



December 2017

SOCIAL SECURITY DISABILITY

Additional Measures and Evaluation Needed to Enhance Accuracy and Consistency of Hearings Decisions

Accessible Version

GAO Highlights

Highlights of [GAO-18-37](#), a report to the Chairman, Subcommittee on Social Security, Committee on Ways and Means, House of Representatives

Why GAO Did This Study

Individuals who do not agree with the initial decision on a claim for Social Security disability benefits can ultimately appeal the decision by requesting a hearing before one of SSA's approximately 1,500 administrative law judges. However, the rate at which these judges have allowed benefits has varied, raising questions about the reasons for this variation. GAO was asked to review aspects of SSA's oversight of judges' decisions.

This report examines (1) to what extent allowance rates vary across administrative law judges, and factors associated with this variation; and (2) the extent to which SSA has processes to monitor the accuracy and consistency of hearings decisions.

GAO developed a statistical model to analyze SSA data on adult disability decisions made by administrative law judges from fiscal years 2007 through 2015, the most current data available at the time of GAO's analysis; reviewed relevant federal laws, regulations, and agency documents; and interviewed SSA officials and chief judges in SSA's 10 regions, as well as officials from organizations representing judges, disability claimants, and claimant representatives.

What GAO Recommends

GAO is making two recommendations, including that SSA systematically evaluate its quality assurance reviews and take steps to reduce or better manage any unnecessary overlap among them. SSA concurred and plans to address them through a comprehensive assessment of its oversight.

View [GAO-18-37](#). For more information, contact Charles Jeszeck at (202) 512-7215 or jeszeck@gao.gov.

December 2017

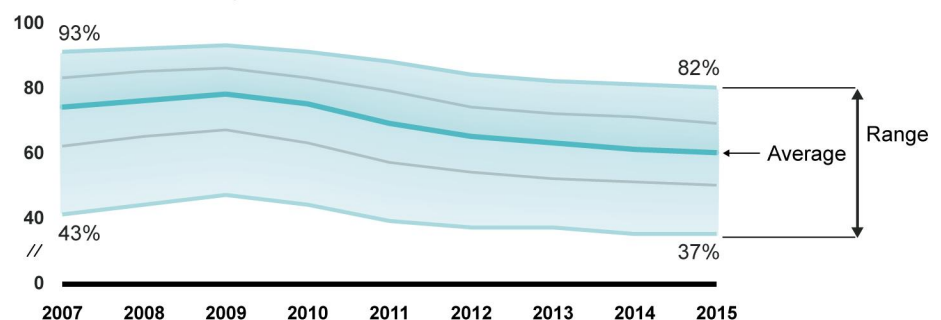
SOCIAL SECURITY DISABILITY

Additional Measures and Evaluation Needed to Enhance Accuracy and Consistency of Hearings Decisions

What GAO Found

Allowance rates—the rate at which Social Security Administration (SSA) administrative law judges allowed disability benefits to be paid when claimants appealed—varied across judges, even after holding constant certain characteristics of claimants, judges, hearing offices, and other factors that could otherwise explain differences in allowance rates. Specifically, GAO estimated that the allowance rate could vary by as much as 46 percentage points if different judges heard a typical claim (one that was average in all other factors GAO analyzed). SSA officials said that this level of variation is not surprising, given the complexity of appeals and judicial discretion. Nonetheless, the variation declined by 5 percentage points between fiscal years 2007 and 2015 (see figure), a change officials attributed to enhanced quality assurance efforts and training for judges. GAO also identified various factors that were associated with a greater chance that a claimant would be allowed benefits. In addition to characteristics related to disability criteria, such as the claimant's impairment and age, GAO found that claimants who had representatives, such as an attorney or family member, were allowed benefits at a rate nearly 3 times higher than those without representatives. Other factors did not appear related to allowance rates, such as the percentage of backlogged claims in a hearing office.

Estimated Allowance Rates across Social Security Administrative Law Judges for Typical Disability Claims, Fiscal Years 2007-2015
Allowance rate (Model adjusted)



Source: GAO analysis of Social Security Administration (SSA) data. | GAO-18-37

Note: The range is the difference in allowance rates for judges at the 95th and 5th percentiles. We also identify the 75th and 25th percentiles. A typical claim had average values on all other factors we analyzed.

SSA has various reviews to monitor the accuracy and consistency of hearings decisions by administrative law judges, but some of these reviews may overlap and SSA has not systematically evaluated them. Specifically, SSA conducts five types of quality assurance reviews of hearings decisions, several of which have similar goals and may look at similar claims. SSA has not evaluated the efficiency or effectiveness of these reviews, despite spending at least \$11 million on them in fiscal year 2016. Moreover, the agency has struggled to sustain all of its quality reviews due to competing priorities—two of the five reviews were curtailed in 2016 because SSA reassigned staff to help expedite claims decisions. By evaluating which quality assurance reviews are most effective and efficient in improving accuracy and consistency, SSA would be better positioned to meet its goals within its resources.

Contents

Letter	1
Background	4
Allowance Rates Vary Across Judges, Even for Typical Claims	14
SSA's Efforts to Monitor Accuracy and Consistency of Hearings- Level Decisions Lack Performance Measures and Have Not Been Evaluated	29
Conclusions	42
Recommendations for Executive Action	43
Agency Comments and Our Evaluation	43
<hr/>	
Appendix I: Objectives, Scope, and Methodology	47
Analysis of Variation in Allowance Rates across Judges and Associated Factors	48
Statistical Model of Variation in Allowance Rates across Judges and Associated Factors	58
Covariates and Subsamples	65
Results	67
Evaluation of SSA's Processes to Monitor Accuracy and Consistency in Hearings Decisions	72
<hr/>	
Appendix II: Comments from the Social Security Administration	74
<hr/>	
Appendix III: GAO Contact and Staff Acknowledgments	79
GAO Contact	79
Staff Acknowledgments	79
<hr/>	
Appendix IV: Accessible Data	80
Data Tables	80
Agency Comment Letter	85

Tables

Table 1: Social Security Administration (SSA) Performance Measures for Disability Decisions, Fiscal Years 2017- 201830
--

Table 2: Comparison of Goals of Social Security Administration (SSA) Quality Reviews of Hearings-level Decisions	37
Table 3: Comparison of Cases Reviewed by Social Security Administration (SSA) Quality Reviews of Hearings-level Decisions	39
Table 4: Social Security Administration (SSA) Spending on Quality Assurance Reviews in Fiscal Years 2015 and 2016 and Number of Cases Reviewed (dollars in millions)	41
Table 5: Distribution of Allowances and Denials, by Program Type and Step in the Social Security Administration's (SSA) Disability Decision-making Process, Fiscal Years 2007-2015	58
Table 6: Estimated Distributions of Social Security Disability Allowance Rates across Judges, Hearing Offices, and Primary Diagnoses	67
Table 7: Sample Distributions and Estimates for Primary Model of Social Security Disability Allowance Rates across Judges	69
Data Table for Highlights figure, Estimated Allowance Rates across Social Security Administrative Law Judges for Typical Disability Claims, Fiscal Years 2007-2015	80
Data Table for Figure 1: Five-Step Determination Process for Social Security Disability Claims	80
Data Table for Figure 2: Social Security Disability Appeals Process	81
Data Table for Figure 3: Social Security Administration (SSA) Requests for Administrative Law Judge (ALJ) Hearings, Case Dispositions, and Pending Cases, Fiscal Years 2010-2016	82
Data Table for Figure 4: Variation in Social Security Administration (SSA) Disability Allowance Rates across Administrative Law Judges, Fiscal Years 2007-2015	82
Data Table for Figure 5: Estimated Variation in Social Security Administration (SSA) Allowance Rates across Judges for Typical Claims, Fiscal Years 2007-2015	82
Data Table for Figure 6: Estimated Variation in Social Security Administration (SSA) Disability Allowance Rates across Hearing Offices for Typical Claims, Fiscal Years 2007-2015	83
Data Table for Figure 7: Social Security Administration (SSA) Disability Allowance Rates for the 25 Most Commonly Recorded Primary Impairments, Fiscal Years 2007-2015	83

Data Table for Figure 8: Four Social Security Administration (SSA) Offices Responsible for Quality Assurance Reviews of Hearings-level Decisions, as of September 2017	84
--	----

Figures

Figure 1: Five-Step Determination Process for Social Security Disability Claims	6
Figure 2: Social Security Disability Appeals Process	8
Figure 3: Social Security Administration (SSA) Requests for Administrative Law Judge (ALJ) Hearings, Case Dispositions, and Pending Cases, Fiscal Years 2010-2016	11
Figure 4: Variation in Social Security Administration (SSA) Disability Allowance Rates across Administrative Law Judges, Fiscal Years 2007-2015	15
Figure 5: Estimated Variation in Social Security Administration (SSA) Allowance Rates across Judges for Typical Claims, Fiscal Years 2007-2015	17
Figure 6: Estimated Variation in Social Security Administration (SSA) Disability Allowance Rates across Hearing Offices for Typical Claims, Fiscal Years 2007-2015	18
Figure 7: Social Security Administration (SSA) Disability Allowance Rates for the 25 Most Commonly Recorded Primary Impairments, Fiscal Years 2007-2015	22
Figure 8: Four Social Security Administration (SSA) Offices Responsible for Quality Assurance Reviews of Hearings-level Decisions, as of September 2017	35

Abbreviations

ACS	American Community Survey
ALJ	Administrative Law Judge
CARES	Compassionate And Responsive Services plan
CDR	Continuing Disability Review
CPMS	Case Processing and Management System
DI	Disability Insurance
DDS	Disability Determination Services
eBB	electronic Bench Book
FPPS	Federal Personnel and Payroll System
HALLEX	Hearings, Appeals, and Litigation Law manual
MI	management information
MSPB	Merit Systems Protection Board
OAO	Office of Appellate Operations
ODAR	Office of Disability Adjudication and Review (as of October 1, 2017, called the Office of Hearings Operations)
OIG	Office of the Inspector General
OPM	Office of Personnel Management
SAA	senior attorney adjudicator
SGA	substantial gainful activity
SSA	Social Security Administration
SSI	Supplemental Security Income

This is a work of the U.S. government and is not subject to copyright protection in the United States. The published product may be reproduced and distributed in its entirety without further permission from GAO. However, because this work may contain copyrighted images or other material, permission from the copyright holder may be necessary if you wish to reproduce this material separately.



December 7, 2017

The Honorable Sam Johnson
Chairman
Subcommittee on Social Security
Committee on Ways and Means
House of Representatives

Dear Mr. Chairman:

The Social Security Administration (SSA) manages two disability benefit programs—Disability Insurance (DI) and Supplemental Security Income (SSI)—that together provide about 16 million Americans with about \$200 billion in benefits annually. Overall, more than 6 percent of the U.S. working-age population receives disability benefits from one or both of these programs, although the rate varies by state (ranging from about 4 percent to more than 12 percent), according to SSA.¹ A claimant who is dissatisfied with the initial decision on his or her application can ultimately appeal at a hearing, where an administrative law judge (ALJ)² reviews the case and any new evidence submitted by the claimant.³ About 30 percent of the claimants who are ultimately allowed benefits are granted benefits at the hearings level or beyond.⁴

For more than 15 years, the number of people applying for disability benefits has generally increased, along with the number of appeals. At the same time, the percentage of appealed claims that have been allowed benefits has declined. However, these rates—known as allowance rates—can vary widely across judges and hearings offices, raising questions about the reasons for this variation.

¹SSA, *Annual Statistical Report on the Social Security Disability Insurance Program*, 2015.

²For readability, we are using the term administrative law judge (ALJ) and judge interchangeably.

³For the purposes of this report, the term “appeal” refers to claimants’ requests for a review of a denied claim to an ALJ, and not any subsequent appeals of ALJ decisions to SSA’s Appeals Council. In most states, claimants must first request a reconsideration of the initial decision at the state Disability Determination Services before appealing to an ALJ. We describe this process in more detail in the background section of this report.

⁴GAO calculations of the proportion of claimants granted benefits at the hearings level or beyond are based on data available in SSA, *Annual Statistical Report on the Social Security Disability Insurance Program*, 2015.

Moreover, reports of improper decisions made by certain judges⁵ and a scheme between a judge and an attorney in one state to commit fraud,⁶ have raised questions about SSA's oversight of ALJs. You asked us to examine allowance rates and SSA's oversight of ALJs. This report examines (1) the extent to which allowance rates vary across administrative law judges, and any factors that are associated with this variation, and (2) the extent to which SSA has processes to monitor the accuracy and consistency of hearings decisions.

To address these objectives, we conducted a statistical analysis of data on adult disability decisions made by administrative law judges from fiscal years 2007 through 2015 and the factors that may be associated with variation in allowance rates. Specifically, we considered the following factors:

- Claimant characteristics relevant to the disability determination process (e.g., age and primary impairment),
- Participants in the process other than claimants (e.g., claimant representatives and medical and vocational experts),
- Judge characteristics (e.g., year of appointment as a judge, and any prior related SSA experience),
- SSA administrative characteristics (e.g., hearing office where the case was decided and whether the hearing was conducted by videoconference), and
- Economic characteristics (unemployment and poverty rates in the claimant's state).

(Appendix I provides more detail on our statistical analysis, including a complete list of the factors we analyzed.)

⁵SSA Office of the Inspector General, "Administrative Law Judges with Both High Dispositions and High Allowance Rates." A-12-14-24092. (Baltimore, Md.: Nov. 14, 2014).

⁶According to the Department of Justice, a retired administrative law judge was sentenced to prison for his role in a scheme, along with an attorney, to fraudulently obtain Social Security disability benefits for thousands of claimants. Department of Justice, "Former Social Security Administrative Law Judge Sentenced to Four Years in Prison for Role in \$550 Million Social Security Fraud Scheme." Aug. 25, 2017. This case was originally reported in a *Wall Street Journal* investigation in 2011. See Damian Paletta, "Disability Claim Judge Has Trouble Saying 'No': Near Perfect Approval Record; Social Security Strained," *The Wall Street Journal*, May 19, 2011.

We assessed the reliability of SSA's administrative data on disability decisions and judge characteristics by (1) performing electronic testing of required data elements, (2) reviewing existing information about the data and the system that produced them, and (3) interviewing agency officials knowledgeable about the data. We determined that the data were sufficiently reliable for the purposes of this report. (Appendix I describes our data in more detail.)

We reviewed relevant federal laws, regulations, and documentation and collected testimonial evidence from SSA officials to describe and evaluate the processes SSA uses to monitor hearings decisions, detect variation, and improve accuracy. We assessed these monitoring efforts against federal internal control standards and our management and evaluation guide for assessing fragmentation, overlap, and duplication in government programs.⁷ We also reviewed SSA's annual performance plans from fiscal years 2006 through 2017 to describe the performance measures the agency has established to improve the accuracy and consistency of its hearings decisions. We evaluated the current performance measures using key attributes of performance measures identified in prior GAO work⁸ and federal internal control standards.⁹

We conducted this performance audit from December 2015 to December 2017 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

⁷GAO, *Fragmentation, Overlap and Duplication: An Evaluation and Management Guide*, [GAO-15-49SP](#) (Washington, D.C.: Apr. 14, 2015).

⁸In a previous report, we developed nine attributes of performance goals and measures based on previously established GAO criteria, as well as relevant federal laws and performance management literature. See GAO, *Tax Administration: IRS Needs to Further Refine Its Tax Filing Season Performance Measures*, [GAO-03-143](#) (Washington, D.C.: Nov. 22, 2002).

⁹GAO, *Standards for Internal Control in the Federal Government*, [GAO-14-704G](#) (Washington, D.C.: Sept. 10, 2014). See, for example, Principles 6, 7, 10, 13, and 16.

Background

The Social Security Administration's (SSA) Disability Insurance (DI) and Supplemental Security Income (SSI) programs are the two largest federal programs providing cash assistance to people with disabilities.

- The DI program, established in 1956, provides monthly payments to working-age adults (and their dependents or survivors) who are unable to work due to a long-term disability.
- The SSI program, established in 1972, is a means-tested income assistance program that provides monthly payments to adults or children who are aged, blind, or have other disabilities and whose income and assets fall below a certain level.
- Individuals with low incomes and assets who also have a sufficient work history may qualify for the DI and SSI programs concurrently. In this case, the individual's SSI payment is generally offset by the amount of the DI payment.

In fiscal year 2016, according to SSA, about 10.8 million disabled workers and their family members received about \$143 billion in DI benefits, and an estimated 8.2 million individuals received almost \$59 billion in SSI benefits (of those, 2.6 million received SSI in addition to DI or Old-Age and Survivors benefits).¹⁰

Disability Criteria

Although DI and SSI have different purposes and target populations, the disability criteria for adults are the same for both programs. To be considered eligible for either program as an adult, a person must have a medically determinable physical or mental impairment that (1) has lasted or is expected to last for at least a continuous period of 1 year or result in death, and (2) prevents them from engaging in any substantial gainful activity (SGA).¹¹

¹⁰SSA, Fiscal Year 2018 Congressional Justification.

¹¹42 U.S.C. §§ 423(d) and 1382c(a)(3)(A). SGA is work activity that involves significant physical or mental activities that is done for pay or profit, regardless of whether profit is realized. In 2017, SSA set SGA as monthly earnings above \$1,950 for blind individuals and \$1,170 for non-blind individuals.

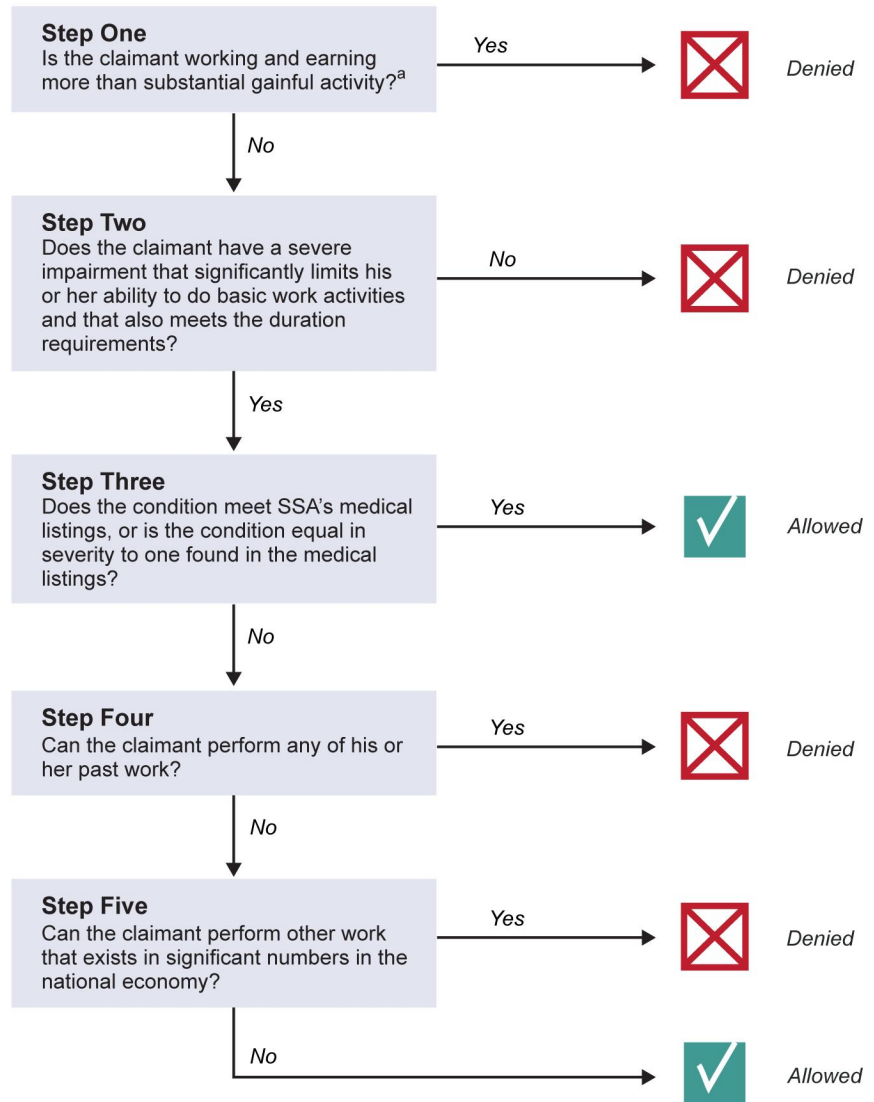
The disability decision-making process includes five sequential steps (see fig. 1).¹² First, SSA determines if a claimant is working and screens out (denies) claimants who earn over a specified amount. Second, SSA determines whether the claimant has an impairment severe enough to significantly limit his or her ability to do basic work activities and expected to last more than 12 months or result in death, and denies claimants who do not meet these criteria. At the third step, SSA determines whether a claimant's impairment meets or is equivalent to an impairment listed in SSA's Listings of Impairments. If a claimant "meets" or "equals" one of the listed impairments, they are allowed benefits. If not, SSA proceeds to the last two steps and assesses whether a claimant, given their impairment, can do their past work (step four) or other work that exists in significant numbers in the national economy (step five).

Over time, more of SSA's disability decisions have been made at the last two steps in the process, which require more judicial discretion than decisions made at steps 1 through 3, according to SSA. In 2000, 29 percent of decisions were made at steps 4 and 5, according to an SSA report. By 2014, nearly half—49 percent—of all decisions were made at these steps.¹³

¹²See 20 C.F.R. §§ 404.1520 and 416.920.

¹³SSA, *Annual Statistical Report on the Social Security Disability Insurance Program, 2015*.

Figure 1: Five-Step Determination Process for Social Security Disability Claims



Source: GAO analysis of Social Security Administration (SSA) information. | GAO-18-37

^aIn 2017, the substantial gainful activity (SGA) threshold was \$1,950 per month for blind individuals and \$1,170 per month for non-blind individuals.

Disability Application and Appeals Process

To apply for benefits, a claimant must file an application online, by telephone, or mail, or in person at a local Social Security office. If field office staff determine that the claimant meets the nonmedical eligibility criteria, they forward the claim to the appropriate state Disability Determination Services (DDS) office. DDS staff—generally a team comprised of disability examiners and medical consultants—review medical and other evidence provided by the claimant, obtaining additional evidence as needed, and make the initial disability determination. In fiscal year 2016, SSA received more than 2.5 million disability claims.¹⁴

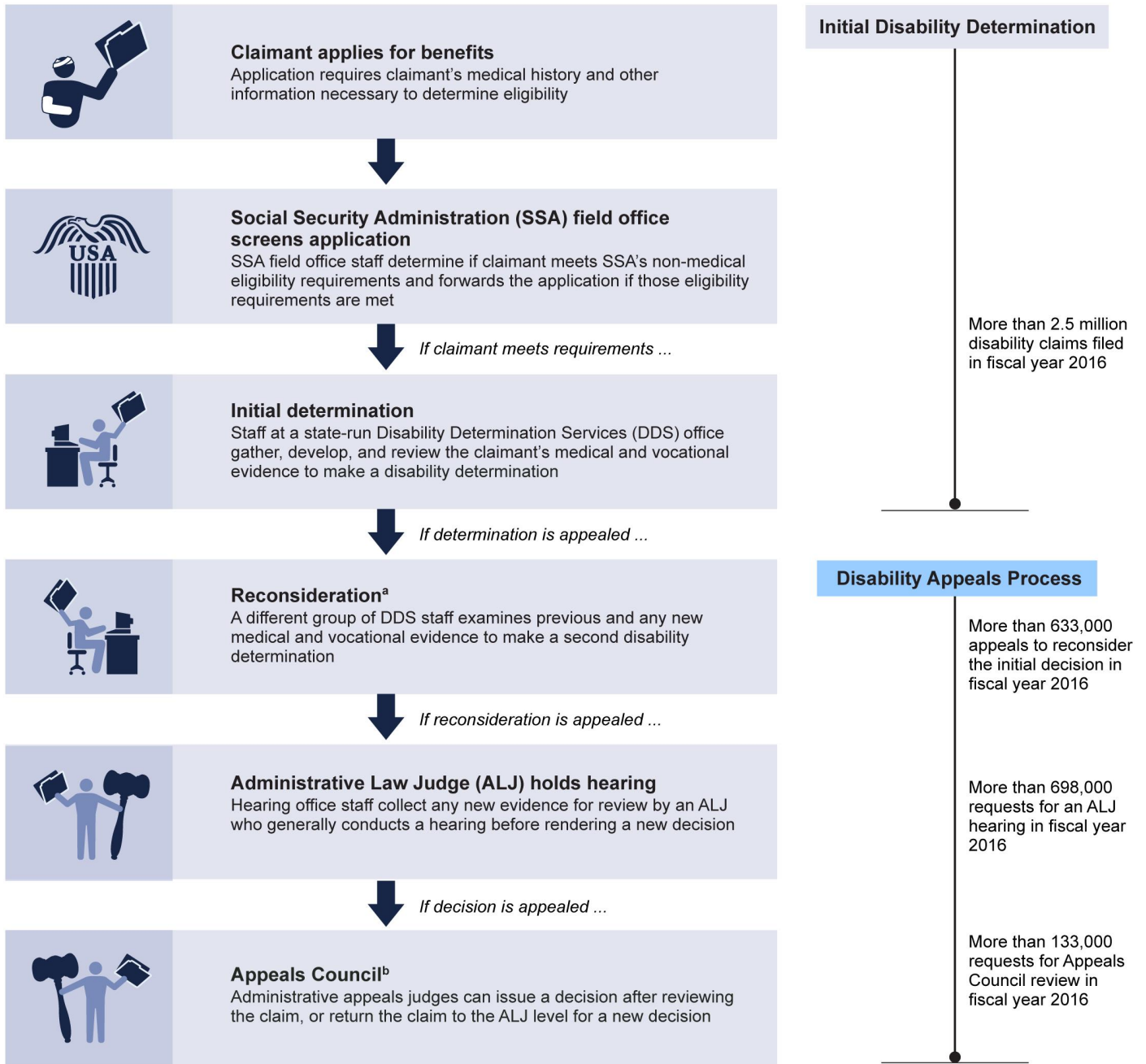
If the claimant is not satisfied with this determination, in most states he or she may request a reconsideration of the decision within the same DDS office. If the claimant is dissatisfied with the reconsideration, he or she may request a hearing before an administrative law judge (ALJ). In one of several initiatives to improve the disability determination process, SSA has eliminated the reconsideration step of the process in 10 states, allowing the claimant to appeal the initial decision directly to an ALJ.¹⁵ In fiscal year 2016, claimants appealed more than 698,000 decisions to the hearings level, and SSA issued more than 637,000 dispositions (including allowances, denials, and dismissals).¹⁶ (See fig. 2).

¹⁴SSA, Fiscal Year 2018 Congressional Justification.

¹⁵The 10 states are Alabama, Alaska, California (Los Angeles North and Los Angeles West Branches), Colorado, Louisiana, Michigan, Missouri, New Hampshire, New York, and Pennsylvania.

¹⁶SSA, Fiscal Year 2018 Congressional Justification.

Figure 2: Social Security Disability Appeals Process



Source: GAO analysis of Social Security Administration (SSA) information. | GAO-18-37

^aIn 1999, SSA began testing a new process, the Disability Redesign Prototype model, in 10 states, which included eliminating the reconsideration step of the administrative review process for disability claims.

^bAfter going through the disability appeals process within SSA, claimants must file any further actions in federal court.

Within SSA's Office of Disability Adjudication and Review (ODAR),¹⁷ there are approximately 1,500 ALJs who are located in 166 hearing offices across the country, as well as at five National Hearing Centers. In general, cases are randomly assigned to ALJs within the area each hearing office serves, in the order in which the requests for a hearing are received.¹⁸ The ALJ reviews the claimant's file, including any additional evidence the claimant submitted after the initial determination, and generally conducts a hearing. At the hearing, the ALJ may hear testimony from the claimant, medical experts on the claimant's medical condition, and vocational experts regarding the claimant's past work and jobs currently available in significant numbers in the national economy. The majority of claimants are represented at these hearings by an attorney or nonattorney representative, such as a professional disability representative, relative, or social worker.

If the claimant is not satisfied with the ALJ decision, he or she may request a review by SSA's Appeals Council, which is the final administrative appeal within SSA.¹⁹ The Appeals Council may grant, deny, or dismiss a request for review. If it agrees to review the case, the Appeals Council may uphold, modify, or reverse the ALJ's decision, or it may remand the case back to the ALJ to hold another hearing and issue

¹⁷As of October 1, 2017, this office is called the Office of Hearings Operations (OHO).

¹⁸As specified in the Hearings, Appeals, and Litigation Law (HALLEX) manual section I-2-1-55, cases are assigned on a rotational basis to ALJs, with the oldest cases receiving priority. According to HALLEX, there are some exceptions to this case assignment process, such as terminal illness cases, or veterans who have a 100 percent permanent and total disability compensation rating.

¹⁹Beyond the Appeals Council, the claimant may appeal to a federal district court. The claimant can continue legal appeals to the appropriate U.S. Court of Appeals and, ultimately, to the Supreme Court of the United States.

a new decision.²⁰ In fiscal year 2016, the Appeals Council reviewed more than 154,000 ALJ decisions and remanded 13 percent of them.²¹

Hearings Backlogs and Processing Times in Recent Years

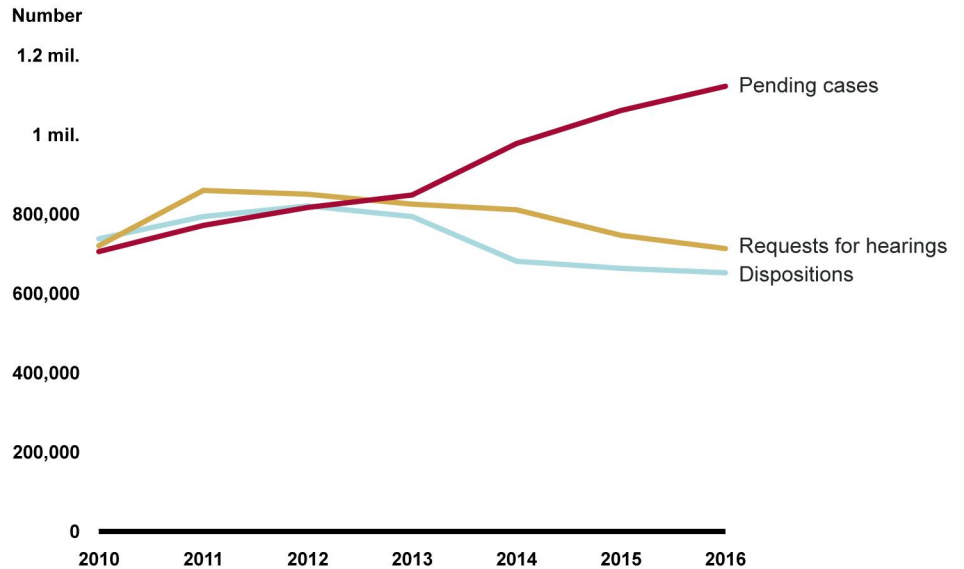
Hearings-level backlogs and processing times have increased between fiscal years 2010 and 2016.²² The number of annual requests for a hearing before an ALJ peaked in fiscal 2011, and declined in each subsequent year, through fiscal year 2016. Despite this decline, SSA has not been able to keep pace with the demand, in terms of dispositions—the number of cases the agency decided or dismissed—in each of those years after 2010 (see figure 3). By the end of fiscal year 2016, SSA reported there were about 1.1 million pending cases. Average processing times for hearings-level decisions also increased during this same time period, from 426 days to 543 days.

²⁰As specified in the Hearings, Appeals, and Litigation Law (HALLEX) manual sections I-2-1-55 and I-3-7-40, remands are generally assigned to the same ALJ who issued the decision or dismissal. However, SSA's Office of the Inspector General (OIG) found that about half the remanded cases in its sample were assigned to a different ALJ than the original ALJ. According to the OIG, this shifts the feedback process away from the ALJ who conducted the hearing and could leave ALJs uninformed of errors they made in deciding cases. SSA OIG, *Oversight of Administrative Law Judge Decisional Quality*, A-12-16-50106 (Baltimore, Md.: March 14, 2017).

²¹SSA, Appeals Council Requests for Review, FY 2016 (For Reporting Purposes: 09/26/2015 through 09/23/2016).

²²In 2015, SSA's OIG found that the agency's efforts to eliminate the hearings backlog have not been effective in eliminating the pending hearings backlog or reducing average processing times; the OIG attributed the lack of effectiveness to four factors: (1) an increase in hearing requests, (2) a decrease in ALJ productivity, (3) a decrease in senior attorney adjudicator decisions, and (4) a recent decrease in the number of available ALJs. In its written comments, SSA noted similar factors, including an increase in hearing requests, an inability to hire sufficient numbers of ALJs and support staff, and years of funding below requested amounts. SSA OIG, *The Social Security Administration's Efforts to Eliminate the Hearings Backlog*, A-12-15-15005 (Baltimore, Md.: Sept. 23, 2015).

Figure 3: Social Security Administration (SSA) Requests for Administrative Law Judge (ALJ) Hearings, Case Dispositions, and Pending Cases, Fiscal Years 2010-2016



Source: GAO analysis of data from SSA's Annual Statistical Supplements, 2013 and 2016. | GAO-18-37

Note: Data from fiscal year 2016 are from SSA's Fiscal Year 2018 Congressional Budget Justification, as data from that year are not yet available from the latest annual statistical supplement, which includes data through fiscal year 2015.

During these years, the number of ALJs declined, along with the number of case dispositions per month. For example, SSA reported it employed 1,356 ALJs in fiscal year 2013, and these judges had an average of 48 case dispositions per month. In fiscal year 2015, 1,265 judges had an average of 44 case dispositions per month.²³ Also during this time period, SSA reduced its reliance on senior attorney adjudicators (SAA) to make fully-favorable, on-the-record decisions (that is, decisions in which a

²³Data from SSA's *Annual Statistical Supplement, 2016*. SSA reported hiring an additional 264 ALJs in fiscal year 2016.

hearing is not necessary because the documentary evidence alone supports a decision that is fully favorable to the claimant).²⁴

According to SSA, its backlog will be eliminated when the national average processing time for a hearing decision is 270 days. In January 2016, SSA issued a plan to achieve this goal by the end of fiscal year 2020.²⁵ However, in its fiscal year 2018 performance plan, SSA set a goal for processing hearings decisions in 600 days (up from a target of 485 days in fiscal year 2010). SSA reported that the increase in average processing times is due to the increase in the number of pending cases. Since SSA generally processes cases in the order in which they are received, they focus on the oldest cases first, which increases the average processing time for closed cases.

Requirements for Hiring, Overseeing, and Disciplining SSA Administrative Law Judges

The role of ALJ was created by the Administrative Procedure Act,²⁶ which was enacted in 1946 to ensure fairness and due process in federal agency proceedings involving rulemaking and adjudications.²⁷ ALJs serve in a number of executive branch agencies, although SSA employs the vast majority. ALJs preside and make decisions at formal adjudicatory proceedings.

One of the primary goals behind the creation of the ALJ position is to ensure that judges can conduct hearings free from influence or coercion

²⁴SSA implemented the Senior Attorney Adjudicators (SAA) Program in November 2007. The program has continued since then and was last extended until February 5, 2018. The number of cases decided by senior attorneys increased in the first years of the program and peaked in fiscal year 2010 with more than 54,000 decisions, before declining in subsequent years. In fiscal year 2014, about 1,800 cases were decided by senior attorneys, according to an SSA OIG report. SSA OIG, *On-the-Record Favorable Decisions Processed at Hearing Offices within 100 Days of Receipt*, A-12-14-14082 (Baltimore, Md.: Jan. 21, 2016).

²⁵SSA, *Leading the Hearings and Appeals Process into the Future: A Plan for Compassionate and Responsive Service* (Baltimore, Md.: January 2016).

²⁶Pub. L. No. 79-404, 60 Stat. 237 (1946). When the APA was enacted in 1946, ALJs were called hearing examiners. This title was changed to administrative law judges in 1978. Pub. L. No. 95-251, 92 Stat. 183 (1978).

²⁷Daniel T. Shedd, *Administrative Law Judges: An Overview*. RL34607. Congressional Research Service. (Washington, D.C.: Apr. 13, 2010).

from the agency.²⁸ Although ALJs are hired by and serve as employees of executive branch agencies like SSA, the Office of Personnel Management (OPM) is responsible for the initial examination, certification for selection, and implementation of the three levels of basic pay of ALJs. As part of its responsibilities, OPM sets the minimum qualifications for ALJs, which are that they generally must be licensed attorneys with a minimum of 7 years of experience in litigation and/or administrative law and pass the competitive examination.

The Administrative Procedure Act gave ALJs qualified decisional independence, with some oversight from agencies. Decisional independence means that ALJs can make decisions independently. Federal law also excludes ALJs from performance evaluations and generally requires that disciplinary actions against ALJs be for good cause established and determined by the Merit Systems Protection Board (MSPB). While ALJs have qualified decisional independence, they must follow their agency's policies and procedures when making decisions. The Administrative Procedure Act also authorized agencies to review ALJ decisions. If SSA determines that an ALJ has not followed its policies and procedures, it can issue a directive to the ALJ to comply and, if that is unsuccessful, bring a disciplinary action before the MSPB.

²⁸Shedd, *Administrative Law Judges*.

Allowance Rates Vary Across Judges, Even for Typical Claims

Allowance Rates Have Varied Across Judges and Hearing Offices in Recent Years, Even After Holding Constant a Range of Factors Relevant to the Appeals Process

Allowance rates varied across administrative law judges from fiscal years 2007 through 2015. We defined the “allowance rate” for each judge as the number of claims in which a judge granted the claimant Disability Insurance (DI) and/or Supplemental Security Income (SSI) benefits divided by the total number of decisions issued by the judge (excluding claims that were dismissed).²⁹ We analyzed about 3.3 million decisions made by administrative law judges on adult Social Security disability appeals over this period.³⁰

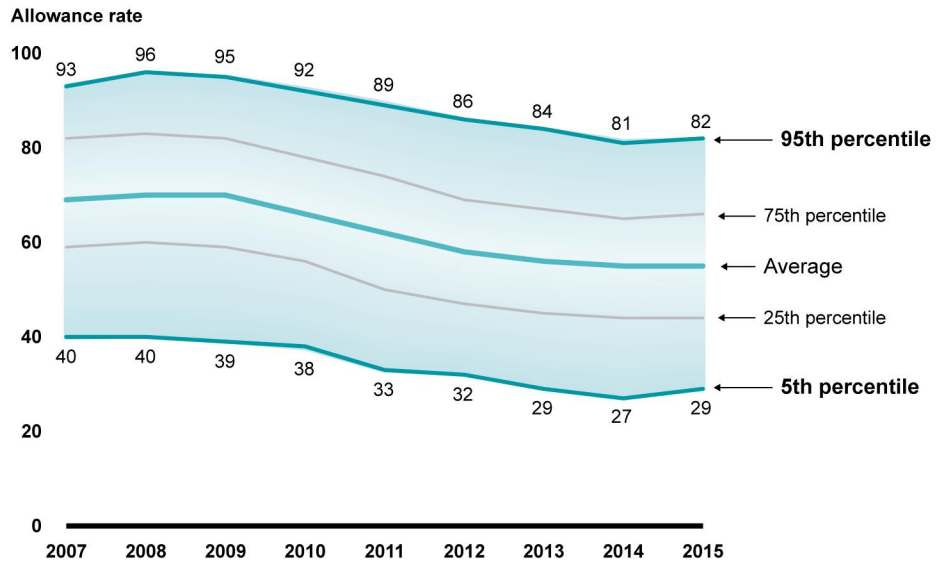
The average allowance rate across judges fell 15 percentage points over this period—from a peak of 70 percent in 2008 to 55 percent in 2015—but the range in allowance rates across judges remained fairly constant (see fig. 4). Specifically, the range—the difference between judges with high allowance rates (those at the 95th percentile) and judges with low

²⁹Claims can be dismissed for reasons not related to the merits of the case, for example, a claimant’s failure to file a timely request or to appear at the scheduled hearing (without good cause). From fiscal years 2007 through 2015, 1,007,526 claims were dismissed. Over this period, between 13 and 18 percent of all dispositions were dismissals. (See appendix I for more details.)

³⁰Our analyses excluded claims that were dismissed, as well as decisions on cases remanded back to a judge from SSA’s Appeals Council; decisions on continuing disability reviews (CDRs), which are reviews SSA periodically conducts to determine if an individual receiving benefits is still disabled; fully favorable decisions made by SSA staff who were not administrative law judges; fully favorable decisions made “on the record” (decisions in which a hearing is not necessary because the documentary evidence alone supports a fully favorable decision); child claims; and non-disability claims. In addition, to ensure that the claims we included were decided by judges with a minimum level of experience, we excluded claims decided within the first year after the judge’s appointment, as calculated by the difference between the date of the judge’s appointment and the date of the decision on each claim. We excluded these decisions to better ensure that variation we identified in allowance rates was not due to the judges’ lack of experience deciding Social Security disability claims. (See appendix I for more details on the criteria we used to select claims for analysis.)

allowance rates (at the 5th percentile)—was 55 percentage points over this period.³¹

Figure 4: Variation in Social Security Administration (SSA) Disability Allowance Rates across Administrative Law Judges, Fiscal Years 2007-2015



Source: GAO analysis of Social Security Administration (SSA) data. | GAO-18-37

Note: We define the “allowance rate” as the number of allowed claims divided by the total number of decisions (claims that are allowed or denied, excluding those that were dismissed) during the period of analysis. Our analyses also excluded decisions on cases remanded back to a judge from SSA’s Appeals Council, decisions on continuing disability reviews (CDRs), fully favorable decisions made by SSA staff who were not administrative law judges, fully favorable decisions made “on the record” (decisions in which a hearing is not necessary because the documentary evidence alone supports a fully favorable decision), child claims, and non-disability claims. In addition, to ensure that the claims we included were decided by judges with a minimum level of experience, we excluded claims decided within the first year after the judge’s appointment, as calculated by the difference between the date of the judge’s appointment and the date of the decision on each claim. We excluded these decisions to better ensure that variation we identified in allowance rates was not due to the judges’ lack of experience deciding Social Security disability claims. See appendix I for more details on the criteria we used to select claims for analysis.

This variation in allowance rates persisted, but fell modestly over time, even when we used multivariate statistical methods to hold constant a variety of factors related to the disability appeals process. These factors included characteristics of claimants, judges, and hearing offices, as well as other factors such as the unemployment rates in a claimant’s state,

³¹Across the middle of the distribution of judges—between judges at the 75th percentile and those at the 25th percentile—the range in allowance rates was 23 percentage points over this period.

that could otherwise explain differences in allowance rates.³² Specifically, for the years 2007 through 2015 combined,³³ our analysis estimated that the allowance rate would vary by 46 percentage points³⁴ for a typical claim, depending on the judge who heard the case.³⁵ For example, we estimated that the allowance rate for a typical claim heard by a judge with low allowance rates would be 42 percent, compared to 88 percent for a judge with high allowance rates. This estimated range fell from 50 percentage points in 2007 to 45 percentage points in 2015 (see fig. 5).³⁶ (Appendix I describes this statistical analysis in more detail.)

³²Our analysis was purely statistical, in that we did not conduct the legal analysis needed to reach conclusions about what legal factors might have affected a judge's decision or whether the decision that was reached in any particular case was correct. Similarly, we are not making any predictions about the correct outcome of future individual decisions. Each case is unique in both its facts and circumstances and must be examined on its own merits. Rather, our multivariate statistical analysis described aggregate associations between various claim characteristics and decisions—specifically, whether allowances were observed to be more or less likely in different circumstances. It did not seek to estimate the causal effects of these characteristics on a judge's decision.

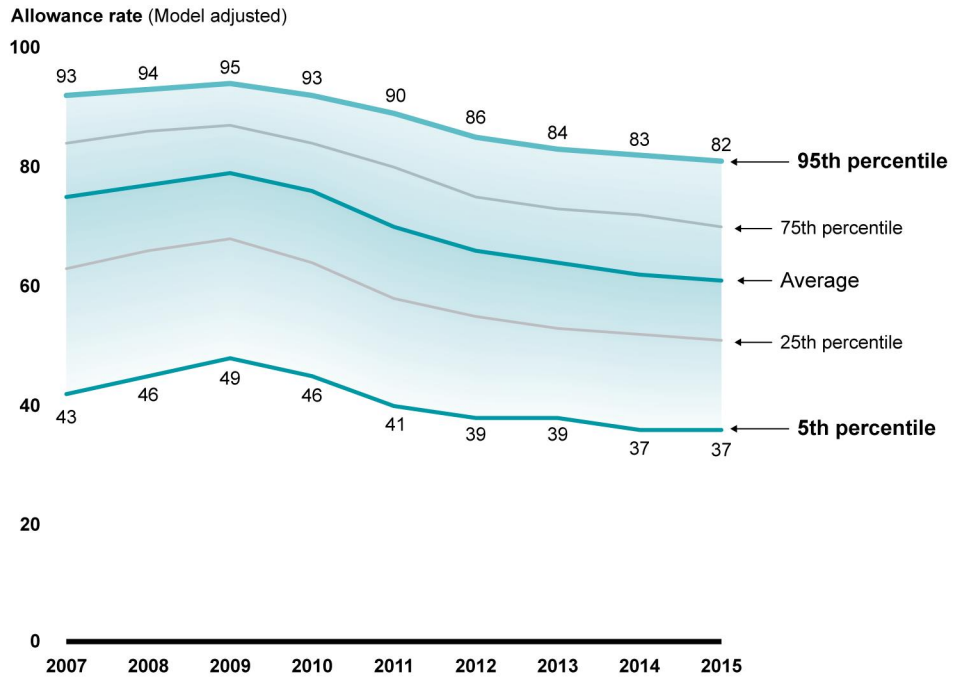
³³Unless otherwise specified, results presented in this report pool cases from all years together.

³⁴Across the middle of the distribution of judges—between judges at the 75th percentile and those at the 25th percentile—the range in allowance rates was 20 percentage points over this period.

³⁵A typical claim had average values on all other factors we analyzed (related to the claimant, judge, other participants in the process, hearing office, and economic characteristics).

³⁶We were unable to determine whether this change was statistically significant (see appendix I).

Figure 5: Estimated Variation in Social Security Administration (SSA) Allowance Rates across Judges for Typical Claims, Fiscal Years 2007-2015



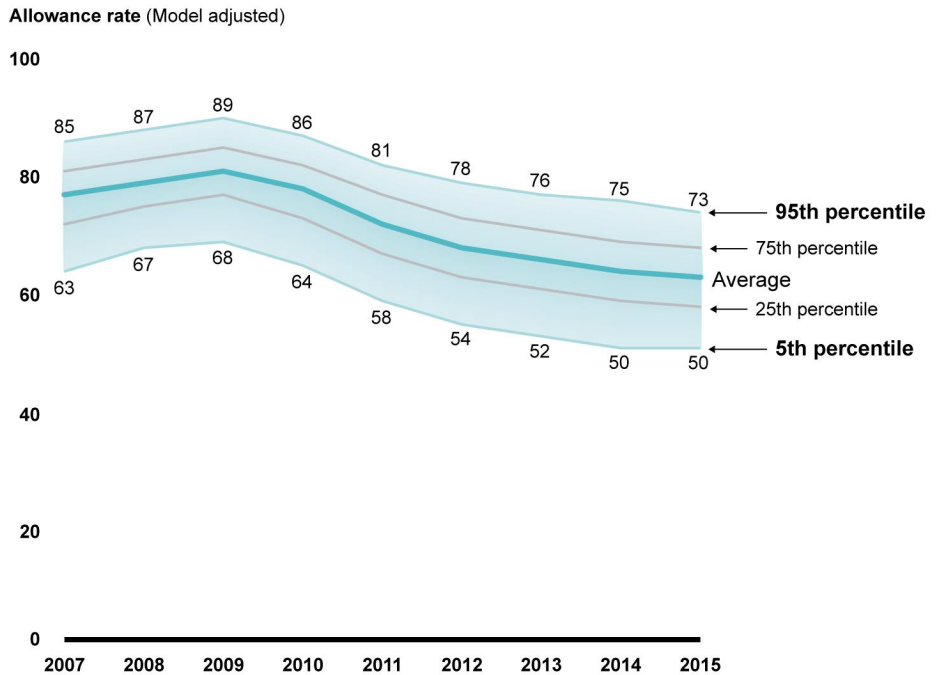
Source: GAO analysis of Social Security Administration (SSA) data. | GAO-18-37

Note: A typical claim had average values on all other factors we analyzed in a multivariate statistical model (related to the claimant, judge, other participants in the process, hearing office, and economic characteristics).

Allowance rates also varied across hearing offices during the same time period, but this variation was considerably smaller than the variation across judges in every year. The estimated range across the entire period was 19 percentage points across hearing offices (see fig. 6),³⁷ compared to a 46 percentage-point estimated range across judges. Accounting for differences in allowance rates across offices ensured that the variation across judges did not reflect characteristics of their offices (such as the types or severity of disability claims received by their offices).

³⁷Across the middle of the distribution of hearing offices—between offices at the 75th percentile and those at the 25th percentile—the range in allowance rates was 8 percentage points over this period.

Figure 6: Estimated Variation in Social Security Administration (SSA) Disability Allowance Rates across Hearing Offices for Typical Claims, Fiscal Years 2007-2015



Source: GAO analysis of Social Security Administration (SSA) data. | GAO-18-37

Note: A typical claim had average values on all other factors we analyzed in a multivariate statistical model (related to the claimant, judge, other participants in the process, hearing office, and economic characteristics).

SSA officials noted that the variation in allowance rates we observed across judges was not surprising, nor was the modest narrowing in this range over time. Administrative law judges usually hear complex appeals that may not be clear-cut allowances or denials. As a result, according to SSA officials, given judges’ decisional independence, different judges could look at cases with similar fact patterns and circumstances and come to different conclusions. At the same time, officials also pointed to several factors potentially related to the modest narrowing in the range of allowance rates. First, they noted that SSA started conducting quality assurance reviews of a random sample of allowances in 2011—previously, such cases were not reviewed. In addition, they said that Social Security’s disability programs and administrative law judges were under increased public and Congressional scrutiny following a high-profile fraud case in 2011 involving a judge and an attorney representative. Further, officials said that the expanding use of electronic case files and data analytics within SSA made it possible for the agency to enhance monitoring of decision-making and share this information with judges.

Finally, while SSA cannot direct judges to decide cases in a particular way, officials suggested that some judges may have “self-corrected” their approach to decision-making, given all of these factors.

Our multivariate analyses had some limitations, but it provides more information than simple comparisons in allowance rates across judges. For example, the SSA data we used for this analysis do not include a measure of the severity of a claimant’s impairment or their remaining ability to work, which could help explain why one claim with a particular impairment was allowed while another was denied. The data also do not include a standardized measure for the nature of claimants’ prior work (such as the skill level or extent of physical labor), which is also relevant for the disability decision. Nevertheless, our multivariate analysis enabled us to compare allowance rates across judges and hearing offices for typical claims.³⁸ In addition, SSA’s practice of assigning cases randomly to judges makes it more likely that the remaining variation we found across judges reflects the unique effect of having a particular judge hear a case, rather than other factors.³⁹ As a result, even though we could not account for all factors that could explain differences in allowance rates, random assignment increases the chances that such factors were similar across all of the cases heard by individual judges.

³⁸The unadjusted range in allowance rates was about 55 percentage points across judges and 29 percentage points across hearing offices, compared to 46 percentage points and 19 percentage points in the multivariate analysis.

³⁹Although there are some exceptions to the random assignment of cases to judges, we analyzed the subset of cases that should have been assigned to judges at random, in accordance with SSA policy. (See appendix I for more information about the random assignment of cases to judges.)

Numerous Factors, Particularly Those Representing SSA's Disability Criteria, Are Associated with Variation in Allowance Rates

Although variation in allowance rates persisted across judges, even after controlling for certain factors, many of the factors we identified had meaningful associations with the chance that a claimant was allowed benefits.⁴⁰ These factors represent criteria in SSA's disability decision-making process, such as the claimant's age, impairment, prior work, and education. We also identified factors that did not have such associations.⁴¹

Claimant Characteristics

Certain claimant characteristics—such as older ages or certain impairments—were associated with higher allowance rates.

- **Age:** Claimants' chances of being allowed benefits increased with age, even holding constant other factors. For example, a 55-year-old claimant was allowed benefits at a rate 4.3 times higher than a typical 35-year-old claimant.⁴² This association is consistent with Social Security's vocational guidelines, which are generally more lenient for older claimants. As part of SSA's five-step process to determine eligibility for adult disability benefits, SSA uses a set of rules to evaluate how a claimant's age, education, and work experience affect their remaining capacity for work. SSA's criteria vary across four primary age groups—45-49, 50-54, 55-59, and 60 and older. The

⁴⁰We defined a "meaningful" association as one that: (1) was statistically distinguishable from 1 at the 95-percent confidence level, and (2) implied at least a 10-percent difference in the odds of an allowance across groups of claimants.

⁴¹We analyzed these factors for a typical claim that had average values on all other factors we analyzed (related to the claimant, judge, other participants in the process, hearing office, and economic characteristics). As noted previously, we did not conduct a legal analysis to determine what legal factors might have affected a judge's decision or whether the decision that was reached in any particular case was correct. Rather, our multivariate statistical analysis described aggregate associations between various claim characteristics and decisions—specifically, whether allowances were observed to be more or less likely in different circumstances. It did not seek to estimate the causal effects of these characteristics on a judge's decision.

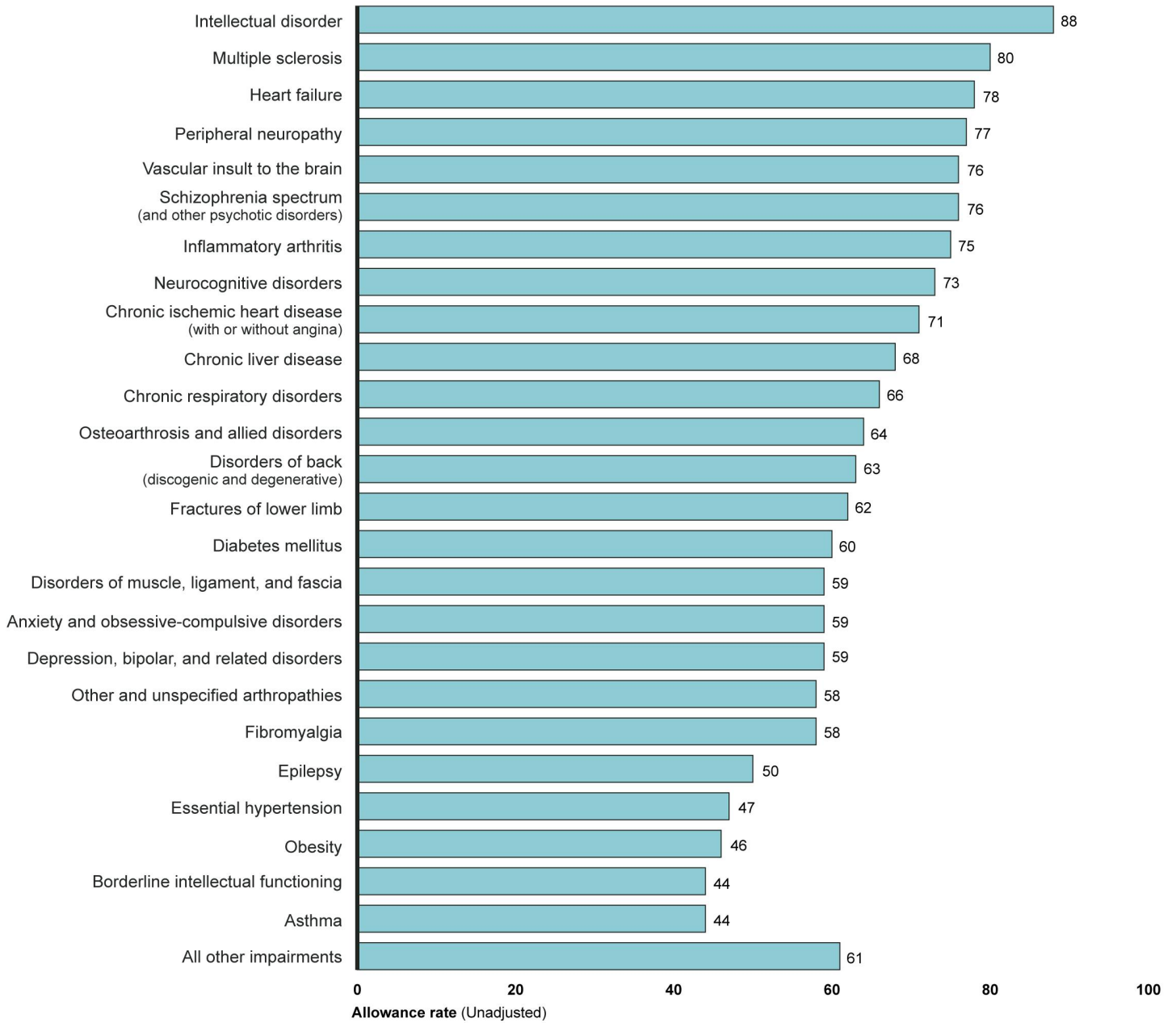
⁴²In this section of the report, we are comparing the odds of an allowance between two or more groups, using an analysis that pools cases across all of the years of our analysis. (See appendix I for more information.)

criteria are less stringent for claimants in older age groups than they are for younger claimants, because the rules assume that individuals at older ages may be less able to transition to other work.

- **Impairment:** Certain impairments were also strongly associated with the chance of being allowed benefits (see fig. 7). For example, claimants with primary impairments⁴³ recorded in SSA's data of heart failure or multiple sclerosis were allowed benefits at rates 4.2 and 5 times higher, respectively, than typical claimants with asthma. From fiscal years 2007 through 2015, the allowance rates for claimants with heart failure or multiple sclerosis were 78 and 80 percent, respectively, compared to 44 percent for asthma.

⁴³The primary impairment, or diagnosis, is the basic condition that the judge determined rendered the person disabled or evidence shows has the most significant effect on the person's ability to work, according to SSA program rules.

Figure 7: Social Security Administration (SSA) Disability Allowance Rates for the 25 Most Commonly Recorded Primary Impairments, Fiscal Years 2007-2015



Source: GAO analysis of Social Security Administration (SSA) data. | GAO-18-37

Note: The figure shows the allowance rates for the 25 most common primary impairment codes that altogether account for 79 percent of all claims; the remaining 21 percent of claims are spread across

227 additional impairment codes. The allowance rates shown above are the raw, or unadjusted, allowance rates for the claimants in our study population.

The primary impairment, or diagnosis, is the basic condition that the judge determined rendered the person disabled or evidence shows has the most significant effect on the person's ability to work, according to SSA program rules. For more information on primary impairments, see SSA's Program Operations Manual System Section DI 26510.015

<https://secure.ssa.gov/apps10/poms.nsf/lnx/0426510015>. For more information on SSA's Listing of Impairments, see <https://www.ssa.gov/disability/professionals/bluebook/listing-impairments.htm>.

- **Critical or terminal case:** Claimants with critical or terminal cases were allowed benefits at a rate 1.4 times higher than a typical claimant without a critical or terminal case. Critical and terminal cases are cases that require special processing, such as a terminal illness or a veteran with a 100-percent permanent and total disability compensation rating.
- **Prior work:** Claimants reporting shorter work histories (4 years or less in the last 15 years before applying for disability benefits) were allowed at a rate 0.8 times as high as a typical claimant with 10 or more years of work history. As expected, given the nature of the work requirements for the DI program, the association with prior work history was stronger for that program than for the SSI program.⁴⁴
- **College education:** Claimants who reported having a college-level education or higher were approved at a slightly higher rate (1.1 times higher) than a typical claimant with a high-school education. SSA officials suggested that this association could be an indirect measure of the severity of a claimant's impairment, a factor for which we did not have data. They said that individuals with higher levels of education often have higher incomes and, therefore, may be less likely to forego their income to apply for disability benefits, were it not for the severity of their disability.
- **Claim type:** DI claimants were allowed at a rate 1.7 times higher than a typical SSI claimant. Across judges, the average allowance rate for DI claimants (67 percent) was higher than for SSI claimants (52 percent) from fiscal years 2007 through 2015, with the allowance rate for claimants applying concurrently for DI and SSI benefits falling in between (58 percent).⁴⁵

⁴⁴Specifically, in order to be eligible for the DI program, individuals must meet requirements for the time they have worked in "covered employment," that is, jobs through which they have paid Social Security taxes. In general, workers over 30 years old must have worked in 5 of the last 10 years before becoming disabled.

⁴⁵These are the raw, or unadjusted, allowance rates for the claimants in our study population.

Other Participants in the Disability Appeals Process

Claimants who had appointed a representative to present their case, or had a medical expert testify at their hearing, were associated with a greater chance of being allowed benefits, but the presence of a vocational expert had the opposite association.

- **Claimant representative:** Similar to findings in our prior work, claimants who had a representative—either an attorney or a nonattorney representative—were allowed at a rate 2.9 times higher than a typical claimant with no representative.⁴⁶ SSA officials stated that representatives may have a screening process for potential clients, and under SSA’s fee structure, representatives are paid only if the claimant is awarded benefits. As a result, representatives may tend to take cases they believe will be successful. Officials also stated that a representative can help the claimant by ensuring that the medical evidence and other records are fully developed and help the claimant present their case at a hearing.⁴⁷ From fiscal years 2007 through 2015, most claimants (77 percent) had an attorney representative, and 12 percent had a nonattorney representative.
- **Expert testimony:** Claimants whose hearings involved testimony from a medical expert were allowed at a rate 1.6 times higher than a typical claimant without a medical expert present.⁴⁸ Medical experts include physicians, psychologists, and other types of medical professionals who provide impartial, expert opinion evidence for an ALJ to consider when making a decision about disability. SSA officials said that the association of medical experts with an increased chance of allowance is expected, given that judges are required to seek the testimony of a medical expert in certain cases, for example, when the judge is considering allowing benefits because the claimant’s impairment may be medically equivalent to one in SSA’s Listing of

⁴⁶Our prior work found that attorney representation increased a claimant’s likelihood of being awarded Social Security disability benefits. See GAO, *SSA Disability Decision-Making: Additional Steps Needed to Ensure Accuracy and Fairness of Decisions at the Hearings Level*, [GAO-04-14](#). (Washington, D.C.: Nov. 12, 2003).

⁴⁷We could not use our analysis to determine whether any of these explanations for the overall effect of having a representative were valid.

⁴⁸When we analyzed a subset of cases decided at the final two steps in the disability decision-making process, the presence of a medical expert did not have a substantial association with allowances. This likely reflects medical experts’ lack of a critical role in decision-making at these steps (compared to earlier steps).

Impairments. In other cases, involving a medical expert is generally at the judge's discretion. From fiscal years 2007 through 2015, 12 percent of decisions involved a medical expert.

The presence of a vocational expert had the opposite effect—claimants with a vocational expert testifying were allowed at a rate 0.8 times as high as claimants without a vocational expert testifying. Vocational experts provide objective, expert opinion evidence to the ALJ, primarily at the last two steps of the disability decision-making process where SSA considers whether claimants can do their prior work or transition to other work available in the national economy.⁴⁹ Although involving a vocational expert is generally at a judge's discretion,⁵⁰ SSA officials said that they were not surprised by this result, because vocational experts are usually called upon at the final two steps in the disability decision-making process. At that point, claimants had already not been allowed benefits at an earlier step because their impairment(s) did not meet or were not equivalent to an impairment in SSA's listings. From fiscal years 2007 through 2015, most hearings (85 percent) involved a vocational expert.

Judge Characteristics

Judges with certain characteristics, such as those appointed in earlier years, were associated with a greater chance of allowing benefits.

- **Appointment cohort:** A claimant whose claim was heard by a judge appointed between 1995 and 1999 was allowed at a rate 1.5 times higher than a typical claim heard by a judge appointed after 2010. SSA officials said that, since 2010, they have changed the way they train and mentor new judges, and introduced new tools to help provide a standardized decision-making template. As a result, SSA officials said, more recently hired ALJs may be more aware of agency policies and procedures.

⁴⁹When we analyzed a subset of cases decided at the final two steps, the effect of having a vocational expert was about the same.

⁵⁰The ALJ must obtain testimony from a vocational expert if the Appeals Council or a court ruling requires it. Otherwise, involving a vocational expert is at the judge's discretion.

Other Factors

Certain characteristics of hearing offices and other factors also were associated with higher chances of allowance. For example:

- **Hearing type:** Claimants whose hearings were held in person were allowed at a slightly higher rate (1.1 times higher) than a typical claimant with a hearing conducted remotely using videoconference technology.⁵¹ This is equivalent to a 2.8 percentage-point higher probability of being allowed benefits for a claimant whose hearing was held in person, compared to an otherwise typical claimant whose hearing was conducted by videoconference. However, we did not seek to estimate the causal impact of videoconferences on allowance rates, and so did not design our analysis to account for all factors that could affect this relationship. Rather, we accounted for the use of videoconferences solely to further ensure that circumstances were similar across the judges and offices we analyzed. Expanding video service delivery is a key goal for SSA, including plans to partner with other agencies, such as the Department of Veterans Affairs, to increase the number of available video hearing sites beyond those already available at hearing offices and the five National Hearing Centers.
- **Year of decision:** Claimants whose appeals were decided in earlier years were associated with a greater chance of being allowed benefits. While this trend is similar to the raw change over time shown in figure 4, our multivariate analysis showed that this change held even for claimants in similar circumstances. For example, claimants who received decisions in 2007 were allowed at a rate 2.0 times higher than a typical claim in 2015. This is consistent with other

⁵¹Our analysis controlled for the type of hearing office that made the decision, including the five National Hearing Centers that conduct hearings exclusively by videoconference. Therefore, this effect applies to both National Hearing Center hearings as well as the approximately 19 percent of cases decided at regular hearing offices in which the hearing was conducted by videoconference.

studies that have found trends of lower allowance rates in recent years.⁵²

Factors Not Associated with Differences in Allowance Rates

Some factors were not meaningfully associated with allowance rates when holding other factors constant.

- **Workload measures:** Workload and productivity measures at the hearing office and judge level were not meaningfully associated with allowance rates. This includes the annual percentage of cases that were backlogged (that is, awaiting a judge's decision for more than 270 days) at each hearing office, as well as the annual number of dispositions (decisions plus dismissals) each judge issued.⁵³ This may suggest that judges' decisions to allow or deny cases are not significantly influenced by the number of cases before them, similar to findings in prior research.⁵⁴
- **Hearing office type:** We found no meaningful differences in allowance rates between similar claims heard at one of SSA's National Hearing Centers or a traditional hearing office, after holding constant other factors (including whether the hearing was held by videoconference). SSA has five National Hearing Centers, which hear cases from across the country by videoconference in order to reduce backlogs in certain hearing offices.
- **Economic characteristics:** The unemployment and poverty rates in the claimant's state at the time of the ALJ decision were not associated with allowance rates. Higher unemployment rates can result in increased applications for Social Security disability benefits

⁵²See Social Security Advisory Board, 2015 Technical Panel on Assumptions and Methods, Report to the Social Security Advisory Board (September 2015). The panel found evidence that the downward trend in DI allowance rates is greater than what is attributable to the business cycle alone, and called for further monitoring and study of the factors that contribute to the decline in allowance rates. See also Robert Nakosteen and Michael Zimmer, "Approval of Social Security Disability Appeals: Analysis of Judges' Decisions," *Applied Economics*, Vol. 46, No. 23 (2014): 2783-2791.

⁵³We considered some additional administrative factors but did not include them in our final model. For example, we did not use judge's average processing time, which we found was moderately correlated with the percent of cases that were backlogged within that judge's hearing office.

⁵⁴Nakosteen and Zimmer, "Approval of Social Security Disability Appeals: Analysis of Judges' Decisions."

because workers with impairments that could qualify them for the program who experience job loss may find it more difficult to become re-employed during periods of high unemployment and apply for benefits.⁵⁵ However, the impact on allowance rates in the research we reviewed is mixed.⁵⁶

⁵⁵Social Security Administration, Office of the Chief Actuary. "Disabled Worker Allowance Rates: Variation Under Changing Economic Conditions," Actuarial Note No. 153 (Baltimore, Md.: August 2013).

⁵⁶One study we reviewed found that unemployment rates at the time of the ALJ decision were positively associated with greater chance of allowance. See Nakosteen and Zimmer, "Approval of Social Security Disability Appeals: Analysis of Judges' Decisions." Other studies have measured unemployment rates at the time the claimant applied for benefits and found unemployment rates were negatively associated with allowance rates. See Social Security Administration, Office of the Chief Actuary. "Disabled Worker Allowance Rates." See also Nicole Maestas, Kathleen J. Mullen, and Alexander Strand. "Disability Insurance and the Great Recession." *American Economic Review*, Vol. 105, No. 5 (2015): 117-182.

SSA's Efforts to Monitor Accuracy and Consistency of Hearings-Level Decisions Lack Performance Measures and Have Not Been Evaluated

SSA Has Timeliness Measures, but Lacks Public Performance Measures for Accuracy and Consistency

SSA has employed a range of efforts to monitor the accuracy and consistency of hearings decisions, but it lacks performance measures to report publicly on these efforts. SSA's current strategic plan includes an objective to "improve the quality, consistency, and timeliness" of its disability decisions; however, all of the hearings-level measures supporting this objective are related to timeliness. In a previous report, we developed nine attributes of performance goals and measures based on previously established GAO criteria, as well as relevant federal laws and performance management literature.⁵⁷ One key attribute states that an agency's suite of performance measures should be balanced to cover various priorities. In addition, each measure should cover a priority such as quality, timeliness, and cost of service. However, because SSA's performance measures do not fully reflect its goals, the overall success of SSA's efforts in this area may be limited.

SSA previously had performance measures related to hearings-level accuracy, which used data from ALJ peer reviews. These measures were discontinued in fiscal year 2009, when the ALJs conducting the reviews were reassigned to hearing cases. By comparison, SSA continues to have a measure for accuracy at the initial decision-making level (see table 1).

⁵⁷GAO, *Tax Administration: IRS Needs to Further Refine Its Tax Filing Season Performance Measures*, [GAO-03-143](#) (Washington, D.C.: Nov. 22, 2002).

Table 1: Social Security Administration (SSA) Performance Measures for Disability Decisions, Fiscal Years 2017-2018

SSA Goal 3: Serve the Public through a Stronger, More Responsive Disability Program

Strategic Objective 3.1:
Improve the quality, consistency and timeliness of disability decisions

Performance measures	Timeliness	Accuracy
Ensure the quality of our decisions by achieving the Disability Determination Services (DDS) decisional accuracy rate for initial disability decisions	Does not have	Has
Increase our ability to provide timely decisions by reducing the percentage of pending Appeals Council requests for review 365 days or older	Has	Does not have-
Improve customer service by reducing the wait time for a hearing decision ^a	Has	Does not have
Complete the budgeted number of initial disability claims ^b	Has	Does not have
Complete the budgeted number of disability reconsideration claims ^b	Has	Does not have
Complete the budgeted number of hearing requests ^b	Has	Does not have
Average processing time for initial disability claims ^b	Has	Does not have
Average processing time for reconsiderations ^b	Has	Does not have
Average processing time for hearing decisions ^b	Has	Does not have

Legend: X=agency has a performance measure; - = agency does not have a performance measure.

Source: GAO analysis of SSA Annual Performance Plan, Fiscal Years 2017-2018. | GAO-18-37

^aAgency Priority Goal

^bBudgeted Workload Measure

SSA officials stated that they have no plans to add new performance measures related to the accuracy and consistency of hearings decisions to the strategic plan. They said that while they collect and monitor a wide variety of workload and performance measures for day-to-day operations, they have to select a few, representative measures that are meaningful to stakeholders and represent agency-wide efforts to achieve its goals. They stated that the current performance measures meet these requirements.

Although SSA officials said the agency does not publicly report performance measures related to the accuracy and consistency of hearings decisions, they said that SSA uses internal performance measures related to hearings decisions. However, these internal measures to monitor quality and consistency of hearings decisions have limitations and are not shared with the public. Regional chief judges—who oversee the hearing offices and judges within each of SSA’s 10 regions—

and others told us that they use a measure known as the “agree rate”⁵⁸ to help monitor the quality of a judge’s decisions. This measure is based on the number of cases that have been appealed to the Appeals Council by the claimant or representative as a request for review. The agree rate reflects the percentage of cases in which the Appeals Council—the final level of appeals within SSA—concluded that the ALJ’s decisions were supported by substantial evidence and contained no error of law or abuse of discretion. However, the agree rate has some limitations. For example, as noted earlier, it does not reflect the accuracy of ALJ decisions that the claimant did not appeal. SSA’s Office of the Inspector General (OIG) found that this measure provided information on less than one-quarter of all ALJ dispositions and it is not representative of the ALJ’s entire workload because it is based only on Appeals Council reviews of appealed cases.⁵⁹ In addition, a March 2017 SSA OIG report found that SSA has not maintained historical data on agree rates, limiting the agency’s ability to analyze agree rate trends.

SSA uses other internal measures to track consistency. For example, SSA developed an internal early monitoring system that tracks 22 metrics of ALJ performance to identify outliers. For example, three of these metrics (average number of dispositions a judge issues per day, agree rate, and allowance rate) have “alarm thresholds” to indicate when an ALJ’s metrics fall outside of a given threshold. Based on these findings, SSA may conduct a focused quality review (a type of quality assurance review) to ensure the judge’s decisions complied with SSA policies, or follow up with the regional chief judge to determine if additional policy guidance or training is needed.

Although these internal measures are helpful for management to monitor and improve accuracy and consistency, without sharing this or similar

⁵⁸Claimants who are dissatisfied with their decision at the hearings level may appeal their cases to the Appeals Council, which is the final step in the administrative appeals process. According to SSA, the Appeals Council reviews the claim to determine whether any legal or procedural errors were committed and all evidence was properly considered. To calculate an agree rate for each ALJ, ODAR tracks the Appeals Council’s decision on every claim. The agree rate is defined as the percentage of cases that were denied review (i.e., cases in which the Appeals Council concluded that the ALJ’s decisions were supported by substantial evidence and contained no error of law or abuse of discretion justifying a remand or reversal) compared to the total number of cases considered by the Appeals Council.

⁵⁹SSA OIG, *Oversight of Administrative Law Judge Decisional Quality*, A-12-16-50106 (Baltimore, Md.: Mar. 14, 2017).

information publicly, SSA lacks accountability for improving the quality of hearings-level decisions. In addition, federal internal control standards state that management should externally communicate the necessary quality information to achieve objectives, including to external stakeholders such as Congress and the public.⁶⁰ Further, given the persistent variation in allowance rates, SSA may be missing an opportunity to provide the public with information on the results of its efforts to improve the accuracy and consistency of disability decisions.

SSA Has Efforts to Monitor and Improve Accuracy and Consistency of Hearings-Level Decisions, but Quality Reviews May Overlap and Have Not Been Systematically Evaluated

SSA provides training and tools to all ALJs and initiates disciplinary actions where needed, as part of its efforts to monitor and improve accuracy and consistency. SSA also conducts multiple quality assurance reviews, but some of these reviews may overlap and SSA has not evaluated them.

Training, Tools, and Policy Guidance

ALJs receive ongoing training and guidance from several sources, including through judicial trainings, mentoring, and policy memorandums. In 2006, SSA implemented a three-phase training program for new ALJs, which includes training on core competencies as well as a formal mentoring program in which new ALJs are paired with experienced ALJs for regular sessions over a nine-month period. Regional managers, judges, and stakeholders we spoke with had positive feedback on the training SSA provides to judges. For example, officials from one stakeholder group told us that they believe training had created more consistency in allowance rates. SSA's chief judge also issues guidance memorandums to clarify policies related to the hearings process. For example, in July 2013, SSA issued a memorandum establishing expectations for the instructions judges provide to decision writers, who are SSA staff who prepare the draft decisions. SSA officials said that they issued the memorandum in response to an ALJ who was providing low-quality instructions to decision writers and SSA realized it had not

⁶⁰GAO, *Standards for Internal Control in the Federal Government*, [GAO-14-704G](#) (Washington, D.C.: Sept. 10, 2014).

provided formal guidance on the topic. In addition, ALJs also receive quarterly continuing education training and have a library of reference materials and on-demand video courses to use as needed.

SSA also uses internal metrics and provides electronic tools to judges to monitor and improve accuracy and consistency. Regional chief judges regularly review management information (MI) reports and develop strategies, such as recommending training, to address identified issues. Beginning in 2011, SSA established an electronic tool called “How MI Doing?”, which allows ALJs to compare their productivity and timeliness metrics to hearing office, regional, or national metrics. The tool also provides data on the agree rate for each judge as well as the hearing office, regional, and national agree rates. Using this tool, judges can also learn the reasons any prior decisions have been remanded, and access on-demand training pertaining to that reason. Regional chief judges we spoke with generally found “How MI Doing?” to be a helpful tool, although SSA does not track judges’ usage and has not formally evaluated its effectiveness. In addition, SSA established the electronic Bench Book (eBB), which is designed to assist users with documenting, analyzing, and making consistent and accurate decisions on hearings-level adult disability cases. However, the SSA OIG recently recommended that SSA evaluate eBB and determine whether to continue it. Regional chief judges we spoke with provided mixed feedback on the use of eBB and its usefulness for ALJs. In fiscal year 2016, nearly 500 ALJs (about one-third) used eBB. In June 2017, SSA officials said that while no formal evaluation of eBB was conducted, they recently received approval to proceed with plans to replace eBB with a similar tool as part of updates to SSA’s case management system.

Disciplinary Action

SSA also addresses identified issues with the accuracy and consistency of hearings decisions by taking disciplinary actions, as needed. SSA can take non-disciplinary or disciplinary action to address performance concerns. Non-disciplinary actions include training and counseling (known as “collegial conversations”). Another non-disciplinary action is a written directive, which SSA can issue to individual judges to improve performance on workload, scheduling or policy compliance.⁶¹ From 2007

⁶¹When ALJ conduct or performance issues arise, federal law precludes SSA from using some tools applicable to the majority of federal employees. ALJs possess a degree of independence under federal law, which provides certain separations of functions that protect them from any influence over the outcome of cases before ALJs.

through 2016, SSA issued about 1,330 such directives. Nearly all (95 percent) were issued to improve timeliness, while about 2 percent were issued to improve policy compliance. If an ALJ's conduct or performance does not change or becomes more egregious, SSA continues with progressive discipline including reprimand or seeking disciplinary action from the Merit Systems Protection Board, such as short- or long-term suspension or removal. From 2007 through 2016, there were 98 reprimands, 34 proposed suspensions, and 16 proposed removals, according to SSA.⁶²

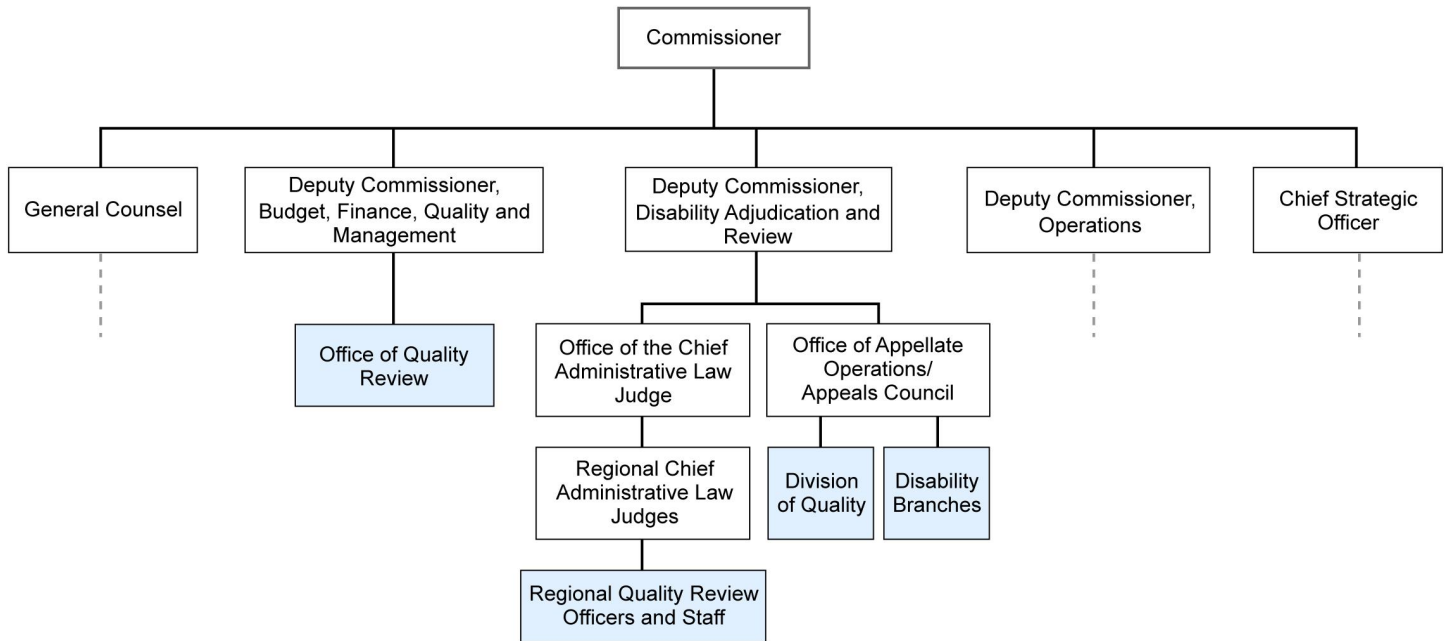
Quality Assurance Reviews

SSA conducts various quality assurance reviews to improve accuracy and consistency. SSA officials stated that the agency has been enhancing its quality review efforts since 2009. Since then, it has added five types of quality assurance reviews that are conducted by three additional offices within SSA (see fig. 8).⁶³

⁶²An ALJ may be removed or disciplined only for good cause established and determined by the Merit Systems Protection Board (MSPB) after an opportunity for a hearing before the Board. In half of these proposed removals (8 of 16), MSPB issued a removal decision. Six judges resigned or retired before MSPB issued a decision, and 2 judges retired pursuant to a settlement agreement.

⁶³SSA did not conduct the Disability Case Review in fiscal years 2016 or 2017. Officials stated that this review will be resumed in fiscal year 2018, and will be called the Hearings Level Review.

Figure 8: Four Social Security Administration (SSA) Offices Responsible for Quality Assurance Reviews of Hearings-level Decisions, as of September 2017



Source: GAO analysis of Social Security Administration (SSA) information. | GAO-18-37

Note: Figure shows selected offices from SSA's organization chart. The four offices whose staff carry out quality assurance reviews of hearings-level decisions are shaded. In addition to the Office of Quality Review, Division of Quality and regional quality review officers and staff, SSA also measures the quality of a judge's decisions through the Appeals Council Disability Branches' ongoing review of appealed cases. This review results in the agree rate, which reflects the percentage of cases in which the Appeals Council concluded that the ALJ's decisions were supported by substantial evidence and contained no error of law or abuse of discretion.

Effective October 1, 2017, SSA created a deputy commissioner-level Office of Analytics, Review and Oversight. The Office of Quality Review and Office of Appellate Operations were folded into this office.

SSA added quality assurance reviews for various reasons. For example, in 2009, SSA's regional staff under the Office of the Chief Administrative Law Judge began conducting regional inline quality reviews, which involve assessing the extent to which hearing office staff are processing cases and preparing them for hearings in accordance with SSA policy, as well as the policy compliance and legal sufficiency of the draft decision. SSA added this review to enhance its reviews of decisions before they are issued, in an effort to reduce remands. Also in 2009, SSA's Office of Quality Review began conducting disability case reviews to provide feedback on decision-making accuracy to ALJs. In addition, in 2010, SSA created the Division of Quality under the Appeals Council, a unit focused on conducting reviews on a regular basis of decisions that claimants did

not appeal.⁶⁴ Prior to 2010, SSA generally only reviewed decisions that claimants appealed through the Appeals Council.

While these quality assurance reviews have somewhat different focuses—for example, some assess aspects of how a case was processed while others review the accuracy of the decisions—they overlap in two key ways. According to prior GAO work, overlap occurs when multiple agencies or programs have similar goals, engage in similar activities or strategies to achieve them, or target similar beneficiaries.⁶⁵ Some of SSA's quality review efforts fit the description of overlap in that they have similar goals and review similar cases. For example:

- **Similar goals:** Several of the reviews have similar goals (see table 2). For example, two of the four entities conducting reviews—the Appeals Council's Division of Quality and staff in SSA's 10 regional offices—both review decisions for policy compliance before those decisions go into effect (known as pre-effectuation reviews). While one review looks at the judge's decision and the other looks at the draft decision prior to the judge's review and approval, according to officials and documents we reviewed, these reviews share similar goals: to guide training and provide feedback to judges. In addition, all the reviews are designed to assess compliance with SSA policy.

⁶⁴The Appeals Council has authority to take own motion review of cases before a hearing decision is effectuated and benefits are paid (pre-effectuation) and it does not identify cases based on the identity of any specific decision maker or hearing office.

⁶⁵GAO, *Fragmentation, Overlap, and Duplication: An Evaluation and Management Guide*, [GAO-15-49SP](#) (Washington, D.C.: April 14, 2015).

Table 2: Comparison of Goals of Social Security Administration (SSA) Quality Reviews of Hearings-level Decisions

Review Entity	Type of Review (Year Established)	Goal or Purpose of Review			
		Policy compliance review ^a	Legal review	Guide training, provide feedback to judges	Prevent or reduce reReview does not include goalwork or remands
Office of Appellate Operations (OAO), Appeals Council, Disability Branches	Requests for review (1940) ^b : Reviews of decisions appealed by claimants. Conducted continuously.	Review includes goal	REVIEW INCLUDES GOAL	REVIEW INCLUDES GOAL	Review does not include goal
Regional Quality Review Officers	Regional inline reviews (2009): Reviews of a random sample of cases in various stages of preparation: Ready for a hearing Draft decision Conducted continuously.	REVIEW INCLUDES GOAL	REVIEW INCLUDES GOAL	REVIEW INCLUDES GOAL	REVIEW INCLUDES GOAL
Office of Quality Review	Disability case reviews (2009): Review of a nationally representative sample of allowance and denial decisions. Conducted annually.	REVIEW INCLUDES GOAL	REVIEW INCLUDES GOAL	REVIEW INCLUDES GOAL	Review does not include goal
OAO, Appeals Council, Division of Quality	<p>Focused quality reviews (2011): Reviews of specific subjects, including but not limited to hearing offices, individual judges, subject matters, policies, or procedures. Conducted biReview does not include goalmonthly (approReview includes goal.).</p> <p>Random preReview does not include goaleffectuation reviews^c (2011): Review of a random sample of favorable hearing decisions. Conducted continuously.</p>	REVIEW INCLUDES GOAL	Review does not include goal	REVIEW INCLUDES GOAL	Review does not include goal
		REVIEW INCLUDES GOAL	REVIEW INCLUDES GOAL	REVIEW INCLUDES GOAL	REVIEW INCLUDES GOAL ^d

Review Entity	Type of Review (Year Established)	Goal or Purpose of Review			
		Policy compliance review ^a	Legal review	Guide training, provide feedback to judges	Prevent or reduce reReview does not include goalwork or remands
	Selective preReview does not include goaleffectuation reviews ^c (2014) Review of a random sample of cases with a common characteristic that increases the likelihood of error. Conducted annually.	REVIEW INCLUDES GOAL	REVIEW INCLUDES GOAL	REVIEW INCLUDES GOAL	REVIEW INCLUDES GOAL ^d

Legend: X = Review includes this goal or purpose; - =Review does not include this goal or purpose.

Source: GAO analysis of SSA documents and interviews with agency officials. | GAO-18-37

^aReviews determine compliance with SSA policy according to the Social Security Act, regulations, Social Security rulings, and Hearings, Appeals, and Litigation Law (HALLEX) manual, according to SSA officials.

^bThis predates the DI and SSI programs. However, according to SSA, in March 1940, an Office of Appeals Council was established and given the authority to review hearings on claims for Old-Age and Survivors Insurance benefits.

^cSSA noted that these reviews are not permitted to identify cases based on the identity of the decision-maker or the identity of the office issuing the decision.

^dSSA noted that these pre-effectuation reviews can sometimes result in remands.

- Similar cases:** SSA’s five quality assurance reviews look at similar cases, and could potentially include the same cases (see table 3). SSA takes some steps to prevent assessing the same claim in multiple quality assurance reviews. Officials told us that, in conducting focused quality reviews (conducted after the decision is final), they exclude cases that were reviewed in a pre-effectuation review. However, they said that the Division of Quality does not know whether cases it has selected were also subject to a regional inline quality review. They said that additional efforts to prevent multiple reviews of a case are manual in nature, and thus there is still the potential for claims to be reviewed more than once. Further, SSA officials said they did not see a need to prevent multiple reviews of a case, in particular, because some reviews are conducted before the decision is final and others are conducted after the decision is final.

Table 3: Comparison of Cases Reviewed by Social Security Administration (SSA) Quality Reviews of Hearings-level Decisions

Review Entity	Type of Review	Cases Reviewed			Before or After Decision is Final?
		Allowances	Denials	Dismissals	
Office of Appellate Operations (OAO), Appeals Council, Disability Branches	Requests for review	Partially reviewed	Reviewed	Reviewed	After
Regional Quality Review Officers	Regional inline reviews	Reviewed	Reviewed	Reviewed	Before
Office of Quality Review	Disability case reviews	Reviewed	Reviewed	Did Not Review	After
OAO, Appeals Council, Division of Quality	Focused quality reviews	Reviewed	Reviewed	Reviewed	After
	Random pre-effectuation reviews	Reviewed	Did Not Review	Did Not Review	Before
	Selective pre-effectuation reviews	Reviewed	Did Not Review	Reviewed	Before

Legend: ● = SSA office reviewed cases; ◐ = SSA office partially reviewed cases; ○ = SSA did not review cases.

Source: GAO analysis of SSA documents and interviews with agency officials. | GAO-18-37

^aAbout 2.6 percent of all requests for review in fiscal year 2016 were of fully or partially favorable decisions.

SSA officials stated that opportunities exist to improve coordination across offices conducting quality assurance reviews. We found that several offices coordinated their work in some cases. For example, SSA’s Division of Quality and Office of Quality Review participate in a multi-office workgroup that addresses such issues as policy compliance across the initial and hearings levels of the disability process. In addition, they have also worked together on several studies, including a one-time quality review of 454 claims that were denied at the initial determination level, but were allowed as fully favorable at the hearings level. The Office of Quality Review also reviews the content of selected training for judges. In addition, the Division of Quality provided some initial input when the regional inline review effort was being designed. Prior GAO work has found that enhanced coordination can help to reduce overlap and improve efficiency.⁶⁶ Effective October 1, 2017, SSA created a new deputy commissioner-level component, the Office of Analytics, Review and Oversight. This agency reorganization moved six oversight offices into the new component, including the Division of Quality and Office of Quality Review. Officials said the new component will create opportunities for improved coordination between these six offices. While this

⁶⁶GAO-15-49SP.

reorganization creates the opportunity for SSA to assess many of its quality assurance reviews, the regional quality review staff will not be included in the new office, and it is too early to tell how this reorganization will help manage the overlap between SSA's various quality assurance reviews.

In addition, SSA has struggled to sustain all of its quality assurance reviews due to competing demands for the staff who perform them. For example, SSA placed regional inline quality reviews on hold in September 2016 and again in December 2016, because officials said that the agency needed staff to complete pending decisions before a change in the medical listings for mental impairments took effect in January 2017. Decisions not completed before the new listings took effect would have to be redone. Also, the Office of Quality Review curtailed its Disability Case Reviews in fiscal year 2016 to help prepare the oldest cases for hearings. As a result, only the Appeals Council's review of appealed ALJ decisions (requests for review) and the Division of Quality's quality assurance reviews were active in 2016.

Even as SSA has added quality assurance reviews, it has not systematically evaluated the efficiency and effectiveness of all the reviews to determine the extent to which they may be overlapping or complementary. We found that reviews conducted by the four entities have resulted in similar findings, raising questions about the efficiency of these reviews. For example, during the same 3-year period (fiscal years 2013 through 2015), quality reviews conducted by all four entities found problems with judges' assessment of a claimant's ability to perform work-related tasks, known as a residual functional capacity assessment. In addition, all four entities found problems with the evaluation of medical opinion evidence.

Moreover, SSA has not conducted a cost-benefit analysis of the five reviews. Officials said that there are no definite plans to do so, although they may consider conducting such an analysis in the future. We found that costs for the quality assurance reviews conducted in fiscal year 2015 were at least \$23.7 million, and in fiscal year 2016 were at least \$11.7 million (see table 4). By evaluating the quality assurance reviews to determine the extent to which each is needed to monitor and improve accuracy and consistency, SSA would be better positioned to meet its goals within its resources.

Table 4: Social Security Administration (SSA) Spending on Quality Assurance Reviews in Fiscal Years 2015 and 2016 and Number of Cases Reviewed (dollars in millions)

	SSA Spending		Cases Reviewed, Fiscal Year 2016 or Most Recent Year Available
	Fiscal year 2015	Fiscal year 2016	
Regional Quality Review Officers: Regional inline reviews	7.2 ^a	5.9 ^a	12,561
Office of Quality Review: Disability Case Reviews	1.7	n/a ^b	600 ^c
Division of Quality: Random and selective pre-effectuation reviews, focused quality reviews	14.8 ^d	5.8 ^d	7,690
Total	23.7	11.7	

Source: Agency documents, cost data and estimates from SSA officials. | GAO-18-37

^aSSA officials stated that SSA does not maintain official cost data for this activity. As a result, these amounts are estimates of payroll costs for the staff responsible for these reviews.

^bSSA did not conduct Disability Case Reviews in fiscal years 2016 or 2017. Officials stated that this review will be resumed in fiscal year 2018, and will be called the Hearings Level Review.

^cData are from fiscal year 2015; the Disability Case Review was not conducted in fiscal year 2016.

^dSSA officials noted that these figures are estimates.

In addition, SSA continues to develop and implement initiatives aimed at improving hearing decisions, without evaluating the potential for overlap with existing quality assurance reviews. For example, as part of its backlog reduction plan known as the Compassionate And Responsive Services (CARES) plan, SSA is using computer algorithms for natural language processing to analyze the text of disability decisions and flag potential errors. Although the agency is piloting this effort in the Appeals Council before expanding it to hearing offices, it did not conduct a cost-benefit analysis. SSA officials said that natural language processing could be used to identify cases for further review, similar to its current selective reviews, and that decision writers could use the tool to conduct their own reviews of their draft decisions. SSA officials said that they do not anticipate much overlap between the use of natural language processing and OAO's pre-effectuation reviews. However, there could be potential for overlap with regional inline reviews, which also review decisions drafted by decision writers. Federal internal control standards state that management should implement control activities through policies. Periodically reviewing policies, procedures, and related control activities for continued relevance and effectiveness in achieving objectives and addressing related risks can help agencies meet this standard.⁶⁷

⁶⁷GAO-14-704G.

Conclusions

SSA's disability programs provide more than \$200 billion in benefits for tens of millions of Americans annually, making it one of the largest components of the nation's social safety net. The hearings and appeals level of the disability decision-making process is particularly important because about one in three people receiving Social Security disability benefits are granted benefits at this level. Given the number of people and the dollars at stake, it is crucial that claimants are treated fairly and their applications are evaluated accurately and consistently across the country, at all levels of the program.

Some of the variation in allowance rates that we found across judges may be expected, given the complexity of the cases and judges' decisional independence. However, the persistent variation we observed over time, even after accounting for various factors that could otherwise explain allowance rates, might warrant additional attention. SSA is rightly focusing on oversight of judges, but our work suggests that the agency's emphasis on timeliness over accuracy in its public metrics and the potential overlap in its quality assurance efforts may offer opportunities for improving the accuracy and consistency of hearing decisions.

First, this amount of variation in allowance rates underscores the need for SSA to measure and hold itself accountable for accuracy and consistency. However, without sharing performance information on the accuracy and consistency of its hearings-level decisions, such as the rate at which the Appeals Council agrees with a judge's decisions, SSA may not be providing the public with adequate information on progress toward its objective to improve the quality, consistency, and timeliness of its disability decisions. Developing a set of performance measures that includes the accuracy and consistency of hearings decisions will help ensure the overall success of the program.

Second, SSA has not systematically considered how each of its quality assurance reviews helps the agency meet its objective to improve the quality of hearings-level decisions. Although the planned consolidation of multiple oversight and quality review offices is a positive step, it will be important for SSA to consider the usefulness of the information yielded by each quality assurance effort, as well as the costs associated with conducting the effort. Evaluating the efficiency and effectiveness of quality assurance activities can help ensure that SSA is using its

resources for maximum benefit toward its objective to improve the quality, consistency, and timeliness of its disability decisions.

Recommendations for Executive Action

We are making the following two recommendations to SSA:

The Commissioner of SSA should develop a set of public performance measures, to include accuracy and consistency, as well as timeliness, of administrative law judges' (ALJ) disability decisions. SSA could consider whether existing quality review or monitoring efforts could provide suitable data for such measures. (Recommendation 1)

The Commissioner of SSA should systematically evaluate the efficiency and effectiveness of its quality assurance reviews and take steps to reduce or better manage any unnecessary overlap among them to ensure strategic use of resources. Such steps could include enhancing collaboration where reviews overlap or only conducting the reviews that are most efficient and effective in achieving agency goals for improving accuracy and consistency of ALJ disability decisions. (Recommendation 2)

Agency Comments and Our Evaluation

In commenting on a draft of this report, SSA agreed with our two recommendations to (1) establish public performance measures for the accuracy and consistency of administrative law judges' decisions, and (2) systematically evaluate its various quality assurance reviews and take steps to reduce or better manage any unnecessary overlap among them. SSA stated that it would address both recommendations as part of a comprehensive assessment and refinement of its oversight roles and processes. SSA made several other comments about one of our conclusions and our analysis of variation in administrative law judge allowance rates, which we discuss below. SSA also provided technical comments, which we incorporated into the report as appropriate.

In its comments, SSA described its evolving oversight activities at the hearings level, including providing policy guidance and training for judges, capturing and utilizing data to gain a better understanding of trends and challenges, and implementing additional oversight review processes, all of which we discussed in our report. SSA's comments acknowledged that

our report describes the steps that the agency has taken to improve oversight, but disagreed with our conclusion that SSA emphasizes timeliness over accuracy. Our final report clarifies that we came to this conclusion based on a review of the performance measures the agency shares with the public in its annual strategic plan and performance reports. As we state in the report, SSA has employed a range of efforts to monitor the accuracy and consistency of hearings decisions, but it lacks performance measures to report publicly on these efforts.

Regarding our analysis of variation in ALJ allowance rates, SSA raised a concern about our finding (on page 26 of the final report) that claimants whose hearings were held in person were slightly more likely (by about 2.8 percentage points) to be allowed benefits than a typical claimant with a hearing held by videoconference. SSA cited its own internal analysis, which found a small (0.6 percentage-point) difference in allowance rates between in-person and videoconference hearings after controlling for a number of factors. It is not surprising, however, that our estimates are somewhat different, since SSA's internal analysis differs from ours in several ways. The primary purpose of our statistical analysis was to isolate variation in allowance rates due to the unique judge or hearing office assigned to each claim. To do this, we developed a multilevel model using 9 years of data that controls for judge, hearing office, and claimant-level factors associated with allowance rates. On the other hand, SSA's analysis was specifically designed to look at the difference in allowance rates between in-person and video hearings. SSA's analysis also covered a shorter, more recent period of time (part of fiscal year 2015, fiscal year 2016, and part of fiscal year 2017), than our study (fiscal years 2007 through 2015). Additionally, the version of the model SSA cited in its comments included hearings held in person or by videoconference only in regular hearing offices, whereas our analysis included hearings held in National Hearing Centers as well as regular hearing offices and controlled for the type of hearing office. These differences notwithstanding, we agree with SSA that the estimated model-adjusted difference in allowance rates between in-person and videoconference hearings in both GAO's and SSA's analyses could potentially be explained by unmeasurable factors.

In addition, SSA noted that our measure of variation in judge decisions focused on allowance rates at the extremes of the distribution. Given that our charge was to explore the extent of variation in allowance rates across judges, we believe it was appropriate and important to measure the range of allowance rates between judges with high allowance rates (at the 95th percentile) and those with low allowance rates (at the 5th

percentile). This would be more conservative than an approach that looks at allowance rates across all judges, including potential extreme values; and more nuanced than an approach that looks at the number of judges whose allowance rates are higher or lower than a given threshold. Further, our analysis shows that unadjusted allowance rates at the 95th percentile declined over the period of our analysis, from a high of 96 percent in fiscal year 2008 to 82 percent in fiscal year 2015. We saw a comparable decline in allowance rates after applying our multivariate model. To provide additional context, our report figures also show the middle of the distribution (the 25th and 75th percentiles), as well as the average allowance rates. We have also added information to our report further describing this middle range.

Finally, SSA noted that our analysis was not weighted by the number of determinations a judge made, suggesting that judges who decided very few claims, for example, could affect the range in allowance rates or the trends. As we show in Appendix I, Table 7, only 2.3 percent of the judges in our study population heard fewer than 250 claims per year. This group of judges had an unadjusted allowance rate of 61.9 percent, very similar to the allowance rate among judges who heard 500-699 claims per year (61.6 percent). Furthermore, the statistical methods we used to estimate the distributions of allowance rates (multilevel models) adjust the estimates for judges with fewer claims by weighting them more heavily toward the overall approval rate.⁶⁸ This mitigates against judges with smaller caseloads, and therefore higher sampling variation, from contributing overestimated allowance rates that might have inflated our estimated variation across judges.

⁶⁸Andrew Gelman and Jennifer Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*. (New York: Cambridge University Press, 2007), 252-260, 309-310.

As agreed with your office, unless you publicly announce the contents of this report earlier, we plan no further distribution until 30 days from the report date. At that time, we will send copies to the appropriate congressional committees, the Commissioner of Social Security, and other interested parties. In addition, the report will be available at no charge on the GAO website at <http://www.gao.gov>.

If you or your staff have any questions about this report, please contact me at (202) 512-7215 or jeszeckc@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. GAO staff who made key contributions to this report are listed in appendix III.

Sincerely yours,

A handwritten signature in black ink that reads "Charles Jeszeck". The signature is written in a cursive style with a large, looping initial "C".

Charles Jeszeck
Director, Education, Workforce and Income Security Issues

Appendix I: Objectives, Scope, and Methodology

Our objectives were to assess (1) the extent to which allowance rates vary across administrative law judges, and any factors that are associated with this variation, and (2) the extent to which the Social Security Administration (SSA) has processes to monitor the accuracy and consistency of hearings decisions.

To answer these objectives, we reviewed SSA policies and procedures related to administrative law judge (ALJ) disability hearings and decisions; manuals and documents describing SSA's case processing systems for each level of SSA's disability decision-making process,¹ and guidance and training provided to judges for making disability decisions. We interviewed SSA officials in several offices within the Office of Disability Adjudication and Review (ODAR),² including the Office of the Chief Administrative Law Judge and the Office of Appellate Operations, as well as conducted semi-structured interviews with Regional Chief Administrative Law Judges in each of SSA's 10 regions. We also observed administrative law judge hearings in one of SSA's five National Hearing Centers, in Falls Church, Virginia, (hearings in these offices are conducted by videoconference), as well as in two of SSA's regular hearings offices, in Washington, D.C., and Seattle, Washington. The purpose of these observations was to gain a better understanding of the hearings process in practice, and to inform our scope and methodology for this study. We selected these sites, which are not generalizable to the population of all hearing offices, for a number of reasons, primarily: (1) to observe hearing offices in different geographic locations and observe both in-person and video teleconference hearings, and (2) to select sites at which a cross-section of cases with different types of disabilities and impairments were available. We attended hearings involving both adult and child claimants with a mix of physical and mental impairments.

This appendix is divided into three parts. The first describes our data sources and analysis of allowance rates across judges and associated

¹For more information on the different levels of SSA disability decision-making, see figure 2.

²As of October 1, 2017, this office is called the Office of Hearings Operations (OHO).

factors, the second describes our multivariate statistical model, and the third describes our work related to our second research objective on SSA's processes to monitor the accuracy and consistency of hearings decisions.

Analysis of Variation in Allowance Rates across Judges and Associated Factors

For this objective, we analyzed data from two primary sources from fiscal years 2007 through 2015: SSA's administrative data systems for the initial and hearings levels of the disability decision-making process, and the agency's personnel data system. We also obtained other SSA administrative data on staffing levels and numbers of pending cases in each hearing office. Finally, we obtained data on state poverty and unemployment rates from the U.S. Census Bureau and the Bureau of Labor Statistics, respectively.

SSA Administrative Data Systems

To analyze information on all adult disability decisions made by administrative law judges from fiscal years 2007 through 2015—the most current data available at the time of our analysis—we compiled claims data from several SSA administrative data systems.³ These data contained information on the outcomes of the disability decisions and the characteristics of claims associated with each decision. Specifically, the information was drawn from the following systems:

- **831 File and Structured Data Repository:** The 831 File pertains to the initial and reconsideration level of the disability determination process, within the state Disability Determination Services (DDS). Data on claimant characteristics we used from this system include the date of the claimant's initial application for benefits and the claimant's self-reported years of education. We also received a limited set of data captured from the claimant's disability application in SSA's electronic case folder system (Structured Data Repository), including

³We received data from fiscal years 2005 through 2006, but those data were incomplete. According to SSA, the Case Processing and Management System (CPMS) came online in 2005 and data are complete as of 2007.

the number of years a claimant reported being employed out of the 15 years before becoming disabled.

- **Case Processing and Management System (CPMS):** This system pertains to the hearings level and was our primary source of information on hearing outcomes, claim, and claimant characteristics. Specifically, this system provided information on claim type (i.e., Disability Insurance, DI; Supplemental Security Income, SSI; or concurrent claim); the outcome of the claim (i.e., dismissed, allowed, or denied) and the date the decision was made; the unique identification number of the administrative law judge (ALJ) who made the decision; whether a medical expert or vocational expert attended the hearing; whether the claimant was represented; the hearing office where the claim was decided and the type of hearing office (i.e., hearing office or National Hearing Center); the claimant's date of birth; the primary impairment at the time of the hearing level; the presence of a secondary impairment; and whether the case was classified as being a critical case—that is, a case requiring special processing, such as a terminal illness.

We used case identifiers to link the information from each of these databases that pertained to each disability decision we analyzed.⁴

Federal Personnel and Payroll System

We obtained data from SSA's Federal Personnel and Payroll System (FPPS) database on all administrative law judges who were employed by the agency at any time during the period from January 1, 2005, through December 31, 2015. We obtained information on each judge, such as their date of appointment as an ALJ and the type of appointment (regular career appointment or non-permanent); service computation date; and prior position titles within SSA, if any.

⁴Specifically, SSA used the claimant's Social Security number to link data from the 831 File and the Structured Data Repository to the hearings-level data in CPMS, and the judge's name and personal identification number to link the CPMS data to the Federal Personnel and Payroll System (FPPS) database. We used the unique case identification code to link the personnel data to the CPMS data. Data provided to GAO had masked Social Security numbers.

Other SSA Administrative Data

We obtained summary-level data, as of January 2017, from SSA on staffing levels (numbers of ALJs, decision writers, and other support staff) at each hearing office for each fiscal year in our study period (fiscal years 2007 through 2015) from SSA's Payroll Operational Data Store system. We also obtained data on the numbers of cases left pending at the end of each fiscal year (including the number of cases pending for more than 270 days). SSA provided those data from a management information report that uses CPMS data.

Economic Conditions Data

We used publically available estimates of state poverty rates for each year in our analysis (calendar years 2007 through 2015) from the U.S. Census Bureau's American Community Survey (ACS).⁵ We considered using estimates at the county level but that approach had limitations. First, we would have been limited to using 3-year or 5-year estimates for all counties, because 1-year ACS estimates are only available for areas with populations of 65,000 or more. Second, the Census Bureau cautions against using estimates for particular time periods that do not align with the periods of its estimates. Although using state-level estimates reduced the geographic precision of the estimates, we gained precision by having annual estimates and the ability to measure potential variation in poverty rates over narrower time intervals.

We also used publically available estimates of state unemployment rates in calendar years 2007 through 2015 from the Bureau of Labor Statistics' Local Area Unemployment Statistics data.⁶ This variable allowed us to control for labor market conditions over time.

⁵Specifically, using the Census Bureau's American FactFinder system, we selected Table S1701 - Poverty Status in the Past 12 Months, filtered by the geography "All States within United States, Puerto Rico, and the Island Areas," and selected 1-year ACS. We filtered the table to include only the variable "percent below poverty level."

⁶We downloaded state-wide, annual average unemployment rates from the BLS Local Area Unemployment Statistics website. This data series is located at: <https://download.bls.gov/pub/time.series/la/la.data.2.AllStatesU>.

Data Reliability

SSA constructed custom files for GAO from several SSA datasets in response to our data requests. We assessed the reliability of the data used in our analyses through electronic testing, analyzing related database documentation, examining the SAS code used by SSA to construct the custom files, and working with agency officials to reconcile discrepancies between the data and documentation that we received. We determined that the 831, Structured Data Repository, and CPMS data on ALJ decisions and claimant characteristics and the FPPS data on ALJ appointments were sufficiently reliable for the purposes of describing the extent of variation in the outcomes of ALJ decisions. We also determined that SSA's data on pending caseloads and ALJ and decision writer staffing, by year and hearing office, were sufficiently reliable for the purpose of describing hearing office characteristics. Finally, we determined that ACS data on state poverty rates and BLS data on state unemployment rates were sufficiently reliable for the purposes of describing these state economic characteristics.

Scope of Analysis

Our analyses of ALJ decisions excluded various types of decisions from the CPMS data because they were out of scope for our research objectives (e.g., child cases, non-disability cases, or cases that were decided by SSA staff who were not ALJs) or were not typically randomly assigned to judges. We selected cases that should have been assigned randomly to judges, according to SSA policy, because that random assignment made it more likely that variation in allowance rates across judges in our multivariate analysis reflects the unique causal effect of having a particular judge hear a case, rather than other factors that also vary across judges. Our exclusion criteria were similar to those used by an internal SSA study of ALJ allowance rates, conducted in 2017.⁷ We excluded cases that were:

- **Dismissed.** Cases can be dismissed for reasons not related to the merits of the case and that are usually beyond the ALJ's control—for example, the claimant's failure to file a timely request or to appear at the scheduled hearing (without good cause), or the claimant's death

⁷That study did not exclude claims decided by senior attorney adjudicators or ALJs with limited experience.

before the hearing. In addition, data on key factors for these cases, such as the claimant's impairment, were missing. From fiscal years 2007 through 2015, 1,007,526 claims (16 percent of all claims) were dismissed.

- **Made “on the record” and not randomly assigned to judges.** While most appeals are decided after an ALJ hearing, ALJs and senior attorney adjudicators (SAA) have the authority to issue on-the-record decisions. These are decisions where a hearing is not necessary because the documentary evidence alone supported a fully favorable decision. SSA has created screening criteria, such as the claimant's age (50 and older) and specific impairments, to help identify possible on-the-record decisions earlier in the process. ALJs and SAAs can also issue on-the-record decisions for cases involving critical need, and claimants and their representatives can request that the ALJ or SAA issue an on-the-record decision. These cases are not assigned randomly to judges. From fiscal years 2007 through 2015, 716,574 claims (11 percent of all claims) were on-the-record decisions, although SSA has issued fewer on-the-record decisions in more recent years.
- **Issued for children.** We excluded claimants younger than 18 at the date of the initial application. We also excluded claimants with missing or invalid age values. From fiscal years 2007 through 2015, 492,158 claims (8 percent of all claims) were for people under 18 or with missing or invalid age values. We excluded child cases from our analysis because they involve different evaluation criteria.
- **Remanded back to a judge from SSA's Appeals Council (or federal court).** These cases represent decisions that were corrected after an order from the Appeals Council or a federal court after the original ALJ's decision. In these cases, judges are often addressing a narrow set of issues identified in the remand order. Remanded cases are also not assigned randomly to judges, since the Appeals Council generally sends them back to the judge who originally issued the decision.⁸ However, SSA's Office of the Inspector General (OIG) in 2017 found that about half of the remanded cases in its sample were

⁸As specified in the Hearings, Appeals, and Litigation Law (HALLEX) manual sections I-2-1-55 and I-3-7-40, remands are generally assigned to the same ALJ who issued the decision or dismissal. However, HALLEX § I-3-7-40 provides that, if an ALJ has already heard the case twice before, the case will be assigned to a different ALJ on remand. In addition, according to HALLEX § I-3-2-25(B), the Appeals Council will ordinarily (but not necessarily) direct that a different ALJ handle a case on remand when it finds an abuse of discretion.

assigned to a different ALJ than the original ALJ.⁹ From fiscal years 2007 through 2015, 293,971 claims (less than 5 percent of all claims) were remands.

- **Made by senior attorney adjudicators who were not administrative law judges.** We excluded decisions made by SAAs.¹⁰ SSA implemented a program in 2007 whereby SAAs located in hearing offices across the country could issue fully favorable on-the-record decisions. According to SSA, this allowed ALJs to focus on cases that are more complex or require a hearing. From fiscal years 2007 through 2015, 227,133 claims (4 percent of all claims) were decided by SAAs.
- **Appeals of continuing disability reviews (CDR).** These cases represent decisions about whether or not to continue benefits for claimants who were previously found eligible for the program. As such, they involve different evaluation criteria. From fiscal years 2007 through 2015, 245,862 claims (4 percent of all claims) were appeals of CDRs.
- **Non-disability cases.** These cases include Social Security retirement and survivor benefit decisions. We excluded such cases because they involve different evaluation criteria from disability claims and represent a small minority of decisions at the hearings level. From fiscal years 2007 through 2015, 25,293 claims (less than 0.5 percent of all claims) were for non-disability cases.
- **Decided by judges with limited experience.** We excluded cases decided by judges within the first year (365 days) after their appointment as an ALJ, as calculated by the difference between their date of appointment and the date of the decision on each claim. We excluded these decisions to help ensure that variation we identified in allowance rates was not due to the judges' more limited experience deciding Social Security disability claims. From fiscal years 2007 through 2015, 574,307 claims (approximately 9 percent of all claims) were decided by judges with limited experience.

In total, our exclusion criteria reduced the number of records analyzed by about half. Specifically, out of a universe of about 6.3 million records, our

⁹SSA OIG, *Oversight of Administrative Law Judge Decisional Quality*, A-12-16-50106 (Baltimore, Md.: March 14, 2017).

¹⁰Decisions made by senior attorney adjudicators are also excluded from SSA's public ALJ disposition data.

study population included about 3.3 million decisions.¹¹ Nevertheless, the overall allowance rate for our study population over fiscal years 2007 through 2015 was 62 percent, very close to the overall allowance rate for the entire population of claims during this period, which was 64 percent.¹²

Calculation of Allowance Rates

We calculated allowance rates by dividing the number of favorable decisions by the total number of decisions (both unfavorable and favorable). We calculated allowance rates for different units of analysis:

- Overall, by program type (Disability Insurance, Supplemental Security Income, and concurrent) and by year and for all years and program types pooled together,
- At the judge level, by year and for all years pooled together, and
- At the hearing office level, by year and for all years pooled together.

When analyzing our data at the case level, we identified whether the case was favorable or unfavorable to the claimant (that is, whether the claimant was allowed benefits or not). We did not include cases that were dismissed in our study population for two reasons. First, as discussed above, these cases can be dismissed for reasons not related to the merits of the case, and without a review of the medical evidence. Second, SSA's data on dismissed cases are limited, partially because cases are dismissed without a review of medical evidence. For example, the impairment code from the hearings-level decision was missing for virtually 100 percent of dismissed cases.

For concurrent claims—those in which an individual is applying for Disability Insurance (DI) and Supplemental Security Income (SSI) benefits—we considered a case an allowance if the claimant was

¹¹Our exclusion categories are not mutually exclusive; for example, a senior attorney adjudicator could issue an on-the-record decision. Therefore, our separate exclusion categories do not sum to the entire number of exclusions.

¹²This 2-percentage-point difference is largely explained by the exclusion of senior attorney adjudicators (SAAs) from the study population. Under SSA's Senior Attorney Adjudicator program, SAAs are only able to issue fully-favorable, on-the-record decisions (that is, decisions in which a hearing is not necessary because the documentary evidence alone supports a decision that is fully favorable to the claimant). The allowance rate among the entire population of claims minus those heard by SAAs was 62 percent over this period.

approved for either or both programs. Our classification of allowances for concurrent cases differs from SSA's usual practice (although a 2017 internal study of ALJ allowance rates used the same method as ours). SSA officials said they usually allow the SSI decision to "control" the overall outcome of the case. That is, SSA classifies a concurrent claim as an allowance if the SSI decision is an allowance, regardless of the outcome for the DI claim. Officials said that they chose this method primarily for convenience. This results in a different classification of some cases in which the SSI claim was denied but the DI claim was allowed. In such cases, the claimant is receiving a benefit as a result of their concurrent disability claim but would be classified in SSA's data as a denial. However, the resulting differences in the number of allowances is very small—less than 4,000 claims over fiscal years 2007 through 2015—and the different definitions did not substantively affect allowance rates in any year.

Random Assignment of Cases to Judges

SSA policy states that cases are generally assigned on a "first in, first out" basis, meaning that cases are assigned to judges in the order in which they are received. Administrative law judges are assigned cases on a rotational basis, with the oldest case in the backlog given to a judge who most recently decided a case. Therefore, as noted in prior research, the initial assignment of cases to judges is random (conditional on applying at a given hearing office at a given time). Judges do not select their cases, nor are claimants able to request another judge after one is assigned.¹³ Claimants are generally assigned to hearing offices based on their ZIP code, although some claimants in hearing offices with higher numbers of pending claims may be transferred to one of SSA's five National Hearing Centers. In those cases, hearings are conducted by videoconference rather than in person, as is traditionally done in SSA's regular hearing offices. However, the claimant may opt out of a video conference hearing within 30 days of receiving a written notice acknowledging the request for a hearing.

There are some exceptions to the "first in, first out" rule, such as cases that are likely to be dismissed or decided on the record (without a

¹³See Eric French and Jae Song, "The Effect of Disability Insurance Receipt on Labor Supply," *American Economic Journal*, Vol. 6 No. 2. (May 2014). This study also notes that the judge initially assigned a case may not be the judge who ultimately decides the case, but the judges were the same in 96 percent of the cases studied.

hearing) and critical cases (including terminal illness cases and veterans who have a 100-percent permanent and total disability compensation rating). However, as discussed previously, we have excluded all major categories of exceptions but critical cases from our analyses, and included a variable to identify critical cases in our analyses.

Assignment of Claims to Steps in SSA's Sequential Disability Decision-Making Process

SSA's disability decision-making process includes five sequential steps, and one part of our analysis was to determine the step at which each decision was made. In consultation with SSA officials, we used a code in CPMS—called the regulation basis code—to assign each claim to a particular step.

Each claim in CPMS has between one and four regulation basis codes, depending on whether the claim was for a single program (DI or SSI), or a concurrent claim for both.¹⁴ We assigned each claim to one of the five steps in SSA's disability decision-making process, based on its regulation basis code. Each regulation basis code is associated with one of five steps. Therefore, if a claim had just one regulation basis code, we assigned it to the corresponding step. If a claim had more than one regulation basis code, we used a series of decision rules to select the most appropriate step. Specifically, claims for a single program have up to two regulation basis codes listed, and we used the code that matched

¹⁴Disability Insurance claims have either one or two regulation basis codes; Supplemental Security Income claims have either one or two regulation basis codes; and concurrent claims have up to four regulation basis codes, i.e. one or two for each program (Disability Insurance or Supplemental Security Income).

the outcome of the case and/or the latest step.¹⁵ We used a similar method for concurrent claims.¹⁶

We found that approximately 19 percent of all allowances occur at step 3, when SSA determines whether a claimant's impairment meets or is equivalent to an impairment listed in SSA's Listing of Impairments. Most (80 percent) of all allowances are made at the final step (step 5), when SSA determines whether the claimant can do any work in the national economy, given the limitations of their impairment and their age, education, and work experience.

Nearly a third (28 percent) of denials are made at step 4, where SSA determines whether the claimant can do their past work, and 62 percent of denials are made at step 5. There are some differences between DI claims and SSI claims in the distribution of allowances and denials over the five steps. SSI allowances occur at step 3 to a greater extent than DI allowances, while SSI denials occur at step 5 to a greater extent than DI denials (see table 5 below).

¹⁵In single program cases, if one of the two regulation basis codes is associated with an allowance and the other regulation basis code is associated with a denial or a dismissal or is not populated, we used the regulation basis code associated with the allowance. In single program cases where both regulation basis codes are associated with an allowance, we selected the regulation basis code associated with the higher (later) step, where step 5 is the highest (latest) step and step 1 is the lowest (earliest) step.

¹⁶In concurrent claims, if benefits are allowed under one program but denied under the other program, we selected from the regulation basis codes associated with the program that was allowed; if the claim has two regulation basis codes for that program, we follow the same assignment rule as we do for single program claims. If both programs are allowed, we selected the program that had the regulation basis code associated with higher (later) step. For claims that are denied we applied a similar decision rule, selecting the regulation basis code associated with the higher (later) step.

Table 5: Distribution of Allowances and Denials, by Program Type and Step in the Social Security Administration’s (SSA) Disability Decision-making Process, Fiscal Years 2007-2015

	All Claims		DI Only		SSI Only		Concurrent Claims	
	Allowances	Denials	Allowances	Denials	Allowances	Denials	Allowances	Denials
Step 1	2	22,601	0	8,516	0	2,370	2	11,715
	0 percent	2 percent	0 percent	2 percent	0 percent	1 percent	0 percent	2 percent
Step 2	11	57,144	0	21,800	0	14,575	11	20,769
	0 percent	5 percent	0 percent	6 percent	0 percent	5 percent	0 percent	4 percent
Step 3	377,711	4	123,097	3	88,258	0	166,356	1
	19 percent	0 percent	16 percent	0 percent	26 percent	0 percent	19 percent	0 percent
Step 4	6	351,658	0	120,477	0	56,614	6	174,567
	0 percent	28 percent	0 percent	34 percent	0 percent	19 percent	0 percent	30 percent
Step 5	1,620,226	770,839	661,667	190,719	246,332	211,555	712,227	368,565
	80 percent	62 percent	83 percent	54 percent	73 percent	71 percent	80 percent	62 percent
Missing/Other ^a	23,417	38,943	9,661	11,852	5,145	11,530	8,611	15,561
	1 percent	3 percent	1 percent	3 percent	2 percent	4 percent	1 percent	3 percent
Total	2,021,373	1,241,189	794,425	353,367	339,735	296,644	887,213	591,178
	100 percent	100 percent	100 percent	100 percent	100 percent	100 percent	100 percent	100 percent

Source: GAO analysis of Social Security Administration (SSA) data. | GAO-18-37

^aWe could not assign a step to 62,360 claims. Approximately 35,000 claims had no regulation basis code listed in the claim level data provided by SSA, while another approximately 27,000 claims had a regulation basis code that was not associated with one of the five steps of the disability determination process.

Statistical Model of Variation in Allowance Rates across Judges and Associated Factors

We developed our multivariate statistical analyses in consultation with GAO statisticians, economists, and social scientists and SSA officials and experts. Our analysis was also informed by a comprehensive review of the literature pertaining to judicial decision-making and, in particular, adjudication for SSA’s disability programs. Specifically, we reviewed more than 90 potentially relevant peer-reviewed academic journal articles, government reports, and nonprofit association and think tank white papers. We selected 39 of these studies or reports for a detailed review of the scope and methodology, key factors or variables used in any empirical analyses, and other relevant findings. We also reviewed relevant SSA Office of the Inspector General (OIG) reports and consulted

with SSA and OIG officials, and reviewed prior GAO reports that modelled judicial outcomes.¹⁷ Our statistical model included variables that are either direct or approximate measures for: (1) claimant characteristics that represent criteria used in the disability decision-making process, (2) judge characteristics, (3) other participants in the decision-making process, (4) SSA administrative characteristics, and (5) economic characteristics of the claimant's state.

Our analysis was purely statistical, in that we did not conduct the legal analysis needed to reach conclusions about what legal factors might have affected a judge's decision or whether the decision that was reached in any particular case was correct. Similarly, we are not making any predictions about the likely or correct outcome of future individual decisions. Each case is unique in both its facts and circumstances and must be examined on its own merits.

Claimant characteristics

- We included factors that represent criteria used in decision-making process, such as the type of claim (DI, SSI, or concurrent) and the claimant's age, years of education (grouped into equivalent levels: less than high school, high school, some college, and college or higher), and primary impairment.

Judge characteristics

- We included factors related to the judge's employment as an ALJ, such as the year appointed as a judge, the type of appointment (whether they had a career or temporary, non-permanent assignment), and any prior work history at SSA (specifically, whether they were an attorney or held another position prior to being appointed as an ALJ).

Other participants in the decision-making process

- We included factors that represent other participants in the decision-making process, such as the claimant's use of an attorney or non-attorney representative, or the testimony of a medical or vocational expert at the hearing. Our prior work has shown, for example, that

¹⁷See, for example, GAO, *Asylum: Variation Exists in Outcomes of Applications Across Immigration Courts and Judges*, [GAO-17-72](#) (Washington, D.C.: Nov. 14, 2016); and [GAO-04-14](#).

claimants who were represented by an attorney or a person who is not an attorney (such as a relative or professional disability representative) were more likely to be allowed disability benefits than claimants who had no representative.¹⁸

SSA administrative characteristics

- We included factors related to SSA’s administration of its disability programs, such as the hearing office in which the claim was decided, whether the claim was heard in one of 10 states that do not have a reconsideration step between the initial state-level Disability Determination Service decision and a hearing before an ALJ,¹⁹ and the percentage of pending cases at the hearing office that were pending for more than 270 days (SSA’s definition of a “backlogged” case).²⁰

Economic characteristics

- Finally, we assessed economic characteristics of the state in which the claimant resided because some prior research suggests that such factors may be associated with disability application and allowance rates.²¹ Specifically, we analyzed:
 - The unemployment rate in the claimant’s state as of the year of each decision in our analysis, from the Bureau of Labor Statistics’ Local Area Unemployment Statistics data. We selected this factor

¹⁸GAO-04-14.

¹⁹The 10 states are Alabama, Alaska, California (Los Angeles North and Los Angeles West Branches), Colorado, Louisiana, Michigan, Missouri, New Hampshire, New York, and Pennsylvania.

²⁰SSA defines the hearings backlog as “the number of pending cases that push the average wait time over 270 days.” SSA considers the backlog eliminated when the national average processing time for a hearing decision is 270 days. We considered some additional administrative factors but did not include them in our final model. For example, we did not use judge’s average processing time, which we found was moderately correlated with the percent of cases that were backlogged within that judge’s hearing office.

²¹See, for example, Robert Nakosteen and Michael Zimmer, “Approval of Social Security Disability Appeals: Analysis of Judges’ Decisions,” *Applied Economics*, Vol. 46. No. 23 (2014): 2783-2791. See also Social Security Administration, Office of the Chief Actuary, “Disabled Worker Allowance Rates: Variation Under Changing Economic Conditions,” Actuarial Note No. 153 (Baltimore, Md.: August 2013).

in order to account for the labor market conditions where claimants live.

- The poverty rate in the claimant's state as of the year of each decision in our analysis, from the Census Bureau's American Community Survey (ACS).

Goals of Analysis

The primary goal of our analysis was to isolate variation in allowance rates due to the unique judge or hearing office assigned to each claim by controlling for multiple factors that could otherwise affect this variation. Some variation in allowance rates across judges and hearing offices could reflect the distribution of other factors that are correlated with allowances. For example, judges who hear disability cases in regions of the country with higher obesity rates—a known risk factor for disability—may appear to have higher allowance rates than those in regions with low obesity rates. Because judges' decisions to allow benefits may be related to this or other factors, simple univariate comparisons of allowance rates across judges may reflect characteristics of the cases that judges hear.

To help isolate the potential unique effects of judges, we used multilevel, multivariate statistical models that held constant various factors that could have been associated with allowance rates. We held constant variables available in SSA and other public data sources that were relevant to the claim appeals process, in order to estimate the amount of potential residual variation across judges.

Statistical Model

The data we assembled had a multilevel structure, with applications for disability benefits clustered within the same judges and hearing offices. Judges were associated with multiple hearing offices, because judges sometimes decided cases in multiple hearing offices during the period of our analysis. For example, judges could travel to more remote sites to hear cases on a part-time basis.

The data and outcome of interest suggested that a multilevel or mixed logistic regression model would adequately reflect the data generation process. We developed a mixed model that represented the grouping

variables—judge, hearing office, and primary diagnosis code²²—with random intercepts, similar to prior research.²³ We modeled group variation with random effects primarily for parsimony. Modeling group variation with fixed effects would have required estimating several thousand explicit parameters, one for each group level, which would have consumed many degrees of freedom. Estimating the amount of variation across groups then would have required interpreting many contrasts between pairs of fixed effect estimates. In contrast, modeling group variation using random effects allowed us to represent the variation with probability distributions and a small number of summary (hyper) parameters, such as the standard deviation of the judge random effect.

Substantively, random effects accurately represented the SSA policy of randomly assigning judges to cases in our study population, using a “first in, first out” method. Moreover, we modeled variation across judges and hearing offices as random, which implies that we seek to make inferences about a larger, hypothetical population of judges and hearing offices that could exist if we replicated the study in the future. This seems appropriate, because the application review process could be repeated across many new judges and hearing offices in the future. We do not seek to make inferences limited to the judges and hearing offices at the particular time we assembled data.

We held constant case, judge, and hearing office characteristics using covariates with fixed parameters. The smaller number of parameters associated with these covariates made a fixed effects approach easier to apply and interpret. We assumed that the covariate effects did not vary across groups, so that only the model’s intercept varied randomly. We had no prior expectation that specific covariate effects should have varied across groups. Moreover, increasing the number of random effects would have increased the complexity of the model and could have made it hard to estimate computationally.

We viewed the covariates primarily as controls for isolating variation across judges and offices. We did not attempt to build a comprehensive model that correctly specified how all of the covariates were causally

²²The primary impairment, or diagnosis, is the basic condition that the judge determined rendered the person disabled or evidence shows has the most significant effect on the person’s ability to work, according to SSA program rules.

²³Javier Meseguer, “Outcome Variation in the Social Security Disability Insurance Program: The Role of Primary Diagnoses.” *Social Security Bulletin*, 73 (2): 39-75.

ordered and related to each other and the probability of an allowance. As a result, our estimates of these parameters may not be consistent with those obtained from a more comprehensive modeling effort, or from analyses designed to estimate the causal effects of particular variables, such as the use of videoconferences. In the body of the report, we present alternative explanations and provide context to avoid interpreting the covariate effects as with a high degree of causal certainty. For example, we note that claims with legal representation may have higher approval rates if representatives accept claims with greater merit and, therefore, a greater chance of compensation. Below, we test alternative model specifications for covariates where the causal ordering may be ambiguous, in order to avoid biasing estimates of the judge and office parameters of primary interest.

Certain variables and parameters were applied across multiple versions of the model (described below). Let Y_i denote the allowance or denial decision for claimant i at any step of the appeals process, with $Y_i = 1$ if the ALJ allowed the claim and 0 otherwise. Each model took a typical hierarchical generalized linear form for a binary outcome:

$$Y_i \sim \text{Binomial}(1, \pi_i)$$

$$\pi_i | X_{ij_o} = g(\eta_i) = g(\alpha + X_{ij_o}\beta + X_{j_o}\lambda + X_o\gamma + \epsilon_j + \epsilon_o + \epsilon_d)$$

$$g(\eta_i) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$$

The probability of allowance, π_i , was a function of covariate vectors measuring characteristics of claims, X_{ij_o} , characteristics of the ALJs assessing those claims, X_{j_o} , and characteristics of the hearing offices where the decision occurred, X_o . Claimants were clustered in $j = \{1, 2, \dots, J\}$ judges, and judges were clustered in $o = \{1, 2, \dots, O\}$ offices. g is the inverse logistic link function.

We included normally distributed random effects, $\epsilon_{(\cdot)}$, for each judge, office, and the claimant's primary diagnosis, indexed by diagnosis codes $d = \{1, \dots, D\}$. Random effects allowed the intercept for each group, $\alpha_{(\cdot)} = \alpha + \epsilon_{(\cdot)}$, to vary around the population average intercept, α , as a function of the group's variance, $\sigma_{(\cdot)}^2$:

$$\epsilon_j \sim N(0, \sigma_j^2)$$

$$\epsilon_o \sim N(0, \sigma_o^2)$$

$$\epsilon_d \sim N(0, \sigma_d^2)$$

To make interpretation and computation easier, we classified all continuous covariates into substantively meaningful categories, and set the omitted reference categories to the sample modes. This transformation implied that the random effect variance, $\sigma_{(\cdot)}^2$, described variation across judges and offices for a claim that had the modal value of all other covariates in the model and sample. The reference claim remained constant across models fitted to different subsamples, in order to make inferences about a claim that was typical for the study population.

The center of the data at the modes, α , may not necessarily correspond to an actual claim. For example, all judges do not practice at the modal hearing office, and the modal age for the study population may not be typical for claims made in the modal office. Nevertheless, rescaling facilitates estimation and interpretation of the model, because all inference can be done on α and $\alpha_{(\cdot)}$ directly, using the random effect variance, $\sigma_{(\cdot)}^2$, without transformation. This allowed us to concisely describe variation in allowance rates for a hypothetical, typical claim in the joint covariate distribution.

In the body of this report, we summarized variation across judges and offices, holding constant other covariates at their sample means, using the estimated distribution of group intercepts scaled in logits:

$$\alpha_{(\cdot)}|\bar{X} \sim N\left(\hat{\mu}, \hat{\sigma}_{(\cdot)}^2\right)$$
$$\hat{\mu} = \hat{\alpha} + \bar{X}_{ijo}\hat{\beta} + \bar{X}_{jo}\hat{\lambda} + \bar{X}_o\hat{\gamma}$$

To describe variation across groups on the probability scale, we estimated the quantiles bounding the middle 50 and 90 percent of the group density on the logit scale and then transformed them using the inverse logistic function, g .²⁴ This reference case does not represent a feasible claim, because the means of the categorical covariates are just the sample proportions. However, this approach complements the centering of the sample, which allows the group variance parameters, $\sigma_{(\cdot)}^2$, to represent variation across groups for a feasible reference case at the sample modes.

²⁴Stephen W. Raudenbush and Anthony S. Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2d ed. (Thousand Oaks, Calif.: Sage Publications, 2002), 297.

Covariates and Subsamples

We fit a sequence of models using different covariates and subsamples, listed below. Fitting several models allowed us to assess how simplifying assumptions, such as ignoring the step at which ALJs made allowance decisions, affected our results. This approach also assessed the stability of estimates across multiple runs of the computational model estimation methods.²⁵ We describe the substantive meaning of the covariates above, and give their exact measurement when reporting results in table 7 below.

- Model 1: Intercepts Only

$$X_{ij0}\beta = 0$$

$$X_{j0}\lambda = 0$$

$$X_o\gamma = 0$$

- Model 2: Add Covariates

$$X_{ij0}\beta = AgeDisp\beta_1 + Educ\beta_2 + PriorWorkYrs\beta_3 + CritTerm\beta_4 + Rep\beta_5 +$$
$$Video\beta_6 + MedExpert\beta_7 + VocExpert\beta_8 +$$

$$NoRecon\beta_9 + SecondDiag\beta_{10} + ClaimType\beta_{11} + YearDecision\beta_{12}$$

$$X_{j0}\lambda = NumPriorDecn\lambda_1 + Cohort\lambda_2 + PriorSSA\lambda_3 + PermTemp\lambda_4$$

$$X_o\gamma = Pending270\gamma_1 + OfficeType\gamma_2 + Unemp\gamma_3 + Poverty\gamma_4$$

(Unemployment and poverty vary at the state level, not at the office level, but we include them with the office covariates for simplicity.)

- Model 3: Claims Decided at Steps 4 or 5

We estimated Model 2 for only those claims decided at steps 4 or 5, according to each claim's Regulation Basis Code. In these last two steps of the sequential disability decision-making process, the judge determines whether claimants retain the ability to perform their past work or other

²⁵We estimated the model using maximum approximated likelihood methods, as implemented in the lme4 package in R 3.4.1. The lme4 package is prone to false positive warnings about convergence and estimation errors, according to its authors. To further assess convergence, the package authors recommend checking for random effect variances that approach zero and replicating all estimates with a different optimizer. We applied these checks when lme4 warned about nonconvergence, but due to computational constraints, replicated estimates with a second optimizer only for the models stratified by year. We confirmed the robustness of all estimates.

work in the national economy, given the limitations of their impairment and their age, education, and work experience. SSA officials provided methods to map these codes to steps of the appeals process. Estimating the model for decisions at steps 4 or 5 allowed all parameters to vary at these steps versus all steps in the pooled sample. For example, diagnosis may be less strongly associated with allowances at step 5 than at step 3, while the claimant's age may be more strongly associated.

- Models 4-6: Stratify by Year of Decision and Claim Type

To assess how the amount of variation across judges and offices has changed over time, we estimated Model 2 separately for each year of decision, claim type, and the cross-classification of these variables. Stratified models allowed all parameters to vary across claim types and years.

- Model 7: Exclude Potentially Endogenous Covariates

We excluded covariates from Model 2 that may not be exogenous to the probability of approval. These include representation by an attorney or other person and the presence of a medical or vocational expert.²⁶ Claims with legal representation may have higher approval rates if representatives tend to accept claims with greater merit and, therefore, a greater chance of compensation. (Representatives typically receive a share of their client's benefits as compensation.) According to SSA officials, medical and vocational experts may be more likely to testify at a hearing, depending on the judge's expected ruling on the case. Although judges generally have discretion about whether to involve medical and vocational experts, judges are required to seek the opinion of a medical expert in certain cases. For example, a judge must have a medical expert provide an opinion if the judge is considering allowing benefits because the claimant's impairment may be medically equivalent to one in SSA's Listing of Impairments. Excluding these covariates avoids potentially biasing estimates of the judge and office parameters of primary interest.

²⁶Computing and time constraints required us to analyze a stratified probability sample of claims for this version of our analysis. We drew the sample from within strata formed by the cross-classification of judges and field offices. For strata with 100 claims or less, we sampled at a 100 percent rate. For strata with more than 100 claims, we sampled at a 20 percent rate or the rates that ensured at least 100 claims per stratum. The sample design caused selection probabilities to vary slightly across claims, depending on whether the claim was linked to a judge and office hearing smaller number of cases. We adjusted for unequal selection probabilities by including judge and office—the variables that determined sample selection—as covariates in our models.

Results

We provide the estimated distributions of allowance rates across judges, hearing offices, and primary diagnoses, holding all other covariates at their means, in table 6 below. Each row in the table lists results for one specification of the model described above. We derived quantiles of the distributions across groups with the data and estimated model parameters, using the methods above. The standard deviations of the allowance rates on the logit scale are explicit parameters in the model and were directly estimated with the fixed coefficients. We used these distributions to describe variation across judges, offices, and diagnoses in the body of this report and in figures, where we interpret the results in more detail.

Table 6: Estimated Distributions of Social Security Disability Allowance Rates across Judges, Hearing Offices, and Primary Diagnoses

Parameter	Raw Allowance Rate	Logit SD	Modeled Allowance Rate Distribution					
			5th Quantile	25th Quantile	Mean	75th Quantile	95th Quantile	
Judges								
Model 1: Intercepts Only	61.9	0.77	35.5	53.8	66.2	76.7	87.4	
Model 2: Add Covariates	61.9	0.72	42.1	59.3	70.2	79.3	88.5	
Model 3: Steps 4 and 5 Only	59.1	0.75	38.9	57.0	68.7	78.5	88.3	
Model 4: Stratified by Year								
2007	70.0	0.88	42.6	63.6	75.9	85.1	93.1	
2008	70.4	0.88	45.8	66.6	78.3	86.8	93.9	
2009	70.0	0.88	48.9	69.3	80.4	88.2	94.6	
2010	67.0	0.84	45.7	65.5	77.0	85.5	93.0	
2011	62.7	0.77	40.8	59.3	71.0	80.5	89.7	
2012	58.4	0.69	39.3	55.9	66.9	76.3	86.3	
2013	56.6	0.65	38.8	54.4	64.9	74.2	84.4	
2014	55.0	0.64	37.5	52.8	63.3	72.7	83.3	
2015	54.8	0.62	37.2	51.9	62.0	71.2	81.8	
Model 5: Stratified by Claim Type								
Concurrent	59.9	0.71	41.4	58.4	69.4	78.5	87.9	
DI	69.2	0.72	48.6	65.5	75.5	83.4	91.0	
SSI	53.3	0.73	31.2	47.8	59.9	70.9	83.1	
Model 6: Stratified by Year and Claim Type								

Appendix I: Objectives, Scope, and Methodology

Parameter	Raw Allowance Rate	Logit SD	Modeled Allowance Rate Distribution					
			5th Quantile	25th Quantile	Mean	75th Quantile	95th Quantile	
Concurrent 2007	68.2	0.87	41.3	62.1	74.7	84.2	92.5	
DI 2007	76.0	0.88	48.1	68.5	79.8	87.8	94.4	
SSI 2007	61.9	0.84	34.3	54.1	67.5	78.6	89.2	
Concurrent 2008	68.6	0.85	44.7	64.9	76.6	85.3	93.0	
DI 2008	76.5	0.87	53.5	72.8	82.8	89.6	95.3	
SSI 2008	62.3	0.88	34.7	55.5	69.3	80.3	90.5	
Concurrent 2009	68.2	0.87	48.1	68.3	79.5	87.5	94.2	
DI 2009	76.3	0.84	56.1	74.4	83.7	90.1	95.4	
SSI 2009	62.4	0.85	36.4	56.7	70.0	80.6	90.5	
Concurrent 2010	65.3	0.83	44.6	64.3	75.8	84.6	92.5	
DI 2010	73.5	0.81	52.2	70.6	80.6	87.8	94.0	
SSI 2010	60.1	0.84	35.7	55.6	68.8	79.5	89.8	
Concurrent 2011	61.0	0.76	40.5	58.7	70.3	79.8	89.2	
DI 2011	70.0	0.75	49.2	66.7	76.8	84.5	91.9	
SSI 2011	55.2	0.79	28.9	46.6	59.8	71.7	84.5	
Concurrent 2012	56.3	0.68	37.6	53.8	64.8	74.4	84.9	
DI 2012	66.5	0.69	47.6	63.9	73.8	81.7	89.7	
SSI 2012	49.9	0.72	27.4	43.1	55.1	66.6	80.0	
Concurrent 2013	54.3	0.64	36.7	51.9	62.4	71.9	82.6	
DI 2013	64.7	0.66	47.0	62.7	72.4	80.3	88.6	
SSI 2013	47.4	0.67	26.4	40.7	51.9	62.8	76.4	
Concurrent 2014	52.4	0.64	35.5	50.5	61.0	70.6	81.7	
DI 2014	63.3	0.64	44.7	60.2	70.0	78.3	87.1	
SSI 2014	45.0	0.65	26.2	40.1	51.0	61.8	75.3	
Concurrent 2015	51.8	0.61	35.2	49.6	59.8	69.2	80.3	
DI 2015	62.9	0.63	44.7	59.9	69.6	77.8	86.7	
SSI 2015	44.5	0.62	24.1	36.7	46.9	57.3	71.0	
Model 7: Endogeneity Check	61.9	0.73	40.1	57.6	68.9	78.3	88.0	
Hearing Offices								
Model 1: Intercepts Only	61.9	0.29	54.8	61.7	66.2	70.5	76.0	
Model 2: Add Covariates	61.9	0.27	60.1	66.3	70.2	74.0	78.7	
Primary Diagnoses								
Model 1: Intercepts Only	61.9	0.72	37.6	54.7	66.2	76.1	86.4	
Model 2: Add Covariates	61.9	0.71	42.5	59.5	70.2	79.2	88.3	

Source: GAO analysis of Social Security Administration (SSA) data. | GAO-18-37

Notes: Entries in the third column are estimates of the standard deviation of disability claim approval rates on the logit scale. Entries in all other columns are approval rates scaled in percentage points. The estimated distributions across groups apply to a claim that has the mean value of all other variables in our multivariate statistical models.

Table 7 below provides estimated odds-ratios of allowances for the factors other than judge, hearing office, and diagnosis in our primary model of ALJ allowance rates (Model 2 above), along with sample distributions and raw allowance rates. We used the primary model to support our findings in the body of this report, where we interpret the results below in more detail. Our model included variables that are measures or approximate measures for (1) claimant characteristics that represent criteria used in the disability decision-making process, (2) judge characteristics, (3) other participants in the decision-making process, (4) SSA administrative characteristics, and (5) economic characteristics of the claimant’s state. The interpretation of the odds ratio for a particular variable depends on whether the variable is a dummy variable or a categorical variable. For dummy variables, a statistically significant odds ratio that is greater/less than 1.00 indicates that claimants with that characteristic are more/less likely to be allowed than claimants without it. For categorical variables, a statistically significant odds ratio that is greater/less than 1.00 indicates that claimants in that category are more/less likely to be allowed than the claimants in the reference category.

Table 7: Sample Distributions and Estimates for Primary Model of Social Security Disability Allowance Rates across Judges

Variable	Sample Distribution (percent)	Raw Allowance Rate (percent)	Model Odds-Ratio (95 percent CI)
Claimant Factors			
Age At Disposition: 18 to 24	3.2	47.3	1.25 (1.23, 1.27)
Age At Disposition: 25 to 44	34.2	53.0	Ref
Age At Disposition: 45 to 49	18.9	58.6	1.46 (1.45, 1.47)
Age At Disposition: 50 to 54	21.6	66.6	2.36 (2.34, 2.38)
Age At Disposition: 55 or older	22.1	76.0	4.26 (4.22, 4.29)
Any Secondary Diagnosis: No	20.2	63.1	0.91 (0.90, 0.92)
Any Secondary Diagnosis: Yes	79.8	61.6	Ref
Critical Or Terminal Case: No	94.6	61.6	Ref
Critical Or Terminal Case: Yes	5.4	66.7	1.40 (1.38, 1.42)
Education: Less than High School	28.8	58.5	0.94 (0.94, 0.95)
Education: High School	47.3	61.7	Ref
Education: Some College	17.7	64.6	1.06 (1.05, 1.06)

Appendix I: Objectives, Scope, and Methodology

Variable	Sample Distribution (percent)	Raw Allowance Rate (percent)	Model Odds-Ratio (95 percent CI)
Education: College+	6.2	69.3	1.12 (1.11, 1.13)
Education: Missing	0.0	73.7	1.69 (1.64, 1.73)
GAO Claim Type: Concurrent	45.5	59.9	Ref
GAO Claim Type: DI	35.2	69.2	1.42 (1.41, 1.42)
GAO Claim Type: SSI	19.3	53.3	0.86 (0.86, 0.87)
Prior Work: 0-4 Years	18.1	68.9	0.77 (0.76, 0.78)
Prior Work: 5-9 Years	32.2	74.5	0.86 (0.86, 0.87)
Prior Work: 10-15 Years	49.8	80.2	Ref
Prior Work: Missing	0.0	52.3	0.23 (0.22, 0.23)
Had Representative: No	11.5	42.8	Ref
Had Representative: Yes	88.5	64.4	2.94 (2.92, 2.97)
Year Decision: 2007	7.1	70.0	1.98 (1.95, 2.02)
Year Decision: 2008	8.3	70.4	2.04 (2.01, 2.08)
Year Decision: 2009	10.0	70.0	2.07 (2.04, 2.11)
Year Decision: 2010	11.7	67.0	1.90 (1.87, 1.93)
Year Decision: 2011	12.9	62.7	1.66 (1.63, 1.69)
Year Decision: 2012	13.8	58.4	1.29 (1.28, 1.31)
Year Decision: 2013	13.3	56.6	1.18 (1.16, 1.19)
Year Decision: 2014	11.8	55.0	1.07 (1.06, 0.08)
Year Decision: 2015	11.1	54.8	Ref
Judge Factors			
Prior Experience: Attorney	18.6	54.8	Ref
Prior Experience: Misc. Not Attorney	0.3	52.9	0.74 (0.57, 0.95)
Prior Experience: Missing/Other	81.1	63.6	1.26 (1.18, 1.34)
Type: Career	99.0	61.9	Ref
Type: Nonpermanent	1.0	65.3	1.26 (1.09, 1.46)
Dispositions per Year: Less than 250	2.3	61.9	1.02 (1.00, 1.05)
Dispositions per Year: 250-499	27.9	60.2	0.97 (0.96, 0.98)
Dispositions per Year: 500-699	58.8	61.6	Ref
Dispositions per Year: 700+	11.0	67.5	1.01 (0.99, 1.02)
Cohort: Before 1965	0.0	76.8	1.07 (0.76, 1.52)
Cohort: 1965-1969	0.2	77.5	1.11 (0.81, 1.53)
Cohort: 1970-1974	1.1	71.3	1.25 (1.03, 1.51)
Cohort: 1975-1979	2.6	68.7	1.09 (0.93, 1.28)
Cohort: 1980-1984	5.2	62.9	0.95 (0.84, 1.07)
Cohort: 1985-1989	5.8	63.9	1.11 (0.98, 1.25)

Appendix I: Objectives, Scope, and Methodology

Variable	Sample Distribution (percent)	Raw Allowance Rate (percent)	Model Odds-Ratio (95 percent CI)
Cohort: 1990-1994	19.0	64.3	0.97 (0.89, 1.05)
Cohort: 1995-1999	21.5	63.2	Ref
Cohort: 2000-2004	15.4	62.0	0.92 (0.85, 1.00)
Cohort: 2005-2009	22.2	59.6	0.81 (0.76, 0.87)
Cohort: 2010 or Later	6.9	51.1	0.66 (0.61, 0.72)
Office Type: National Hearing Center	3.7	52.6	0.99 (0.80, 1.21)
Office Type: Other	0.2	50.7	0.75 (0.59, 0.96)
Office Type: Regular	96.2	62.3	Ref
Other Participants			
Medical Expert: No	87.6	60.8	Ref
Medical Expert: Yes	12.3	69.9	1.56 (1.54, 1.58)
Vocational Expert: No	14.8	70.9	1.25 (1.23, 1.26)
Vocational Expert: Yes	85.2	60.3	Ref
SSA Administrative Factors			
Video Conference: No	77.6	63.2	Ref
Video Conference: Yes	22.4	57.1	0.88 (0.87, 0.88)
Video Conference: Missing	0.0	86.9	4.13 (3.57, 4.79)
Office Cases Pending 270 or more days: <20 percent	8.8	60.4	0.95 (0.94, 0.97)
Office Cases Pending 270 or more days: 20-29 percent	20.3	61.8	0.98 (0.97, 0.99)
Office Cases Pending 270 or more days: 30-39 percent	26.4	60.5	0.96 (0.95, 0.97)
Office Cases Pending 270 or more days: 40-49 percent	24.5	62.5	Ref
Office Cases Pending 270 or more days: 50-59 percent	13.8	63.1	1.00 (0.98, 1.01)
Office Cases Pending 270 or more days: 60 percent +	6.2	64.0	1.03 (1.01, 1.04)
Office Cases Pending 270 or more days: Missing	0.0	68.1	1.16 (1.12, 1.20)
State and Economic Factors			
Had Reconsideration Stage: No	27.6	62.3	0.97 (0.96, 0.98)
Had Reconsideration Stage: Yes	72.4	61.7	Ref
Poverty Rate at Disposition: Less Than 13 percent	16.2	64.9	1.09 (1.07, 1.11)
Poverty Rate at Disposition: 13 to 15 percent	33.4	62.3	1.02 (1.01, 1.03)
Poverty Rate at Disposition: 16 to 18 percent	40.1	60.5	Ref
Poverty Rate at Disposition: 19 percent or More	10.3	61.4	0.95 (0.94, 0.97)
Unemployment at Disposition: Less than 5 percent	11.6	64.6	0.99 (0.97, 1.00)
Unemployment at Disposition: 5 to 6 percent	29.4	59.4	Ref
Unemployment at Disposition: 7 to 8 percent	32.2	60.2	1.01 (1.00, 1.02)
Unemployment at Disposition: 9 percent or More	26.8	65.4	1.04 (1.03, 1.05)
Unemployment at Disposition: Missing	0.0	57.3	0.68 (0.59, 0.79)

Source: GAO analysis of Social Security Administration (SSA) data. | GAO-18-37

Notes: The dependent variable is 1 if the claimant was allowed and 0 if the claimant was not allowed.
n = 3,211,517

Evaluation of SSA's Processes to Monitor Accuracy and Consistency in Hearings Decisions

For objective 2, we reviewed relevant federal laws, regulations, and documentation, and collected testimonial evidence from SSA officials to describe and evaluate the processes that SSA uses to monitor hearing decisions, detect variation, and improve accuracy and consistency. We interviewed SSA officials at different levels, including officials at headquarters, regional, DDS, and field office levels. We reviewed documents such as SSA's Hearings, Appeals, and Litigation Law (HALLEX) manual, policy memoranda issued by the Chief Administrative Law Judge, monitoring and quality assurance reports, user manuals and guides for electronic tools, SSA OIG reports, and descriptions of processes that are under development. We assessed these monitoring efforts against federal internal control standards and our management and evaluation guide for assessing fragmentation, overlap, and duplication in government programs.²⁷ We also reviewed SSA's annual performance plans from fiscal year 2006 through fiscal year 2017 to identify performance measures the agency has established to improve the accuracy and consistency of its hearings decisions. We evaluated the current performance measures using key attributes of performance measures used in prior GAO work²⁸ and federal internal control standards.²⁹ In addition to interviews with agency officials, as described above, we also interviewed officials from organizations representing judges, disability claimants, and representatives to obtain their

²⁷GAO, *Fragmentation, Overlap and Duplication: An Evaluation and Management Guide*, [GAO-15-49SP](#) (Washington, D.C.: April 14, 2015).

²⁸In a previous report, we developed nine attributes of performance goals and measures based on previously established GAO criteria, as well as relevant federal laws and performance management literature. See GAO, *Tax Administration: IRS Needs to Further Refine Its Tax Filing Season Performance Measures*, [GAO-03-143](#) (Washington, D.C.: Nov. 22, 2002).

²⁹GAO, *Standards for Internal Control in the Federal Government*, [GAO-14-704G](#) (Washington, D.C.: Sept. 10, 2014). See, for example, Principles 6, 7, 10, 13 and 16.

perspectives on SSA's efforts to monitor and improve accuracy and consistency.

Appendix II: Comments from the Social Security Administration



SOCIAL SECURITY
Office of the Commissioner

November 3, 2017

Mr. Charlie Jeszeck
Director, Education, Workforce, and Income Security Issues
United States Government Accountability Office
441 G Street, NW
Washington, DC 20548

Dear Mr. Jeszeck:

Thank you for the opportunity to review the draft report, "SOCIAL SECURITY DISABILITY: Additional Measures and Evaluation Needed to Enhance Accuracy and Consistency of Hearings" (GAO-18-37). Please see our enclosed comments.

If you have any questions, please contact me at (410) 965-9704. Your staff may contact Gary S. Hatcher, Senior Advisor for the Audit Liaison Staff, at (410) 965-0680.

Sincerely,

A handwritten signature in blue ink that reads "Stephanie Hall".

Stephanie Hall
Acting Deputy Chief of Staff

Enclosure

SOCIAL SECURITY ADMINISTRATION BALTIMORE, MD 21235-0001

COMMENTS ON THE GOVERNMENT ACCOUNTABILITY OFFICE'S DRAFT REPORT, "SOCIAL SECURITY DISABILITY: ADDITIONAL MEASURES AND EVALUATION NEEDED TO ENHANCE ACCURACY AND CONSISTENCY OF HEARINGS DECISIONS" (GAO-18-37)

Thank you for the opportunity to comment on the draft report. SSA manages one of the largest adjudicative systems in the world. At the hearing level alone, nearly 1,600 administrative law judges (ALJ) adjudicate more than 600,000 cases each year. The scope of our work is immense, and the work we do matters for the millions of seniors, people with disabilities, children, and widows and widowers who count on us for the fair and timely determination of their claims.

This audit seeks to address two broad issues: (1) the extent to which ALJ decisional outcomes vary across this broad group of adjudicators and the factors associated with the variation; and (2) the processes that the agency has put in place to monitor the accuracy and consistency of these decisions. As GAO acknowledged, there is complexity in the appeals process, and our ALJs enjoy qualified decisional independence. In recent years, the agency has implemented modern methods and technologies and utilized the full range of its regulatory authority to improve the oversight of hearing decisions.

Our oversight continues to evolve. We have increased the availability of policy guidance and feedback to the ALJs, improved training, captured and utilized data to gain a more comprehensive understanding of adjudication trends and challenges, and implemented additional oversight review processes that provide the agency and the ALJs with a more complete understanding of their decisional output. We conduct quarterly continuing education training sessions for ALJs and decision writers to address adjudication and policy issues, and the Chief ALJ issues memos about expectations and preparing quality decision writing instructions. For several years, various agency components have conducted oversight processes. Recent organizational changes have positioned us to evaluate and revise these reviews. Advanced technologies also offer an opportunity to address these issues in new ways.

Using data captured from the Appeals Council (AC) reviews – both request for review and the Division of Quality's samples for own motion review – the agency worked with the Administrative Conference of the United States to explore opportunities for policy modifications on compelling issues that cause difficulty for correct policy implementation by SSA adjudicators as well as federal courts. The AC also improved training for adjudicators. The training has proven so effective the AC received the W. Edwards Deming Outstanding Training Award in 2011 and again in 2015.

As GAO noted, in recent years, we have supplemented the long-standing AC oversight of the hearings process. These reviews include the inline quality review done by hearing-level regional quality review staff, which provide feedback to decisions writers prior to the issuance of ALJ decisions. Another quality assurance initiative performed by the Office of Appellate Operations in collaboration with the Office of the Chief ALJ is the post-effectuation focused review of both favorable and unfavorable decisions. No corrective action is taken on individual cases; rather, the program provides immediate feedback from a review of multiple cases. Subjects have

included the policy compliance of individual ALJs' work, claimant representatives, specific medical sources, and agency policies. As noted by GAO, the primary purpose of these reviews is to evaluate policy compliance and to provide individual feedback and training to ALJs. Our efforts to improve the quality and policy compliance of ALJ decisions have centered on our strong commitment to ensuring that the ALJs have access to timely feedback, training, and readily available and understandable policy guidance.

This oversight is working as evidenced by the decreased percentage of cases that have required corrective action by the AC, which results in fewer cases that require a new hearing and decision and, thus, better public service. The rate at which the AC grants review of denied claims to take corrective action or remand has also decreased, and we have seen improvement in the "agree rate." While we recognize that the agree rate is not a comprehensive quality measure, we have used this measure internally to help give an ALJ a simple guide with which to understand the increase or decrease in the percentage of cases in which the AC granted review of his or her appealed decisions over time.

We appreciate that GAO recognizes we have made progress in eliminating decision-making inconsistencies. There are now far fewer outliers as we provide adjudicators with better training and tools to support policy-compliant decision-making. This quality enhancement has resulted in significant public service improvements, which SSA's Office of the Inspector General has recognized in several reports.

We continue to build on our ability to issue timely, policy-compliant decisions. For example, we recently developed INSIGHT, a natural language processing and machine-learning tool that extracts text from a written hearing decision and analyzes the language for policy compliance, as well as consistency with other parts of the decision. INSIGHT processes this review in a matter of seconds and generates a report of its findings, highlighting potential problems. The report provides the adjudicator with convenient hyperlinks to problematic decisional language and links to policy guidance. INSIGHT identifies issues that contribute to remands at the AC-level and generates data about cases reviewed, which we can address with future training. The AC is currently using INSIGHT, and several hearing sites are testing it, with an eye toward full implementation.

Historically, several agency components have had some role in providing review and oversight responsibilities for hearing decisions. During the course of this audit, the Acting Commissioner of Social Security announced an organizational realignment effective October 1, 2017 to improve coordination of oversight of the disability adjudication process. The realignment integrates offices with complementary missions, including several of those identified in the current audit, that have a role in the oversight of ALJ decisions. These components now reside in the Office of Analytics, Review, and Oversight.

Given the aforementioned oversight we have in place to address the accuracy and consistency of ALJ decisions, we disagree with GAO's conclusion that "the agency's emphasis (is) on timeliness over accuracy." Memos issued by the Chief ALJ in December 2007 and March 2010 and repeated congressional hearing testimony do not support this conclusion. The referenced memos and testimony emphasize that, while ALJs should be aiming to release 500 - 700

dispositions per year, those dispositions should also be legally sufficient. Our commitment to the accuracy of hearings decisions is evidenced by our decision to devote significant resources to conducting multiple types of decisional quality review and our ongoing ALJ training programs.

We also have concerns with GAO's assertion on page 23 of the report related to hearing type. We appreciate GAO noting that this study did not seek to estimate the causal impact of video teleconferences on allowance rate and was not designed to account for all factors that could affect this relationship. As we discussed with GAO, our own internal analysis of the impact of video teleconferencing accounted for all but 0.6 percent of the difference in allowance rate between in-person and video teleconference hearings. Our study controlled for over 30 factors, but because we do not have data on all factors that might influence allowance rates, even the negligible difference we identified cannot be causally linked to the use of video teleconference. We are in the process of updating our internal analysis with full fiscal year 2017 data.

Lastly, we note that the methodology used to measure the variance of ALJ decisions focused primarily on ALJ allowance rates at the extremes of the distribution. We believe the report should highlight that currently there are far fewer ALJs with allowance rates at the extremes than there were previously. Additionally, the percentile ranking of ALJs by allowance rate appears to be un-weighted. If the ALJs were weighted in the distribution by the number of determinations, that would alter the 95th and 5th percentiles. This weighting could have a significant effect on the range and the trend in size of the range, due to the changing numbers of outlier judges.

Below are our responses to the recommendations. We also provided technical comments at the staff level.

Recommendation 1

Develop a set of public performance measures, to include accuracy and consistency, as well as timeliness, of administrative law judges' (ALJ) disability decisions. SSA could consider whether existing quality review or monitoring efforts could provide suitable data for such measures.

Response

We agree to consider establishing additional performance measures as part of our comprehensive assessment and refinement of the various oversight roles and processes, as discussed in our response to recommendation 2.

Recommendation 2

Systematically evaluate the efficiency and effectiveness of its quality assurance reviews and take steps to reduce or better manage any unnecessary overlap among them to ensure strategic use of resources. Such steps could include enhancing collaboration where reviews overlap or only conducting the reviews that are most efficient and effective in achieving agency goals for improving accuracy and consistency of ALJ disability decisions.

Response

We agree with GAO that our next step is a comprehensive assessment and refinement of the various oversight roles and processes. We are actively identifying opportunities to enhance our processes and agree with GAO that quality review should be coordinated at all levels of the claims process. Additionally, while our various efforts to ensure more policy-compliant ALJ decisions are working, this oversight is evolutionary.

Appendix III: GAO Contact and Staff Acknowledgments

GAO Contact

Charles Jeszeck, (202) 512-7215 or jeszeckc@gao.gov

Staff Acknowledgments

In addition to the individual named above, Erin M. Godtland, Assistant Director; Rachael Chamberlin, Analyst-in-Charge; Dana Hopings, LaToya King, Stephen Komadina, Rhiannon Patterson, and Jeff Tessin made significant contributions to the report. In addition, Daniel Bertoni, Deborah Bland, David Chrisinger, Melinda Cordero, Holly Dye, Bill Egar, Alex Galuten, Benjamin Licht, Serena Lo, Mimi Nguyen, Samuel Portnow, Sheila McCoy, and Shana Wallace made valuable contributions.

Appendix IV: Accessible Data

Data Tables

Data Table for Highlights figure, Estimated Allowance Rates across Social Security Administrative Law Judges for Typical Disability Claims, Fiscal Years 2007-2015

Fiscal year	5th percentile	25 th percentile	Average	75th percentile	95th percentile
2007	43%	64%	76%	85%	93%
2008	46%	67%	78%	87%	94%
2009	49%	69%	80%	88%	95%
2010	46%	65%	77%	85%	93%
2011	41%	59%	71%	81%	90%
2012	39%	56%	67%	76%	86%
2013	39%	54%	65%	74%	84%
2014	37%	53%	63%	73%	83%
2015	37%	52%	62%	71%	82%

Data Table for Figure 1: Five-Step Determination Process for Social Security Disability Claims

Step	Right arrow	Down arrow
Step One Is the claimant working and earning more than substantial gainful activity? ^a	Yes Denied	No Go to Step 2
Step Two Does the claimant have a severe impairment that significantly limits his or her ability to do basic work activities and that also meets the duration requirements?	No Denied	Yes Go to Step 3
Step Three Does the condition meet SSA's medical listings, or is the condition equal in severity to one found in the medical listings?	Yes Allowed	No Go to Step 4
Step Four Can the claimant perform any of his or her past work?	Yes Denied	No Go to step 5

Step	Right arrow	Down arrow
Step Five	Yes	No
Can the claimant perform other work that exists in significant numbers in the national economy?	Denied	Allowed

Source: GAO analysis of Social Security Administration (SSA) information. | GAO-18-37

Data Table for Figure 2: Social Security Disability Appeals Process

Phase	Step
<p>Initial Disability Determination</p> <p>More than 2.5 million disability claims filed in fiscal year 2016</p>	<p>Claimant applies for benefits</p> <p>Application requires claimant’s medical history and other information necessary to determine eligibility</p> <p>Social Security Administration (SSA) field office screens application</p> <p>SSA field office staff determine if claimant meets SSA’s non-medical eligibility requirements and forwards the application if those eligibility requirements are met</p> <p>If claimant meets requirements proceed to next step “Initial determination”</p> <p>Initial determination</p> <p>Staff at a state-run Disability Determination Services (DDS) office gather, develop, and review the claimant’s medical and vocational evidence to make a disability determination</p> <p>If determination is appealed proceed to next step “Reconsideration”.</p>
<p>Disability Appeals Process</p> <p>More than 633,000 appeals to reconsider the initial decision in fiscal year 2016</p> <p>More than 698,000 requests for an ALJ hearing in fiscal year 2016</p> <p>More than 133,000 requests for Appeals Council review in fiscal year 2016</p>	<p>Reconsideration ^a</p> <p>A different group of DDS staff examines previous and any new medical and vocational evidence to make a second disability determination</p> <p>If reconsideration is appealed proceed to next step, “Admin Law judge hearing”.</p> <p>Administrative Law Judge (ALJ) holds hearing</p> <p>Hearing office staff collect any new evidence for review by an ALJ who generally conducts a hearing before rendering a new decision</p> <p>If decision is appealed proceed to next step “Appeals Counsel”</p>

Phase

Step

Appeals Counsel ^b

Administrative appeals judges can issue a decision after reviewing the claim, or return the claim to the ALJ level for a new decision

Source: GAO analysis of Social Security Administration (SSA) information. | GAO-18-37

Data Table for Figure 3: Social Security Administration (SSA) Requests for Administrative Law Judge (ALJ) Hearings, Case Dispositions, and Pending Cases, Fiscal Years 2010-2016

Fiscal year	Requests for hearings	Dispositions	Pending cases
2010	720,161	737,616	705,367
2011	859,514	793,563	771,318
2012	849,869	820,484	816,575
2013	824,989	793,580	847,984
2014	810,715	680,963	977,736
2015	746,300	663,129	1,060,907
2016	699,000	637,000	1,122,000

Data Table for Figure 4: Variation in Social Security Administration (SSA) Disability Allowance Rates across Administrative Law Judges, Fiscal Years 2007-2015

Fiscal year	5 th percentile	25 th percentile	Average	75 th percentile	95 th percentile
2007	40	59	69	82	93
2008	40	60	70	83	96
2009	39	59	70	82	95
2010	38	56	66	78	92
2011	33	50	62	74	89
2012	32	47	58	69	86
2013	29	45	56	67	84
2014	27	44	55	65	81
2015	29	44	55	66	82

Data Table for Figure 5: Estimated Variation in Social Security Administration (SSA) Allowance Rates across Judges for Typical Claims, Fiscal Years 2007-2015

Fiscal year	5 th percentile	25 th percentile	Average	75 th percentile	95 th percentile
2007	43%	64%	76%	85%	93%
2008	46%	67%	78%	87%	94%
2009	49%	69%	80%	88%	95%
2010	46%	65%	77%	85%	93%

Fiscal year	5th percentile	25th percentile	Average	75th percentile	95th percentile
2011	41%	59%	71%	81%	90%
2012	39%	56%	67%	76%	86%
2013	39%	54%	65%	74%	84%
2014	37%	53%	63%	73%	83%
2015	37%	52%	62%	71%	82%
2007	42%	59%	70%	79%	88%

Data Table for Figure 6: Estimated Variation in Social Security Administration (SSA) Disability Allowance Rates across Hearing Offices for Typical Claims, Fiscal Years 2007-2015

Fiscal year	5 th percentile	25 th percentile	Mean	75 th percentile	95 th percentile
2007	63	71	76	80	85
2008	67	74	78	82	87
2009	68	76	80	84	89
2010	64	72	77	81	86
2011	58	66	71	76	81
2012	54	62	67	72	78
2013	52	60	65	70	76
2014	50	58	63	68	75
2015	50	57	62	67	73

Data Table for Figure 7: Social Security Administration (SSA) Disability Allowance Rates for the 25 Most Commonly Recorded Primary Impairments, Fiscal Years 2007-2015

Impairment	Allowance rate (Unadjusted)
Intellectual disorder	88
Multiple sclerosis	80
Heart failure	78
Peripheral neuropathy	77
Vascular insult to the brain	76
Schizophrenia spectrum (and other psychotic disorders)	76
Inflammatory arthritis	75
Neurocognitive disorders	73
Chronic ischemic heart disease (with or without angina)	71
Chronic liver disease	68
Chronic respiratory disorders	66
Osteoarthritis and allied disorders	64

Impairment	Allowance rate (Unadjusted)
Disorders of back (discogenic and degenerative)	63
Fractures of lower limb	62
Diabetes mellitus	60
Disorders of muscle, ligament, and fascia	59
Anxiety and obsessive-compulsive disorders	59
Depression, bipolar, and related disorders	59
Other and unspecified arthropathies	58
Fibromyalgia	58
Epilepsy	50
Essential hypertension	47
Obesity	46
Borderline intellectual functioning	44
Asthma	44
All other impairments	61

Data Table for Figure 8: Four Social Security Administration (SSA) Offices Responsible for Quality Assurance Reviews of Hearings-level Decisions, as of September 2017

- 1) Commissioner
 - a) General Counsel
 - b) Deputy Commissioner, Budget, Finance, Quality and Management
 - i) Office of Quality Review
 - c) Deputy Commissioner, Disability Adjudication and Review
 - i) Office of the Chief
 - ii) Administrative Law Judge
 - iii) Regional Chief
 - iv) Administrative Law Judges
 - (1) Regional Quality Review Officers and Staff

-
- v) Office of Appellate Operations/Appeals Council
 - (1) Division of Quality
 - (2) Disability Branches
 - d) Deputy Commissioner, Operations
 - e) Chief Strategic Officer

Source: GAO analysis of Social Security Administration (SSA) information. | GAO-18-37

Agency Comment Letter

Text of Appendix II: Comments from the Social Security Administration

Page 1

November 3, 2017

Mr. Charlie Jeszeck

Director, Education, Workforce, and Income Security Issues United States Government Accountability Office

441 G Street, NW Washington, DC 20548

Dear Mr. Jeszeck:

Thank you for the opportunity to review the draft report, "SOCIAL SECURITY DISABILITY: Additional Measures and Evaluation Needed to Enhance Accuracy and Consistency of Hearings" (GAO-18-37). Please see our enclosed comments.

If you have any questions, please contact me at (410) 965-9704. Your staff may contact Gary S. Hatcher, Senior Advisor for the Audit Liaison Staff, at (410) 965-0680.

Sincerely,

Stephanie Hall

Acting Deputy Chief of Staff

Page 2

COMMENTS ON THE GOVERNMENT ACCOUNTABILITY OFFICE'S DRAFT REPORT, "SOCIAL SECURITY DISABILITY: ADDITIONAL MEASURES AND EVALUATION NEEDED TO ENHANCE ACCURACY AND CONSISTENCY OF HEARINGS DECISIONS" (GAO-18-37)

Thank you for the opportunity to comment on the draft report. SSA manages one of the largest adjudicative systems in the world. At the hearing level alone, nearly 1,600 administrative law judges (ALJ) adjudicate more than 600,000 cases each year. The scope of our work is immense, and the work we do matters for the millions of seniors, people with disabilities, children, and widows and widowers who count on us for the fair and timely determination of their claims.

This audit seeks to address two broad issues: (1) the extent to which ALJ decisional outcomes vary across this broad group of adjudicators and the factors associated with the variation; and (2) the processes that the agency has put in place to monitor the accuracy and consistency of these decisions. As GAO acknowledged, there is complexity in the appeals process, and our ALJs enjoy qualified decisional independence. In recent years, the agency has implemented modern methods and technologies and utilized the full range of its regulatory authority to improve the oversight of hearing decisions.

Our oversight continues to evolve. We have increased the availability of policy guidance and feedback to the ALJs, improved training, captured and utilized data to gain a more comprehensive understanding of adjudication trends and challenges, and implemented additional oversight review processes that provide the agency and the ALJs with a more complete understanding of their decisional output. We conduct quarterly continuing education training sessions for ALJs and decision writers to address adjudication and policy issues, and the Chief ALJ issues memos about expectations and preparing quality decision writing instructions. For several years, various agency components have conducted oversight processes. Recent organizational changes have positioned us to evaluate and revise these reviews. Advanced technologies also offer an opportunity to address these issues in new ways.

Using data captured from the Appeals Council (AC) reviews - both request for review and the Division of Quality's samples for own motion

review - the agency worked with the Administrative Conference of the United States to explore opportunities for policy modifications on compelling issues that cause difficulty for correct policy implementation by SSA adjudicators as well as federal courts. The AC also improved training for adjudicators. The training has proven so effective the AC received the W. Edwards Deming Outstanding Training Award in 2011 and again in 2015.

As GAO noted, in recent years , we have supplemented the long-standing AC oversight of the hearings process. These reviews include the inline quality review done by hearing-level regional quality review staff, which provide feedback to decisions writers prior to the issuance of ALJ decisions . Another quality assurance initiative performed by the Office of Appellate Operations in collaboration with the Office of the Chief ALJ is the post-effectuation focused review of both favorable and unfavorable decisions. No corrective action is taken on individual cases; rather, the program provides immediate feedback from a review of multiple cases. Subjects have

Page 3

included the policy compliance of individual ALJs' work, claimant representatives, specific medical sources, and agency policies. As noted by GAO, the primary purpose of these reviews is to evaluate policy compliance and to provide individual feedback and training to ALJs. Our efforts to improve the quality and policy compliance of ALJ decisions have centered on our strong commitment to ensuring that the ALJs have access to timely feedback, training, and readily available and understandable policy guidance.

This oversight is working as evidenced by the decreased percentage of cases that have required corrective action by the AC, which results in fewer cases that require a new hearing and decision and, thus, better public service. The rate at which the AC grants review of denied claims to take corrective action or remand has also decreased, and we have seen improvement in the "agree rate." While we recognize that the agree rate is not a comprehensive quality measure, we have used this measure internally to help give an ALJ a simple guide with which to understand the increase or decrease in the percentage of cases in which the AC granted review of his or her appealed decisions over time.

We appreciate that GAO recognizes we have made progress in eliminating decision-making inconsistencies. There are now far fewer

outliers as we provide adjudicators with better training and tools to support policy-compliant decision-making. This quality enhancement has resulted in significant public service improvements, which SSA's Office of the Inspector General has recognized in several reports.

We continue to build on our ability to issue timely, policy-compliant decisions. For example, we recently developed INSIGHT, a natural language processing and machine-learning tool that extracts text from a written hearing decision and analyzes the language for policy compliance, as well as consistency with other parts of the decision. INSIGHT processes this review in a matter of seconds and generates a report of its findings, highlighting potential problems. The report provides the adjudicator with convenient hyperlinks to problematic decisional language and links to policy guidance. INSIGHT identifies issues that contribute to remands at the AC-level and generates data about cases reviewed, which we can address with future training. The AC is currently using INSIGHT, and several hearing sites are testing it, with an eye toward full implementation.

Historically, several agency components have had some role in providing review and oversight responsibilities for hearing decisions. During the course of this audit, the Acting Commissioner of Social Security announced an organizational realignment effective October 1, 2017 to improve coordination of oversight of the disability adjudication process. The realignment integrates offices with complementary missions, including several of those identified in the current audit, that have a role in the oversight of ALJ decisions. These components now reside in the Office of Analytics, Review, and Oversight.

Given the aforementioned oversight we have in place to address the accuracy and consistency of ALJ decisions, we disagree with GAO's conclusion that "the agency's emphasis (is) on timeliness over accuracy." Memos issued by the Chief ALJ in December 2007 and March 2010 and repeated congressional hearing testimony do not support this conclusion. The referenced memos and testimony emphasize that, while ALJs should be aiming to release 500 – 700

Page 4

dispositions per year, those dispositions should also be legally sufficient. Our commitment to the accuracy of hearings decisions is evidenced by our decision to devote significant resources to conducting multiple types of decisional quality review and our ongoing ALJ training programs.

We also have concerns with GAO's assertion on page 23 of the report related to hearing type. We appreciate GAO noting that this study did not seek to estimate the causal impact of video teleconferences on allowance rate and was not designed to account for all factors that could affect this relationship. As we discussed with GAO, our own internal analysis of the impact of video teleconferencing accounted for all but 0.6 percent of the difference in allowance rate between in-person and video teleconference hearings. Our study controlled for over 30 factors, but because we do not have data on all factors that might influence allowance rates, even the negligible difference we identified cannot be causally linked to the use of video teleconference. We are in the process of updating our internal analysis with full fiscal year 2017 data.

Lastly, we note that the methodology used to measure the variance of ALJ decisions focused primarily on ALJ allowance rates at the extremes of the distribution. We believe the report should highlight that currently there are far fewer ALJs with allowance rates at the extremes than there were previously. Additionally, the percentile ranking of ALJs by allowance rate appears to be un-weighted. If the ALJs were weighted in the distribution by the number of determinations, that would alter the 95th and 5th percentiles. This weighting could have a significant effect on the range and the trend in size of the range, due to the changing numbers of outlier judges.

Below are our responses to the recommendations. We also provided technical comments at the staff level.

Recommendation 1

Develop a set of public performance measures, to include accuracy and consistency, as well as timeliness, of administrative law judges' (ALJ) disability decisions. SSA could consider whether existing quality review or monitoring efforts could provide suitable data for such measures.

Response

We agree to consider establishing additional performance measures as part of our comprehensive assessment and refinement of the various oversight roles and processes, as discussed in our response to recommendation 2.

Recommendation 2

Systematically evaluate the efficiency and effectiveness of its quality assurance reviews and take steps to reduce or better manage any unnecessary overlap among them to ensure strategic use of resources. Such steps could include enhancing collaboration where reviews overlap or only conducting the reviews that are most efficient and effective in achieving agency goals for improving accuracy and consistency of ALJ disability decisions.

Page 5

Response

We agree with GAO that our next step is a comprehensive assessment and refinement of the various oversight roles and processes. We are actively identifying opportunities to enhance our processes and agree with GAO that quality review should be coordinated at all levels of the claims process. Additionally, while our various efforts to ensure more policy-compliant ALJ decisions are working, this oversight is evolutionary.

GAO's Mission

The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's website (<http://www.gao.gov>). Each weekday afternoon, GAO posts on its website newly released reports, testimony, and correspondence. To have GAO e-mail you a list of newly posted products, go to <http://www.gao.gov> and select "E-mail Updates."

Order by Phone

The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's website, <http://www.gao.gov/ordering.htm>.

Place orders by calling (202) 512-6000, toll free (866) 801-7077, or TDD (202) 512-2537.

Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.

Connect with GAO

Connect with GAO on [Facebook](#), [Flickr](#), [LinkedIn](#), [Twitter](#), and [YouTube](#).
Subscribe to our [RSS Feeds](#) or [E-mail Updates](#). Listen to our [Podcasts](#).
Visit GAO on the web at www.gao.gov and read [The Watchblog](#).

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Website: <http://www.gao.gov/fraudnet/fraudnet.htm>
E-mail: fraudnet@gao.gov
Automated answering system: (800) 424-5454 or (202) 512-7470

Congressional Relations

Orice Williams Brown, Managing Director, WilliamsO@gao.gov, (202) 512-4400,
U.S. Government Accountability Office, 441 G Street NW, Room 7125,
Washington, DC 20548

Public Affairs

Chuck Young, Managing Director, youngc1@gao.gov, (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, DC 20548

Strategic Planning and External Liaison

James-Christian Blockwood, Managing Director, spel@gao.gov, (202) 512-4707
U.S. Government Accountability Office, 441 G Street NW, Room 7814,
Washington, DC 20548