantial and unexplained variation also exists within states.[84]

Even within the same SSA office, grant rates vary wildly. Looking at dispositions from September 29, 2012 through December 28, 2012 only at SSA Regional Offices starting with the letter A is telling: In the Akron office, Judge Thomas Ciccolini awarded 203 of 239 dispositions for a grant rate of 85% while Judge Barbara Sheehe awarded 48 of 103 for a rate of 47%; in Albuquerque, Judge Susan A. Burke awarded 134 of 144 for a rate of 93% while Judge Ben Willner awarded 56 of 112 for a rate of 50%; in Alexandria, Judge Michael Walder awarded 118 of 174 for 68% and Judge Lawrence Ragona awarded 29 of 101 for 29%. The biggest differential in the "A" offices belonged to Atlanta, with Judge Leroy Bryant awarding 72 of 81 for 89% while Judge Dale Glendening had 7 of 37 for 19%. Given that cases are to be assigned randomly within each office,[85] these differences are hard to fathom. It is no wonder the

80. Charles T. Hall, *"Secret ALJ" Policy on Its Last Legs*, SOC. SEC. NEWS (Feb. 14, 2013, 10:23 AM), http://socsecnews.blogspot.com/2013/02/secret-alj-policy-on-its-last-legs.html.

81. For the inconsistency of asylum determinations, see Ramji-Nogales et al., *supra* note 29.

82. Meseguer, *supra* note 21, at 62.

83. KATHY A. RUFFING, CTR. ON BUDGET & POLICY PRIORITIES, GEOGRAPHICAL PATTERN OF DISABILITY RECEIPT LARGELY REFLECTS ECONOMIC AND DEMOGRAPHIC FACTORS (2015).

84. *Id.*

85. OFFICE OF THE INSPECTOR GEN., *supra* note 68, at 9.

claimant community cares so deeply about knowing which ALJ will hear the claim prior to the day of the hearing.[86]

To get a better sense of the inconsistency, we compiled a data set of 1509 ALJs. Data were available from 1129 ALJs in FY 2009, 1256 ALJs in FY 2010, and 1360 ALJs in FY 2011, comprising a total of 3745 yearly data points.[87] Three separate data sets were created representing monthly, yearly, and overall average statistics for each ALJ. Outcome variables included the percentage of dispositions with favorable outcomes (either fully favorable or partially favorable)[88] and the number of dispositions conducted in a time period.

A. Statistical Analysis of ALJ Variances

1. Distribution of Favorable Rates

On average, the yearly favorable rate was 56%. The distribution covered a wide range of favorable rates (Standard Deviation ("SD")=15%), with 95% of the rates falling between 26% and 85%.[89] At the lower end of the distribution, the favorable rate was less than half of the average rate, while those at the top of the distribution had favorable rates over 50% higher than the average. The lowest and highest favorable rates (4% and 98%, respectively) very nearly spanned the full range of possible values.

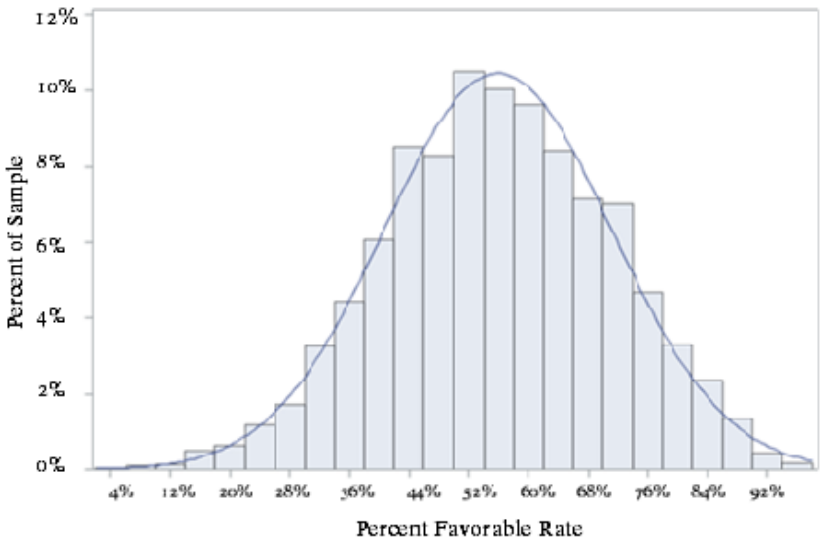
86. *See* Hall, *supra* note 80.

87. Our initial work helped predicate a report by the Administrative Conference of the United States, recommending a number of reforms in administration of the Social Security disability system. *See* Krent & Morris, *supra* note 18.

88. Outcomes were initially classified into four categories: (1) fully favorable; (2) partially favorable; (3) unfavorable; and (4) dismissed. Because the rate of partially favorable decisions was small and did not show trends that were distinct from those of fully favorable decisions, the two were combined for this analysis.

89. A similar range was found when considering only fully favorable dispositions, where 95% of the rates were between 20% and 82%.

FIGURE 1: DISTRIBUTION OF YEARLY FAVORABLE RATE
(FULLY OR PARTIALLY FAVORABLE)



Favorable rates have been steadily declining over time. The mean favorable rate dropped from 59% in FY 2009 to 57% in FY 2010 and 53% in FY 2011, while the variance in favorable rates has been relatively stable over time. This decline occurred across the entire distribution of favorable rates, although there tended to be less change at the lower end of the distribution. Table 1 demonstrates the change at three points of the distribution. Similar changes were observed at the midpoint of the distribution and the 90th percentile,[90] while change was smaller at the 10th percentile.

Corresponding to the drop in the mean favorable rate, the number of ALJs with extremely high rates has been declining. As shown in Table 2, the percentage of ALJs with favorable rates over 80% has dropped by half, from 9% in 2009 to 4% in 2011. At the other end of the distribution, the percentage of ALJs with extremely low favorable rates has been relatively stable—in essence, 5% in FY 2009 and 6% in FY 2011.

90. The 90th percentile is the favorable rate which separates the top 10% of scores from the bottom 90%. Thus, the 90th percentile represents scores at the top of the distribution, while the 10th percentile represents scores at the bottom of the distribution.

FIGURE 2: AVERAGE FAVORABLE RATE OVER TIME[91]

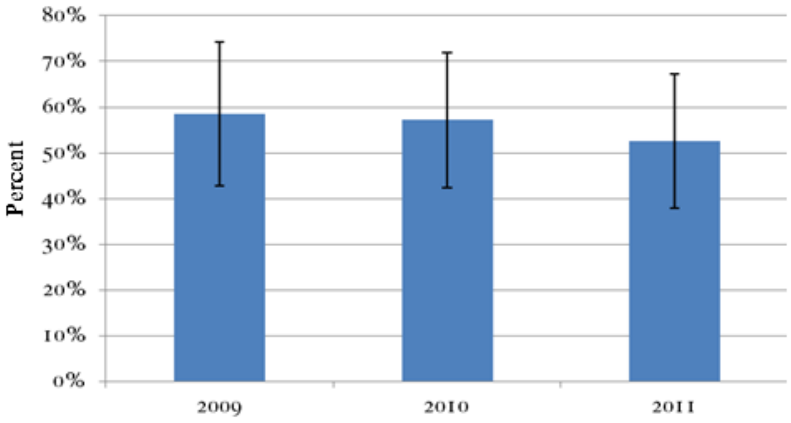


TABLE 1: CHANGE IN FAVORABLE RATES OVER TIME

Percentile	2009	2010	2011
10th	38%	38%	34%
Median	59%	57%	52%
90th	79%	76%	72%

TABLE 2: PERCENTAGE OF ALJS WITH EXTREME FAVORABLE RATES (FULLY FAVORABLE PLUS PARTIALLY FAVORABLE)

	Percent of ALJs below 20% favorable	Percent of ALJs below 30% favorable	Percent of ALJs above 70% favorable	Percent of ALJs above 80% favorable
2009	1%	5%	25%	9%
2010	1%	4%	21%	7%
2011	2%	6%	13%	4%

91. Error bars indicate one SD above and below the average.

B. Explanation of Variability in Allowance Rate

In our study of ALJ determinations, we did not come to any conclusion as to *why* there was such a great variety in allowance rate among ALJs. Although ALJs themselves have different practice backgrounds and seniority, SSA has not released such information to the public, and we could not map the allowance rate against any such variables.

We did, however, note one correlation that may be of concern to policymakers. There was a strong correlation between the percentage of claims allowed and the total number of claims decided. Anecdotally, ALJs routinely informed us that so-termed "outlier judges"—those who adjudicated the most cases—routinely granted claims against the government because written opinions supporting grants took so much less time than denials.[92] Granted claims do not get appealed, and therefore ALJs could cut corners in all cases awarding benefits.[93] Although decision writers draft the opinions, ALJs must first specify the analysis for the decisions writers to follow and then edit the drafts to ensure conformity with their views. In other words, ALJs could help reduce the backlog of those awaiting hearings by granting a higher percentage of claims. Moreover, ALJs with greater number of dispositions were perceived to curry favor with management, whether resulting in praise or preferential transfer options.[94] Of course, some ALJs might grant more claims than others due to sympathy for the claimants or because of the particular characteristics of claims assigned to them. The SSA Inspector General previously reported that ALJs with high disposition rates were more likely to decide cases in favor of claimants.[95]

To test this hypothesis, we examined the correlation between the grant rate for each ALJ and the average number of dispositions per year. We found a small but statistically significant correlation, $r=.15$, $p<.001$. Although significant, this correlation is modest, implying that the number of dispositions can account for about 2% of the variance in favorable rates. Thus, when considering the entire distribution of ALJs, the data do not support a general trend where higher distribution rates correspond to substantially higher favorable rates.

Rather than examine the entire distribution, we then focused on ALJs at the extremes of the distribution. If those ALJs with extremely high disposition rates also tend to have high favorable rates, this might

92. *See also* Ass'n of Admin. Law Judges v. Colvin, 777 F.3d 402, 404 (7th Cir. 2015) (commenting on same assertion).

93. Interestingly, Judge Kozinski has noted that the increase in workload for federal court judges pressures them into giving short shrift to small cases. Kozinski, *supra* note 3, at 1097.

94. *See also* Issa, *supra* note 68, at 35–45.

95. Office of the Inspector Gen., *supra* note 68, at 7.

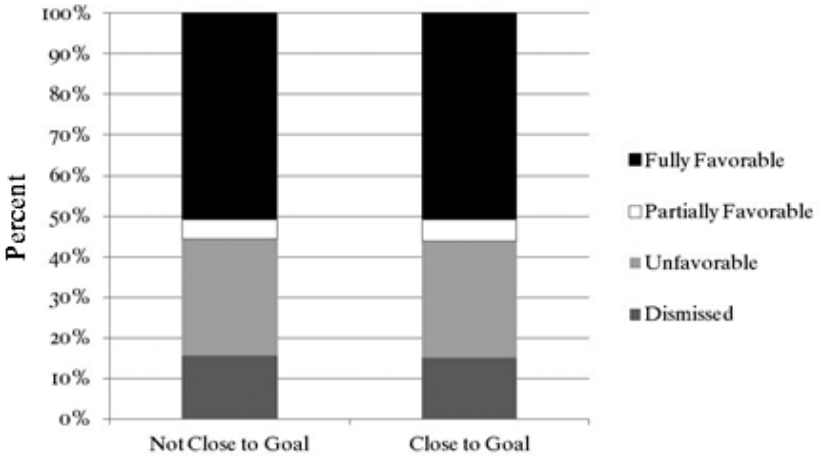
support the idea that favorable decisions are being used to boost disposition frequencies.

We defined the top 1% of yearly disposition frequencies as those with 1079 or more dispositions in a fiscal year. This consisted of twenty-two ALJs, six of whom were outliers in all three years, four were outliers in two years, and twelve were outliers in a single year. Comparison of the top 1% to other ALJs indicates that outliers issued fewer partially favorable decisions (2% vs. 5%), fewer unfavorable decisions (16% vs. 29%), and more dismissed dispositions (23% vs. 15%). The difference in fully favorable rates (58% vs. 50%) was not statistically significant. Further, there was a trend of higher fully favorable rates among those with two or three years in the top 1% (65% fully favorable) compared to those in the top 1% for no or one year (50% fully favorable).

Outliers can also be identified in terms of the decision outcomes. For this analysis, we combined fully favorable and partially favorable decisions to identify those ALJs who tended to have atypically high or low grant rates. The top 1% of favorable rates (those issuing fully or partially favorable decisions in over 89% of cases) consisted of twenty-five ALJs, fifteen with one year in the top 1%, seven with two years in the top 1%, and three with three years in the top 1%. Groups defined by number of years in the top 1% of favorable rates tended to differ in the number of dispositions issued, $F(3741)=17.32$, $p<.01$. ALJs with three years of high favorable rates tended to issue a high number of dispositions per year ((SD)=360, Mean ("M")=866), as did ALJs with two years of high favorable rates (M=724, SD=338). Those with a single year in the top 1% issued a similar number of dispositions (M=557, SD=220) to those who were never in the top 1% (M=537, SD=177).

The bottom 1% of favorable rates (those under 19%) consisted of twenty-four ALJs, fourteen with one year in the bottom 1%, five with two years in the bottom 1%, and five with three years in the bottom 1%. There was no consistent pattern in the number of dispositions issued. The number of dispositions tended to be below average for ALJs with one year in the bottom 1% (M=405, SD=158) and those with three years in the bottom 1% (M=417, SD=87). However, those with two years in the bottom 1% (M=530, SD=81) were similar to those who were never in the bottom 1% (M=541, SD=181).

FIGURE 3: DISPOSITION OUTCOME RATES FOR ALJS CLOSE TO YEARLY DISPOSITION GOAL



The correlation between number of cases decided and allowance percentage led the House of Representatives recently to invite ALJs with high allowance rates to a committee hearing. The House Oversight and Government Reform Committee issued a report in early 2014 summarizing that ALJs with large numbers of dispositions had summarily granted benefits at clips above eighty-five percent, potentially costing the taxpayers billions in lifetime benefits.[96] One of the judges testified that he had received a letter from superiors at SSA appreciative of his speedy turnover of cases.[97]

In short, the percentage of disability claims awarded by ALJs differs markedly from region to region and from ALJ to ALJ. No immediate explanations arise other than the predisposition of the ALJ, although we determined—as have others[98]—that there was a correlation between the number of dispositions and claims granted, at least for so-termed outlier judges.

III. DISPOSITIONS AT THE FEDERAL COURT LEVEL

Unlike with ALJs, no record of adjudications is available for district court judges or magistrates on the SSA website. We therefore sought to

96. Press Release, House Oversight & Gov't Reform Comm., Oversight Report Finds Disability Programs' Rubber-Stamping Judges Waste Billions (June 10, 2014) (on file with author).

97. Dave Boucher, *Report: Charleston Judge "Rubber Stamped" Disability Claims*, CHARLESTON GAZETTE-MAIL (June 10, 2014), http://www.wvgazettemail.com/article/20140610/DM01/140619970.

98. *See* Ass'n of Admin. Law Judges v. Colvin, 777 F.3d 402, 404 (7th Cir. 2015); *see also* OFFICE OF THE INSPECTOR GEN., *supra* note 68.

build a database of district court judge and magistrate decisions in disability cases to determine, first, if a similar variation in disposition rates exists and, second, whether we could then attempt to correlate any variation in rate with particular attributes of the judges and magistrates, whether seniority or political leaning.

A. The Data

To generate a database, we analyzed approximately 10,750 cases in the federal courts between 2010 and 2012 seeking review of SSA's administrative decisions denying benefits. Although the evidentiary basis in SSDI and SSI decisions differ, we decided that combining the categories for coding purposes was appropriate because the provisions' definitions of disability are the same and many cases cited to both provisions.

1. District Court Judges

The Federal Judicial Center, the "research and education agency" of the federal judiciary, maintains a publicly accessible database on all current and former Article III judges. This database contains extensive biographical, demographic, educational, and professional information on each judge and is available for download through the Federal Judicial Center's Database Export tool.[99] We narrowed this set to the following variables for each judge: circuit, district, state, appointment year (to calculate years on the bench[100]), gender, race, birth year, and political party affiliation of the appointing president.

We then added eight categories of relevant prior work experience: private practice, three concerning government work (prosecution, criminal defense, and civil), military, academia, public interest, and judicial (whether at the state or administrative level). Reasoning that a minimum of two years' experience in a field is needed to influence an individual's worldview significantly, we indicated such information for each judge.

The data used for our analysis consisted of 5581 decisions by 342 judges,[101] with each judge providing between 5 and 144 decisions (M=16.3, Median=11, SD=16.1). The sample of judges consisted of 261 men and 81 women. The ethnic composition of judges was 87% Caucasian, 10% African American, and 3% Latino. Age in 2010 ranged from forty-one to ninety-eight, with a median age of sixty-two (M=63, SD=9.9).

99. *History of the Federal Judiciary*, Fed. Judiciary Ctr., http://www.fjc.gov/history/home.nsf/page/export.html (last visited Feb. 8, 2016).

100. Years on the bench was defined as of 2010, regardless of the year in which a decision was reached.

101. We were able to match 668 judges to outcomes from 6281 cases (M=9.3 cases per judge, Median=5). A large number of judges had only one or two decisions, which provided too little information to compute a remand rate. Therefore, only judges with at least five decisions were retained for further analysis.

Judges had from zero to forty-seven years of experience, with a median of twelve years (M=13.3, SD=9.1).

2. Magistrate Judges

In contrast to information on the district judges, no comprehensive public database exists for federal magistrate judges. In part, the comparably higher rate of turnover for federal magistrates makes such a task more difficult, but even a thorough search of academic, subscription-based services yields little to no information for many magistrates currently on the bench. Our search started with compiling a list of all magistrates serving for any period of time between 2010 and 2012. Through web research using search engines likes Google News and Google Scholar, open source wiki sites like Judgepedia,[102] law schools' career services materials, and biographical pages from federal district court websites, we identified 639 magistrate judges fitting our criteria. The vast majority of these were full-time magistrates (eight-year appointments), although many were part-time magistrates (four-year appointments) or recalled magistrates (appointments of retired magistrates for up to five years, based on the district's needs).

Next, we created a database for the magistrates using the same variables as the district court judge set, except for the political party affiliation field (as magistrates are appointed by federal district court judges rather than the President). To gather these variables, we utilized media sources gleaned through web searches (outlined above), as well as subscription services like the *Almanac of the Federal Judiciary* and the *West Legal Directory*. To a much lesser extent, we used the *Judicial Yellow Book* and the *American Bench*. Additionally, we relied on data from the Just the Beginning Foundation, a nonprofit organization dedicated to diversity in the federal judiciary, which maintains statistics on African American, Asian American, Latino American, Native American, and female federal judicial officials.[103] For judges whose birth years were not ascertainable, we used years since their undergraduate graduation, where available, plus twenty-three years as a proxy for age.

This analysis of magistrate decisionmaking was based on 7632 rulings from 280 magistrates,[104] with between 5 and 259 cases per magistrate (M=27.3, Median=18, SD=40.0). The sample consisted of 195

102. *Judgepedia*, BALLOTPEDIA, http://ballotpedia.org/Judgepedia (last visited Feb. 8, 2016).

103. *Statistics on the Integration of the Federal Judiciary*, JUST BEGINNING FOUND., http://www.jtb.org/index.php?src=gendocs&ref=StatisticsontheIntegrationoftheFederalJudiciary (last visited Feb. 8, 2016). The Foundation's lists are current through October 2012.

104. We were able to match 393 magistrates to outcomes from 7885 cases; however, a large number of magistrates had only one or two decisions, which provided too little information to compute remand rates. Therefore, only magistrates with at least five decisions were retained for further analysis.

men and 85 women. Ethnic composition of the magistrates was 62% Caucasian, 5% African American, 2% Latino, 1% Asian American, and 29% unknown. The average age was 55.9 years (Median=58, SD=9.3).

3. District Court Characteristics

Several features of the district in which the magistrates sat were coded in our analysis. The population of each district was obtained from the 2011 U.S. Census population estimates. Because the distribution of population across districts was positively skewed, we used the natural log of the population as the index of district size. The proportion of the population receiving disability benefits was obtained from Social Security Administration 2011 Annual Statistical Report on the Social Security Disability Insurance Program.[105] Each state was coded as Democrat-leaning, Republican-leaning, or neutral, using data compiled from Gallup polls in 2012.[106] A state was coded as identifying with a party if the percent of the population supporting that party was at least five percent greater than the percent supporting the opposing party.

4. Cases

We experimented with a variety of searches in WestlawNext and Lexis Advance to determine which terms and database would yield the most relevant results. Lexis Advance contained many times more (unpublished) SSDI and SSI opinions than WestlawNext, so we further refined our search terms on that database. A search using the words "SSI" or "SSDI" and "ALJ" and "Astrue" in federal jurisdictions from 2010 through 2012 yielded 11,264 hits.[107] As we progressed through the compilation process (described below), we recorded only cases that received a ruling on the merits for the first time between 2010 and 2012.[108] Thus, we eliminated costs and fees proceedings, rulings on motions for reconsideration, and cases with history predating 2010. Additionally, we did not track appeals to the circuit courts.[109] This process resulted in a

105. OFFICE OF RET. & DISABILITY POLICY, SOC. SEC. ADMIN., ANNUAL STATISTICAL REPORT ON THE SOCIAL SECURITY DISABILITY INSURANCE PROGRAM, 2011 (2012), https://www.socialsecurity.gov/policy/docs/statcomps/di_asr/2011/index.html.

106. Lydia Saad, *In the U.S., Blue States Outnumber Red States, 20 to 12*, GALLUP (Jan. 30, 2013), http://www.gallup.com/poll/160175/blue-states-outnumber-red-states.aspx#2.

107. Michael J. Astrue served as the Commissioner of the Social Security Administration from 2007 to 2013, covering the whole period of this study. *Press Office*, SOC. SEC. ADMIN., https://www.ssa.gov/news/press/factsheets/astrue.htm (last visited Feb. 8, 2016).

108. Occasionally, magistrates' opinions take the form of letters addressed to counsel. These letters are generally structured like opinions and contain a closing line stating that, despite their informal nature, they should be filed as opinions. We treated these as opinions and included them in the database.

109. We chose not to track appellate decisions because there were an insufficient number to obtain a representative sampling from each appellate judge.

final set of 10,743 cases.[110] The Federal Judicial Center has reported that almost four times as many SSDI and SSI cases were resolved during that period.[111] Many likely were settled or voluntarily dismissed, and others likely were not reported in the Lexis Advance database. We have no reason to doubt that the 10,743 cases analyzed are representative of all denials appealed to the courts.

5. Database Structure

To index the cases, we built a database with the following fields: circuit, district, case name and citation, case year, magistrate judge name, report and recommendation ("R&R"), magistrate judge's ruling, name and citation for the case's subsequent history, subsequent history year, district court judge name, R&R adoption, and the district court judge's ultimate ruling. The ultimate concern of the study—determining what factors, if any, contribute to a plaintiff's success in convincing a federal court to rethink the denial of benefits by an administrative agency—led to two decisions about coding the adjudication data. First, we disregarded the distinction between sentence four and sentence six of 42 U.S.C. § 405(g), as the details of the case outcome were not as important as whether the appeal was affirmed or not.

Second, again to reflect the up or down outcome for the plaintiff, we combined reversals and remands. In cases where the magistrate or district court judge reversed or remanded the ALJ's ruling, we coded this as "Rev/Rem," reflecting a victory for the plaintiff.[112] "Affirm" is used for complete affirmations, while "Rev/Rem in Part" is used for all partial rulings. Because the number of decisions involving Rev/Rem in Part was quite small (less than two percent of rulings), these outcomes were combined with Rev/Rem rulings for the purpose of the analyses described below. For magistrate opinions, the database indicates whether the ruling is a final adjudication (under 28 U.S.C. § 636(c)) or an R&R requiring approval by a district judge. In cases where a district judge reviewed a magistrate's R&R, the database indicates whether the district judge adopted, adopted in part, or completely rejected the R&R.

6. Database Compilation

To ensure accuracy, a research team of assistants coded several matched sample sets and compared discrepancies. Once all of the team members were coding the cases similarly, we narrowed the search results

110. The subsequent history of the cases in the database is current through November 2013.

111. *Judicial Business of the United States Courts*, U.S. CTS., http://www.uscourts.gov/Statistics/JudicialBusiness/archive.aspx (last visited Feb. 8, 2016).

112. Of course, such plaintiffs might not have received benefits upon remand, but they nonetheless succeeded in forcing another look at their cases. As mentioned previously, most remands result in an award of benefits. *See* U.S. GOV'T ACCOUNTABILITY OFFICE, *supra* note 77.

by circuit, sorted them chronologically, and proceeded to record the data for each case. Cases appearing in the initial search hits often contained links to prior or subsequent case history that met our study's criteria; we included these in the database as well.

For cases that went first to a district court judge, we recorded the judge's final ruling. For cases that went first to a magistrate judge, the process was bifurcated. If the magistrate, upon consent of the parties, issued an adjudication under 28 U.S.C. § 636, we recorded the magistrate's final ruling. If the magistrate issued an R&R (the most common of the three possibilities), we coded the R&R ruling, tracked that R&R to the district judge, and coded both the final ruling and the R&R adoption or rejection by the district court judge. Occasionally, the research team encountered an R&R with no locatable district court judge adopting opinion, or a district court judge adopting opinion with no locatable R&R; for the sake of completeness, the database contains these rare cases.

B. Variability in Remand Rates

On average, district court judges issued partial or full remands or reversals in 40% of cases (Median=38%). The remand rate varied considerably across judges, ranging from 0% to 100%, with a standard deviation of 23%. Seventeen of the judges (5%) issued no remands, and three judges (1%) remanded all of their cases. The number of cases reviewed by these judges tended to be slightly lower than in the full dataset (five to twenty-one cases, with about nine cases per judge on average); nevertheless, the probability of these results occurring due to chance, if the remand rate were the same for all judges, was extremely small.

The distribution of remand rates is shown in Figure 4. Even ignoring the extreme cases (the 5% in the extreme upper and lower tails of the distribution), remand rates ranged from 0% to 85%. The remand rate at the 75th percentile of the distribution (57% remanded) was almost three times the remand rate at the 25th percentile (20% remanded).

The results for magistrate decisions by magistrates were similar. On average, magistrates issued full or partial remands or reversals in 40% of cases The median was 50%. The remand rate for magistrates ranged from 0% to 100%, with a standard deviation of 21.4%. Ignoring the most extreme 5% of the distribution, remand rates ranged from 8% to 86%. Twenty-five percent of magistrates had remand rates below 22%, while the top 25% remanded over 56% of cases. In other words, the remand rate at the 75th percentile (56% remanded) was 2.5 times higher than the 25th percentile (22% remanded). The distribution of magistrate remand rates is shown in Figure 5.

FIGURE 4: DISTRIBUTION OF DISTRICT COURT JUDGE REMAND RATES

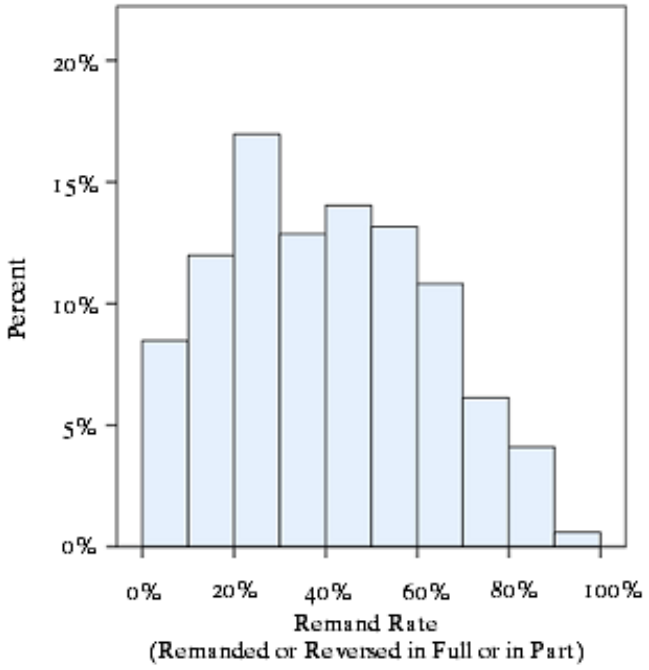
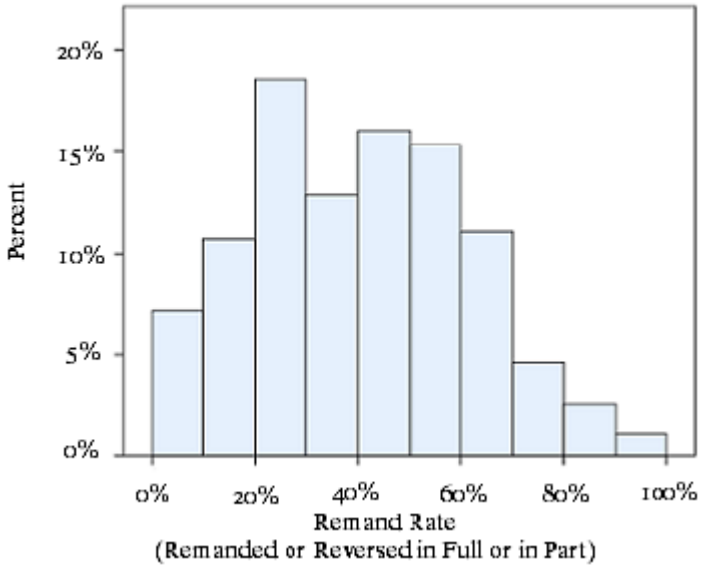


FIGURE 5: DISTRIBUTION OF MAGISTRATE JUDGE REMAND RATES



Given the concerns expressed regarding inconsistencies in ALJ adjudication outcomes, we compared the variability of outcomes in the current study to the variability of ALJ decisions previously noted. We recognize that ALJs see more cases and act as initial decisionmakers, as opposed to magistrates and district court judges who serve in an appellate capacity. Nevertheless, the comparison, however inexact, serves as a rough indication of the magnitude of the inconsistency.

Table 5 summarizes the distribution of judge and magistrate remand rates in relation to ALJ allowance rates.[113] Comparison of the means reveals a substantial difference, with noticeably lower favorable decision rates for judges and magistrates than for ALJs. As noted above, this likely reflects differences in the nature of the two types of decisions and the degree of prior screening of cases reviewed.

Of greater importance for the current study is the comparison of standard deviations, which show a substantially larger spread among judges and magistrates than among ALJs. The variability of judges' remand rates (SD=.23) was about 50% higher than the variability of ALJ allowance rates (SD=.15).

An examination of the distribution reveals that the greater variability among judges represents more density at the lower end of the distribution. If one ignores a few outliers at the upper and lower end of the distribution (excluding the top and bottom 2.5%), ALJ allowance rates ranged from 26% to 85%. For district court judges and magistrates, the upper end of the distribution was similar, with the highest affirmance rate around 85%. In contrast, the lower tail of the distribution for district court judges and magistrates reached down to 0%, substantially lower than the 26% lower bound for ALJs.

A similar picture appears if one considers the percentage of decisionmakers with extreme decision rates. A sizable number of district court judges had high remand rates (12% of district court judges remanded over 70% of cases), and was similar to the rate of high allowance rates among ALJs (13% of ALJs had allowance rates over 70%). A substantial number of district court judges also had very low remand rates. Over one-third of district court judges had remand rates below 30% and 21% had a rate below 20%. In contrast, only 6% of ALJ allowance rates were below 30% and only 2% were below 20%.

In summary, both district court judges and magistrates demonstrated substantial variability in remand rates. Moreover, the inconsistencies observed in district court decisions were substantially greater than that observed in ALJ allowance rates. We were surprised by the degree of

113. ALJ statistics were taken from an analysis of ALJ dispositions from 2009 to 2011. KRENT & MORRIS, *supra* note 18.

inconsistency reflected by the divergent remand rates in district court decisionmaking.

TABLE 3: DISTRIBUTION OF FAVORABLE DECISION RATES

Descriptive Statistics	ALJ Allowance Rate	District Judges Reversal/Remand Rate	Magistrates Reversal/Remand Rate
Mean	56%	40%	40%
SD	15%	23%	21%
Range	4%–98%	0%–100%	0%–100%
Range Excluding the Most Extreme 5%	26%–85%	0%–85%	0%–84%

TABLE 4: PERCENT WITH EXTREME DECISION RATES[114]

Below 20% Favorable	2%	21%	18%
Below 30% Favorable	6%	37%	36%
Above 70% Favorable	13%	12%	8%
Above 80% Favorable	4%	6%	3%

C. SOURCES OF VARIANCE—DISTRICT COURT JUDGES AND MAGISTRATES

Given the substantial variability in remand rates, we searched for variables that might explain these discrepancies. Explanatory variables might be identified from a variety of sources including characteristics of the judges, characteristics of their environment, and characteristics of the judicial circuit and district in which they sit. In order to identify which sources had the biggest impact on outcomes, we partitioned the observed variance into the variance associated with each source. This type of analysis does not identify specific explanatory variables, but rather helps to identify which sources are likely to yield such explanatory variables. That is, it asks whether the characteristics of the judges or the context of the cases more significantly impact the ultimate outcome.

The data on remand rates have a hierarchical structure. Cases are nested within judges, who are nested within districts, which are nested

114. Favorable decisions refer to full or partial allowance by ALJs, and reversals or remands by judges and magistrates. ALJ statistics were obtained from KRENT & MORRIS, *supra* note 18. ALJ means, SDs, and ranges were computed from all decisions issued in FYs 2009 through 2011. Percent of ALJs with extreme decision rates represent FY 2011.

within circuits. The relative contribution of different sources can be expressed in terms of the intraclass correlation coefficient ("ICC"), which indicates the proportion of the total variance that is attributable to each level of the hierarchy. The variance in remand rates was partitioned into three sources: variance across judges, variance across district courts, and variance across circuit courts.[115]

Remand rates for district court judges were fairly stable over time. The overall remand rate did not change significantly over the three years examined: 38% in 2010; 43% in 2011; and 41% in 2012. Among the 121 judges with enough data to compute yearly remand rate in at least two years, there was substantial consistency across time (ICC=.59). For the majority of judges, their highest yearly remand rate was no more than twenty percent greater than their lowest rate. Among judges who were in the top quarter of the distribution in one year, fifty-four percent of them were in the top quarter for two or more years. Less consistency was observed at the lower end of the distribution. Of those in the lower quarter of the distribution in at least one year, thirty percent were in the bottom quarter for multiple years. These results support the idea that many judges view cases through a lens that tends to favor one of the parties.

Somewhat surprisingly, demographic characteristics of the judges in general did not correlate with rulings. Remand rates were similar for male (M=.43, SD=.20) and female judges (M=.39, SD=.24), and the difference was not statistically significant. The value of the T test with 340 degrees of freedom (hereinafter denoted "t(340)"), was 1.30 with probability ("p")=.20. Remand rates were similar for Caucasian (M=.29, SD=.23), African American (M=.44, SD=.23), and Latino (M=.46, SD=.18) judges. These differences were not significant, as indicated an F-ratio with 2 and 338 degrees of freedom (hereinafter, denoted "F(2, 339)") of 1.29. Remand rates were not correlated with the judge's age, as indicated by a correlation coefficient "r" of .02, which was not significant ("ns"). We also found no evidence that decisions were influenced by the judge's experience (r's ranged from -.04 to .05). We had imagined that views toward claimants might be correlated with either prior litigation experience in general or prior work for the government in particular. Remand rates also were not correlated with years on the bench, r=-.04, ns. We had thought it possible that a judge's view of the disability system might differ dependent on his or her seniority. With respect to ideology, judges appointed by a Democratic president tended to have slightly higher remand rates (M=.42, SD=.22,

115. Yearly remand rates were computed for each year in which a judge issued at least five rulings. The resulting data consisted of 414 yearly remand rates from 232 district court judges, and 476 yearly remand rates from 160 magistrates.

Sample size ("N")=147) than judges appointed by a Republican president (M=.38, SD=.24, N=195), t(340)=1.94, p=.05, but the difference is not that significant. Ideology as reflected in the politics of the appointing president, therefore, has had only modest impact on disability decisionmaking.

Magistrate judge demographics similarly had little effect on remand rates. Remand rates were similar for male (M=.40, SD=.21) and female (M=.40, SD=.23) magistrates, t(277)=-0.11, ns. Remand rates were slightly lower for African American (M=.32, SD=.20) than Caucasian (M=.39, SD=.22) magistrates,[116] but the difference was not statistically significant, t(186)=1.32, ns. Nor were remand rates correlated with magistrate age, r(N=232)=-.04, ns.

Magistrate experience also was not related to outcomes. Neither were remand rates correlated with years on the bench, r(N=269)=-.09, ns, or specific types of prior experience (r's ranged from -.02 to .09). For magistrates as well as district court judges, individual characteristics played no obvious role in explaining the substantial difference in remand rates.[117]

Moreover, we also examined contextual characteristics of the sixty-eight districts for which sufficient data were available to compute remand rates. Remand rates were not significantly correlated with the population of the district (r=.08) for either judges or magistrates. We thought it possible that judges and magistrates in smaller towns or closer to rural areas might tend to be more sympathetic to claimants. We also wondered whether judges and magistrates in areas where a significant number of individuals received disability benefits might tend, at the margin, to be more skeptical of claimants. A slight negative correlation was found between the remand rate and the proportion of the state population receiving disability benefits (r=-.29, p=.02), again for both judges and magistrates. That is, remand rates tended to be lower in states with higher disability rates. Finally, we checked if remand rates differed in Democrat-leaning, Republican-leaning, and neutral states, t(50.7) =1.05, ns.

116. Comparisons with other ethnic groups were not possible due to an insufficient number of magistrates.

117. A separate set of analyses was conducted on cases where the magistrate issued an R&R to the district judge. Data were available from 3534 cases where the magistrate provided R&R, which represented 46% of the magistrate decisions. In the vast majority of cases (94.7%), the district judge adopted the magistrate's recommendation. Individual adoption rates were computed for 156 magistrates who provided an R&R on at least five cases. Adoption rates ranged from 50% to 100%. The adoption rate was 100% for over half of the magistrates (53%), and for 78% of magistrates, over 90% of their recommendations were adopted. The final decisions for cases where the magistrate provided an R&R (42% reversed or remanded) were similar to cases that did not involve an R&R (40% reversed or remanded). Examining only the sixty magistrates who had at least five cases of each type (R&R and non-R&R), we found that remand rates tended to be higher when making an R&R (45% remanded) than when not making an R&R (33% remanded), t(59) = 3.73, p<.001.

Perhaps the judicial decisionmakers would be swayed because of the political sentiment surrounding them, but no such correlation was found.

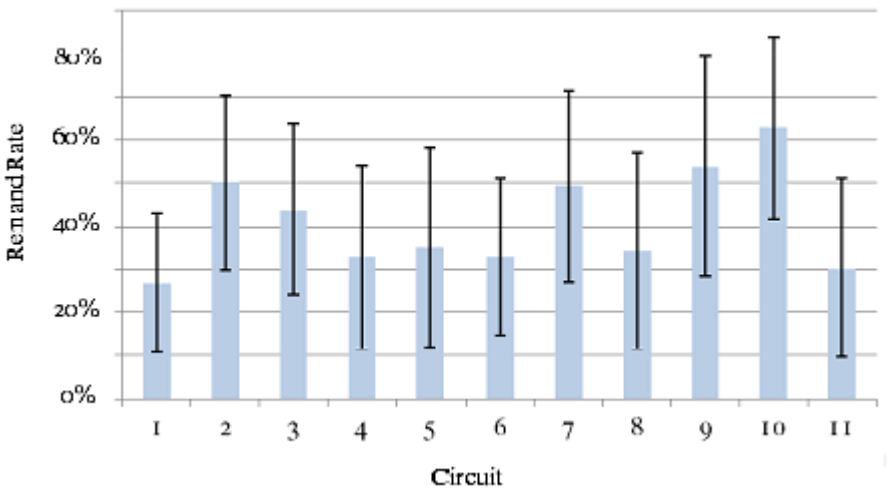
Thus, despite the substantial inconsistencies in remand rates among district court judges and magistrates, we were not able to explain the variability by dint of ideology, demographic factors, or geographic location. Whatever prompts the different decisionmaking, therefore, cannot be tied to the same sociological and political factors that previous studies have focused on to explain variability in judging.[118] Perhaps in low salience cases, which plainly include Social Security disability cases, political and sociological factors play no role.

D. Variation in Judicial Circuits

Although our study could not explain inconsistencies in decisionmaking by focusing on characteristics particular to each judge, including ideology and location, the picture changed dramatically when considering the judicial circuit in which the judge or magistrate sits. Remand rates differed significantly across circuits, $F(10,331)=6.91$, $p<.001$. Differences across circuits accounted for a sizable portion (17%) of the variance in remand rates. As shown in Figure 6, several circuits had average remand rates around 30% while others had average remand rates over 50%. Remand rates from district court judges in the Seventh, Ninth, and Tenth Circuits almost doubled the rates within the First and Fourth Circuit. Knowing the judicial circuit in which the appeal was filed was the single most significant factor we could trace in predicting whether there would be a remand.

118. *See supra* notes 10–11.

FIGURE 6: AVERAGE REMAND RATE FOR DISTRICT COURT JUDGES BY CIRCUIT[119]



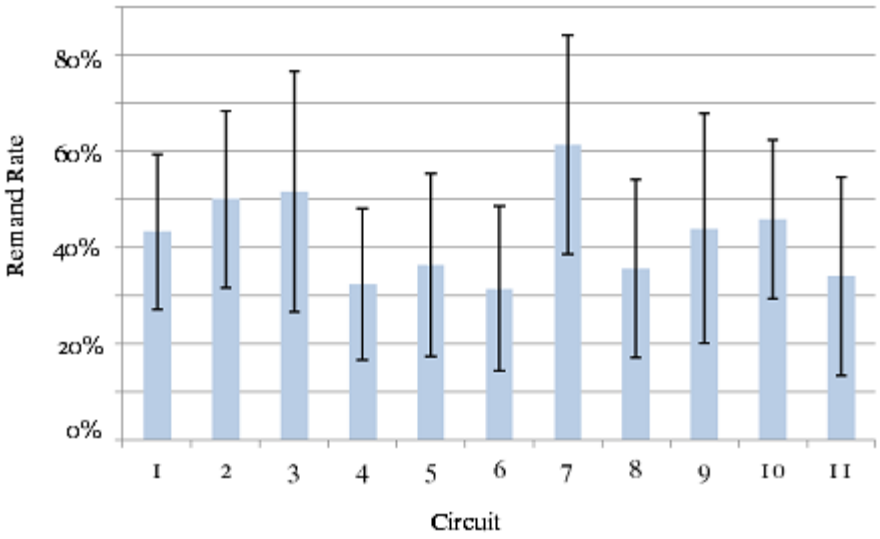
Magistrate remand rates as well differed significantly across circuits, $F(10,269)=5.2$, $p<.001$. As with district court judges, the circuit in which a magistrate sits accounted for a substantial proportion (16%) of the variability across magistrates. As shown in Figure 7, several circuits showed remand rates around 30%, while the remand rate for the Seventh Circuit was over 60%. In circuits in which district court judges had high remand rates, magistrate rates were high as well, and magistrates and district court judges matched for lower remand rates. Thus, the remand rate in the Tenth Circuit for magistrates was significantly higher than for those within the Fourth Circuit.[120]

This striking finding can be explained in several ways. First, to some extent, doctrine matters, and we suggest below one doctrinal split in particular that likely is correlated with higher remand rates. But, the doctrinal differences likely can only explain a sliver of the observed variances. We also posit that the "culture" within the circuit plays a significant part. That is, appellate judges in some circuits have set the tone that ALJ decisions might be so poorly crafted that remands are appropriate.

119. Error bars indicate one SD above and below the average.

120. A Government Accountability Office study based on remand and reversal rates for the 2005 year reveals similar variability from circuit to circuit. *See* U.S. GOV'T ACCOUNTABILITY OFFICE, *supra* note 77, at 15.

FIGURE 7: AVERAGE REMAND RATE FOR MAGISTRATES BY CIRCUIT[121]



With respect to doctrine, we were not able to identify conventional splits in the circuit courts on Social Security disability issues that could affect large groups of claimants. Nonetheless, at least one split exists that is consistent with the spread in remand rates circuit by circuit. Courts have disagreed whether an ALJ's failure to cite a listed impairment that arguably matches a claimant's condition requires a remand even if other SSA evidence suggests there is no disability. The Third Circuit in *Burnett v. Commissioner of Social Security Administration*[122] and the Tenth Circuit in *Clifton v. Chater*[123] held that such a failure automatically warrants remand. The Eighth Circuit, in contrast, has concluded that such an omission is not necessarily fatal.[124] Perhaps coincidentally, the Eighth Circuit has a considerably lower remand rate than the other two circuits.

We are convinced that the tone set by appellate courts affects the remand rate far more than doctrinal splits. There is a strong correlation between remand rate and those circuit courts that have been particularly hostile to SSA decisionmaking. That skepticism might in turn cue district court judges and magistrates to be more stringent in reviewing claimants' appeals.

121. Error bars indicate one SD above and below the average.
122. 220 F.3d 112, 119–20 (3d Cir. 2000).
123. 79 F.3d 1007, 1009 (10th Cir. 1996).
124. Senne v. Apfel, 198 F.3d 1065, 1067 (8th Cir. 1999).

Consider, for instance, the "credit-as-true" rule embraced by the Ninth Circuit. There, courts will remand for a court-ordered award of benefits when the ALJ's effort to discredit the treating source (that is, the physician) is deemed insufficient.[125] Ninth Circuit courts credit the treating source testimony and remand for an award of benefits where: (1) the ALJ has failed to provide legally sufficient reasons for rejecting such testimony; (2) there are no outstanding issues that must be resolved before a determination of disability can be made; and (3) it is clear from the record that the ALJ would be required to find the claimant disabled were the testimony credited.[126] In effect, the credit-as-true rule deprives an ALJ of a second opportunity to reweigh testimonial evidence or correct any errors in his or her initial opinion with respect to the treating source rule. As a result, claimants who are not disabled may receive benefits.

To illustrate, in *Folio v. Astrue,* the district court reversed the agency's decision and remanded for an award of benefits because the treating physician's opinion stated that Folio was "permanently disabled."[127] However, the same source opinion asserted that Folio's "medical problems are stable;" that Folio is "fully aware of medication side effects;" and that "[h]is prognosis is fair to good, depending largely on how he controls his Diabetes."[128]

Alongside the treating physician's ambiguous opinion, the record contained two examining non-treating opinions that were inconsistent with a conclusion of "permanent" disability, and so the ALJ relied on those opinions to contravene the treating source opinion. The ALJ also determined that the claimant had exaggerated his symptoms both in the hearing and to the treating physician. Upon finding that the ALJ did not adequately articulate a rejection of Folio's treating physician's opinion, the district court ordered an award of benefits relying on the treating source opinion's conclusion that Folio was "permanently disabled."[129] The court order awarding benefits deprived the ALJ of an opportunity to provide additional articulation of the reasons why he denied "controlling weight" to the treating physician's opinion. Indeed, the ALJ had also

125. Benecke v. Barnhart, 379 F.3d 587, 594 (9th Cir. 2004) ("Because the ALJ failed to provide legally sufficient reasons for rejecting Benecke's . . . treating physicians' opinion[], we credit the evidence as true.").

126. Smolen v. Chater, 80 F.3d 1273, 1292 (9th Cir. 1996); *see also* Brewes v. Comm'r of Soc. Sec. Admin., 682 F.3d 1157 1164–65 (9th Cir. 2012) (concluding further proceedings were not necessary and remanding for an award of benefits where "[t]he *complete* record show[ed] that Brewes [was] likely to miss multiple days of work per month [and] [t]he vocational expert testified that a person with Brewes' characteristics who would miss that much work was not employable.").

127. No. CV 06-2700-PHX-EHC, 2008 WL 3982972, at *4 (D. Ariz. Aug. 20, 2008).

128. *Id.*

129. *Id.* at *8.

determined that the claimant exaggerated his symptoms both in the hearing and to the treating physician.

In *Young v. Commissioner of Social Security Administration*[130] the Commissioner conceded that the ALJ improperly discredited a treating source, but sought remand for further proceedings to permit the ALJ to make additional findings with respect to the disabling effects of the claimant's mental condition. Claimant's physician opined that the claimant had marked limitation in concentration, persistence, and pace.[131] If the medical evidence were credited as true it would have pointed to an emotional disorder.[132] However, establishing an emotional disorder or a dysthymic disorder generally is not "*per se* disabling."[133] Instead of remanding for determination of the disabling effects of the claimant's impairment, however, the court remanded for a calculation of benefits.[134] Application of the credit-as-true rule, where evidence might have supported either outcome, effectively supplants the judgment of the ALJ for that of the reviewing court.[135]

The credit-as-true rule sets a tone that ALJ decisionmaking is not to be trusted. One judge opened her opinion by noting "[t]his matter is now nearly fifteen years old and has a record that is nearly 1,000 pages."[136] The opinion continued by elaborating upon the long procedural history of the case before beginning any analysis of the issues. The court ultimately found that the ALJ erred by discrediting three treating source opinions without clear and convincing reasons.[137] Given the Ninth Circuit's credit-as-true precedent, the judge terminated the proceedings and ordered an immediate payment of benefits.[138] An understandable frustration with delay might sway judges to award immediate benefits rather than remanding for further fact finding. At the end of a separate opinion, a judge concluded, "[i]n light of the extensive delay in Plaintiff's application for benefits, the Court invokes its discretion and remands this

130. No. CV 09-744-KI, 2010 WL 360776, at *1 (D. Or. Sept. 13, 2010).

131. *Id.*

132. *Id.* at *3.

133. *See* Sample v. Schweiker, 694 F.2d 639, 642 (9th Cir. 1982).

134. *Young*, 2010 WL 360776, at *4.

135. Matney v. Sullivan, 981 F.2d 1016, 1019 (9th Cir. 1992) ("The trier of fact and not the reviewing court must resolve conflicts in the evidence, and if the evidence can support either outcome, the court may not substitute its judgment for that of the ALJ.") (citing Richardson v. Perales, 402 U.S. 389, 400 (1971)); *see also* Agnew-Currie v. Astrue, 875 F. Supp. 2d 967, 974 (D. Ariz. 2012) (holding that credit-as-true rule was discretionary because it otherwise would result in payment of benefits where not due under the statute).

136. Feskens v. Astrue, 804 F. Supp. 2d 1105, 1111 (D. Or. 2011).

137. *Id.* at 1122.

138. *Id.* at 1122–23.

case for the payment of benefits . . . . Further delays at this point would be unduly burdensome on Plaintiff."[139]

The credit-as-true rule leads to reversal as opposed to remand, so it is not directly responsible for the Ninth Circuit's higher than average remand rate. Nonetheless, it signals that review of agency decisionmaking should be strict, and district court judges and magistrates evidently follow that tone in finding that ALJ decisions lack sufficient persuasive value to uphold the Secretary's denial of disability.

Consider, as well, the Seventh Circuit's professed frustration with the quality of ALJ decisionmaking. ALJs labor under a target of issuing 500 to 700 decisions a year. They give instructions to decision writers to draft their decisions and often review the decision writers' efforts in a cursory fashion. The result can be poorly reasoned decisions. The ALJs might have reached the right result, but not in a coherent judge-like manner. Indeed, SSA itself has concluded that much ALJ decisionmaking reflects more of a gestalt-type approach than the rigorous analysis that typically marks agency and district court decisions.[140]

For instance, in *Bjornson v. Astrue*[141] Judge Richard Posner castigated the agency for its crude efforts in writing decisions: "Reading the administrative law judge's opinion, we first stubbed our toe on a piece of opaque boilerplate . . . where . . . the opinion states . . . 'the claimant's statements concerning the intensity, persistence and limiting effects of these symptoms are not credible . . . .'"[142] The Seventh Circuit's opinion related that the ALJ's "'template' is a variant of one that this court (and not only this court) had criticized previously . . . . The statement by a trier of fact that the witness's testimony is 'not *entirely* credible' yields no clue as to what weight the trier of fact gave the testimony."[143] Indeed, "[s]uch boilerplate language fails to inform us in a meaningful, reviewable way of the specific evidence the ALJ considered in determining that the claimant's complaints were not credible."[144] The agency's regulations provide that "an individual's statements about the intensity and persistence of pain or other symptoms or about the effect the symptoms have on . . . her ability to work may not be disregarded

139. D'Angelo v. Astrue, No. CV 06-03055-PHX-EHC, 2007 WL 4617186, at *9 (D. Ariz. Dec. 27, 2007).

140. *Judge Gerald Ray Addresses Disability Adjudication Consistency*, EXECUTIVE DIRECTOR'S BROADCAST (ODAR Off. of App. Operations), Jan. 13, 2012, at 6 ("People frequently use heuristics, or rules of thumb, to form a mental framework to simplify consideration of issues . . . . However, use of inartfully crafted heuristics to view a problem . . . may result in framing issues too narrowly, overreliance on incomplete information, and limiting consideration of options . . . .").

141. 671 F.3d 640 (7th Cir. 2012). *Bjornson* followed a line of similar decisions. *See, e.g.*, Martinez v. Astrue, 630 F.3d 693 (7th Cir. 2011); Spiva v. Astrue, 628 F.3d 346 (7th Cir. 2010).

142. *Bjornson*, 671 F.3d at 644.

143. *Id.* at 645.

144. *Id.* (quoting Hardman v. Barnhart, 362 F.3d 676, 679 (10th Cir. 2004)).

solely because they are not substantiated by objective medical evidence."[145] An ALJ disbelieving a claimant's testimony must therefore justify that belief as opposed to presenting the conclusion. In *Bjornson,* itself, the Seventh Circuit concluded that the ALJ's decision, though manifesting skepticism about the claimant's testimony of pain, "failed to build a bridge between the medical evidence (along with Bjornson's testimony, which seems to have been fully consistent with that evidence) and the conclusion that she is able to work . . . ."[146]

The problem with much ALJ decisionmaking is that while the decisions tell us why the ALJ believes that the claimant is not disabled, they do not provide sufficient reasoning to allow meaningful review.[147] Conversely, a recent study by SSA's Office of the Inspector General examined 275 sample *grant* decisions from so-termed outlier ALJs, that is, those with grant rates in excess of eighty-five percent.[148] The SSA study concluded that 216 of the 275 decisions had "quality issues," and of those SSA's Division of Quality remanded half back to the ALJ for correction.[149] ALJ decisions granting benefits as well suffer from inadequate explanation.

Indeed, of greater salience, SSA's Office of General Counsel ("OGC") in a significant number of cases declines to represent the agency decision in court, and instead, seeks a consent order with the

---

145. Soc. Sec. Admin., Policy Interpretation Ruling Titles II and XVI: Evaluation of Symptoms in Disability Claims: Assessing the Credibility of an Individual's Statements, SSR 96-7P, 1996 WL 374186, at *1 (July 2, 1996).

146. 671 F.3d at 649.

147. Similar issues arise in application of the treating source rule, pursuant to which ALJs must give special deference to the opinions of treating physicians. The regulations provide in pertinent part:

> Generally, we give more weight to opinions from [the claimant's] treating sources, since these sources are likely to be the medical professionals most able to provide a detailed, longitudinal picture of [the claimant's] medical impairment(s) and may bring a unique perspective to the medical evidence that cannot be obtained from the objective medical findings alone or from reports of individual examinations, such as consultative examinations or brief hospitalizations. If [SSA] find[s] that a treating source's opinion on the issue(s) of the nature and severity of [the claimant's] impairment(s) is well-supported by medically acceptable clinical and laboratory diagnostic techniques and is not inconsistent with the other substantial evidence in [the claimant's] case record, [SSA] will give it controlling weight . . . .

20 C.F.R. § 416.927(c)(2) (2015). The problem is that there are often multiple treating physicians and a number of non-treating physician opinions as well, and reconciling them can be quite difficult. ALJs have struggled to assign the proper weight to each medical opinion offered, and courts have scrutinized excessively the justifications that ALJs proffer to discredit the treating physician opinion. Too little of the analysis has centered on the pivotal issue of disability itself. The complexity of the inquiry provides ample room for district court judges and magistrates to reach whichever result they prefer.

148. Office of the Inspector Gen., Soc. Sec. Admin., A-12-14-24092, Administrative Law Judges with Both High Dispositions and High Allowance Rates 2 (2014).

149. *Id.* at 4–5.

claimant to remand the case back to the agency.[150] Typical reasons include inadequate ALJ analysis of the material in the file, contradictions within the ALJ decision itself, or new evidence presented.[151] The attorneys in the OGC believe that such decisions will not survive judicial review as written.[152] The fact that OGC itself refuses to defend as written roughly fifteen percent of the cases appealed[153] strongly illustrates the problem of poorly written opinions.[154] For appellate judges, such sloppiness is an affront to the judicial process, and the opinion in *Bjornson* and many more[155] signal to district court judges and magistrates to resist boilerplate and require ALJ decisionmaking, rushed though it is, to adhere to basic tenets of judicial opinion writing.[156] This is not to suggest that judges in other jurisdictions would disagree with Judge Posner's analysis, but they might have more faith that the ALJs' gestalt-like analysis of the claims satisfies the substantial evidence test.

Although decisions in other circuits have, at times, manifested skepticism of ALJ decisionmaking in the disability context, only the Tenth and Seventh Circuits have chastised ALJs for use of boilerplate language *per se*. In contrast, some courts outside those circuits have critiqued the *Bjornson* formulation. For instance, in *Jones v. Commissioner of Social Security*,[157] the district court noted that the Sixth Circuit did not embrace *Bjornson* and counseled against a broad reading. Perhaps not surprisingly, the Sixth Circuit has a far more modest remand rate than

150. From FY 2006 to FY 2011, the numbers of voluntary requests for remand were: 2229 in FY 2011; 2419 in FY 2010; 2403 in 2008; 2496 in FY 2007; and 2763 in FY 2006. No information was available for FY 2009 due to a change in software. Those numbers represent roughly fifteen percent of all cases appealed. KRENT & MORRIS, *supra* note 18, at 9 (referring to follow-up email to discussion with Jeff Blair of the Office of General Counsel on Sept. 14, 2012).

151. *Id.*

152. *Id.*

153. *Id.*

154. The same sloppiness is even more pronounced in decisions *granting* benefits given that those decisions generally are shorter and ALJs know that such decisions will not be subject to judicial review. A recent investigation by Republican staffers on the Senate Permanent Subcommittee on Investigation focused on 300 randomly selected agency grants. It concluded that ALJs' reasoning in roughly twenty-five present of the cases "failed to properly address insufficient, contradictory or incomplete evidence." PERMANENT SUBCOMM. ON INVESTIGATIONS, S. COMM. ON HOMELAND SEC. & GOVERNMENTAL AFFAIRS, 112TH CONG., SOCIAL SECURITY DISABILITY PROGRAMS: IMPROVING THE QUALITY OF BENEFIT AWARD DECISIONS 3–4 (2012).

155. *See, e.g.*, Hardman v. Barnhart, 362 F.3d 676, 679 (10th Cir. 2004) (criticizing SSA use of boilerplate "almost without regard to whether the boilerplate paragraph has any relevance to the case").

156. Judge Posner's critique of ALJ sloppiness continues. *See, e.g.*, Ass'n of Admin. Law Judges v. Colvin, 777 F.3d 402, 404–05 (7th Cir. 2015) (analogizing ALJs to assembly line workers whose quality of work will deteriorate as the pressure to decide more cases mounts); *see also* Goins v. Colvin, 764 F.3d 677 (7th Cir. 2014) (ordering a remand just two weeks after oral argument based on numerous weaknesses in ALJ opinion).

157. No. 2:12-cv-110, 2012 U.S. Dist. LEXIS 155228, at *17–22 (S.D. Ohio Sept. 26, 2013).

the Seventh, Ninth, or Tenth.[158] Similarly, district courts in the Fifth Circuit have distanced themselves from *Bjornson.* The district court in *Kiefer v. Commissioner of Social Security Administration*,[159] for instance, explained that the "Fifth Circuit will remand an ALJ's decision where consideration of the claimant's subjective complaints of pain is wholly excluded. However, the Fifth Circuit will uphold a decision where the ALJ acknowledges the pain, cites the regulation, summarizes evidence *relevant to some of the . . . factors.*"[160] The court did not appear to be a stickler for precise analysis. And, another district court within the Fifth Circuit stressed in *Morris v. Colvin*[161] that the "Court is mindful of the boilerplate language used by the ALJ, but the deference this Court must give to the Commissioner requires that it take the ALJ's findings, as long as supported by substantial evidence, at face value." The remand rate for courts in the Fifth Circuit, as in the Sixth, is much lower than in the Seventh Circuit.

In contrast, district courts in the Second Circuit have cited *Bjornson* approvingly,[162] and the Second Circuit's remand rate is high as well. *Bjornson* does not purport to alter the doctrine by which the agency decisions are reviewed, but its tone—and the tone of similar decisions—evidently made a difference. In short, the circuit in which the district court judge and magistrate sits is the factor that best explains the observed statistical inconsistency in remand rates. That variability to a modest degree might stem from doctrinal differences from circuit to circuit—that is, doctrine matters. But, we conclude that the remand rate differences derive substantially from appellate court criticism of the structure of ALJ determinations and insistence that SSA ALJ decisions be analyzed just like any other agency decision. The tone is set at the top, and the statistical analysis lends credence to the impact that such statements have on subordinate decisionmakers within the Article III judicial system.

We do not make the claim that a circuit court's culture is a perfect explanatory variable. Some district court judges and magistrates value their independence, recognize that SSDI and SSI decisions are not likely to be appealed, and find for or against claimants as they deem fit. Indeed,

158. *See also* Bailey v. Comm'r of Soc. Sec., No. 5:13-cv-01672, 2014 U.S. Dist. LEXIS 90935 (N.D. Ohio May 28, 2014). Even the Tenth Circuit has distanced itself to some extent from *Bjornson*. *See* Romero v. Colvin, 563 F. Appx. 618 (10th Cir. 2014).

159. No. 4:13-CV-330, 2014 U.S. Dist. LEXIS 80429 (E.D. Tex. June 13, 2014).

160. *Id.* at *12 (emphasis added).

161. No. 6:12-2692-JFA-KRM, 2013 U.S. Dist. LEXIS 186067, at *38 (D.S.C. Dec. 2, 2013) (quoting Craven v. Astrue, 9:11–cv–01674–RBH, 2013 WL 1282022 at *5 (D.S.C. Mar. 26, 2013).

162. *See, e.g.*, Lumpkin v. Colvin, No. 3:2cv1817(DJS), 2014 U.S. Dist. LEXIS 112490 (D. Conn. Aug. 13, 2014); Ferguson v. Colvin, No. 1:12-cv-0333(MAT), 2014 U.S. Dist. LEXIS 109996 (W.D.N.Y. Aug. 8, 2014).

we found significant inconsistency from district to district even within particular circuits.

Of the total variance in remand rates, nearly half (43%) occurred between districts. District remand rates ranged from as low as 12% in the Western District of Missouri to as high as 77% in the Western District of Washington.[163] About 10% of districts had remand rates below 20% and 10% had rates over 60%. About one-third of this between district variance could be attributed to differences across circuits.[164] That is, there was substantially more consistency among districts within a circuit than between districts from different circuits.

The general pattern is illustrated by a comparison of the Sixth and Ninth Circuits. Although remand rates were generally higher in the Ninth Circuit than in the Sixth Circuit, there were also sizable differences within these circuits. In the Sixth Circuit, remand rates ranged from 18% in the Middle District of Tennessee and 19% in the Western District of Michigan to 46% in Southern District of Ohio. In the Ninth Circuit the remand rate was as low as 21% in the Southern District of California and as high as 77% in the Western District of Washington. Despite these discrepancies, remand rates tended to be somewhat consistent across districts within many of the circuits. For example, the Third Circuit ranged from 36% to 53%, and the Tenth Circuit ranged from 60% to 69%. A similar pattern of circuit and district level effects was observed for magistrates. Remand rates for magistrates correlated strongly with those of district court judges in the same district, $r=.69$, $p<.001$. Thus, taken as a whole, the different remand rates from circuit to circuit strongly suggest that many district court judges and magistrates follow cues from the appellate courts.

CONCLUSION

In essence, our study reveals that law broadly understood might affect judging more than judicial ideology and demographic factors, at least in low salience cases such as SSDI and SSI adjudication. The wide disparity in remand rates among hundreds of district court judges and magistrates can be explained more by the tone set by appellate judges in

163. Some of the observed differences could be due to chance, owing to the small number of decisions available for some districts. For about a quarter of the districts, rates were computed from fifteen or fewer decisions. Still, even among districts with larger sample sizes, remand rates were as low as 12% in the Western District of Missouri with fifty-one decisions from four judges, or 13% in the Eastern District of Virginia from nine judges in 113 decisions. The remand rate of 77% for the Western District of Washington was based on 213 decisions from ten judges.

164. The variance in remand rates was partitioned into differences across circuits (14%), differences across districts within a circuit (29%), and differences among judges within a district (57%). The total variance across districts, comprised of both between-circuit and within-circuit components, accounted for 14% + 29% = 43% of the variance. Focusing on the variance among districts, 14% / 43% = 33% could be attributed to differences between circuits.

each circuit than by the political preferences or vocational experience of the jurists. Perhaps, at the end of day, the results should not be surprising, because appellate leadership, if not review, can make a difference.

Congress and SSA must assess how to reshape ALJ decisionmaking so that ALJs can satisfy the demands on judicial technique that Article III judges expect from agency administrative decisions. Perhaps ALJs need more time to write adequate decisions; perhaps ALJs need to coordinate more effectively with decision writers. Or, perhaps judicial review of facts is not appropriate in the disability context.[165] But, for the near future, the remand rate overall will remain high, and particularly so in the Tenth, Ninth, and Seventh Circuits, as a result of the direction that those circuit court judges have given to district court judges and magistrates in their circuits.

165. For years after its founding, Congress provided no judicial review for denial of veterans' benefits claims, and now provides only for limited review of factual issues. *See, e.g.*, James D. Ridgway, *The Splendid Isolation Revisited: Lessons from the History of Veterans' Benefits Before Judicial Review*, 3 VETERANS L. REV. 135 (2011).

***