

Reforming Administrative Legal Review in Disability Insurance

By Mark J. Warshawsky and Ross A. Marchand

Reprinted from *Tax Notes*, October 5, 2015, p. 139

Reforming Administrative Legal Review in Disability Insurance

By Mark J. Warshawsky and
Ross A. Marchand

Mark J. Warshawsky is a senior research fellow at the Mercatus Center at George Mason University and a visiting fellow at the MIT Center for Finance and Policy. Ross A. Marchand is an MA fellow at the Mercatus Center.

In this article, Warshawsky and Marchand update and deepen Warshawsky's prior examination of the administrative law judge system for the Social Security disability programs, reviewing recent comprehensive reports from academic and government sources and analyzing judicial decisions over a longer period. They estimate that failings by the ALJ system will lead, on net, to payments of more than \$72 billion to disability claimants over their lifetimes through likely unwarranted benefit awards given from 2005 through 2014. They argue that although recent administrative changes have curbed some of the worst excesses, permanent reform is needed to lock in the changes and to enhance their positive effects.

The Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) programs have many financial and structural problems. In a recent editorial, we detailed problems with the system of benefit claim appeals, presided over by administrative law judges, and we proposed several solutions.¹ Because the editorial, based on an earlier study by Mark J. Warshawsky,² was published in *The Wall Street Journal*, it garnered attention from the public, the media, and others. We therefore thought it worthwhile to continue, update, and deepen our examination of the topic, adding reviews of some recent comprehensive reports — econometric, analytical, and based on case studies — from academic and government sources.

¹Mark J. Warshawsky and Ross A. Marchand, "Disability Claim Denied? Find the Right Judge," *The Wall Street Journal*, Mar. 8, 2015.

²Warshawsky, "Administrative Problems With Social Security Disability Programs: Some Solutions," *Bloomberg Pension & Benefits Daily*, Apr. 2, 2012.

Also, we undertook our own empirical analysis of judicial decisions over a longer period that includes more recent data.

It is difficult to come to a definitive conclusion based on any one study, report, or analysis, but when varied sources reach similar conclusions through different methods and approaches, they build a compelling case. Here we find such a case, in which serious failings by the ALJ system — at a time of large claim backlogs — have led, on net, to large losses for taxpayers. We estimate that more than \$72 billion will be paid to claimants over their lifetimes through likely unwarranted disability benefit awards given by ALJs over the 10-year period from 2005 through 2014. Recent public scrutiny and administrative changes have curbed some of the worst excesses, but serious problems with the benefit claim appeals process remain. Moreover, the problems' original magnitude could easily return when public and management attention moves elsewhere or when political pressure builds again to reduce the claim backlog. Therefore, serious permanent administrative reform is needed to lock in the recent changes and to build on them to enhance their positive effects. The appropriate time to do so is now, as the disability insurance program is expected to be insolvent in less than two years.

Background

The Social Security Administration (SSA) manages two large federal disability programs: SSDI and SSI. As of 2014 about 19.4 million individuals receive about \$200 billion annually in benefits through these two programs. Individuals enrolled in SSDI for two years are also automatically enrolled in Medicare, which costs taxpayers about \$80 billion a year. SSI recipients are eligible for Medicaid immediately.

When an individual applies for disability benefits, the case is initially decided by state employee examiners in Disability Determination Services (DDS). There is also an automatic pre-effectuation review — an internal review of a decision before it is finalized and communicated to the claimant — of 50 percent of DDS allowance decisions. In 40 states and in most of California, an applicant who is denied benefits may appeal to a different reviewer in the same office. The SSA, which oversees each state's DDS, claims that there are few errors in the original adjudication of these decisions. If the second reviewer denies benefits, the applicant may

appeal to an ALJ.³ If the ALJ then awards disability benefits, the decision is final because the government cannot appeal it. But if the ALJ denies benefits, the individual may appeal to the SSA Appeals Council and, in a civil case, to the several levels of the federal courts. In total, there are at least five levels of review for a disability benefits applicant. Any error by a state adjudicator or an ALJ against an applicant is fixable, whereas an error by either against a taxpayer is not.

According to economists David Autor and Mark Duggan, the average lifetime disability benefit, including cash and the value of healthcare, is about \$300,000.⁴ During the 1980s, the disability programs experienced loosened eligibility requirements and increased benefits. Autor and Duggan find that the result of those changes was a rise in the proportion of people who reported work limitations and a concomitant decrease in the employment of men with disabilities.⁵ This trend has continued since their study. General survey indicators of disability rates in the working-age population are flat or declining.⁶ SSA economists Till von Wachter, Jae Song, and Joyce Manchester find that rejected applicants who are young and who have low-mortality-reported disabilities show a relatively strong attachment to the labor force.⁷

According to George Washington University law professor Richard Pierce, reversals of SSA denials by ALJs have increased significantly from 1970 to

the present, primarily because ALJs have been granting benefits to applicants with less severe mental illnesses and pain than ALJs considered sufficient to qualify for disability benefits in the past.⁸ In recent years, about 60 to 70 percent of ALJ rulings in disability benefit appeals have been in the claimant's favor. ALJs have a greater incentive to award benefits than to deny them because denials are subject to judicial appeal and must be fully documented, which takes longer than the decisions and drafting of approvals.⁹

The approval rate has fallen recently to about 54 percent. This decline is perhaps a reaction by the agency and the ALJs to the negative publicity arising from investigative reports in *The Wall Street Journal* about ALJs who held for claimants in virtually all their cases.¹⁰ The reports highlight Judge David Daugherty from the Office of Disability Adjudication and Review (ODAR) in Huntington, West Virginia. He heard thousands of cases over the years and uniformly granted favorable judgments. Many of the claimants in these cases were represented by a single law firm. There have also been strong congressional inquiries, a series of hearings, and administrative reforms by former Social Security Commissioner Michael Astrue. Another likely factor for the recent decline may be a lagged business-cycle effect involving the increased number of claims made during the Great Recession that were motivated mainly by unemployment rather than by personal disability and were therefore clearly subject to denial. Astrue's administrative reforms will be described in more detail below.

Another piece of relevant historical background is the backlog of disability claim hearings that have been pending for more than 270 days. As Figure 1 shows, the backlog reached more than 400,000 cases in 2008. There was strong political pressure on the SSA to reduce it, which the agency did over the next few years, but then the backlog began to build again in 2012. In 2014 it was again at more than 400,000 cases, and by the second quarter of fiscal 2015, the backlog reached 500,000 cases.

The SSDI trust fund is scheduled to go bankrupt by late 2016, when benefits will have to be cut by 19

³The SSA employs almost 1,400 ALJs to adjudicate about 700,000 cases a year. These independent judges effectively have lifetime tenure and are charged with conducting impartial *de novo* hearings and making decisions on appealed agency determinations. ALJs can be removed from office only through a lengthy and costly process conducted by the Merit Systems Protection Board, and removals are rare. The 1946 Administrative Procedure Act governs administrative adjudication. Under the Constitution, ALJs are Article I judges, which the Supreme Court has recognized as "functionally comparable" to Article III trial judges. ALJs are not supervised by anyone engaged in agency investigative or prosecutorial functions; are immune from liability for judicial acts; and are exempt from performance ratings, evaluations, and the receipt of bonuses. Their salaries are not set by the agency. For this and further information, see Judge Thomas P. McCarthy, "Respect Administrative Law Judges," letter to *The Wall Street Journal*, Aug. 28, 2014.

⁴Autor and Duggan, "Supporting Work: A Proposal for Modernizing the U.S. Disability Insurance System," Center for American Progress and the Hamilton Project paper (2010).

⁵Autor and Duggan, "The Rise in the Disability Rolls and the Decline in Unemployment," 118 *Q. J. of Econ.* 157-206 (2003).

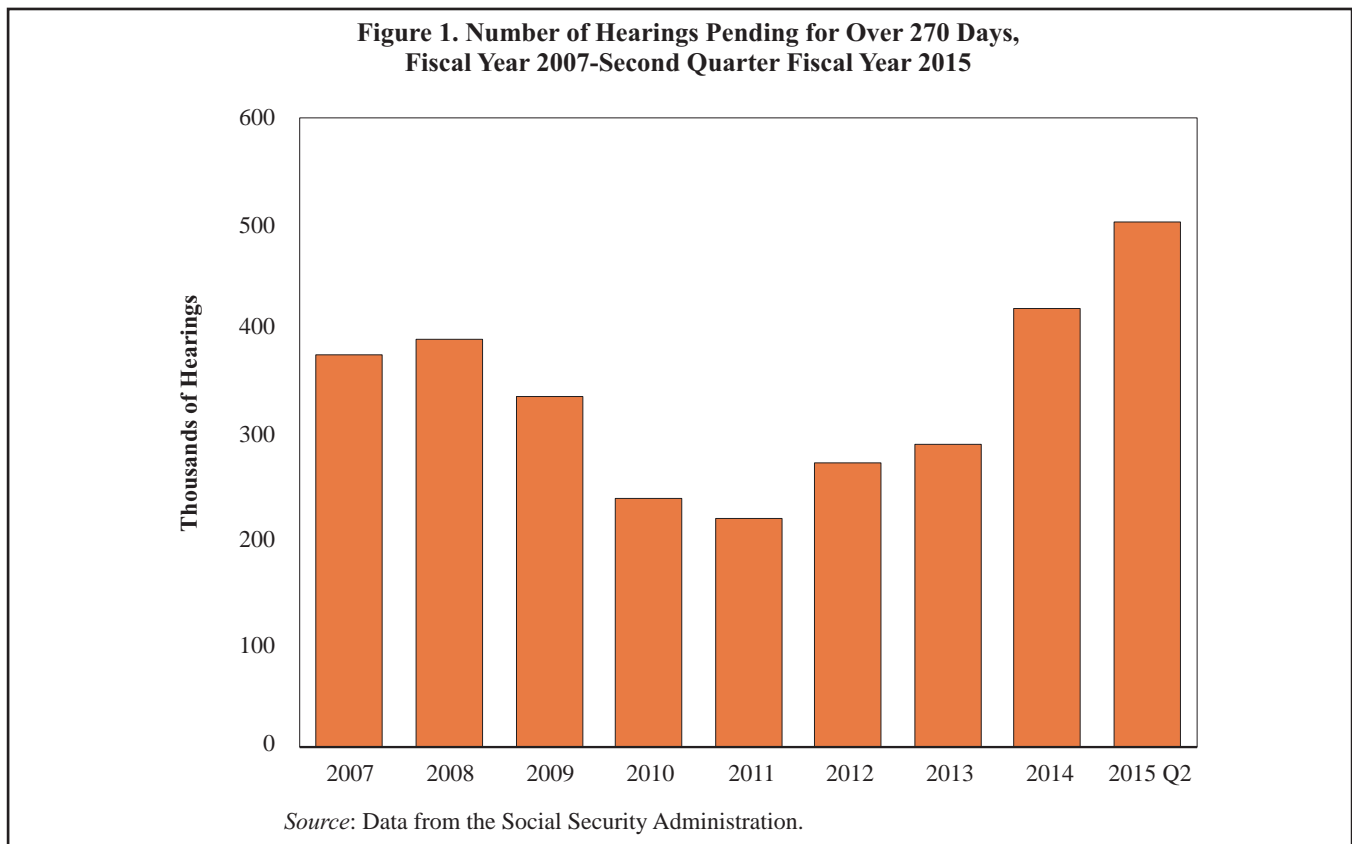
⁶See Warshawsky and Marchand, "Modernizing the SSDI Eligibility Criteria: A Reform Proposal That Eliminates the Outdated Medical-Vocational Grid," Mercatus Center working paper, Apr. 2015.

⁷Wachter, Song, and Manchester, "Trends in Employment and Earnings of Allowed and Rejected Applicants to the Social Security Disability Insurance Program," 101 *Am. Econ. Rev.* 3308 (2011).

⁸Pierce, "What Should We Do About Administrative Law Judge Disability Decision Making?" George Washington University Public Law and Legal Theory paper 573 (2011).

⁹Randy Frye and Marilyn Zahm, interview by Steve Kroft, "Easier to Approve a Disability Case Than Deny It?" *60 Minutes*, CBS, Oct. 6, 2013; and e-mail from the SSA Office of Inspector General (OIG) to staff of the House Oversight and Government Reform Committee.

¹⁰See, e.g., Damian Paletta, "Disability-Claim Judge Has Trouble Saying 'No,'" *The Wall Street Journal*, May 19, 2011.



percent to all SSDI beneficiaries so the program will remain solvent and within its cash flows.¹¹

An Important Econometric Study

Economists Robert Nakosteen and Michael Zimmer conducted an econometric analysis of decisions by ALJs based on data from cases that came before about 1,500 judges for the fiscal years ending September 2010 through September 2012.¹² They examined the data for patterns in both the approval rate and volume of decisions rendered. They used information about the judges' genders, numbers of years of judicial experience, and numbers of decisions rendered, as well as information about the state unemployment rate, an indicator of political makeup for the states in which the judges presided.

Nakosteen and Zimmer show basic statistical evidence of an upward drift in approvals by ALJs as a function of decision volume; judges deciding

many cases trend toward leniency. They find that the mean length of judicial experience is 31 years and that experience is positively related to approval rates. They also find that judges tend to be lenient in environments of relatively high joblessness and in the presence of a Democratic governor. There is no difference in approval rates between genders among the judges. Nakosteen and Zimmer also find that the general disposition toward leniency declined from 2010 through 2012, and although they are not certain of the cause, they emphasize, as mentioned above, the impact of the negative publicity from the series of *Wall Street Journal* investigative articles. Finally, they find econometric evidence that there is a tendency for lenient judges to take large caseloads and that this tendency has grown more pronounced over time.

Analytical Studies

One criticism of some studies examining ALJs is that they focus on lenient judges. However, a 2014 SSA Office of Inspector General (OIG) report examining low-approval judges concludes that the remand and reversal rate (used by the OIG as a quality performance proxy) for these ALJs is, on average,

¹¹SSA board of trustees, "The 2014 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds," House Doc. 113-39 (July 2014).

¹²Nakosteen and Zimmer, "Approval of Social Security Disability Appeals: Analysis of Judges' Decisions," 46(23) *Applied Economics* 2783 (2014).

about the same as that of the general ALJ population.¹³ The report also finds that eight of the 12 lowest-allowance judges “decided fewer cases than the average of their peers.”¹⁴ This finding is telling, given that the SSA regards ALJ decision count as a strong inverse predictor of decisional quality. After examining data from the ODAR, the SSA concluded in a 2012 internal memo that there exists a “strong relationship between production levels and decisional quality on allowances. As ALJ production increases, the general trend for decisional quality on allowances falls.”¹⁵ Thus, we can say from the relationship between decision volume and allowance rate established in the Nakosteen and Zimmer econometric analysis that high-allowance judges have lower decisional quality than low-allowance judges. Legal academics Harold Krent and Scott Morris find that the number of years spent by ALJs in the top 1 percent of allowance rates strongly predicts dispositional volume.¹⁶ This conclusion, coupled with the OIG’s findings on low-allowance ALJ dispositional volume, suggests a large quality gap between high- and low-outlier judges.

In another report, the OIG uses quality review data to compare the accuracy of affirmative ALJ decisions with that of rejection decisions over the 2009-2010 period based on randomly selected cases.¹⁷ Examining 1,022 denials and allowances, the report finds that the approval decisions are 40 percent more likely to garner “disagreeable” ratings by the ODAR’s post-effectuation review process. This finding is hardly surprising, given that hearing-stage applicants have already been rejected at the first two rounds of application by trained examiners at DDS. The SSA conducts pre-effectuation quality reviews of randomly selected DDS decisions to determine whether these examiners reached a sound decision. In fiscal 2010 through 2014, the DDS accuracy rate consistently remained higher than 97 percent. By stating that DDS workers are nearly perfect in their decisions during the first two stages of determination, the SSA is implying that subsequent appeals to ALJs should have low

allowance rates. The existence and persistence of high-approval outlier judges, then, are a cause for concern.

Finally, an OIG report on judicial motivation finds that ALJs are deciding cases on considerations other than what the law and regulations allow, with the ALJs being inappropriately influenced by the unemployment conditions in the local economy or by personal considerations, such as their past occupations or political views.¹⁸

Case Studies

In 2014 the House Oversight and Government Reform Committee released an indictment of “rubber-stamping disability judges.”¹⁹ The committee relied on SSA internal investigations of ALJs with high disposition counts and award rates. One was Charles Bridges, the hearing office chief ALJ for Harrisburg, Pennsylvania. Despite awarding benefits without holding hearings in 7,000 cases and being repeatedly criticized by the SSA for poor decisional quality, he still enjoyed a full caseload, at least through 2014.²⁰ ALJ Harry Taylor, repeatedly accused by colleagues of conducting sloppy work and sleeping on the job, decided nearly 70 percent of his cases without a hearing and denied awards to only 6 percent of claimants. Despite being handed a 14-day suspension for misconduct by the SSA and being recommended for another in 2013, he was still serving on the bench, at least through 2014.²¹

A second committee report on the matter recommended capping the number of annual dispositions at 600,²² consistent with SSA research showing decisional quality decline for ALJs taking more than 617 dispositions in a given year. The report also calls for a prioritization of resources devoted to ALJ decisional review; it states that no more ALJs should be hired until review capacity increases fivefold. More boldly, the committee concludes that judges found to incorrectly apply disability law should be removed and reinstated only upon completion of an observed, compliant trial period. Critically, the committee proposal would make high

¹³SSA OIG, “Subsequent Appellate Actions on Denials Issued by Low-Allowance Administrative Law Judges,” A-12-13-13084 (July 2014).

¹⁴*Id.*

¹⁵SSA, “Memo on Production Levels and Decision Quality,” Sept. 2012.

¹⁶Krent and Morris, “Inconsistency and Angst in District Court Resolution of Social Security Disability Appeals,” Chicago-Kent College of Law Research Paper No. 2014-30 (2014).

¹⁷SSA OIG, “The Social Security Administration’s Review of Administrative Law Judges’ Decisions,” A-07-12-21234 (Mar. 2012).

¹⁸SSA OIG, “Oversight of Administrative Law Judge Workload Trends,” A-12-11-01138 (Feb. 2012).

¹⁹Oversight and Government Reform Committee, “Systemic Waste and Abuse at the Social Security Administration: How Rubber Stamping Disability Judges Cost Hundreds of Billions of Taxpayer Dollars,” June 10, 2014.

²⁰*Id.* at 5-6.

²¹*Id.* at 6-7.

²²Oversight and Government Reform Committee, “Misplaced Priorities: How the Social Security Administration Sacrificed Quality for Quantity in the Disability Determination Process,” Dec. 2014.

allowance rates sufficient to warrant further investigation of an ALJ, which could result in dismissal.²³

We see two problems with those recommendations. First, it is challenging to remove an ALJ for performance through the existing administrative process. Astrue was able to remove only three or four judges, despite considerable efforts, and those cases involved blatant violations.²⁴ Second, the SSA has claimed, with some justification, that it cannot legally focus on judges with high approval rates, per se. That being said, more resources should be devoted to the review and analysis of ALJ decisions.

Our Extended Empirical Analysis

In our new study, as in Warshawsky's 2012 study,²⁵ we calculate the net cost to taxpayers, per year and in total, of presumptively wrong decisions. We assume a cost of \$300,000 per case. We consider whether a decision is presumptively wrong based on a combination of fixed approval or denial rate numbers and twice the standard deviations in that year of all judges' decisions on both the approval and the denial sides of the adjudication distribution. More specifically, we consider as presumptively in error the decisions of all judges with approval rates higher than 80 percent or beyond two standard deviations of the median on the right side of the distribution, as well as all judges with approval rates below 20 percent or beyond two standard deviations on the left side. The addition of outliers in terms of standard deviations (a relative measure) and not just fixed numbers accounts for the natural movement of the average ALJ's performance as a result of the business-cycle effect. Two standard deviations represents behavior on the edges. The fixed numbers of 80 percent and 20 percent are equidistant from a 50 percent approval rate, a bit lower than the average of past decades but consistent with the 80 percent mark indicated by a prominent regional chief ALJ, Jasper Bede, as a red flag for problems in adjudication.²⁶

Table 1 shows the average approval rates of ALJs and the standard deviations. Table 2 shows the proportion of ALJs with approval rates that are two standard deviations above and below the mean, and Table 3 shows the proportion of ALJs from 2005

through 2014 with approval rates greater than 80 percent and lower than 20 percent. Note that approval rates are computed after taking out case dismissals, which are usually administrative actions and do not indicate substantive adjudication. We consider standard deviations (a relative measure) and percentage rates (a fixed measure), in both directions, as indicators of worrisome outliers. One standard deviation contains 66 percent of the ALJs around the mean. Figure 2 shows the distribution of judges by allowance rate from 2005 through 2014.

Table 1. Average Allowance Rates of Administrative Law Judges, 2005-2014

| Year | Average Approval Rate (percentage) | Standard Deviation (percentage) |
|------|------------------------------------|---------------------------------|
| 2005 | 71.3 | 16.0 |
| 2006 | 70.6 | 16.3 |
| 2007 | 70.6 | 16.5 |
| 2008 | 69.6 | 16.5 |
| 2009 | 70.5 | 17.7 |
| 2010 | 67.1 | 16.3 |
| 2011 | 62.2 | 16.7 |
| 2012 | 57.2 | 16.2 |
| 2013 | 55.0 | 15.5 |
| 2014 | 53.7 | 15.1 |

Source: Authors' calculations based on data from the Social Security Administration.

Table 2. Percentage of Administrative Law Judges With Approval Rates Two Standard Deviations Above and Below the Mean, 2005-2014

| Year | Judges 2+ SD Above Mean (percentage) | Judges 2+ SD Below Mean (percentage) |
|------|--------------------------------------|--------------------------------------|
| 2005 | 0.00 | 3.20 |
| 2006 | 0.00 | 3.30 |
| 2007 | 0.00 | 3.50 |
| 2008 | 0.09 | 3.70 |
| 2009 | 0.00 | 3.70 |
| 2010 | 1.50 | 2.80 |
| 2011 | 2.20 | 2.50 |
| 2012 | 2.10 | 2.10 |
| 2013 | 2.40 | 1.70 |
| 2014 | 2.90 | 2.50 |

Source: Authors' calculations based on data from the Social Security Administration.

Several things are evident from these tables and from Figure 2. The average allowance rate dropped from 71 percent in 2005 to 54 percent in 2014, while the standard deviation dropped from 18 percent in 2009 to 15 percent in 2014. The proportion of ALJs with allowance rates more than two standard deviations above the mean increased from 0 to about

²³*Id.* at 49-51.

²⁴House Ways and Means Social Security Subcommittee and the House Committee on the Judiciary Subcommittee on the Courts, Commercial, and Administrative Law, "Statement of Michael J. Astrue, Commissioner, Social Security Administration," 112th Cong., 1st Sess., July 11, 2011.

²⁵Warshawsky, *supra* note 2.

²⁶Oversight and Government Reform Energy Policy, Health Care, and Entitlements Subcommittee, "Continuing Oversight of the Social Security Administration's Management of Federal Disability Programs," 113th Cong., 1st Sess., Nov. 19, 2013.

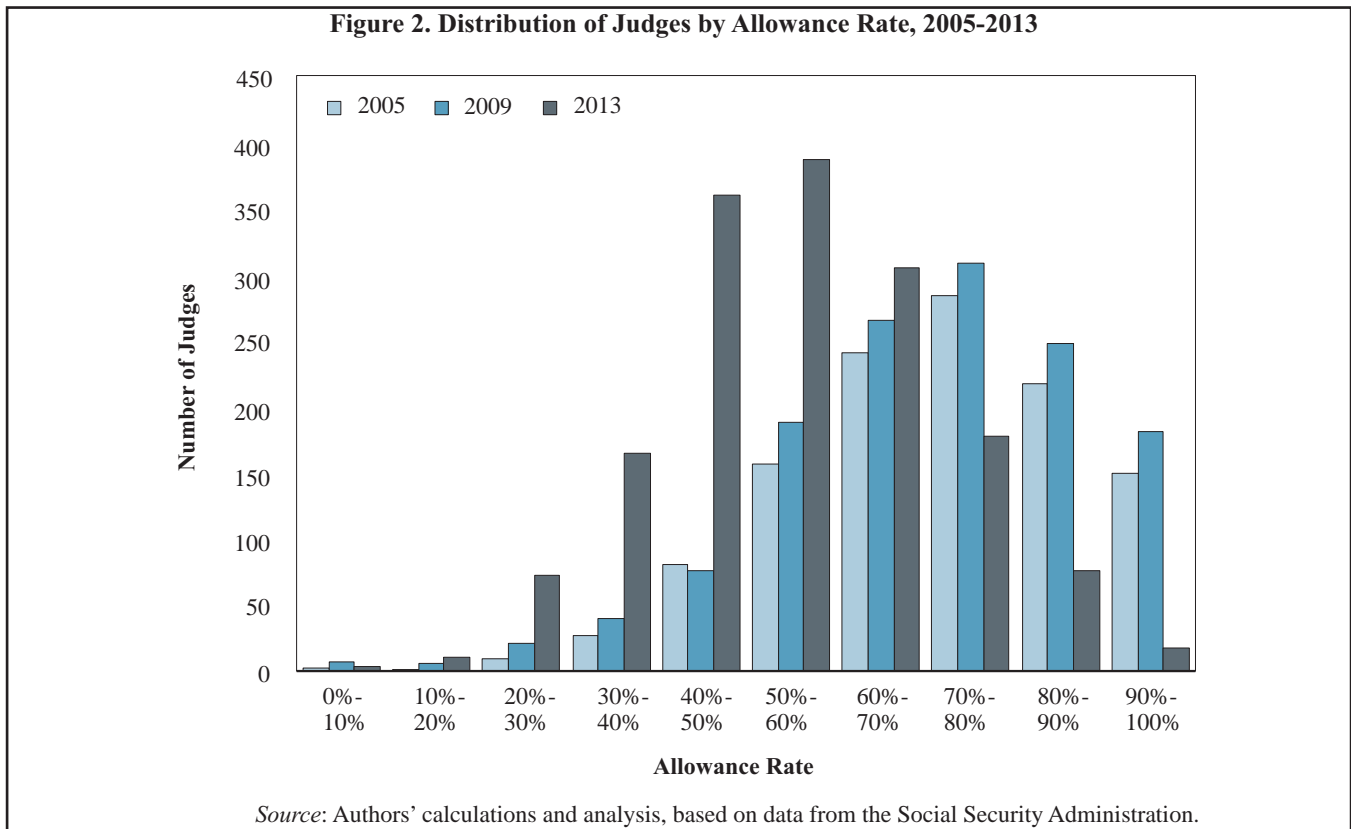


Table 3. Percentage of Administrative Law Judges With Approval Rates of 80 Percent or Higher and 20 Percent or Lower, 2005-2014

| Year | Percentage of Judges With 80 Percent or More Approval | Percentage of Judges With 20 Percent or Less Approval |
|------|---|---|
| 2005 | 31.4 | 0.3 |
| 2006 | 31.3 | 0.6 |
| 2007 | 31.4 | 0.8 |
| 2008 | 28.5 | 0.8 |
| 2009 | 32.0 | 1.0 |
| 2010 | 23.3 | 0.5 |
| 2011 | 16.1 | 0.7 |
| 2012 | 8.6 | 1.2 |
| 2013 | 5.9 | 0.9 |
| 2014 | 4.9 | 1.0 |

Source: Authors' calculations based on data from the Social Security Administration.

3 percent from 2005 through 2014, while the proportion with rates more than two standard deviations below the mean has bounced around somewhat but has dropped overall from a high of almost 4 percent to around 2 percent. By this relative measure of “outlierness,” the number of generous ALJs has increased over time. By contrast, when considering a fixed measure of outlierness, the proportion of ALJs with allowance rates in excess of 80 percent has declined significantly, from

31 percent in 2005 to 5 percent in 2014, while the proportion with allowance rates of less than 20 percent has increased slightly, from 0.3 percent to 1 percent. Overall, the distribution of ALJ allowance has become less skewed and more symmetrical over time.

The proportion of cases decided by ALJs with approval rates exceeding 80 percent (not shown) has dropped even more than the proportion of judges, from nearly 34 percent to 4.5 percent, because the number of cases decided by high-approval judges has declined proportionately over the 2005-2014 period. Among the low-approval judges, the change in the proportion of cases is not significant because low-approval judges have always decided a smaller number of cases. The aggregate number of cases has increased from about 430,000 in 2005 to about 520,000 in 2009 and to 640,000 in 2013. The distribution among judges became more symmetrical around the mean from 2009 to 2013, although it still skews to the right.

Table 4 translates the indications of presumptively wrong decisions in both directions into dollar terms, on net, representing losses to taxpayers from the SSDI trust fund and the general fund of the Treasury (for SSI). The annual loss was more than \$10 billion in 2009 and declined to almost \$1.5 billion in 2014. Over the entire 2005-2014 period, the loss to taxpayers has come to more than \$72 billion.

Figure 3. Regression Analysis of Administrative Law Judge Approval Rates

| Source | SS | df | MS | Number of obs = 13219 | | |
|----------|------------|-------|------------|-----------------------|--------|--|
| Model | 68.9878261 | 11 | 6.27162055 | F(11, 13207) = | 243.59 | |
| Residual | 340.032724 | 13207 | .025746401 | Prob > F = | 0.0000 | |
| Total | 409.02055 | 13218 | .030944209 | R-squared = | 0.1687 | |
| | | | | Adj R-squared = | 0.1680 | |
| | | | | Root MSE = | .16046 | |

| AllowanceRate | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] | |
|-------------------|-----------|-----------|--------|-------|----------------------|-----------|
| SDofAllowanceRate | -.4728226 | .0296069 | -15.97 | 0.000 | -.5308564 | -.4147889 |
| Decisions | .0000523 | 8.00e-06 | 6.54 | 0.000 | .0000367 | .000068 |
| year1 | 0 | (omitted) | | | | |
| year2 | -.0072872 | .006784 | -1.07 | 0.283 | -.0205848 | .0060104 |
| year3 | -.0073604 | .0067725 | -1.09 | 0.277 | -.0206354 | .0059146 |
| year4 | -.0165027 | .006867 | -2.40 | 0.016 | -.029963 | -.0030425 |
| year5 | -.0039769 | .0065499 | -0.61 | 0.544 | -.0168156 | .0088618 |
| year6 | -.0380327 | .0065085 | -5.84 | 0.000 | -.0507904 | -.0252751 |
| year7 | -.0874753 | .0063926 | -13.68 | 0.000 | -.1000057 | -.0749449 |
| year8 | -.1370428 | .0063076 | -21.73 | 0.000 | -.1494066 | -.1246791 |
| year9 | -.1583018 | .0063586 | -24.90 | 0.000 | -.1707656 | -.1458379 |
| year10 | -.1687925 | .006448 | -26.18 | 0.000 | -.1814314 | -.1561535 |
| _cons | .7264799 | .0062145 | 116.90 | 0.000 | .7142987 | .7386612 |

Source: Authors' analysis based on SSA data.

Those numbers are not discounted to present value and represent the value of future benefits. This can be improved when resources and attention are devoted to reforming program management. At the same time, the structural and financial state of the disability program is dire, and while improvements are being made, the continuing high rate of presumptive mistakes adds billions to a growing deficit.

We next conduct a simple least squares regression analysis of the annual approval rates of ALJs, considering only judges with at least three years of experience. The underlying data come from the SSA. In particular, we want to see if judges' approval rates are related to the standard deviations of their approval rates from 2005 through 2014, to the number of decisions they have made annually, and to the years of the decisions. We represent the model in equation form as follows:

$$Y = B_0 + B_1X_1 + B_2X_2 + B_iX_i + u,$$

where Y is the judge's approval rate, 2005-2014; X_1 is the standard deviation of the judge's approval

Table 4. Annual Net Cost of Presumption Mistakes of High- and Low-Approval Administrative Law Judges (billions of dollars)

| Year | High Approval | Low Approval | Sum |
|------|---------------|--------------|------|
| 2005 | 10.4 | (1.2) | 9.1 |
| 2006 | 11.3 | (1.6) | 9.8 |
| 2007 | 11 | (1.5) | 9.5 |
| 2008 | 10.1 | (1.5) | 8.6 |
| 2009 | 11.9 | (1.6) | 10.3 |
| 2010 | 10.7 | (1.5) | 9.2 |
| 2011 | 8.4 | (1.3) | 7.1 |
| 2012 | 5.1 | (1) | 4.1 |
| 2013 | 3.6 | (0.7) | 3 |
| 2014 | 2.4 | (1) | 1.4 |

Source: Authors' calculations.

Note: All numbers are rounded to the nearest tenth.

rate, 2005-2014; X_2 is the annual number of decisions made by the judge; and X_i are time dummy variables for 2005-2014. Figure 3 shows the results.

Several interesting results arise from this analysis. First, the higher the standard deviation of an

ALJ's past decisions, or the more variability in an ALJ's decision-making, the lower the judge's current approval rate. In other words, the more consistently a judge decides in one direction, the higher the judge's approval rate. This is a disconcerting finding because, given the randomness by which cases are supposed to be assigned to judges by the SSA, we should expect variability from year to year in the judge's approval rates. But high-approval judges are high-approval judges year in and year out, whereas low-approval judges are not; their decisions vary more over the years. We also see a positive coefficient on the number of cases decided, another concerning result. The more cases a judge decides, the higher the approval rate. This outcome should not occur in a well-functioning review system. Finally, we see significant and increasingly negative coefficients on the dummy variables for the last five years, as the economy has improved and as SSA administrative reforms, described later, have begun to have an influence.

Further Circumstantial Evidence

The agency rules and their administration affect the overall award rates and the incentive to file an appeal. An interesting and relevant issue is the December 2014 bankruptcy of Binder & Binder, the largest law firm specializing in SSDI claim appeals. The bankruptcy has been attributed to the tightening of administrative procedures, which lowered award rates and hence lowered payments to third-party representatives, especially attorneys. At the same time, the number of ALJ decisions fell noticeably.²⁷

Recommendations

Astrue instituted several administrative reforms in response to the problems with the ALJ system. He hired and trained a record number of new ALJs, drawn from fresh candidate lists, and encouraged some long-serving ALJs to retire. He limited the number of cases that could be heard by any ALJ each year to 1,000, which was later reduced to about 800. To limit claimant and representative abuses in the adjudication process, he allowed only one application for benefits per worker in the system at a time. He set up a more rigorous method of rotating the cases among judges in response to clear signs of judicial collusion with attorneys. Also, he began a program of random review of ALJ decisions, including pre-effectuation reviews of allowances.

Astrue's reforms were good and necessary, but they do not go far enough. Losses to taxpayers

continue despite the recent improvements. Moreover, as public attention moves elsewhere, and if there are renewed demands to fix the claim backlog — because the backlog has recently grown — his changes could easily be undone because administrators face heavy political pressures to expedite and shortcut responsible processes. Future administrators could reverse reforms, intentionally or unintentionally, as bad habits slip back into the system. Astrue's program to increase accountability through random reviews of ALJ decisions and to increase judicial turnover by hiring new judges and encouraging a few to retire (prominently, Daugherty) should therefore be made permanent and reinforced. In particular, Congress should institute 15-year term limits for judges to ensure that fresh legal minds are joining the pool of judges and to prevent it from becoming stale and unresponsive to legal criteria and requirements. A term of a decade and a half is long enough to insulate judges from administrative and political pressures and prevent undue political influence. Also, it should not be at the agency's discretion to conduct a statistically valid number of pre-effectuation reviews on ALJ allowances; that should be required by statute, as it is at the DDS level.

The system faces a large backlog of cases, likely made worse by strategic claimants, for example, those who file serial claims in the hope of eventually getting a lenient judgment at the initial or the appeals level. Congress can limit this gamesmanship by allowing only one application per claimant in a three-year period. This would reduce the number of claims and ensure that those made are more serious and substantive, thereby reducing the backlog. On-the-record decisions should not be allowed; these are difficult for the agency to later evaluate for program eligibility in continuing disability reviews because the documentation and hence the rationale of the original determinations are so poor. As we mentioned earlier, because judges must marshal more documentation for a denial than for an approval, they have an incentive to grant benefits to keep the system chugging along. The agency can fix this problem by further limiting the number of cases each judge must decide to 500 — that is, about 2¼ cases per working day, a reasonable number in light of the complexity of disability adjudication.

The system is further complicated by the so-called three-hat rule, under which the judge must advocate for the claimant, advocate for the government (that is, the taxpayer), and render unbiased judgment. Even if a claimant has legal counsel, the judge must still advocate on the claimant's behalf. This rule must be eliminated. Most claimants — 85 percent — now have third-party representatives,

²⁷Sara Randazzo and Paletta, "Social Security Disability Firm Binder & Binder Prepares for Possible Chapter 11," *The Wall Street Journal*, Dec. 12, 2014.

most of whom are experienced and are paid if they win the case. They can be expected to represent their clients well. Moreover, these professionals should be held responsible for getting supporting materials into court expeditiously and completely so the record can be closed in a timely manner before the hearing is held.

Conclusions

In Lawrence Summers's "The Scientific Illusion in Empirical Macroeconomics," the celebrated academic articulates a standard of evidence often ignored in economics.²⁸ Summers brushes aside "statistical pyrotechnics," noting that "physicists do not compete to find more and more elaborate ways to observe falling apples." He advocates instead for an agenda based on "pragmatic economic work," in which "many different types of data are examined" and "no single test is held out as decisive." Although strong results from one method can indicate a likely problem, examining multiple strands of ALJ data from the previous decade enables us to conclude definitively that there are systemic problems in the allowance process.

Using case studies of outlier judges, quality review data from the SSA, and empirical work pertaining to the iron triangle of dispositional volume, allowance rate, and decisional quality, we conclude that outlier high-allowance judges are deviating from the law by over-providing benefits. This tendency carries large economic consequences; removing both the most and least "generous" judges would have saved taxpayers more than \$72 billion based on awards given from 2005-2014. But even this large amount fails to capture the tremendous opportunity cost that comes with keeping capable workers out of the labor force, which will continue to rise as worker productivity grows over time. By reducing the number of applications, further capping the yearly number of cases heard by ALJs, ending lifetime ALJ tenure, devoting greater time and resources to conducting quality reviews, and ending judicial advocacy for claimants, we can restore sustainability and integrity to a troubled claims appeals process.

²⁸Summers, "The Scientific Illusion in Empirical Macroeconomics," 93(2) *Scandinavian J. of Econ.* 129 (1991).

Let the news find you.

Customized e-mail alerts. Let the tax news and analysis you need find you fast with our customized e-mail alerts. Simply profile your needs, by code section, jurisdiction, document type, subject, search terms, and more, and we'll get you what you want, when you want it. It's one part of a great personalized service.

Visit taxanalysts.com today.

taxanalysts[®]

Tax Notes Today[®] ▲ State Tax Today[®] ▲ Worldwide Tax Daily[®]