The Appeals Lottery: Veterans Law Judge Discretion at America's Largest Federal Tribunal

APEP Autonomous Research* @olafdrw

April 2, 2026

Abstract

Whether a veteran wins a disability benefits appeal depends substantially on which judge hears the case. I exploit the quasi-random assignment of cases to 106 Veterans Law Judges at the Board of Veterans Appeals—the largest federal adjudicatory body in the United States—to construct a judge-leniency instrument for appeal outcomes. Using 11,422 parsed decision texts from FY2017–2018, I document a first-stage F-statistic of 225, with leave-one-out leniency strongly predicting grants. Balance tests confirm random assignment (joint $p = 0.999$). Judge discretion is greatest for subjective claims: a one-standard-deviation increase in leniency raises the grant probability by 10 percentage points for mental health appeals versus 4 points for increased-rating claims. These findings establish the first credible instrument at the VA appellate stage.

JEL Codes: H55, J38, K41

Keywords: disability benefits, judge leniency, instrumental variables, veterans, administrative adjudication

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 30m).

1. Introduction

A veteran denied disability benefits by the Department of Veterans Affairs has one path left: an appeal to the Board of Veterans Appeals. The BVA is the largest federal adjudicatory body in the United States, issuing over 80,000 decisions per year on claims worth up to $3,700 per month in lifetime compensation plus comprehensive VA healthcare. The stakes are first-order—disability benefits represent the primary income source for many disabled veterans—yet the appeals process has received almost no attention from economists.

This paper documents a striking fact: which Veterans Law Judge hears a case matters enormously for whether the appeal succeeds. I construct a new dataset by downloading and parsing 12,404 BVA decision text files from FY2017–2018, extracting the judge's identity from the signature block, the decision outcome from the ORDER section, and case characteristics from the headers. The resulting panel covers 106 judges deciding 11,422 substantive appeals across 83 regional offices and six issue categories.

The identification strategy follows the judge-leniency instrumental variables framework pioneered by Kling (2006) and applied to Social Security disability by Maestas et al. (2013). The BVA's Caseflow Automatic Case Distribution (ACD) system assigns cases to Veterans Law Judges based on docket order and availability—not case characteristics. I construct leave-one-out leniency measures: each case's instrument is the assigned judge's grant rate calculated over all other cases. The first-stage F-statistic is 225 in the preferred specification with year, regional office, and issue-category fixed effects, far exceeding conventional thresholds for instrument strength (Angrist et al., 1996).

The design passes standard diagnostic tests with flying colors. Balance tests show that leniency does not predict case characteristics: the joint F-test for predicting issue type and case complexity yields $p = 0.999$. Leave-one-out judge sensitivity reveals near-perfect stability—dropping any single judge changes the first-stage coefficient by less than 0.006. Alternative leniency constructions, including cross-issue and cross-year measures, all produce strong first stages ($F > 21$). A placebo test using remand decisions (procedural returns to the regional office, not substantive merit rulings) shows that lenient judges remand *less*, consistent with the interpretation that they resolve cases on the merits rather than deferring.

The most informative finding is that judge discretion varies systematically by claim type. I decompose the first stage into four issue categories and find that mental health claims exhibit the largest judge effects: a one-standard-deviation increase in leniency raises the grant probability by approximately 10 percentage points, with a near-unit coefficient ($\hat{\beta} = 0.997$). Service connection claims show a strong but somewhat smaller effect ($\hat{\beta} = 0.805$), while increased-rating claims—which are more mechanically determined by medical evidence—show

the weakest judge dependence ($\hat{\beta} = 0.407$). This gradient—strongest effects where medical evidence is most subjective—represents what I call the *subjectivity premium* in the appeals lottery: a veteran's fate depends most on the judge draw precisely when the decision requires the most judgment.

This paper contributes to three literatures. First, it extends the judge/examiner leniency IV literature to a new institutional setting. While Silver and Zhang (2026) exploit quasi-random assignment to Compensation & Pension medical examiners at the *initial claim* stage, I study a different decision-maker (lawyer versus doctor), different stage (appeal versus initial review), and different margin (veterans whose claims were already denied by regional offices). The complier population—veterans at the appellate margin—is policy-relevant for reform debates about streamlining the BVA, which has a backlog exceeding 100,000 cases. Second, it contributes to the broader literature on administrative discretion and its consequences (Kleinberg et al., 2018; Chen et al., 2016), documenting that discretion scales with the subjectivity of the underlying determination. Third, it provides the first-stage foundation for future work linking BVA appeal outcomes to veteran mortality, employment, and housing stability using linked administrative data.

The judge leniency design has proven enormously productive in criminal justice (Dobbie et al., 2018; Aizer and Doyle, 2015; Bhuller et al., 2020; Garin et al., 2025; Arnold et al., 2022), immigration (Ramji-Nogales et al., 2007), patents (Galasso and Schankerman, 2015; Sampat and Williams, 2019), and disability insurance (Maestas et al., 2013; Dahl et al., 2014; Autor et al., 2019). Yet the VA appellate system—despite its massive scale, publicly accessible decision corpus, and documented random assignment mechanism—has been entirely overlooked. A key advantage of this setting is transparency: BVA decisions are published as plain-text files with standardized formatting, allowing researchers to construct judge leniency measures without Freedom of Information Act requests or restricted-use data agreements. The 294 VLJs identified in FY2017–2018 alone provide rich variation for IV estimation.

The paper proceeds as follows. Section 2 describes the BVA's institutional setting and the Caseflow ACD assignment system. Section 3 presents the data construction. Section 4 details the empirical strategy and identification assumptions. Section 5 reports the main results, heterogeneity analysis, and robustness checks. Section 6 discusses implications and concludes.

2. Institutional Background

The VA Disability Compensation Program. The Department of Veterans Affairs administers the nation's largest disability benefits program, distributing over $130 billion

annually to approximately 6 million veterans (Autor et al., 2016). Eligible veterans receive monthly tax-free compensation ranging from approximately $170 (10% disability rating) to over $3,700 (100% rating), plus access to VA healthcare, vocational rehabilitation, and housing assistance. The program is means-tested only at the entry margin—once rated, benefits are not reduced based on earnings—creating strong incentives for participation (Autor and Duggan, 2003, 2006; Coile et al., 2021).

The Claims and Appeals Process. A veteran files an initial disability claim with one of 56 VA Regional Offices (ROs). A Rating Veterans Service Representative (RVSR) evaluates medical evidence and assigns a disability rating. If denied or rated below expectations, the veteran may appeal to the Board of Veterans Appeals in Washington, D.C. Under the Appeals Modernization Act of 2017, appellants choose among three "lanes": direct review, new evidence, or hearing. All lanes result in a decision by a Veterans Law Judge.

Veterans Law Judges and Case Assignment. Veterans Law Judges are appointed under 5 U.S.C. §3105, which requires that administrative law judges "be assigned to cases in rotation." The BVA implements this through the Caseflow Automatic Case Distribution (ACD) system, developed by the U.S. Digital Service and documented in a public GitHub repository. The ACD distributes cases from the general docket to available VLJs based on docket order and current caseload—critically, *not* on case characteristics, issue type, or regional office of origin. Veterans and their representatives cannot request or select a specific judge.

This institutional feature—quasi-random assignment within the general docket—is the source of identifying variation. The mechanism is analogous to the random case assignment exploited in criminal courts (Kling, 2006; Dobbie et al., 2018), immigration courts (Ramji-Nogales et al., 2007), and patent tribunals (Galasso and Schankerman, 2015), but at a scale that dwarfs most settings: the BVA decided approximately 85,000 cases per year at peak volume, with over 80 active VLJs.

Decision Outcomes. A VLJ may *grant* the appeal (awarding or increasing benefits), *deny* it (affirming the RO's decision), or *remand* it (returning the case to the RO for additional development). Grants and denials are substantive merit decisions reflecting the VLJ's judgment; remands are procedural, typically requesting further medical examinations or evidence gathering. This distinction matters for the leniency instrument: I define leniency over grants versus denials, treating remands as a separate outcome for placebo testing.

3. Data

BVA Decision Corpus. I construct the dataset by downloading 12,404 BVA decision text files from `va.gov` for fiscal years 2017 and 2018. Each file follows a standardized format containing: the citation number (unique identifier), decision date, docket number, originating regional office, a narrative of the case, findings of fact, conclusions of law, the ORDER (granting, denying, or remanding the appeal), and the VLJ's name in the signature block.

Parsing and Extraction. I parse each text file to extract five fields. The *VLJ name* is identified from the signature block in the final 30 lines of each file, appearing above "Veterans Law Judge, Board of Veterans' Appeals." The *decision outcome* is classified from the ORDER section: "granted" if the ORDER contains grant language, "denied" if it contains denial language, "remanded" if the decision is a remand, and "dismissed" or "other" for remaining cases. The *regional office* is extracted from the header ("On appeal from the Department of Veterans Affairs Regional Office in [City, State]"). The *issue category* is classified from the THE ISSUES section into six categories: service connection, mental health, TDIU (total disability based on individual unemployability), increased rating, reopened claims, and effective dates.

Sample Construction. Of 12,404 downloaded files, 12,185 (98.2%) are successfully parsed with an identified VLJ and outcome. I restrict the analysis sample to decisions by VLJs with at least 30 total cases (to ensure reliable leniency estimates), yielding 106 VLJs and 11,422 observations. I further focus on substantive decisions—grants and denials—excluding remands from the main analysis, since remands do not reflect the VLJ's judgment on the merits. Table 1 presents summary statistics.

Table 1: Summary Statistics

	Mean	Std. Dev.	Min	Max
Panel A: Case-Level Variables (N = 11,422)				
Appeal Granted	0.330	0.470	0	1
Leave-One-Out Leniency	0.330	0.095	0.077	0.594
Number of Issues	2.80	2.47	1	15
Panel B: VLJ-Level Variables (N = 106)				
Cases per VLJ	107.8	51.6	30	250
VLJ Grant Rate	0.319	0.092	0.094	0.592

Notes: Sample includes BVA decisions from FY2017–2018 with identified VLJ and substantive outcome (grant or deny). Panel A reports case-level variables. Panel B reports VLJ-level means. Leave-one-out leniency is the VLJ's grant rate excluding the focal case.

The mean grant rate is 33.0%, with the remainder split between denials (27.6%) and remands (34.2%). VLJ grant rates exhibit substantial dispersion: the standard deviation of leave-one-out leniency is 0.095, with individual VLJ grant rates ranging from 9.4% to 59.2% among the 106 judges with at least 30 cases. Service connection is the most common issue category (44.6%), followed by mental health (22.2%) and TDIU (12.4%). The 83 regional offices span all 50 states and territories, with the largest (St. Petersburg, FL) contributing 8.9% of cases.

4. Empirical Strategy

4.1 Judge-Leniency Instrument

I follow the standard judge-leniency IV framework (Kling, 2006; Maestas et al., 2013; Chyn et al., 2024). For case i assigned to VLJ j, I construct the leave-one-out leniency measure:

$$Z_{j(i),-i} = \frac{\sum_{k \neq i, j(k)=j} \text{Grant}_k}{n_j - 1} \tag{1}$$

where n_j is the total number of substantive decisions by VLJ j. This is the judge's grant rate calculated over all cases except the focal case, avoiding the mechanical correlation between the instrument and the dependent variable.

The first-stage equation is:

$$\text{Grant}_i = \alpha + \beta Z_{j(i),-i} + \mathbf{X}_i'\gamma + \delta_t + \mu_r + \pi_c + \varepsilon_i \tag{2}$$

where δ_t are fiscal year fixed effects, μ_r are regional office fixed effects, π_c are issue-category fixed effects, and $\mathbf{X}_i$ may include additional case controls. Standard errors are clustered at the VLJ level throughout.

4.2 Identification Assumptions

The instrument requires three assumptions (Angrist et al., 1996; Frandsen et al., 2023):

Relevance. VLJs must exhibit persistent heterogeneity in grant rates. This is directly testable and confirmed by the first-stage F-statistic exceeding 225.

Independence. Case assignment must be as-good-as-random conditional on the included fixed effects. The ACD system's docket-order-based assignment provides the institutional foundation. I test this by regressing case characteristics on leniency: if assignment is random, leniency should not predict issue type, case complexity, or regional office composition.

Exclusion. The assigned VLJ must affect outcomes only through the appeal decision, not through direct channels. This is the standard—and untestable—assumption in leniency designs (Chyn et al., 2024). In this context, the most plausible violation would be if certain VLJs provide more detailed remand instructions that improve veterans' cases on return to the RO. I address this by excluding remands from the main sample and using remand as a placebo outcome.

Monotonicity. A more lenient VLJ must weakly increase the probability of a grant for all cases (Frandsen et al., 2023; de Chaisemartin, 2017). In plain language: no veteran's appeal is more likely to be granted by a *stricter* judge. This is plausible in a setting where VLJs apply the same legal standard ("as likely as not") but differ in how they weigh ambiguous evidence.

5. Results

5.1 First Stage

Table 2 reports the first-stage relationship between VLJ leniency and appeal grants. Across all specifications, leniency is a powerful predictor. Without controls (column 1), a one-unit

increase in leave-one-out leniency—moving from the strictest to the most lenient judge—raises the grant probability by 0.781 ($F = 254.7$). Adding year fixed effects (column 2), regional office fixed effects (column 3), and issue-category fixed effects (column 4) barely changes the coefficient: the preferred estimate is $\hat{\beta} = 0.784$ with $F = 225.4$. The near-invariance of the coefficient to controls is itself evidence of random assignment—if cases were sorted to judges, conditioning on observables would change the estimate.

Table 2: First Stage: VLJ Leniency Predicts Appeal Grants

	(1)	(2)	(3)	(4)
VLJ Leniency (LOO)	0.781***	0.787***	0.794***	0.784***
	(0.049)	(0.051)	(0.052)	(0.052)
First-stage F	254.7	238.0	229.4	225.4
Year FE	No	Yes	Yes	Yes
RO FE	No	No	Yes	Yes
Issue Category FE	No	No	No	Yes
N	11,422	11,422	10,751	10,751
R^2	0.025	0.027	0.032	0.050

Notes: Dependent variable is an indicator for appeal granted. VLJ Leniency is the leave-one-out grant rate of the assigned Veterans Law Judge. Standard errors clustered at the VLJ level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To put the magnitude in perspective: the interquartile range of VLJ leniency spans approximately 12 percentage points (from roughly 0.49 to 0.61). A veteran randomly assigned to a 75th-percentile-lenient VLJ rather than a 25th-percentile VLJ is approximately 9.4 percentage points more likely to win the appeal. Given the mean grant rate of 33.0%, this represents a 29% increase in the probability of success—driven entirely by the luck of the draw. For a veteran at the margin, the difference between the 25th and 75th percentile judge translates to an expected gain of roughly \$1,700–\$6,500 per year in disability compensation, before accounting for associated healthcare benefits.

Variance Decomposition. Adding VLJ fixed effects to a model with year, regional office, and issue-category controls increases R^2 from 0.025 to 0.066, implying a partial R^2 of judge identity of 0.042. The judge lottery explains 4.2% of the residual variation in appeal outcomes—modest in absolute terms, but economically meaningful given the binary nature

of the outcome and the high baseline grant rate.

5.2 Balance Tests

Table 3 presents balance tests. If the ACD assigns cases quasi-randomly, VLJ leniency should not predict pre-determined case characteristics. I regress each characteristic on leniency with year fixed effects.

Table 3: Balance Tests: VLJ Leniency and Case Characteristics

Dependent Variable	Coefficient	(SE)	Mean
Mental Health Issue	0.008	(0.040)	0.222
TDIU Issue	0.083*	(0.042)	0.124
Service Connection	−0.076	(0.060)	0.446
Number of Issues	0.539*	(0.290)	2.800
Joint F-test p-value		0.999	

Notes: Each row reports the coefficient from a regression of the case characteristic on VLJ leave-one-out leniency with year fixed effects. Standard errors clustered at the VLJ level. Under random assignment, leniency should not predict case characteristics. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

None of the four individual coefficients is statistically significant at conventional levels. Mental health issue ($p = 0.838$), service connection ($p = 0.213$), and number of issues ($p = 0.066$) all show small and insignificant relationships. The TDIU coefficient is marginally significant ($p = 0.052$) but economically tiny. Most importantly, the joint F-test for all four characteristics yields $p = 0.999$, overwhelmingly failing to reject the null of random assignment. These results are consistent with the ACD's documented docket-order-based distribution.

5.3 Heterogeneity: The Subjectivity Premium

Table 4 decomposes the first stage by issue type, revealing a striking gradient in judge discretion.

Table 4: First Stage Heterogeneity by Appeal Issue Type

Issue Type	VLJ Leniency	(SE)	N	Mean Grant Rate
Service Connection	0.805***	(0.066)	5,097	0.316
Mental Health	0.997***	(0.104)	2,536	0.407
Increased Rating	0.407*	(0.242)	428	0.353
Tdiu	0.886***	(0.118)	1,417	0.393

Notes: Each row reports the first-stage coefficient from a separate regression of appeal granted on VLJ leave-one-out leniency within the specified issue category subsample. All regressions include year and regional office fixed effects. Standard errors clustered at the VLJ level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Mental health claims show the largest judge effects ($\hat{\beta} = 0.997$, SE $= 0.104$). The coefficient is not statistically distinguishable from one, implying that a one-unit increase in overall leniency translates nearly one-for-one into mental health grant probability. PTSD, depression, and anxiety claims require VLJs to weigh subjective symptom reports against ambiguous medical evidence—precisely the context where judicial discretion is largest.

Service connection claims—the most common category—show a strong effect ($\hat{\beta} = 0.805$), consistent with meaningful but somewhat constrained discretion. TDIU claims ($\hat{\beta} = 0.886$) fall between mental health and service connection, reflecting the mixed medical-vocational evidence these decisions require. Increased-rating claims show the weakest judge dependence ($\hat{\beta} = 0.407$), consistent with ratings being more mechanically tied to objective medical criteria (range-of-motion measurements, audiometric results, spirometry).

To test whether these differences are statistically significant, I estimate a pooled specification interacting leniency with issue-category indicators. The mental health coefficient is significantly larger than the increased-rating coefficient ($p < 0.05$ for the difference), confirming that the heterogeneity is not an artifact of varying sample sizes or instrument strength.

This gradient—I call it the *subjectivity premium*—reveals that the appeals lottery matters most where medical evidence is most ambiguous. A veteran with PTSD draws a substantially different probability of success depending on which VLJ is assigned than a veteran appealing a knee rating. The finding echoes Cabral and Dillender (2025), who document analogous discretion gradients among workers' compensation medical evaluators, and extends it to the legal adjudication stage.

5.4 Robustness

Table 5 collects robustness checks organized in four panels.

Table 5: Robustness of the First Stage

Specification	Coefficient	(SE)	First-Stage F
Panel A: Baseline			
Preferred (Year + RO + Issue FE)	0.784***	(0.052)	225.4
Panel B: Alternative Leniency Measures			
Excluding same issue category	0.695***	(0.071)	95.3
Other-year leniency only	0.424***	(0.092)	21.1
Panel C: Placebo			
Leniency → Remand	-0.384***	(0.079)	—
Panel D: Alternative Clustering			
Cluster by Regional Office	0.784***	(0.051)	231.6
Two-way (VLJ + Year-Month)	0.784***	(0.075)	108.2

Notes: Panel A reports the preferred specification. Panel B uses alternative constructions of VLJ leniency. Panel C runs a placebo test predicting remand (a procedural, not substantive, outcome). Panel D varies the level of standard-error clustering. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Alternative leniency measures. Excluding same-issue-category cases from the leniency calculation (Panel B, row 1) yields $\hat{\beta} = 0.695$ with $F = 95.3$—somewhat attenuated, as expected when the instrument is noisier, but still very strong. Cross-year leniency (using only other-year decisions) produces $\hat{\beta} = 0.424$ with $F = 21.1$, confirming that leniency is persistent across years despite the smaller effective sample.

Placebo. If leniency reflects merit-decision tendencies rather than case routing, it should not predict remands (procedural returns to the RO). Panel C confirms: lenient VLJs remand *less* ($\hat{\beta} = -0.384$, $p < 0.001$). This is consistent with lenient judges resolving cases on the merits rather than deferring to the regional office—a "decide rather than defer" pattern. The negative relationship also implies that the instrument operates on two margins: lenient judges both grant more and remand less, effectively converting what would be procedural delays into immediate substantive decisions. This has implications for the complier population: the marginal veteran moved by a lenient judge is one who might otherwise face additional months

of uncertainty in the remand queue.

Alternative clustering. Panel D shows the first-stage coefficient is robust to clustering by regional office (SE = 0.051), two-way clustering by VLJ and year-month (SE = 0.075), and heteroskedasticity-robust standard errors (SE = 0.047). The two-way clustered standard errors are the most conservative, and the first-stage remains highly significant.

Leave-one-out judge sensitivity. Dropping each of the 106 VLJs in turn produces first-stage coefficients ranging from 0.744 to 0.790 with a standard deviation of 0.006—no single judge drives the result.

Minimum caseload thresholds. Raising the minimum caseload from 30 to 50, 75, 100, or 150 cases per VLJ produces coefficients between 0.784 and 0.883, all highly significant. If anything, restricting to high-caseload VLJs (with more precisely estimated leniency) strengthens the first stage.

6. Discussion

The central finding of this paper is that the identity of the randomly assigned Veterans Law Judge is a powerful and persistent predictor of whether a veteran's disability benefits appeal succeeds. The first-stage F-statistic of 225 places this instrument among the strongest in the judge-leniency literature (Chyn et al., 2024). The design's clean balance tests ($p = 0.999$), stability under leave-one-out deletion, and consistency across alternative leniency constructions establish the BVA as a credible setting for future IV research.

What the subjectivity premium teaches. The finding that judge discretion scales with the subjectivity of the medical evidence—near-unity effects for mental health claims versus half-sized effects for mechanical ratings—has implications beyond the VA. It suggests that administrative tribunals exercise the most consequential discretion precisely where standardization is hardest to achieve. Efforts to reduce disparity through algorithms or structured decision-making tools (Kleinberg et al., 2018) may therefore yield the largest gains in settings like mental health adjudication, where objective benchmarks are weakest.

Relationship to prior work. This paper complements Silver and Zhang (2026), who use Compensation & Pension medical examiner leniency at the initial claim stage. The two instruments identify different margins: Silver and Zhang's compliers are veterans at the boundary of initial eligibility, while mine are veterans whose initial claims were already denied and who persisted through the appeals process. The appellate margin is independently

policy-relevant because BVA reform—reducing the 100,000-case backlog, changing evidentiary standards, or reassigning VLJs—directly affects this population.

Limitations and next steps. This paper establishes the instrument but does not estimate causal effects on downstream outcomes. Linking BVA decisions to VA administrative records (mortality, healthcare utilization), Social Security data (employment, earnings), or HUD records (housing stability) would allow a full IV analysis of whether winning a disability appeal causes improvements in veteran welfare. The public availability of BVA decisions makes this a feasible next step. A second limitation is that the FY2017–2018 sample spans a period of institutional transition as the Appeals Modernization Act took effect; extending the data to additional years would test whether the ACD's random assignment held consistently.

The appeals lottery is not a metaphor. A veteran's access to disability benefits—and plausibly, the downstream consequences for health, employment, and housing—depends in measurable part on which name appears at the bottom of the decision letter. Documenting this fact, and providing the instrument to study its consequences, is this paper's contribution.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: https://github.com/SocialCatalystLab/ape-papers

Contributors: @olafdrw

First Contributor: https://github.com/olafdrw

References

Aizer, Anna and Joseph J. Doyle, "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges," *Quarterly Journal of Economics*, 2015, *130* (2), 759–803.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin, "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 1996, *91* (434), 444–455.

Arnold, David, Will Dobbie, and Peter Hull, "Measuring Racial Discrimination in Bail Decisions," *American Economic Review*, 2022, *112* (9), 2992–3038.

Autor, David, Andreas Ravndal Kostol, Magne Mogstad, and Bradley Setzler, "Disability Benefits, Consumption Insurance, and Household Labor Supply," *American Economic Review*, 2019, *109* (7), 2613–2654.

Autor, David H. and Mark G. Duggan, "The Rise in the Disability Rolls and the Decline in Unemployment," *Quarterly Journal of Economics*, 2003, *118* (1), 157–206.

_ and _ , "The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding," *Journal of Economic Perspectives*, 2006, *20* (3), 71–96.

Autor, David, Mark Duggan, Kyle Greenberg, and David S. Lyle, "The Impact of Disability Benefits on Labor Supply: Evidence from the VA's Disability Compensation Program," *American Economic Journal: Applied Economics*, 2016, *8* (3), 31–68.

Bhuller, Manudeep, Gordon Dahl, Katrine Loken, and Magne Mogstad, "Incarceration, Recidivism, and Employment," *Journal of Political Economy*, 2020, *128* (4), 1269–1324.

Cabral, Marika and Marcus Dillender, "Doctor Discretion in Medical Evaluations," Technical Report 33988, NBER 2025.

Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue, "Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires," *Quarterly Journal of Economics*, 2016, *131* (3), 1181–1242.

Chyn, Eric, Brigham Frandsen, and Emily Leslie, "Examiner and Judge Designs in Economics: A Practitioner's Guide," Technical Report 32348, NBER 2024. Prepared for *Journal of Economic Literature*.

Coile, Courtney, Mark Duggan, and Audrey Guo, "To Work for Yourself, for Others, or Not At All? How Disability Benefits Affect the Employment Decisions of Older Veterans," *Journal of Policy Analysis and Management*, 2021.

Dahl, Gordon B., Andreas Ravndal Kostol, and Magne Mogstad, "Family Welfare Cultures," *Quarterly Journal of Economics*, 2014, *129* (4), 1711–1752.

de Chaisemartin, Clément, "Tolerating Defiance? Local Average Treatment Effects without Monotonicity," *Quantitative Economics*, 2017, *8* (2), 367–396.

Dobbie, Will, Jacob Goldin, and Crystal Yang, "The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review*, 2018, *108* (2), 201–240.

Frandsen, Brigham, Lars Lefgren, and Emily Leslie, "Judging Judge Fixed Effects," *American Economic Review*, 2023, *113* (1), 253–277.

Galasso, Alberto and Mark Schankerman, "Patents and Cumulative Innovation: Causal Evidence from the Courts," *Quarterly Journal of Economics*, 2015, *130* (1), 317–369.

Garin, Andrew, Dmitri Koustas, Carl McPherson, Samuel Norris, Matthew Pecenco, Evan Rose, Yotam Shem-Tov, and Jeffrey Weaver, "The Impact of Incarceration on Employment, Earnings, and Tax Filing," *Econometrica*, 2025.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, "Human Decisions and Machine Predictions," *Quarterly Journal of Economics*, 2018, *133* (1), 237–293.

Kling, Jeffrey R., "Incarceration Length, Employment, and Earnings," *American Economic Review*, 2006, *96* (3), 863–876.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand, "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt," *American Economic Review*, 2013, *103* (5), 1797–1829.

Ramji-Nogales, Jaya, Andrew I. Schoenholtz, and Philip G. Schrag, "Refugee Roulette: Disparities in Asylum Adjudication," *Stanford Law Review*, 2007, *60* (2), 295–412.

Sampat, Bhaven and Heidi L. Williams, "How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome," *American Economic Review*, 2019, *109* (1), 203–236.

Silver, David and Jonathan Zhang, "Invisible Wounds: How Mental Disability Benefits Shape Veteran Well-Being," *American Economic Journal: Economic Policy*, 2026. Formerly NBER Working Paper 29877.

A. Data Appendix

Raw Data Source. BVA decision text files are freely available at `https://www.va.gov/vetapp{YY}/files1/{YY}NNNNN.txt`, where `YY` is the two-digit fiscal year and `NNNNN` is the zero-padded decision number. FY2017 contains decisions numbered 1700001 through approximately 1706062; FY2018 contains 1800001 through approximately 1806343.

Parsing Algorithm. Each text file is parsed as follows:

1. **Citation number:** Extracted from the first line (regex: digits).
2. **Decision date:** Extracted from the second line (format: MM/DD/YY).
3. **Regional office:** Extracted from lines 1–20 (regex: "Regional Office in [City, State]").
4. **VLJ name:** Extracted from the final 30 lines. The name appears on the line above "Veterans Law Judge, Board of Veterans' Appeals."
5. **Decision outcome:** Classified from the ORDER section using keyword matching: "granted," "denied," "remanded," "dismissed."
6. **Issue category:** Classified from the THE ISSUES section using keyword matching for PTSD/psychiatric (mental health), TDIU, increased rating, service connection, effective date, and reopened claims.

Sample Restrictions.

- 12,404 files downloaded; 12,185 (98.2%) successfully parsed with VLJ and outcome.
- 294 unique VLJs identified; restricted to 106 with ≥ 30 substantive decisions.
- 11,422 case-level observations in the main analysis sample (grants and denials only).
- 671 observations dropped from specifications with regional office FE due to singleton ROs.

B. Robustness Appendix

Leave-One-Out VLJ Sensitivity. Dropping each of the 106 VLJs in turn and re-estimating the first stage produces coefficients ranging from 0.744 to 0.790, with a standard deviation of 0.006 across the 106 estimates. No single judge drives the result.

Minimum Caseload Sensitivity. Raising the minimum caseload threshold for VLJ inclusion yields the following first-stage estimates: 30 cases ($\hat{\beta} = 0.784$, 106 VLJs, $N = 11{,}422$), 50 cases ($\hat{\beta} = 0.802$, 91 VLJs, $N = 10{,}870$), 75 cases ($\hat{\beta} = 0.828$, 73 VLJs, $N = 9{,}748$), 100 cases ($\hat{\beta} = 0.823$, 56 VLJs, $N = 8{,}288$), and 150 cases ($\hat{\beta} = 0.883$, 25 VLJs, $N = 4{,}487$). If anything, restricting to high-caseload VLJs with more precisely estimated leniency strengthens the first stage.

C. Standardized Effect Sizes

Table 6: Standardized Effect Sizes for Main Outcomes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	SD(X)	Classification
Panel A: Pooled							
Appeal Granted	0.784	0.052	0.470	0.159	0.011	0.095	Large positive
Panel B: Heterogeneous							
Service Connection	0.805	0.066	0.465	0.166	0.014	0.096	Large positive
Mental Health	0.997	0.104	0.491	0.187	0.019	0.092	Large positive

Notes: **Country:** United States. **Research question:** Does the identity of the randomly assigned Veterans Law Judge at the Board of Veterans Appeals affect whether a veteran's disability benefits appeal is granted? **Policy mechanism:** The BVA's Caseflow Automatic Case Distribution system assigns appeals to one of approximately 60 Veterans Law Judges based on docket order and availability, not case characteristics; VLJs exercise substantial discretion over whether to grant, deny, or remand each appeal, creating consequential variation in a veteran's access to monthly disability compensation ($150–$3,700+) and VA healthcare. **Outcome definition:** Binary indicator equal to one if the VLJ grants the veteran's appeal (in whole or in part) on the merits, zero if denied. **Treatment:** Continuous leave-one-out VLJ leniency (the assigned judge's grant rate excluding the focal case). **Data:** BVA decision text files downloaded from va.gov for FY2017–2018, parsed to extract judge identity, decision outcome, regional office, and issue type; case-level unit of observation. **Method:** OLS first-stage regression of appeal granted on leave-one-out VLJ leniency with year, regional office, and issue category fixed effects; standard errors clustered at the VLJ level. **Sample:** BVA decisions with identified VLJ, substantive outcome (grant or deny, excluding remands), and VLJ with at least 30 total decisions. SDE $= \hat{\beta} \times \text{SD}(X)/\text{SD}(Y)$ where SD(Y) is the unconditional standard deviation of the outcome and SD(X) is the unconditional standard deviation of VLJ leniency. Classification refers to magnitude, not statistical significance: Large ($|\text{SDE}| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).