The Equilibrium Impacts of Appeals:
Evidence from Social Security Disability Programs

Colin Gray
MIT
July 2019

Abstract
Many administrative and legal processes involve the option to appeal an initial decision. This provides an additional opportunity to demonstrate eligibility for a program, innocence in a criminal case, or fitness for residency in a country. However, researchers and policymakers know very little about the effects of adding or subtracting stages of appeal. On one hand, fewer stages of appeal means fewer opportunities to demonstrate eligibility, which may decrease overall allowances, administrative costs, and processing times. On the other hand, if applicants and adjudicators anticipate future appeals and infer information from past appeals, eliminating an appeals stage may induce initial adjudicators to increase allowances or encourage applicants to pursue higher levels of appeal. After illustrating the key theoretical parameters dictating these impacts, I study a 1999 reform to the Social Security disability adjudication process in which an intermediate appeals stage was eliminated for ten states. In line with the latter model, I find evidence that eliminating this appeals stage increased allowances onto the program and saved little in terms of processing time and administrative costs due to these dynamic responses. In practice, appeals are not simply extra chances to demonstrate eligibility: they are tools for switching between equilibria whose effects depend on a set of key behavioral parameters.¹

¹ I would like to thank Jason Brown and Ted Horan for making my work at the Social Security Administration possible, as well as Jae Song, Mark Sarney, Alexander Strand, and Patrick McGuire for their support and assistance.
I. Introduction

The option to repeatedly appeal an adjudicative decision is common in administrative and legal institutions. Whether to prove fitness for residency in a country (Norris (2017)), innocence in a criminal trial (Stith (1990)), or eligibility for a social insurance program (French and Song (2014)), an individual who is denied this status may request another decision from a higher adjudicative level. These appeals are typically asymmetric: individuals may appeal denials but typically they and the agency do not appeal allowances. Moreover, information is incomplete: adjudicators do not know everything about applicants, and applicants face uncertainty in whether they would be allowed at each stage. Adding or subtracting an appeals stage therefore alters incentives and information in complex ways, leading to theoretically ambiguous impacts. Moreover, because it is rare for major institutions to change their appeals processes, researchers and policymakers know very little about the empirical impacts of appeals. In the context of Social Security disability adjudication, this paper shows that eliminating an opportunity to appeal increased allowances and had muted effects on other key outcomes due to adjudicator and applicant responses. An appeal is not simply an additional chance for eligibility, but results in a new equilibrium dictated by key features of the adjudication structure.

I first illustrate why adding a stage of appeal has ambiguous theoretical impacts. I build a simple model in which adjudicators at different stages receive independent, imperfect signals of applicant eligibility. When adjudicators make all decisions independently based on the information they receive, adding an appeals stage will necessarily increase allowances, administrative costs, and processing times ceteris paribus. However, when

The research reported herein was performed pursuant to grant DRC12000002 from the US Social Security Administration (SSA) funded as part of the Disability Research Consortium. The opinions and conclusions expressed are solely those of the author and do not represent the opinions or policy of SSA, any agency of the Federal Government, or NBER. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the United States Government or any agency thereof.
adjudicators have knowledge of previous denials and anticipate future appeals, the effect on allowances may be substantially muted. Two natural extensions can result in a negative relationship between the number of appeals and total allowances: if adjudicators at different stages behave differently, or if applicants selectively persist as a function of underlying eligibility. Either of these extensions, when combined with dynamic adjudicator behavior, lead to theoretically ambiguous effects of additional appeals on allowances, administrative costs, and processing times. The effects of removing appeals is similarly ambiguous.

I then document the empirical effects of removing an appeals stage in the context of disability programs administered through the Social Security Administration (SSA). In October 1999, SSA offices in ten states eliminated the first appeals stage from their adjudication process. While the policy change coincided with another change in the adjudication process (the so-called Claimant Conference), this second change was quickly wound down and entirely eliminated by mid-2002. Using a difference-in-difference method, I provide evidence that eliminating the first stage of appeals increased total allowances by 3 percentage points (∼6%) despite providing fewer opportunities to demonstrate eligibility. This occurred through two measurable channels: adjudicators increased initial allowances (perhaps in anticipation of fewer subsequent appeals stages) and applicants who were not discouraged by a denial in the first appeal were more likely to pursue a higher appeals stage with higher allowance rates. While administrative costs and processing times still fell on net with the elimination of this appeal, the responses of applicants substantially muted these savings. These channels support the theoretical model, highlighting that the response of forward-looking adjudicators and the conditional persistence of applicants can reverse the impact of appeals in practice.

This work relates to a small body of literature on administrative law and adjudication processes (Mashaw (1985); Norris (2017)). It also relates to a larger body of theory

---

2 Formally, the elimination of reconsideration was a test rather than a permanent reform. The test was periodically renewed and held continuously for the subsequent two decades.
focused on information aggregation, especially with regard to jury voting (Feddersen and Pesendorfer (1998); Li et al. (2001)) and political economy (Besley (2007)). As in my model, these papers highlight how equilibrium effects are often not simple aggregations of static effects when aggregating noisy signals.

The paper proceeds as follows. Section 2 illustrates the theoretical effects of instituting a new appeals process, showing that the comparative statics from a simple model of adjudication break down when dynamic considerations are added. Section 3 gives background on the SSA appeals process and the data sets I use. Section 4 shows that eliminating an opportunity to appeal in the SSA disability adjudication process actually increased allowance rates, and led to only modest decreases in average processing times and administrative costs due to the channels highlighted in the theoretical model. Section 5 concludes.

II. Theory

This following model illustrates how the addition or removal of an appeals stage may affect allowance rates, processing times, and administrative costs. I first consider a model in which applicants and adjudicators make static decisions and do not infer additional information from other stages of the appeals process. I then consider a model in which applicants and adjudicators make dynamic decisions that take other stages into account. The effects of adding or subtracting an appeals stage in the dynamic model are muted relative to the static model. I then show that either of two further extensions – heterogeneity in sequential stages or selective attrition from the appeals process – make the effects of additional appeals theoretically ambiguous. These models highlight that the way in which agents deal with asymmetric appeals and incomplete information determines the effects of adding or removing appeals. While applicants could strategically re-apply instead of appealing, this may not help them in practice because cases can be linked by Social Security.
**Model Setup**

An adjudicator at each stage $s$ receives a signal of the severity of an applicant $i$’s limitation. I denote this signal $\hat{d}_{si}$, and assume that it is an unbiased but imperfect signal of underlying severity $d_i$:

$$\hat{d}_{si} = d_i + \epsilon_{si}$$

Let $\epsilon_{si} \sim G$ for some unbounded, mean-zero distribution $G$. The adjudicator must decide whether to allow ($A_{si} = 1$) or deny ($A_{si} = 0$). She gets a benefit $\beta d_i$ from an applicant being allowed, but there is some cost $\kappa$ for each allowance that she gives. This very simple model captures two ideas: (1) adjudicators have incomplete information about a given case, and (2) adjudicators want applicants with severe limitations to be allowed but try to avoid having unreasonably high overall allowance rates. Numerous extensions, such as penalties for having a decision overturned or heterogeneity in adjudicator opinions, would affect the parameters $\beta$ and $\kappa$ in this model without loss of generality.

I allow for applicants to exit the appeals process, where the probability of applicant $i$ appealing to the subsequent adjudication stage is given by the function $\rho(d_i)$. This function captures both the cost of appeals, and the ways in which applicants infer information from past denials. For example, applicants may always persist ($\rho(d_i) = 1 \forall d_i$); may persist sometimes in a way uncorrelated with their impairment ($\rho(d_i) = \rho < 1 \forall d_i$); or may gradually update their beliefs about eligibility and therefore persist as a function of their private information ($\rho'(d_i) > 0$ or $\rho'(d_i) < 0$).

**Static Model**

I first consider a model in which each adjudicator makes the decision to allow or deny an application using only their signal $\hat{d}_{si}$. In this case, an adjudicator at a given stage $s$ faces the following payoffs:

$$u_s(A_{si}) = \begin{cases} \beta d_i - \kappa & \text{if } A_{si} = 1 \\ 0 & \text{if } A_{si} = 0 \end{cases}$$
Each adjudicator sets a cutoff rule based on their signal \( \hat{d}_i \), where they allow if \( f \)\( f \):

\[
\beta E[d_i|\hat{d}_i] - \kappa \geq 0
\]
\[
\iff \hat{d}_i \geq \frac{\kappa}{\beta}
\]

(1)

The following proposition shows that, in this model, adding an additional appeals stage necessarily increases allowance rates, administrative costs, and wait times.

**Proposition 1:** When adjudicators make decisions in isolation of one another, adding an additional identical appeals stage increases allowance rates, increases administrative costs, and increases processing times.

*Proof:* Denote the probability of allowance at the union of all stages as \( \Pr(A_i^U) \). When there is one stage, this probability is:

\[
\Pr_1(A_i^U|d_i) = \Pr(A_{1i} = 1|d_i) = \Pr \left( \hat{d}_{1i} \geq \frac{\kappa}{\beta} | d_i \right) = \Pr \left( \epsilon_{1i} \geq \frac{\kappa}{\beta} - d_i \right) = 1 - G \left( \frac{\kappa}{\beta} - d_i \right)
\]

(2)

When there are two stages, the overall probability of allowance becomes:

\[
\Pr_2(A_i^U|d_i) = \Pr(A_{1i} = 1|d_i) + \rho(d_i)\Pr(A_{2i} = 1|A_{1i} = 0, d_i)
\]
\[
= \Pr \left( \hat{d}_{1i} \geq \frac{\kappa}{\beta} | d_i \right) + \rho(d_i)\Pr \left( \epsilon_{1i} \geq \frac{\kappa}{\beta} - d_i \right) = \Pr \left( \epsilon_{1i} \geq \frac{\kappa}{\beta} - d_i \right) + \rho(d_i)\Pr \left( \epsilon_{1i} \geq \frac{\kappa}{\beta} - d_i \right) = \left( 1 - G \left( \frac{\kappa}{\beta} - d_i \right) \right) \left( 1 + \rho(d_i) G \left( \frac{\kappa}{\beta} - d_i \right) \right) > 1 - G \left( \frac{\kappa}{\beta} - d_i \right)
\]

Inductively, for any given number of stages \( S \),

\[
\Pr_S(A_i^U|d_i) = \left( 1 - G \left( \frac{\kappa}{\beta} - d_i \right) \right) \left\{ \sum_{s=0}^{S-1} \rho(d_i)^s G \left( \frac{\kappa}{\beta} - d_i \right)^{s-1} \right\} > \Pr_{S-1}(A_i^U|d_i)
\]
To understand the effects of appeals on administrative costs and processing times, note that the addition of an appeals stage \(s\) will induce \(\{1 - Pr_{T-1}(A_{i}^u | d_{i})\} \rho (d_{i})^{s-1}\) percent of applicants to pursue that appeals stage. Denoting processing time \(T_{s}\) and administrative cost \(C_{s}\), the addition of an appeals stage increases average processing time by \(\{1 - Pr_{T-1}(A_{i}^u | d_{i})\} \rho (d_{i})^{s-1} T_{s}\) and average administrative costs by \(\{1 - Pr_{T-1}(A_{i}^u | d_{i})\} \rho (d_{i})^{s-1} C_{s}\).

QED

**Dynamic Model**

Next, I consider a model in which adjudicators take the structure of the adjudication process into account when making a determination. In order to clearly illustrate the behavior of adjudicators in the model, I assume that applicants do not persist as a function of underlying eligibility: \(\rho (d_{i}) = \rho\) for all \(d_{i}\). I also assume that the adjudicative process goes from one stage to two stages. I relax both assumptions later.

I first consider allowance rates. In a dynamic model, two forces push allowance rates downwards relative to the static model. First, later adjudicators may shade their expectations of severity downwards given their signal, due to the fact that the applicant was denied at the first stage. Second, earlier adjudicators may defer the most difficult cases to the subsequent adjudicator, which (given the asymmetry of appeals) involves denying more cases.

To show that dynamic considerations mute the effect of a second adjudicative stage on allowance rates, it suffices to show that allowance rates in the dynamic model are lower at both stages relative to the static model. Because the dynamic and static model are identical when there exists only one stage, this implies that adding a second stage has a muted effect on total allowances in the dynamic model relative to the static model. The following proposition establishes this fact.
Proposition 2: Allowance rates at both the first and second adjudicative stage are lower in a model where adjudicators take other adjudicative stages into account.

Proof: First, I evaluate the cutoff rule for the adjudicator in the second stage, conditional on the first stage adjudicator following her own cutoff rule \( \{A_{1} = 1 \Leftrightarrow \hat{d}_{2i} \geq \psi\} \) for some value \( \psi \). The second stage adjudicator then follows the cutoff rule

\[
\beta E[d_i|\hat{d}_{2i}, A_{1i} = 0] - \kappa \geq 0 \\
E[(\hat{d}_{2i} - \epsilon_{2i})|\hat{d}_{2i}, d_{1i} < \psi] \geq \frac{\kappa}{\beta} \\
\hat{d}_{2i} - E[\epsilon_{2i}|\hat{d}_{2i}, d_{1i} < \psi - \epsilon_{1i}] \geq \frac{\kappa}{\beta} \\
\hat{d}_{2i} - E[\epsilon_{2i}|\hat{d}_{2i}, \hat{d}_{2i} - \epsilon_{2i} < \psi - \epsilon_{1i}] \geq \frac{\kappa}{\beta} \\
\hat{d}_{2i} - E[\epsilon_{2i}|\hat{d}_{2i}, \epsilon_{2i} > \hat{d}_{2i} - \psi + \epsilon_{1i}] \geq \frac{\kappa}{\beta} \tag{3}
\]

Since the expectations term is censored from below, we know that \( E[\epsilon_{2i}|\hat{d}_{2i}, \epsilon_{2i} > \hat{d}_{2i} - \psi + \epsilon_{1i}] > 0 \) and therefore allowance requires \( \hat{d}_{2i} > \frac{\kappa}{\beta} \). Intuitively, the second adjudicator may infer that \( d_{i} \) is actually lower than the signal \( \hat{d}_{2i} \) that she received based on the fact that the first adjudicator denied the application, and therefore she sets a higher cutoff. I let \( E[\epsilon_{2i}|\hat{d}_{2i}, \epsilon_{2i} > \hat{d}_{2i} - \psi + \epsilon_{1i}] \equiv E[\epsilon_{2i}|\hat{d}_{2i}, A_{1i} = 0] \) for notational simplicity.

I next use this cutoff rule to study the probability of allowance at the second stage conditional on rejection and appeal, which will enter into later derivations. From the second-stage adjudicator’s cutoff rule, we have

\[
\Pr(A_{2i} = 1|A_{1i} = 0, d_{i}) = \Pr\left(\hat{d}_{2i} - E[\epsilon_{2i}|\hat{d}_{2i}, A_{1i} = 0] \geq \frac{\kappa}{\beta}|d_{i}\right) \quad \text{Without specifying the distribution G, this probability does not have a simple closed form solution. However, one can still use properties of this expression to make useful conclusions. Specifying a function } q(x) \equiv x - E[\epsilon_{2i}|x, A_{1i} = 0], \text{ the fact that } q(x) < x \quad \forall \quad x \text{ implies that } q^{-1}(x) > x \quad \forall \quad x. \text{ Rewriting the expression above then yields}
\]

\[
\Pr(A_{2i} = 1|A_{1i} = 0, d_{i}) = \Pr\left(q(\hat{d}_{2i}) \geq \frac{\kappa}{\beta}|d_{i}\right) \\
= \Pr\left(\hat{d}_{2i} \geq q^{-1}\left(\frac{\kappa}{\beta}\right)|d_{i}\right) \\
= 1 - G\left(q^{-1}\left(\frac{\kappa}{\beta}\right) - d_{i}\right) < 1 - G\left(\frac{\kappa}{\beta} - d_{i}\right) \tag{4}
\]

where the last inequality follows from \( q^{-1}\left(\frac{\kappa}{\beta}\right) > \frac{\kappa}{\beta} \). Equation (4) yields a lower probability of second-stage allowance in the dynamic model relative to equation (2) in the static model.
Now consider the adjudicator in the first stage \((s = 1)\). With the existence of a second stage, there is some chance that an initially denied applicant would still ultimately be allowed. The problem facing the first adjudicator is therefore:

\[
u_t(A_{1t}) = \begin{cases} 
\beta d_t - \kappa & \text{if } A_{1t} = 1 \\
\beta \rho d_t \Pr(A_{2t} = 1 | A_{1t} = 0, d_t) & \text{if } A_{1t} = 0
\end{cases}
\]

Taking expectations, the cutoff rule becomes to allow if:

\[
\hat{d}_{1t} - \rho \beta \left[ \frac{\beta}{\rho} \right] \geq \frac{\kappa}{\beta}
\]

Note that the expectations term is a positive function of \(\hat{d}_{1t}\), implying that the new cutoff rule requires \(\hat{d}_{1t} > \frac{\kappa}{\beta}\).

Relative to equation (1), \(\hat{d}_{1t}\) must therefore be higher to satisfy (5). Intuitively, a second stage provides another opportunity for an accurate determination, but only if the first adjudicator does not allow the case.

Again, without specifying a form for the distribution \(G\), equation (5) does not yield a closed form cutoff for \(\hat{d}_{1t}\). This prevents me from writing a general closed form solution \(P_r^S(A_t^u | d_t)\) for the dynamic model. However, the preceding equations show that cutoffs at each stage are higher in the dynamic model relative to the static model, implying that allowance rates are lower.

QED

In this model, the addition of a second stage increases administrative costs and wait times by more in the dynamic model relative to the static model. This occurs because fewer applicants are allowed at the initial stage, and so more applicants appeal. Appendix A illustrates that the same forces muting the response in allowance rates for a two-stage model also act in a three-stage model. While I do not provide a formal inductive proof, these same principles extend further in processes with many appeals stages.
Adding Heterogeneous Stages

The theory above has established that adding uniform appeals with no dynamic behavior increases allowance rates, administrative costs, and wait times, while adding uniform appeals with dynamic behavior results in smaller increases in allowance rates and larger increases in administrative costs and wait times. A simple extension to this model can result in ambiguous effects for all three outcomes: heterogeneity in adjudicators at different stages.

Proposition 3: In a model of dynamic adjudicators in which adjudicator parameters differ at each stage, adding an early adjudication stage may increase or decrease allowance rates.

Proof: Suppose that adjudicators at stage two differ from those at stage one, such that \( \frac{k_2}{\beta_2} < \frac{k_1}{\beta_1} \), and consider the effect of adding the first appeals stage to the adjudicator process. With only stage two, the derivation in Section 2.2 yields the probability of allowance \( Pr(A_{1i} = 1 | d_i) = 1 - G \left( \frac{k_2}{\beta_2} - d_i \right) \)

Now consider adding the first stage to the appeals process. By the derivation in Section 2.3, the second stage allowance rate follows the expression \( Pr(A_{2i} = 1 | A_{1i} = 0, d_i) = \rho \left( 1 - G \left( q^{-1} \left( \frac{k_2}{\beta_2} \right) - d_i \right) \right) \) while the first stage cutoff satisfies the expression \( \hat{d}_{1i} - E \left[ \rho \left( \hat{d}_{1i} - \epsilon_{1i} \right) \left\{ 1 - G \left( q^{-1} \left( \frac{k_2}{\beta_2} \right) + \epsilon_{1i} - \hat{d}_{1i} \right) \right\} \right] \geq \frac{k_2}{\beta_2} \). While overall allowance rates do not have a simple closed-form solution, two counterexamples demonstrate that the effect of adding the first appeals stage on overall allowances is ambiguous. On one hand, consider a series of models in which \( \frac{k_1}{\beta_1} \rightarrow \frac{k_2}{\beta_2} \). Bringing the objective functions of the two stages arbitrarily close brings this model arbitrarily close to the model in Section 2.3, in which adding an appeals stage increased allowances by a finite amount. On the other hand, consider a series of models in which persistence becomes arbitrarily low: \( \rho \rightarrow 0 \). In this case, the first-stage cutoff approaches \( \frac{k_1}{\beta_1} \) and the probability of allowance at any stage approaches \( 1 - G \left( \frac{k_1}{\beta_1} - d_i \right) \). Since \( \frac{k_2}{\beta_2} < \frac{k_1}{\beta_1} \) by assumption, we have \( 1 - G \left( \frac{k_1}{\beta_1} - d_i \right) < 1 - G \left( \frac{k_2}{\beta_2} - d_i \right) \) implying that the allowance rate has fallen with the addition of the first appeals stage. Conversely, eliminating this stage could increase the allowance rate.

QED
When appeals stages are heterogeneous, the administrative cost and wait time implications are also ambiguous. Following the extreme examples from the proof, one could have a two-stage adjudicative process that looks like two versions of the second stage \( \frac{\kappa_1}{\beta_1} \rightarrow \frac{\kappa_2}{\beta_2} \), resulting in cost and processing time increases demonstrated in Section 2.4. Alternatively, one could have a process that looks like a single-stage process with only the first stage \( \rho \rightarrow 0 \). In this case, if the first stage is less expensive or time-consuming than the second stage, then adding an additional stage decreases costs and processing time. Depending on these sets of parameters – in particular the response of adjudicators and the persistence of applicants – the effects of adding an initial adjudicative stage could go in either direction for all three outcomes.

**Adding Selective Appeals**

As shown above, the existence of dynamic adjudicators and heterogeneous appeals stages is sufficient to make the effects of adding an early adjudication stage theoretically ambiguous. However, it is worth noting that an alternative extension can generate the same theoretical ambiguity: selective attrition from appeals.

Formally, I relax the assumption \( \rho(d_i) = \rho \). Leaving the function \( \rho(d_i) \) unspecified allows for the possibility that applicants update their likelihood of eligibility in a Bayesian fashion, either accurately or inaccurately. It could also reflect a difference in persistence correlated with health. This may be an important channel in practice: it is common for applicants to wait 1-2 years for a hearing, over which time their health could substantially deteriorate. I again focus on a two-stage process, and show that allowance rates could be larger or smaller at each stage in the dynamic model relative to the static model.

**Proposition 4:** When adjudicators take other adjudicative stages into account and applicants selectively persist as a function of underlying eligibility, the effects of adding an appeals stage on allowance rates are theoretically ambiguous.
Proof: Let \( P_s \) indicate whether an applicant would have persisted to stage \( s \) conditional on being denied at stage \( s - 1 \). By analogous derivations to Section 2.3, equation (3) becomes

\[
\hat{d}_{2i} - E[\epsilon_{2i}|\hat{d}_{2i}, A_{1i} = 0, P_{1i} = 1] \geq \frac{\kappa}{\beta}
\]

while equation (5) becomes

\[
\hat{d}_{1i} - E\left[\{\hat{d}_{1i} - \epsilon_{1i}\}\left\{1 - G\left(q_p^{-1}\left(\frac{\kappa}{\beta}\right) + \epsilon_{1i} - \hat{d}_{1i}\right)\right\}\right] \geq \frac{\kappa}{\beta}
\]

where \( q_p(x) \equiv x - E[\epsilon_{2i}|x, A_{1i} = 0, P_{1i} = 1] \).

These equations do not have closed form solutions with specifying the function \( \rho(d_i) \) and the distribution of the error term \( G \). However, note that the conditional expectation \( E[\epsilon_{2i}|x, A_{1i} = 0, P_{1i} = 1] \) plays a critical role in both expressions. Intuitively, the second adjudicator may infer substantial information from the fact that an applicant persisted to the second stage, depending on the function \( \rho(d_i) \).

To show why allowances may be higher or lower in this model relative to the static model, consider two extreme cases. First, suppose that \( d_i \) is distributed bimodally, with half of the mass at \( \left(\frac{\kappa}{\beta} - \phi\right) \) for some \( \phi \) and the other half at \( \left(\frac{\kappa}{\beta} + 2\phi\right) \). Suppose further that the persistence function was piecewise constant, such that

\[
u_s(A_{si}) = \begin{cases} 1 & \text{if } d_i = \frac{\kappa}{\beta} + 2\phi \\ 0 & \text{if } d_i = \frac{\kappa}{\beta} - \phi \end{cases}
\]

With a single stage, over one-half of applicants are allowed, since with uniform noise over one-half of \( \hat{d}_{si} \) fall above \( \frac{\kappa}{\beta} \). However, when there exists a second stage, the second adjudicator will let in exactly those individuals with \( d_i = \frac{\kappa}{\beta} + 2\phi \). Since allowing any applicant who would not be let in at the second stage would result in a negative outcome for the first adjudicator, the first adjudicator lets in zero applicants. In equilibrium, one-half of applicants persist and are allowed – yielding an allowance rate lower than the case of a single stage.

At the other extreme, suppose that \( \rho(d_i) \) is arbitrarily flat. As the function flattens, the model converges to the model in 2.3, in which an additional adjudication stage can increase allowance rates.

QED

Note that the two models above are closely related: while the former assumes that adjudicators at different stages have different parameters for external reasons, the latter
suggests a micro-foundation for why different adjudicators may have different ultimate allowance rates. I do not empirically distinguish between these two models.

Depending on the distributions and parameters of the model, the effect of adding an additional appeal on total allowances, processing times, and administrative costs is theoretically ambiguous. If adjudicators use as much information as possible and either stages differ or applicants selectively exit, then the results inferred from the simpler static model may be of the wrong magnitude or even the wrong sign. Moreover, these effects largely depend on two theoretical objects: the extent to which adjudicators are willing to modify allowance rates \( \left( \frac{\kappa}{\beta} \right) \) and the degree of applicant persistence \( \left( \rho(d_i) \right) \).

### III. Empirical Background & Data

To illustrate the importance of these theoretical channels in practice, I focus on a quasi-experiment using data on the Social Security Administration (SSA) disability appeals process. This section describes the steps involved in SSA disability applications and appeals, as well as the administrative data that I use for this study.

The two largest disability insurance programs in the U.S. are adjudicated by the Social Security Administration. Social Security Disability Insurance (SSDI) provides cash benefits and Medicare to eligible individuals with a work history who are under age 65. Supplemental Security Income (SSI) provides smaller cash benefits and Medicaid to eligible individuals with low income and asset levels. Combined, the two programs paid cash benefits totaling over $180 billion to over 10 million individuals (Deshpande (2016); Social Security Administration Office of Retirement and Disability Policy (2018)). While the two programs have different financial eligibility criteria, the process to determine whether an applicant is medically and vocationally eligible is virtually identical.

In most states, applicants go through an initial adjudication procedure and multiple (optional) stages of appeal. In the initial stage, applicants submit paperwork and medical records to a local SSA field office. If the field office determines that the applicant is financially eligible (i.e. has a sufficient work history for SSDI or has sufficiently low income/
assets for SSI), then the field office submits the application to a state-administered Disability Determination Services (DDS) office for a medical determination. A DDS examiner studies these medical records with the assistance of adjudicative guidelines and in-office medical and vocational experts, and documents details of the case for SSA’s internal system using Form 831. The examiner allows the award if she concludes that the applicant is unable to engage in Substantial Gainful Activity ($1,220 per month in 2019) due to an impairment lasting more than one year or resulting in death. An applicant who is allowed is immediately eligible for cash benefits, often including back pay, and is eligible for health benefits either immediately (SSI) or after a 5 month waiting period (SSDI). In the late 1990s, an initial adjudication took about 100 days on average and about one-third of applicants were initially allowed.

If the applicant is denied at the initial stage, they have the option to appeal. Traditionally, the first stage of appeal is a reconsideration, in which the applicant may request that another DDS examiner assess their case. To do so, the denied applicant mails a simple form to SSA within 60 days of their initial denial, and submits any additional new evidence regarding their impairments. In the late 1990s, reconsiderations took about 150 days to process on average. While about one-third of all applicants requested a reconsideration (one-half of all initially denied applicants), in the administrative data only 4 − 5% of all applicants were allowed at reconsideration (yielding a reconsideration allowance rate below 15%).

If an applicant is again denied at reconsideration, they may appeal to an Administrative Law Judge (ALJ). This is a longer process, in which a judge working for SSA studies the case and gives the applicant an opportunity to submit new evidence. Typically, applicants employ a legal representative at this stage. About one-quarter of ALJ decisions did not require a full hearing, including dismissals and “on the record” decisions performed by either an ALJ or an attorney adjudicator. However, the majority of ALJ requests did result in a full hearing. In the period I study (1996-2006), approximately 20% of all initial applicants request an ALJ hearing and around 13% are allowed by an ALJ, yielding an ALJ allowance rate of about two-thirds. However, applying to an ALJ is costly: in the late 1990s it required an additional wait of almost 300 days on average, during which period any gainful
employment would reduce the chances of allowance. The passage of time could strengthen or weaken a case for allowance, given actual and presumed changing employment opportunities and the possible progression of a disability. While applicants denied at the ALJ can technically appeal to an additional appeals stage through the SSA Appeals Council, less than 3% of all applicants would reach this stage in a typical filing year (Social Security Advisory Board (2012)).

These institutions provide a natural setting to study the effects of appeals. In addition to being relatively transparent and observable through administrative data sources, there exists a useful historical quasi-experiment. On October 1, 1999, ten states’ DDS offices were converted to a Prototype Model, representing around 20% of incoming applications nationwide. For all new applications filed after that date in these states, DDS would implement three new procedures:

1. DDS gave examiners Single Decision Maker (SDM) authority, letting them make some specific medical determinations without a medical consultant’s signature;

2. DDS implemented the Claimant Conference (CC) policy, in which a denied applicant had the right to discuss their case with the examiner who made the determination by phone;

3. DDS eliminated the reconsideration step, so that initially denied applicants would appeal directly to an ALJ.

Many states volunteered for small-scale trials of new disability adjudication procedures throughout the mid-1990s, and which states decided to implement the Prototype model was largely based on idiosyncratic factors related to DDS and agency leadership. Figure 1 shows the states that implemented the Prototype model in red, along with states

3 States systematically selecting into Prototype status would threaten identification of causal effects if states facing different counterfactual trends selected into treatment. Differences in levels of variables would not be a concern, since all point estimates are normalized to the office level using fixed effects. While selection on trends remains a theoretical possibility, it would likely be evident in pre- and post-trends. The point estimates in Figures 3-8 are very stable, suggesting that this is not a major concern.
mentioned in Appendix C that tested the SDM policy independently during the late 1990s (so-called *SDM II* states) and states with no policy changes. California and New York are *Partial Prototype* states: in California only DDS offices in the Los Angeles area implemented the Prototype model, while in New York the Prototype model was rolled out in stages from October 1999 through April 2001.

I am able to credibly isolate the effect of eliminating reconsiderations from these two other policies. First, Appendix C provides evidence that expanding SDM authority did not have economically meaningful impacts on allowance rates, processing times, or administrative costs. While the SDM policy could have had meaningful impacts on specific subgroups or in some offices, it does not appear to explain the average treatment effects I see across the whole sample. Second, after the Claimant Conference was implemented in October 1999, DDS officials reported a sharp increase in processing times (Morton (2018)). As a result, states begin experimenting with less formal versions of the Claimant Conference throughout 2001, and the Commissioner of Social Security announced the formal elimination of the Claimant Conference in June 2002. While the Claimant Conference could have retained a temporary influence over examiner behavior after that point, the high turnover of DDS examiners suggests that such an effect would fade quickly. I interpret the remaining long-run differences – which turn out to be very stable over many subsequent years – as the effect of eliminating reconsiderations.

In order study these policy changes, I combine three administrative data sets. First, Form 831 provides information on the universe of DI and SSI applications that received a DDS medical determination from the early 1990s onwards. By matching applications based on SSN, adjudicative level, and the date of the DDS decision, I link initial decisions and reconsiderations for the same application. Second, the Office of Hearings and Appeals Case Control System (OHACCS) legacy system records ALJ appeals through the 1990s into the mid-2000s, including key dates in the hearings process and ultimate ALJ decisions as used in French and Song (2014). Third, the Case Processing Management System (CPMS) records similar information on ALJ appeals from the mid-2000s onwards. Harmonizing OHACCS and CPMS yields a consistent series of ALJ appeals information. By merging this appeals information onto Form 831, I am able to track initial applications, reconsideration
appeals, and ALJ appeals consistently since the early 1990s. Appendix B discusses the data construction process in more detail. Table 1 shows the characteristics of the sample in Non-Prototype and Prototype offices, respectively, for applications filed in 1998. While these numbers move slightly over my extended sample period, they corroborate the descriptions above.

**IV. Empirical Results**

This section provides evidence that the elimination of reconsiderations led to a meaningful increase in allowance rate, as well as a modest change in administrative costs and processing times, due to the dynamic effects described in Section 2. Rather than the decrease in allowances implied by the static model, a difference-in-difference method detects a clean 3pp increase in combined allowances from all three adjudicative stages. As suggested by the dynamic model, this discrepancy can be accounted for in the data through two channels: initial adjudicators substantially increased allowance rates when the subsequent reconsideration step was unavailable, and applicants persisted to the ALJ appeals stage at a higher rate (where a high fraction were allowed). While eliminating an appeals step reduced administrative costs and processing times on net, it did so by much less than it would have absent these dynamic responses.

**Empirical Method**

The unit of analysis for this study is the field office-state combination. While most field offices send the vast majority of cases to a single state DDS, field offices near state borders may send some applications to neighboring state DDS offices. I code applications as Prototype applications if they are processed by states following the Full Prototype model (see Figure 1). I code applications processed by state DDS offices following the Partial Prototype model (New York and California) as Prototype if they are handled by a field office that switched to the Prototype model on October 1, 1999. Field offices that switched later (parts of New York) are excluded from the analysis. To simplify exposition, these state-field office combinations will be referred to as “offices”. Unless otherwise specified,
my results exclude SSI child cases and focus on adult cases. This does not affect the key results of the study, and allows me to focus on a sample with simpler and more uniform regulations over time. Later robustness checks report results for SSI child cases.

I first show that the number and composition of applications to SSA disability programs did not change discontinuously in the years of key policy changes. Figure 2 shows the total number of applications relative to 1996 with a medical determination in each year. The solid vertical line marks the year that the Prototype policy was implemented (1999) while the dashed vertical line marks the year that the claimant conference was eliminated (2002). To also make this figure relevant for Appendix C and Appendix D, I split non-prototype offices into those that rolled out only the Single Decision Maker model in the late 1990s (so-called SDM II states) and those with no such policy changes. The figure does not reveal any sharp changes in relative applicant flows around 1999 or 2002 between any of these office groups. This also provides evidence that most applicants were not strategically re-applying instead of pursuing an appeal: while rejected applicants often do apply again at a later date, there may be little incentive to apply in lieu of an appeal because cases can by linked by Social Security. Similarly, Appendix Figure A3 does not reveal stark changes in composition around these policy changes.

Given the lack of selection into applications, I estimate the causal impacts of the claimant conference and the elimination of reconsiderations on applicant outcomes using a difference-in-difference method. Each of the subsequent figures studies an outcome $Y_{ict}$ at the level of the DDS office $i$, the claim type $c$ (DI, SSI, or concurrent), and the calendar quarter $t$. The left panel shows the time series of raw averages $\bar{Y}_{it}$ for Prototype and non-Prototype offices. The right panel shows the coefficients $\beta_t$ from the regression

$$Y_{ict} = \mu_i + \mu_t + \sum_{t} \beta_t 1(Prototype)_t + \gamma X_{ict} + e_{ict}$$

Observations $\{i, c, t\}$ are weighted by the number of applications, and standard errors are clustered at the office level. $\beta_t$ is normalized to give 1998Q1 a coefficient of zero. The vector includes a rich set of controls: the fraction of applications associated with each body system; the fraction of applications in each 10-year age bin; and a binned, rolling measure
of the average state unemployment rate over the past year separately interacted with claim types (DI, Concurrent, SSI). When considering processing times, I mitigate the influence of outliers by winsorizing the top and bottom 1% of applications and ignoring hearings that occur more than two years after the previous decision date.

To illustrate this method and to demonstrate that my data accurately capture key policy changes, Figure 3 shows the fraction of DDS decisions that proceeded to a reconsideration stage in each calendar quarter. While about one-third of applicants typically proceed to a reconsideration, this number drops to essentially zero in Prototype states at the end of 1999. Similarly, Figure 4 shows stark evidence of the implementation and repeal of the Claimant Conference policy. Processing times for an initial decision moved in tandem across Prototype and non-Prototype states during the late 1990s and the mid-2000s. However, for the brief period in which the Claimant Conference was in place, processing times spiked by over 20 days. This spike decreased as states began experimenting with new ways to expedite the Claimant Conference, and was completely eliminated when the Commissioner announced the formal repeal of the policy.

**Effects on Allowance Rates**

Perhaps the most consequential and theoretically interesting outcome to study in this context is the effect on overall allowance rates: did the elimination of an appeals stage decrease or increase the unconditional allowance rate?

Figure 5 provides strong evidence that the elimination of reconsiderations increased overall allowance rates by about 3 percentage points (∼6%). While the allowance rate in Prototype offices was typically 3pp lower than in non-Prototype offices during the 1990s, this relative value jumped up sharply by 5pp when the Prototype policy was implemented. While this initial jump captures both the elimination of reconsiderations and the Claimant

---

4 Monthly state unemployment rates are available through BLS Local Area Unemployment Statistics. Note that claim type fixed effects are captured by the interaction of claim type and state unemployment rate bins.

5 The number of reconsiderations should not fall to exactly zero. For example, individuals who move states mid-application may receive a reconsideration from another state.
Conference, the latter policy was quickly wound down – after which there remained a stable and very persistent relative increase of $3pp$.

While lasting effects from the Claimant Conference policy could explain some of this $3pp$ difference in theory, there is substantive evidence against this explanation. On the applicant side, the lack of changes in application flows or characteristics (Figures 2 and A3) suggest that the existence of the Claimant Conference or its aftermath was not affecting applicants’ decisions to apply for benefits. On the DDS side, an effect driven by examiner memory would likely fade quickly due to high turnover of DDS examiners. Annual examiner turnover averaged 13% nationwide in the early 2000s, with one-quarter of DDS offices experiencing turnover greater than 20% (GAO (2004)). By the mid-2000s, it is likely that approximately half of DDS examiners had never experienced the Claimant Conference policy. Yet the $3pp$ coefficient is remarkably stable into the late 2000s. While I cannot rule out all possible channels of institutional memory, the very stable effect size suggests a consistent incentive throughout this period, rather than the long-term aftermath of a short-lived policy.

This $3pp$ increase contrasts starkly with the prediction of the static model in Section 2.2. In the late 1990s, over 4% of applicants were allowed at the reconsideration stage (Table 1). Without that opportunity to appeal, a reasonable observer might expect that eliminating reconsiderations resulted in a decrease in allowances exceeding $4pp$. The reasons for this $7pp$ discrepancy can be traced to at least two independent reactions: that of initial examiners, and that of denied applicants.

**Reaction of Initial Examiners**

The first driver of higher overall allowances is the fact that initial allowance rates increased sharply after the elimination of reconsiderations. Figure 6 shows that relative rates of initial allowance jumped suddenly in late 1999, rising almost $10pp$ in Prototype offices. As the Claimant Conference was wound down and ultimately eliminated, initial allowance rates remained elevated in Prototype offices by about $4pp$. Again, this effect is persistent
for many subsequent years, suggesting the existence of a persistent incentive rather than decaying institutional memory. This persistent jump in initial allowance rates was large enough to make up for the loss of reconsideration allowances: the relative percent of applicants allowed at either the initial or reconsideration stages is almost identical in the late 1990s and the early 2000s, even though Prototype offices did not have the reconsideration stage in the early 2000s (Figure A4). These regression coefficients are especially pronounced for DI applicants (who have higher allowance rates anyways) and are substantially smaller for SSI child cases (Figure A5). Mapping this finding to Section 2, this supports a strong dynamic reaction with a high value \((\frac{k}{\beta})\).

**Reaction of Applicants**

The second driver of higher overall allowance rates is the fact that more applicants appealed to ALJs after the elimination of reconsiderations. ALJs have historically had much higher rates of allowance than reconsideration stages, so this increase in ALJ appeals increased allowances. I measure the fraction of applicants pursuing hearings using the frequency of dispositions, which include decisions by a judge and cases that are dismissed for any reason (e.g. SSI ineligibility, no shows). Figure 7 suggests that eliminating reconsiderations increased the fraction of applicants receiving an ALJ disposition by 5pp. An alternative measure is the fraction who complete a full hearing, which excludes all dismissed cases and excludes cases decided “on the record” (i.e. without an in-person hearing). Figure A6 shows that the fraction completing a full hearing rose by 2.5pp. Most of these applicants were allowed, either on-the-record or at a hearing, resulting in an increase in allowances by ALJ around 2.5pp (Figure 8). This is consistent with the model in Section 2.4 and a reasonably low degree of persistence \((\rho)\).

Both sets of results give evidence that the channels illustrated in the dynamic model are of first order importance in this setting. The sharp and persistent relative changes in allowance rates suggest that eliminating reconsiderations increased overall allowances on
the order of 3\(pp\), rather than decreasing allowances as a static model would predict. Two particular channels – the forward-looking behavior of early adjudicators \(\left(\frac{\kappa}{\beta}\right)\) and the incomplete persistence of applicants \(\rho\) – receive direct support from the data.

**Effects on Administrative Costs and Processing Times**

This quasi-experiment can inform two additional outcomes of programmatic interest: the administrative costs of the adjudication system, and the total time that an applicant spends in the combined adjudication process.

While not directly observable in my setting, the effects of eliminating an appeals stage on program administrative costs can be estimated through a simple formula. Recent testimony suggests that a typical reconsideration costs $666 to administer, while a typical ALJ appeal costs $2,771 (Morton 2018). From the results above, about 1/3 of applicants in Prototype states would have received a reconsideration, but instead an additional 5\(pp\) received an ALJ decision. Using these (modern) cost numbers with the (historical) take up numbers, I calculate a change in per application administrative costs of

\[
0.05(2771) - 0.33(666) \approx -80
\]

While these numbers should be taken as rough estimates, they suggest that the net cost savings from eliminating a stage of appeal ($80) were approximately one-third of the gross cost savings ($222) due to the higher propensity of applicants to pursue expensive ALJ appeals.

Finally, a similar formula suggests that the elimination of reconsideration decreased average wait times modestly in this context. Historically, reconsiderations took about 150 days while hearings took closer to 300 days (Table 1). The net change in average wait times would therefore be approximately

\[
0.05(300) - 0.33(150) \approx -35 \text{ days}
\]
Again, the net effect of eliminating reconsiderations appears to be substantially smaller than the gross effect (50 days) given applicant reactions. To corroborate this estimate, I directly estimate the effect of eliminating reconsiderations on the number of days from initial filing to the last recorded decision (initial, reconsideration, of ALJ) in Figure A7. While the figure is not nearly as clean as previous findings, likely due to intertemporal variation in ALJ backlogs adding noise to the series’, it provides suggestive evidence of a small decrease in wait times on the order of 15 – 20 days.

These calculations highlight the importance of considering incomplete applicant persistence when considering adjudicative policy. In the context of SSA disability programs, eliminating an intermediate appeal induced a substantial number of applicants to pursue more expensive and longer ALJ appeals. The cost savings from eliminating this stage were substantially reduced by this reaction, and with a large enough reaction could have been reversed.

V. Conclusion

Despite how ubiquitous appeals are across adjudicative and legal procedures, researchers have little evidence on their effects. Given the two key properties of most appeals – that appeals are asymmetric and information is incomplete – an appeal is not simply another chance at allowance. If adjudicators consider the sequence of appeals at each stage, an additional appeal results in a new equilibrium with theoretically ambiguous effects. I establish the theoretical importance of examiner reactions \((\kappa \beta)\) and applicant persistence \((\rho(d_i))\) in determining the sign of these effects. I then illustrate the importance of these factors in a case study of SSA disability appeals. Instead of reducing overall allowances, the elimination of an intermediate appeal appears to have increased allowances by 3ppp while having only small impacts on administrative costs and processing times. In practice, the extent to which adjudicators are forward-looking and the degree of persistence by applicants are of first-order importance in the design of adjudicative processes.
References


This figure marks Prototype states (red) that implemented the Prototype model for all DDS claims filed on or after 1 October 1999; Partial Prototype states (orange) where some offices implemented the Prototype model on 1 October 1999 while other either delayed or did not implement the model; SDM II states (pink) that implemented Single Decision Maker authority gradually from 1997-1999; and states with no systematic policy changes over this period (tan).

This figure shows the annual aggregate number of total applications (DI, SSI, and concurrent) relative to 1996 for three groups of DDS offices: Prototype offices (red), SDM II offices (blue), and offices with no policy changes (green).
**Figure 3**

This figure shows the fraction of applicants with a medical determination that received a DDS reconsideration in each office-quarter (left) and the associated quarterly regression coefficients (right), excluding child SSI cases. In the right figure, points show regression coefficients relative to 1998Q1 from a regression at the office-quarter level, and error bars show 95 percent confidence intervals clustered at the office level. The regression controls for the fraction of applications associated with each body system, the fraction in each 10-year age bin, and a binned rolling measure of the average state unemployment rate over the past year interacted with claim types.

**Figure 4**

This figure shows the average number of days (winsorized) from filing to an initial DDS decision in each office-quarter (left) and the associated quarterly regression coefficients (right), excluding child SSI cases. In the right figure, points show regression coefficients relative to 1998Q1 from a regression at the office-quarter level, and error bars show 95 percent confidence intervals clustered at the office level. The regression controls for the fraction of applications associated with each body system, the fraction in each 10-year age bin, and a binned rolling measure of the average state unemployment rate over the past year interacted with claim types.
Figure 5

This figure shows the allowance rate at any appeals stage (initial, reconsideration, ALJ) in each office-quarter (left) and the associated quarterly regression coefficients (right), excluding child SSI cases. In the right figure, points show regression coefficients relative to 1998Q1 from a regression at the office-quarter level, and error bars show 95 percent confidence intervals clustered at the office level. The regression controls for the fraction of applications associated with each body system, the fraction in each 10-year age bin, and a binned rolling measure of the average state unemployment rate over the past year interacted with claim types.

Figure 6

This figure shows the allowance rate at the initial stage in each office-quarter (left) and the associated quarterly regression coefficients (right), excluding child SSI cases. In the right figure, points show regression coefficients relative to 1998Q1 from a regression at the office-quarter level, and error bars show 95 percent confidence intervals clustered at the office level. The regression controls for the fraction of applications associated with each body system, the fraction in each 10-year age bin, and a binned rolling measure of the average state unemployment rate over the past year interacted with claim types.
Figure 7

This figure shows the fraction of applications eventually receiving an ALJ disposition (including on-the-record and dismissed cases) in each office-quarter (left) and the associated quarterly regression coefficients (right), excluding child SSI cases. In the right figure, points show regression coefficients relative to 1998Q1 from a regression at the office-quarter level, and error bars show 95 percent confidence intervals clustered at the office level. The regression controls for the fraction of applications associated with each body system, the fraction in each 10-year age bin, and a binned rolling measure of the average state unemployment rate over the past year interacted with claim types.

Figure 8

This figure shows the fraction of applications allowed by an ALJ in each office-quarter (left) and the associated quarterly regression coefficients (right), excluding child SSI cases. In the right figure, points show regression coefficients relative to 1998Q1 from a regression at the office-quarter level, and error bars show 95 percent confidence intervals clustered at the office level. The regression controls for the fraction of applications associated with each body system, the fraction in each 10-year age bin, and a binned rolling measure of the average state unemployment rate over the past year interacted with claim types.
### Table 1

<table>
<thead>
<tr>
<th></th>
<th>Non-Prototype</th>
<th>Prototype</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (1000s)</td>
<td>1,103</td>
<td>294</td>
</tr>
<tr>
<td>DI Only (%)</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>DI + SSI (%)</td>
<td>32</td>
<td>28</td>
</tr>
<tr>
<td>SSI Only (%)</td>
<td>33</td>
<td>37</td>
</tr>
<tr>
<td>Average Age (Years)</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>Musculoskeletal (%)</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Mental (%)</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>Allowed at Any Stage (%)</td>
<td>52</td>
<td>49</td>
</tr>
<tr>
<td>Allowed at Initial (%)</td>
<td>36</td>
<td>32</td>
</tr>
<tr>
<td>Had Recon (%)</td>
<td>33</td>
<td>36</td>
</tr>
<tr>
<td>Allowed at Recon (%)</td>
<td>5.1</td>
<td>4.6</td>
</tr>
<tr>
<td>Had ALJ Decision (%)</td>
<td>18</td>
<td>20</td>
</tr>
<tr>
<td>Allowed by ALJ (%)</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Days to Initial</td>
<td>99</td>
<td>97</td>
</tr>
<tr>
<td>Days from Initial to Recon</td>
<td>145</td>
<td>145</td>
</tr>
<tr>
<td>Days from Recon to ALJ</td>
<td>278</td>
<td>272</td>
</tr>
</tbody>
</table>

Table shows the composition of applications with a medical determination in 1998 across offices that would soon be Prototype or Non-Prototype offices. Table excludes SSI child cases, which constitute about 20% of all cases.
Appendix A: Multi-Stage Dynamic Models

This appendix shows that the principles from Proposition 2, which led to muted responses of additional appeals on allowance rates in the dynamic model, also extend to models with more stages. I consider a model with three stages, again letting \( \rho(d_i) = \rho \).

The third (terminal) adjudicator’s cutoff rule satisfies:

\[
\beta E[d_i|\tilde{d}_{3i}, A_{3i} = 0, A_{2i} = 0] - \kappa \geq 0
\]

\[
\tilde{d}_{3i} \geq \frac{\kappa}{\beta} + E[\epsilon_{3i}|\tilde{d}_{3i}, A_{3i} = 0, A_{2i} = 0]
\]

which implies

\[
Pr(A_{3i} = 1|A_{1i} = 0, A_{2i} = 0, d_i) = 1 - G\left(q_3^{-1}\left(\frac{\kappa}{\beta}\right) - d_i\right) < 1 - G\left(q_3^{-1}\left(\frac{\kappa}{\beta}\right) - d_i\right)
\]

where \( q_3(x) \equiv x - E[\epsilon_{3i}|\tilde{d}_{3i}, A_{3i} = 0, A_{2i} = 0] \). The inequality comes from the censoring argument in Section 2.3, which implies that \( E[\epsilon_{3i}|\tilde{d}_{3i}, A_{3i} = 0, A_{2i} = 0] > E[\epsilon_{3i}|\tilde{d}_{3i}, A_{3i} = 0] \) and therefore \( q_3^{-1}(x) > q_3^{-1}(x) \). Relative to the second adjudicator in a two-stage process, the third adjudicator in the three-stage process gets additional information from the fact that the applicant was denied a second time. The second adjudicator’s cutoff rule becomes

\[
\beta E[d_i|\tilde{d}_{2i}, A_{1i} = 0] - \kappa \geq \rho \beta E[d_i Pr(A_{3i} = 1|A_{1i} = 0, A_{2i} = 0, d_i)|\tilde{d}_{1i}]
\]

\[
\tilde{d}_{2i} - E[\epsilon_{2i}|\tilde{d}_{2i}, A_{1i} = 0] = \rho \left[\beta E[d_i Pr(A_{3i} = 1|A_{1i} = 0, A_{2i} = 0, d_i)|\tilde{d}_{1i}] - E[\epsilon_{2i} - \tilde{d}_{2i}]\right] \geq \frac{\kappa}{\beta}
\]

This expression looks like equation (3) but has an additional term on the left-hand side that is decreasing in \( \tilde{d}_{2i} \). The cutoff \( \tilde{d}_{2i} \) in this case is even higher than that in equation (3), since the second adjudicator now gleans information about \( d_i \) from the first adjudicator’s denial and anticipates the existence of a third adjudicator. The first adjudicator’s cutoff rule takes on a complex form that does not lend itself to a closed-form solution:

\[
\beta E[d_i|\tilde{d}_{1i}] - \kappa \geq \rho \beta E[d_i Pr(A_{2i} = 1|A_{1i} = 0, A_{2i})|\tilde{d}_{1i}] + \rho \beta E[d_i Pr(A_{3i} = 1|A_{1i} = 0, A_{2i} = 0, d_i)|\tilde{d}_{1i}]
\]

\[
\tilde{d}_{1i} \geq \rho E[d_i Pr(A_{2i} = 1|A_{1i} = 0, d_i)|\tilde{d}_{1i}] - \rho E[d_i Pr(A_{3i} = 1|A_{1i} = 0, A_{2i} = 0, d_i)|\tilde{d}_{1i}] \geq \frac{\kappa}{\beta}
\]

Just as the cutoff was higher in (5) than in (1), this cutoff \( \tilde{d}_{1i} \) is even larger due to the first adjudicator anticipating two subsequent adjudicators. While the existence of three or more stages adds substantial complexity to the equations governing the model, the principles remain the same.
In a static model, adding more stages of adjudication will necessarily increase allowance rates. However, in a dynamic model both the early and later adjudicators may decrease their allowance rates with the addition or more adjudicators, resulting in muted effects of a new appeal on overall allowance rates.

Appendix B: Details of Data Construction

This section explains the data construction process in detail. Most data are available in reliable form from the mid-1990s through the late 2000s. Most results focus on the decade from 1996-2006 in order to provide a reasonable, symmetric window around the policy changes I study, although slight extensions of the time window in either direction do not meaningfully change any results.

First, I append all Form 831 records satisfying a small set of conditions: each record is for a primary claimant pursuing either a disabled or blind case under Title II (DI) or Title XVI (SSI). I consider a claim to be concurrent if a Title II and a Title XVI application have the same SSN, DDS decision date, and adjudication level. When considering filing dates or demographic variables for concurrent applications, I typically use the entry for the DI application. My main sample keeps only the first initial decision for a given SSN, filing date, claim type, and adjudication level. This drops about 1% of all rows and does not meaningfully affect my results.

Second, I split this the sample into initial and reconsideration decisions. I re-merge these two samples together by associating a reconsideration decision with an initial decision if the two decisions shares an SSN, do not indicate conflicting claim types, and have the temporally closest decision dates of all such pairs.

Third, I append information on ALJ appeals from OHACCS and CPMS. I keep all regular Title II and Title XVI applications (e.g. excluding Continuing Disability Reviews) in both samples. I drop duplicate hearings from the early 2000s, when OHACCS was phasing out of use and CPMS was phasing in. I then merge hearings data onto Form 831 data using a similar procedure as I used to merge initial and reconsideration decisions. In particular, I associate an ALJ decision with an initial claim if it shares an SSN, is not for a conflicting claim type, and is the first appeal request date since the last available (initial or reconsideration) decision date.
The resulting file includes a row for every initial application with a medical decision, and follows these applications as they move to reconsideration and ALJ appeals stages. I winsorize all wait time variables at the top and bottom 1%.

**Appendix C: Effects of Single Decision Maker Authority**

This section provides evidence that the single decision maker policy (SDM) had small effects on allowances and processing times. This allows me to estimate the effect of eliminating reconsiderations without serious concerns about confounding due to SDM policies. To estimate these effects, I compare states that gradually rolled out the SDM policy (so-called SDM II states) to states that experienced no policy changes in the period 1997-1999. Since the SDM policy excludes children applying for SSI, I exclude this group from the subsequent analysis. This method does not rule out all possible impacts of SDM, particularly on certain subsets of applicants, but assuages concerns over a large average treatment effect that could provide an alternative explanation of my core results.

I run regressions of the same form specified in Section 4.1, except I omit Prototype offices and focus on offices adopting SDM. I also focus on the late 1990s, when SDM II states were gradually rolling out that policy. The regression takes the form:

\[
Y_{ict} = \mu_i + \mu_t + \sum_t \alpha_t 1(SDM)_i + \gamma X_{ict} + e_{ict}
\]

Because SDM rolled out gradually in treated states – mostly in 1997 and 1998 – a treatment effect of SDM at the initial DDS level would be detectable as a gradual change in allowance rates or processing times. Figures A1 and A2 show that neither of these trends seem to hold cleanly in the data. Figures A3 and A4 run similar regressions that include later stages of appeal. SDM does not appear to have affected average allowance rates at any stage of the process. There is some evidence that SDM could have increased the time of deliberation at later stages due to a lack of medical evidence, although the impact on average overall processing times is small (< 2 weeks). Readers who wish to adjust other results for this possibility can add an additional drop of -10 days on Figure A7, bringing this figure closer to the processing time estimate of -35 days on page 23.
**Figure A1**

This figure shows the allowance rate of initial DDS decision for SDM II states vs. No Change states in each office-quarter (left) and the associated quarterly regression coefficients (right), excluding child SSI cases. In the right figure, points show regression coefficients relative to 1998Q1 from a regression at the office-quarter level, and error bars show 95 percent confidence intervals clustered at the office level. The regression controls for the fraction of applications associated with each body system, the fraction in each 10-year age bin, and a binned rolling measure of the average state unemployment rate over the past year interacted with claim types.

**Figure A2**

This figure shows the average number of days (winsorized) from filing to an initial DDS decision for SDM II states vs. No Change states in each office-quarter (left) and the associated quarterly regression coefficients (right), excluding child SSI cases. In the right figure, points show regression coefficients relative to 1998Q1 from a regression at the office-quarter level, and error bars show 95 percent confidence intervals clustered at the office level. The regression controls for the fraction of applications associated with each body system, the fraction in each 10-year age bin, and a binned rolling measure of the average state unemployment rate over the past year interacted with claim types.
Figure A3

This figure shows the allowance at any stage for SDM II states vs. No Change states in each office-quarter (left) and the associated quarterly regression coefficients (right), excluding child SSI cases. In the right figure, points show regression coefficients relative to 1998Q1 from a regression at the office-quarter level, and error bars show 95 percent confidence intervals clustered at the office level. The regression controls for the fraction of applications associated with each body system, the fraction in each 10-year age bin, and a binned rolling measure of the average state unemployment rate over the past year interacted with claim types.

Figure A4

This figure shows the average number of days (winsorized) from filing to the last recorded decision for SDM II states vs. No Change states in each office-quarter (left) and the associated quarterly regression coefficients (right), excluding child SSI cases. In the right figure, points show regression coefficients relative to 1998Q1 from a regression at the office-quarter level, and error bars show 95 percent confidence intervals clustered at the office level. The regression controls for the fraction of applications associated with each body system, the fraction in each 10-year age bin, and a binned rolling measure of the average state unemployment rate over the past year interacted with claim types.
Appendix D: Additional Figures

Figure A3

This figure shows the fraction of applicants with musculoskeletal or mental impairments (left) and the fraction aged over 50 at the time of filing (right) across three groups of DDS offices: Prototype offices (red), SDM II offices (blue), and offices with no policy changes (green).

Figure A4

This figure shows the allowance rate at the initial or reconsideration stages in each office-quarter (left) and the associated quarterly regression coefficients (right), excluding child SSI cases. In the right figure, points show regression coefficients relative to 1998Q1 from a regression at the office-quarter level, and error bars show 95 percent confidence intervals clustered at the office level. The regression controls for the fraction of applications associated with each body system, the fraction in each 10-year age bin, and a binned rolling measure of the average state unemployment rate over the past year interacted with claim types.
This figure shows the coefficient on initial allowance rates from quarterly regressions, run separately for each claim type. The regression controls for the fraction of applications associated with each body system, the fraction in each 10-year age bin, and a binned rolling measure of the average state unemployment rate over the past year interacted with claim types.

Figure A6

This figure shows the fraction of applications with a completed hearing date (not on-the-record nor dismissed) in each office-quarter (left) and the associated quarterly regression coefficients (right), excluding child SSI cases. In the right figure, points show regression coefficients relative to 1998Q1 from a regression at the office-quarter level, and error bars show 95 percent confidence intervals clustered at the office level. The regression controls for the fraction of applications associated with each body system, the fraction in each 10-year age bin, and a binned rolling measure of the average state unemployment rate over the past year interacted with claim types.

Figure A7
This figure shows the average days from filing to the last recorded decision (winsorized) in each office-quarter (left) and the associated quarterly regression coefficients (right), excluding child SSI cases. In the right figure, points show regression coefficients relative to 1998Q1 from a regression at the office-quarter level, and error bars show 95 percent confidence intervals clustered at the office level. The regression controls for the fraction of applications associated with each body system, the fraction in each 10-year age bin, and a binned rolling measure of the average state unemployment rate over the past year interacted with claim types.