



Administrative Conference of the United States

**DATA APPENDIX:
A STUDY OF SOCIAL SECURITY LITIGATION IN THE
FEDERAL COURTS**

Final Data Appendix: July 28, 2016

Jonah Gelbach
University of Pennsylvania

David Marcus
University of Arizona

This data appendix was prepared for the consideration of the Administrative Conference of the United States. The opinions, views and recommendations expressed are those of the authors and do not necessarily reflect those of the members of the Conference or its committees, except where formal recommendations of the Conference are cited.

We used several data sources for the quantitative aspects of this report.

1. Data from the Office of General Counsel (OGC)

The first source was provided by SSA and includes records providing, *inter alia*, the following information about cases:

- State and U.S. district court divisional office name where the case was filed.
- A numerical case number that is part of the docket number as kept by each district court pursuant to the rules of the Administrative Office of the U.S. Courts (“AO”).
- Information on whether the case was remanded or whether the agency’s denial was affirmed in full.

These data were drawn from a database administered by OGC, and we will refer to this data source as the OGC data set. Because it provides information on case outcomes in a single file, the OGC data set was the primary one we used concerning the results of disability appeals. We dropped cases in the OGC data for which we could not identify either the year the case was filed, the district court in which it was filed, or the case number assigned to the case. The remaining data set had the following numbers of cases by year filed:

- 2010: 12,097 cases
- 2011: 12,848 cases
- 2012: 14,945 cases
- 2013: 11,536 cases

These cases were drawn from 85 of the 94 U.S. district courts. The districts unrepresented in this set of cases were: Southern District of Alabama; Guam; Eastern District of Louisiana; Middle District of Louisiana; Northern Mariana Islands; Utah; Virgin Islands; Eastern District of Wisconsin; and Western District of Wisconsin.

2. Data from the Office of Disability Adjudication and Review (ODAR)

The agency also provided us with information from the Office of Disability Adjudication and Review (“ODAR”). These data arrived in several different files.

One of these contained information on what the agency calls “New Court Cases”. As the agency explained to us, these are “the result of new civil lawsuit filings by claimants seeking Social Security or Supplemental Security Income benefits after denial through the Appeals Council level.”¹ This information in this file included a case’s district court docket number, the judicial district in which it was filed, the year in which it was filed, the hearing office that handled the claim at the ALJ level, and several other variables we requested but ultimately determined we did not need to or could not use for this report. The New Court Cases data file did not contain information on how cases were resolved.

¹ See file “ACUS Data Description v1.0.doc” provided as part of the second data feed from the agency (henceforth the “Data Description file”), at 1.

A second file included information on what the agency calls “Court Remand Cases”. These are “claims for benefits returned to the Social Security Administration by the courts after court review for further administrative action by SSA.”² By design, the Court Remand Cases file does not contain information on cases that were not remanded. To use this file to determine remand rates would require both (i) merging cases in this file to cases in the Court Remand Cases file, and (ii) knowing the status of those cases in the New Court Cases file that did not successfully merge.

While the Court Remand Cases file contains information that should in principle allow merging with cases in the New Court Cases file, we have no way of knowing whether cases that do not match are still pending or were affirmed. Especially for later years in the data, this problem left us concerned about the accuracy—and thus the usefulness—of using the two ODAR data files for most aspects of the quantitative analysis. In addition, the “CLMT_ID” variable that was present in both files did not uniquely identify cases.³

We did use the New Court Cases data file for some specialized purposes in Part IV.B. First, this file identifies the hearing office through which each district court appeal traveled, so we were able to use it to determine the fraction of cases filed in district courts that came from each hearing office. We used such information to link variables on ALJ productivity and hearing office award rates to districts for purposes of our second-step regression analysis. Second, we used information from this file in our qualitative analysis in Part IV.D, to determine the fraction of cases in which claimants were represented in two districts described therein.

SSA also provided us with information on the number of RVRs issued by district court and year. We used these counts, together with information on the number of decisions in each district and year, to generate an RVR rate.

3. Docket Information

The second data source is docket reports originally drawn from the Public Access to Court Electronic Records (“PACER”) system by Westlaw, and then provided to one of us pursuant to a contract agreement with Thomson Reuters, owner of Westlaw. Among other things, these docket reports provide:

- (i) The name of the U.S. district court where the case was filed, and, often, the name of the office where the case was filed.
- (ii) The full case docket number kept by the AO, which includes a number corresponding to the code for each district court divisional office.
- (iii) The PACER “nature of suit” code. Social Security disability appeal cases have the value 863 or 864 for this field.⁴
- (iv) Text indicating whether any party was represented or proceeded *pro se*.
- (v) The name of the Article III judge and/or magistrate judge assigned to the case.

² Data Description file, at 1.

³ We reached out to the agency for help with these files but were told that the IT employee who constructed the data files had changed jobs.

⁴ See <https://www.pacer.gov/documents/natsuit.pdf>.

For our quantitative investigation of judge-specific variables in Part IV.A, we used this information to match case data from docket reports to the OGC data set described above. There were 51,445 cases in the OGC data that were filed between 2010 and 2013 and for which we believed we were able to identify the district court in which the case were filed. Of these, 2,995 cases had docket sequence numbers equal to the docket sequence number for at least one other case filed in the same year and, from what we could tell, district court. We dropped these 2,995 “duplicate” cases, leaving us with 48,450 uniquely identified cases in the OGC data for filing years 2010-2013. Using our dockets data, we were able to match a judge to 46,078 of these 48,450, which is a match rate of 95.1%. Thus we were able to match a judge to just over 95% of cases in the OGC data that we sought to match.⁵

4. Other Data Used

The sources of other data we used are as follows:

- (a) **District-level caseload pressures.** We measured the importance of caseload pressures with two variables. The first is the share of all civil cases in the district that are disability appeals. We obtained data on both the numerator and denominator of this variable from several years of the Statistical Tables For The Federal Judiciary’s Table C-3, “U.S. District Courts—Civil Cases Commenced, by Nature of Suit and District”. The second was the number of pending cases—whether civil or criminal—per congressionally approved Article III judgeship. We obtained both the numerator and denominator of this Federal Court Management Statistics tables titled “Comparison of Districts Within the Circuits”.
- (b) **Judicial trust in the federal government.** We obtained data on the share of criminal sentences in a district in which judges sentenced criminal defendants to more lenient terms than given by federal sentencing guidelines. These data are available in the “Individual Offender Datafiles”, which we downloaded from <http://www.ussc.gov/research-and-publications/commission-datafiles#individual>.
- (c) **District-level variables constructed in an effort to measure the performance of hearing offices within the district.** We used ODAR data posted on the SSA website to compile hearing office-level values of the award rate and the number of dispositions per day per ALJ. We downloaded this information from https://www.ssa.gov/appeals/DataSets/archive/archive_data_reports.html. We then had to match the information provided to the hearing office provided in cases in the ODAR data. To do so we undertook the following steps, which we illustrate with the hearing office award rate variable we constructed:

⁵ We note that many more cases in the dockets data did not match a case in the OGC data (there were a total of 62,129 cases in the dockets data, for a match rate of $46,078/62,129=0.742$ —just under 75%). There are two reasons why this is not surprising. First, the dockets data included both terminated and still-open civil actions, whereas our OGC data include only terminated cases. Second, as noted above we were not able to identify the district for all cases in the OGC data.

1. For each judicial district, we calculated the share of cases in the ODAR data that were listed as being assigned to each hearing office.
2. We then used the shares from step 1 to calculate the weighted average of the 2010 hearing office-level award rate among cases filed in each judicial district.

For example, suppose that in a given district, there were 100 cases filed in hearing office 1 and 300 cases filed in hearing office 2. Then the step-1 weight for hearing office 1 would be 0.25 (100 divided by (100+300)), while the step-2 weight would be 0.75 (300 divided by (100+300)). If the 2010 award rates were 10% and 20% in hearing offices 1 and 2, respectively, then the judicial district's weighted average of the 2010 award rate would be $(0.25 \times 10\% + 0.75 \times 20\%)$, which is 17.5%.

We then matched the resulting judicial district-level information to our data set of district-level remand rates based on the OGC data.

- (d) Labor market conditions in the years preceding a disability appeal.** We obtained data on state employment from the website of the Bureau of Labor Statistics; our data come from the file <http://download.bls.gov/pub/time.series/sm/sm.data.55.TotalNonFarmStateWide.All>. Annual estimates for state-level population were downloaded in two files from the Bureau of the Census website: <http://www.census.gov/popest/data/intercensal/state/tables/ST-EST00INT-01.csv> and <http://www.census.gov/popest/data/state/totals/2013/tables/NST-EST2013-01.csv>.
- (e) Salaries of lawyers.** Data on lawyer salaries for 2014 are available for download at <http://www.bls.gov/oes/special.requests/oesm14st.zip>.
- (f) Urbanization.** Information about county-level urbanization and population was downloaded from http://www.ers.usda.gov/datafiles/RuralUrban_Continuum_Codes/ruralurbancodes2013.xls and then the county-to-district relationship (data were available at <http://www.justice.gov/ust/locate-your-judicial-district/Alaska-Wyoming#KY>) was used to form district-level measures of average county urbanization.