Econometrica, Vol. 92, No. 1 (January, 2024), 79–110

DESIGNING DISABILITY INSURANCE REFORMS: TIGHTENING ELIGIBILITY RULES OR REDUCING BENEFITS?

ANDREAS HALLER
Department of Economics, Norwegian School of Economics

STEFAN STAUBLI
Department of Economics, University of Calgary, CEPR, IZA, and NBER

JOSEF ZWEIMÜLLER
Department of Economics, University of Zurich, CEPR, CESifo, and IZA

This paper develops a sufficient statistics framework for analyzing the welfare effects of disability insurance (DI). We derive social-optimality conditions for the two main DI policy parameters: (i) eligibility rules and (ii) benefit levels. Applying this framework to two restrictive DI reforms in Austria, we find that tighter DI eligibility rules triggered higher fiscal cost savings and lower insurance losses. Hence, tighter DI eligibility rules dominate DI benefit reductions in scaling back the Austrian DI system.

KEYWORDS: Disability insurance, screening, benefits, policy reform.

1. INTRODUCTION

DESPITE IMPROVING HEALTH, HIGHER MATERIAL LIVING STANDARDS, and less physically demanding working conditions, the number of disability insurance (DI) recipients has risen rapidly over the past decades in most OECD countries. The increasing financial burden of DI programs has led many governments to implement DI reforms aiming explicitly at reducing the DI program inflow and DI expenditures. While restrictive DI reforms lessen the financial burden for taxpayers, they also impose utility losses on individuals with a disability. The welfare consequences ultimately depend on how DI reforms address this incentive-insurance trade-off.

DI systems insure individuals against income losses due to a long-lasting medical impairment. Such impairments are often difficult to verify (as for mental health or musculoskeletal problems). Therefore, a formal DI application process with a medical assessment is integral to any DI system. Since a DI application is costly to a worker, her decision to apply for DI will depend on the DI benefit level and the likelihood of a favorable dis-

Andreas Haller: andreas.haller@nhh.no
Stefan Staubli: sstaubli@ucalgary.ca
Josef Zweimüller: josef.zweimueller@econ.uzh.ch

We thank Tabea Bucher-Koenen, Raj Chetty, Richard Disney, Thomas Hoe, Lucija Muehlenbachs, Timothy Moore, Matthew Notowidigdo, Luigi Pistaferri, Philippe Ruh, Florian Scheuer, Johannes Spinnewijn, Alexander Strand, Conny Wunsch, and seminar participants at Erasmus University Rotterdam, University of Amsterdam, University of Bonn, University of Manitoba, University of Melbourne, Universitat Pompeu Fabra, University of Salzburg, University of Zurich, CEPR Labour Economics Symposium, CEPR Public Economics Symposium, CEPR/NBER Aging/Health workshop, NBER Summer Institute, workshop on "Family, Aging, Social Insurance" in Bergen, and the VfS Population Economics meeting in Basel for helpful comments. This research was supported by the U.S. Social Security Administration through Grant #1DRC12000002-04 to the National Bureau of Economic Research as part of the SSA Disability Research Consortium. The findings and conclusions expressed are solely those of the authors and do not represent the views of SSA, any agency of the Federal Government, or the NBER. All remaining errors are our own.

ability determination. Thus, there are two main policy parameters through which a DI reform can affect the DI inflow: DI benefits and DI eligibility rules.

In this paper, we develop a comprehensive sufficient statistics approach to assess the social optimality of existing DI systems.[1] Our approach builds upon previous theoretical work by Diamond and Sheshinski (1995). It characterizes the socially optimal DI system by an optimal DI benefit level and an optimal DI eligibility rule (modeled as a disability threshold above which the government considers an applicant deserving of DI). The resulting optimality conditions highlight the set of sufficient statistics needed to assess the optimality of a given DI system. An advantage of our framework is that we can directly compare the welfare effects of the two (rather different) policy parameters. If a government wants to curb DI expenditures, our framework will reveal whether it should better reduce DI benefits or set stricter DI eligibility rules.

While optimal DI benefits have been studied in previous work and generally follow the logic of the Baily–Chetty formula for optimal UI, our analysis of optimal DI eligibility rules is new.[2] It turns out that the welfare analysis of the stricter rules is different from the one of lower benefits. The reason is imperfect screening. While a DI benefit reduction affects all—including the most disabled—DI recipients, stricter DI eligibility rules affect only DI applicants who are screened out. Stricter rules may affect only few applicants, but their insurance losses may be substantial. In other words, the two policies affect individuals differentially, and their welfare effects become a question of targeting: Who loses how much? Answering this question is tricky because we cannot observe the true disability level of screened-out applicants.

Our analysis aims to compare the welfare consequences of DI benefits and DI eligibility rules. To accomplish this aim, we need to compare their relative insurance losses to their relative incentive costs. We capture the incentive costs of the respective instrument by a "fiscal multiplier," the total fiscal cost savings of a DI reform relative to the "mechanical" fiscal effect (the hypothetical cost savings absent any change in behavior). Any difference between total and mechanical cost savings is due to the DI reform distorting individual behavior (DI applications and labor supply).[3] Evaluating the relative insurance losses of stricter eligibility and lower benefits is challenging because they affect individuals with different intensities. We make progress on this front by deriving upper and lower bounds of the relative insurance losses in terms of marginal utilities of consumption. These bounds do not depend on individuals' disability levels which we cannot observe. Furthermore, extending Fadlon and Heien Nielsen (2019), we show that the relative insurance-loss bounds can be inferred from spousal labor supply responses.

The second main contribution of our paper is the empirical implementation of our sufficient statistics framework. To obtain reduced-form estimates of workers' responses

[1]The sufficient statistics approach has been pioneered by Chetty (2006a) in the context of unemployment insurance (UI). Chetty (2006a) derived empirically implementable formulas to assess the optimality of existing UI systems, building on earlier theoretical work by Baily (1978).

[2]Meyer and Mok (2019) and Deshpande, Gross, and Su (2021) studied optimal DI benefits with a sufficient statistics approach. Ball and Low (2014) estimated the effect of DI on consumption in the UK to infer the insurance value of DI benefits.

[3]The concept of the fiscal multiplier to measure overall program costs closely relates to the fiscal externalities estimated by Lee, Leung, O'Leary, Pei, and Quach (2021) and Hendren and Sprung-Keyser (2020). Lee et al. (2021) estimated the fiscal externality of UI benefit reforms. The fiscal externality is the behavioral fiscal effect relative to the mechanical fiscal effect. What we refer to as the fiscal multiplier is 1+fiscal externality. Hendren and Sprung-Keyser (2020) used the concept of the Marginal Value of Public Funds ("MVPF"), which is the willingness to pay for a policy divided by the net cost to the government. Our multiplier is a way to measure the net cost of a policy.

to changes in DI benefits and DI eligibility rules, we exploit two Austrian DI reforms and estimate their effects using population data from Austrian administrative DI registers. The first DI reform was implemented in 2003 and changed the pension formula, resulting in substantially lower DI benefit levels for some individuals but less so for others. The quasi-experimental variation in benefit levels over time and across individuals allows us to identify the causal effect of DI benefits. The second DI reform was implemented in 2013 and tightened DI eligibility rules. Specifically, the reform raised the "relaxed screening age" (RSA), the critical age above which not only medical but also vocational factors are taken into account in the DI assessment process. Once workers reach the RSA, access to DI benefits becomes much easier, and DI award rates increase strongly. Before 2013, the RSA was 57. The 2013 DI reform increased it step-wise to age 60. We focus on the first (and second) RSA hike, from age 57 to age 58 (and 59). The RSA increase allows identifying the causal effect of stricter rules through a comparison of cohorts: The older (control) cohort still faces the lenient pre-reform DI rules with the RSA at age 57. In contrast, the younger (treated) cohorts are subject to tight DI rules at age 57, and the RSA only applies from age 58 (or 59) onward.

Our reduced-form estimates reveal that both DI reforms generated significant behavioral responses, substantially lowering DI program costs. We estimate a fiscal multiplier of reduced DI benefits of around 1.4. That is, reducing DI benefits by one Euro creates an additional 40 Cents in cost savings because fewer individuals apply for DI benefits. We find even stronger behavioral responses for stricter eligibility rules and estimate fiscal multipliers between 2.0 and 2.5.

Estimating the fiscal multipliers associated with the two DI policy instruments requires separate estimates of the total fiscal cost savings (the numerator of the multiplier) and the mechanical cost savings (the denominator). In the case of lower DI benefits, estimating the fiscal multiplier is straightforward. The behavioral fiscal effect can be estimated using standard program evaluation techniques, while the mechanical fiscal effect can be directly calculated by applying the known benefit cut to the control group. Estimating the fiscal multiplier of stricter DI eligibility rules is more tricky. Standard program evaluation methods allow us to estimate the effect of tighter rules on total fiscal cost savings. However, splitting the total effect into its mechanical and behavioral components is difficult. The reason is that, in the data, we cannot distinguish marginal applicants (who abstain from applying in response to tighter eligibility rules) from always applicants (who do not change their application behavior). We argue—and provide supportive evidence—that we can estimate the mechanical DI cost savings (resulting from lower DI award rates for always applicants) from the group of previously rejected DI applicants. Workers in this group have already applied for DI under the strict pre-RSA rule, indicating a sufficiently severe disability. To the extent that their re-application behavior is not affected by RSA rules, fiscal cost reductions in response to the RSA increase are entirely mechanical. Based on this strategy, we estimate a fiscal multiplier between 2.0 and 2.5. This estimate turns out robust to alternative specifications and subsamples.

To estimate bounds on the insurance losses of stricter DI eligibility rules relative to lower DI benefits, we follow Fadlon and Heien Nielsen (2019) and exploit spousal labor supply responses. The upper bound of relative insurance losses depends on the marginal utilities of consumption among marginal entrants (those who no longer receive DI under the stricter rules) relative to the marginal utilities of all DI recipients. This bound captures the intuition that lower DI benefits affect all DI recipients, while more stringent rules only affect individuals who no longer qualify for DI. Following Fadlon and Heien Nielsen (2019), we show that the marginal utilities of consumption can be expressed in terms of spousal labor supply responses to a DI shock. Empirically, we document that stricter eligibility rules are associated with higher spousal earnings among DI entrants. Interpreted through the lens of Fadlon and Heien Nielsen's (2019) approach, this finding implies that the marginal DI recipient (i.e., DI entrant) values DI less than the average DI recipient. Consequently, the insurance loss of stricter eligibility is smaller than the loss of lower benefits. The exact implementation of the insurance loss bounds also depends on ρ, the curvature of the spousal disutility of labor. Based on estimates from the literature, we set $\rho = 0.6$ and find that the relative insurance losses of stricter eligibility rules versus lower benefits lie between 0.09 and 0.50.

Taken together, our empirical results suggest a clear ranking for Austrian DI policies: stricter DI eligibility rules dominate reduced DI benefits as a policy tool for rolling back the DI program. They generate higher fiscal savings and impose smaller insurance losses on affected workers.

Our paper contributes to the theoretical and empirical literature on the labor market and welfare effects of DI programs (for reviews, see Bound and Burkhauser (1999); Low and Pistaferri (2020)). Unlike existing sufficient statistics approaches, which focus predominantly on (marginal) changes in transfers, we highlight the welfare implications of variations in eligibility criteria and compare them to changes in transfers. Previous studies exploring the incentive-insurance trade-off in DI programs have relied on structural models (Benitez-Silva, Buchinsky, and Rust (2004), Bound, Stinebrickner, and Waidmann (2010), Autor, Kostøl, Mogstad, and Setzler (2019)). Low and Pistaferri (2015) structurally estimated a life-cycle framework to assess optimal DI benefits and DI eligibility criteria for the United States.[4] Autor et al. (2019) studied spousal labor supply responses to DI receipt of marginal appellants by exploiting random assignment to administrative law judges and used their reduced-form evidence to estimate a structural model. Our sufficient statistics analysis complements these structural approaches.

Our study also contributes to the empirical DI literature using reduced-form methods (Parsons (1991), Gruber and Kubik (1997), Autor and Duggan (2003)). We add to this literature by providing novel evidence on the causal effect of DI eligibility rules using population data from social security records and DI applications. Finally, our paper contributes to the literature exploring the determinants of DI application behavior and labor supply (Gruber (2000), Campolieti (2004), Mullen and Staubli (2016), Deshpande and Li (2019), Garcia-Mandicó, García-Gómez, Gielen, and O'Donnell (2020), Godard, Koning, and Lindeboom (2022)).[5]

The paper is organized as follows. The next section presents a model of disability insurance and derives formulas for optimal disability eligibility and benefits. Section 3 describes the data and the institutional background. Sections 4 and 5 present the empirical results

[4]In their analysis, Low and Pistaferri (2015) emphasized the importance of type I errors (rejections of deserving DI applicants) and provided evidence on the welfare effects of this margin in the U.S. context (see also the discussion in Low and Pistaferri (2020)). The advantage of our approach is that it infers the welfare effects directly from behavioral responses to DI reforms and does not rely on information on classification errors. The disadvantage is that we cannot explicitly address policies that improve the screening accuracy.

[5]An important strand of the DI literature studies the impact of DI receipt on labor force participation by comparing accepted and rejected DI applicants (Bound (1989); Chen and van der Klaauw (2008); Maestas, Mullen, and Strand (2013), French and Song (2014); Autor et al. (2019)). We do not directly study the labor supply of accepted versus rejected DI applicants. Still, changes in labor force participation to stricter eligibility rules are reflected in our program cost estimates. We also do not study outflow from DI (Campolieti and Riddell (2012); Borghans, Gielen, and Luttmer (2014); Moore (2015)) or earnings of DI recipients (Kostol and Mogstad (2014); Gelber, Moore, and Strand (2017); Ruh and Staubli (2019); and Kostøl and Myhre (2021)). However, in our analysis, these responses also enter the fiscal multiplier.

on stricter DI eligibility rules and lower DI benefit levels. Section 6 estimates the fiscal multipliers of the two policy instruments. Section 7 estimates the bounds on the relative insurance losses of stricter eligibility versus lower benefits exploiting spousal labor supply responses. Section 8 concludes.[6]

2. THEORETICAL FRAMEWORK

In this section, we theoretically explore how the two main DI policy parameters—strictness of DI eligibility rules and level of DI benefits—affect social welfare, labor supply, and application behavior of potential DI claimants.[7] We focus on a static model based on Diamond and Sheshinski (1995). It highlights the main trade-offs and intuition. Online Appendix A.2 extends the analysis to a general dynamic setting, showing that the same trade-offs appear.

Setup

A disability shock θ is modeled as a random draw from a continuous distribution $F(\theta)$. If the disability is not very severe (θ is small), the agent works and enjoys utility $u(c^w) - \theta$, where $c^w = w - \tau + A$ is the after-tax labor income, $w - \tau$, plus other income A. If the disability is severe (θ is large), the agent applies for DI benefits. Applying causes disutility ψ, capturing the extensive medical checks and bureaucratic hassle associated with the DI assessment process. With probability $p(\theta)$ the application is accepted, where $p'(\theta) > 0$. If the application is accepted, the agent claims DI benefits b, and gets utility $v(c^b) - \psi$, where $c^b = b + A$. If the application is rejected, the applicant either resumes work and gets utility $u(c^w) - \theta - \psi$, or the applicant claims welfare benefits $z < b$ and gets utility $v(c^z) - \psi$, where $c^z = z + A$. No disutility or uncertainty is associated with claiming welfare benefits.

DI Applications and Labor Supply

An agent prefers working over claiming welfare benefits if her disability is $\theta < \theta^R \equiv u(c^w) - v(c^z) > 0$, that is, the utility of claiming welfare benefits falls short of the utility of working. Hence, θ^R is the disability of the "marginal welfare recipient." The "marginal DI applicant" who is indifferent between filing a DI application and remaining employed has disability

$$\theta^A = u(c^w) - v(c^b) + \frac{\psi}{p(\theta^A)}. \tag{1}$$

Panel (a) of Figure 1 characterizes the outcome of an agent's DI application choice. It draws the probability of a DI award $p(\theta)$ against θ and indicates the disability cutoffs θ^A and θ^R. Agents with a disability $\theta < \theta^A$ remain employed. Agents with a disability $\theta \geq \theta^A$ apply for DI; if rejected, those with a disability $\theta \in [\theta^A, \theta^R)$ return to work, while those with $\theta \geq \theta^R$ go on welfare benefits.[8]

[6]See Haller, Staubli, and Zweimüller (2024) for the Online Appendix. Additional Supplementary Material may be found in Haller, Staubli, and Zweimüller (2023).

[7]By increasing the "strictness of DI eligibility rules," we mean any policy that reduces the likelihood of a DI benefit award for an applicant with a given disability. Low and Pistaferri (2015) and Diamond and Sheshinski

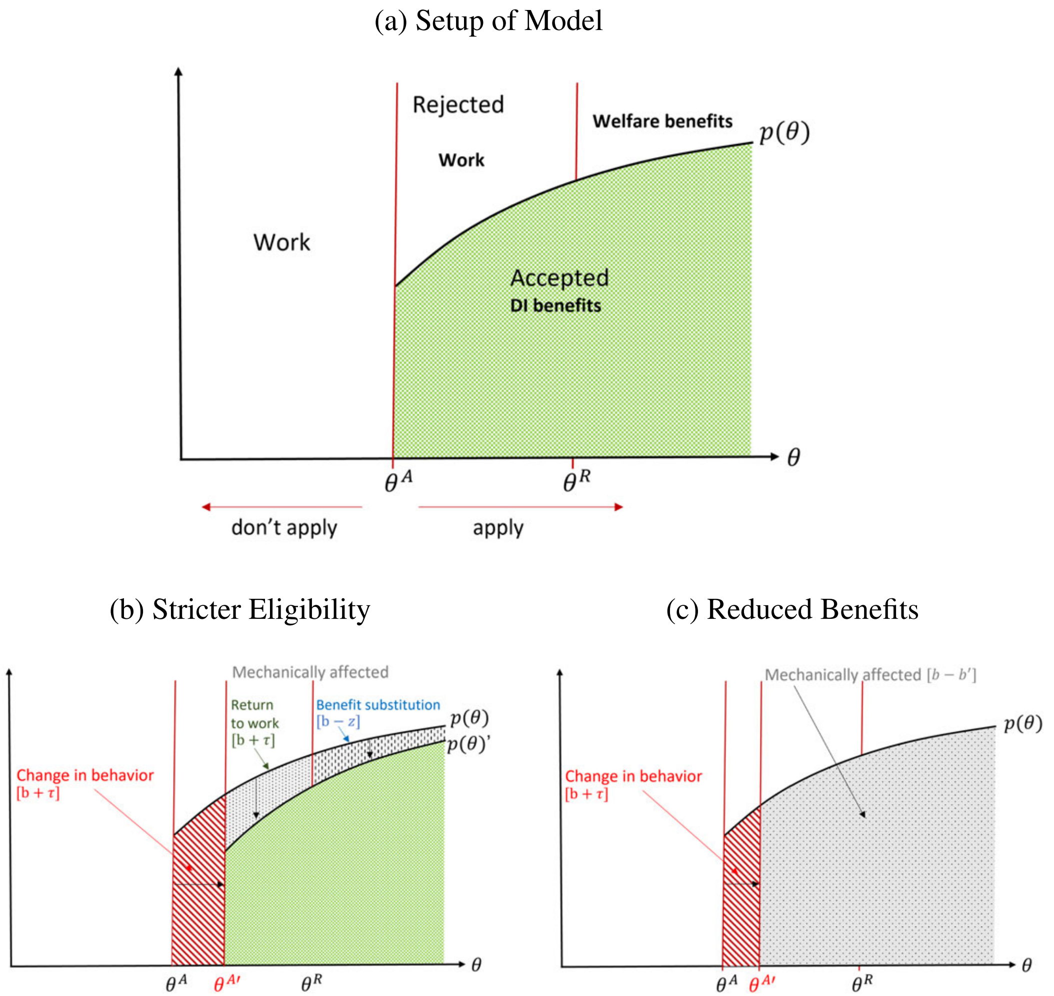

FIGURE 1.—Graphical representation of static model. Notes: Panel (a) illustrates the basic setup. Individuals are characterized by disability level θ and can choose whether to work, apply to DI, or leave the labor force and consume social welfare benefits. The award process to DI is noisy and individuals are awarded DI with probability $p(\theta)$. We assume that $p(\theta)$ is weakly increasing in θ. This captures that (i) it is difficult to assess the true disability level of an individual and (ii) the assessment contains nonetheless some valuable information on the true disability level. The marginal DI applicant is denoted by θ^A and individuals with $\theta \geq \theta^A$ apply to DI. The marginal welfare benefits type is denoted by θ^R and individuals with $\theta \geq \theta^R$ will go on welfare benefits if they are rejected. Panel (b) illustrates the effects of stricter eligibility criteria. Stricter criteria shift down the award probability curve. How much the curve is shifted downwards at each θ depends on how the stricter rules affect the award probabilities at different disability levels. The area between the two award probability curves is the mechanical effect on DI entry. A fraction of the mechanically rejected applicants returns to work (dotted area) with a fiscal impact of $[b + \tau]$. The other fraction substitutes DI benefits with welfare benefits (vertically dashed area) with a fiscal impact of $[b - z]$. Stricter eligibility criteria also change application behavior (shift the marginal applicant to the right). The area under the $p(\theta)$-curve between θ^A and $\theta^{A'}$ measures the impact of the change in application behavior on DI entry. The behavioral fiscal effect is this diagonally striped area multiplied by its fiscal impact, $[b + \tau]$. Panel (c) illustrates the effects of reduced benefits. Lower benefits make DI less attractive and reduce applications (diagonally striped area) with a fiscal impact of $b + \tau$. Lower benefits affect all DI recipients mechanically (dotted area). The mechanical fiscal effect of lower benefits is the dotted area multiplied by the difference in benefits $b - b'$.

DI Policy Instruments

We now assess the welfare effects of two policy instruments that characterize any DI system: the level of DI benefits b and the strictness of DI eligibility rules θ^*. While the role of DI benefits is straightforward and poses no major conceptual problems, the role of DI eligibility rules needs further discussion. The inherent difficulty of the DI assessment process is that the true disability θ is the agent's private information. For this reason, a DI applicant has to undergo a disability assessment process, which delivers an estimate of her disability to the government. Formally, the government observes $s = \theta + e(\theta)$, where s is a noisy signal, θ is the applicant's true disability, and $e(\theta)$ is the noise. The strictness of DI eligibility rules, which is under the direct control of the government, can be captured by a critical value of s, call it θ^*. DI applications with $s \geq \theta^*$ are accepted, while applications with $s < \theta^*$ are rejected. The acceptance probability can then be written as $p(\theta; \theta^*)$.[9] In what follows, we consider the case where the government can change θ^* but takes the signal as given.

Welfare Effects of DI Reforms

We follow the literature assuming that the government maximizes a utilitarian social welfare function subject to a government budget constraint:

$$W(\theta^*, b) = V(\theta^*, b) + \lambda\big(G(\theta^*, b) - \bar{G}\big), \tag{2}$$

where $V(\theta^*, b)$ denotes the aggregate indirect utility function, $G(\theta^*, b)$ denotes the total fiscal revenue, and $\bar{G}$ is an exogenous revenue constraint. $V(\theta^*, b)$ is given by

$$\begin{aligned} V(\theta^*, b) = {} & \int_0^{\theta^A} \big(u(c^w) - \theta\big)\, dF(\theta) + \int_{\theta^A}^{\theta^R} \big(1 - p(\theta; \theta^*)\big)\big(u(c^w) - \theta\big)\, dF(\theta) \\ & + \int_{\theta^A}^{\infty} p(\theta; \theta^*) v(c^b)\, dF(\theta) + \int_{\theta^R}^{\infty} \big(1 - p(\theta; \theta^*)\big) v(c^z)\, dF(\theta) \\ & - \int_{\theta^A}^{\infty} \psi dF(\theta). \end{aligned} \tag{3}$$

It sums up the utility levels of the various agents: the working healthy (first term), the rejected DI applicants resuming work (second term), the DI recipients (third term), and the welfare benefit recipients (fourth term). The fifth term captures the utility losses from filing a DI application. The total fiscal revenue $G(\theta^*, b)$ is the tax revenue from working

(1995) used the terminology "strictness of screening" and "disability standard," respectively. We discuss the formal definition of strictness below.

[8]Equation (1) and its graphical representation in Figure 1 apply if $\theta^A < \theta^R$, that is, a marginal applicant returns to work if her DI application is rejected. This assumption is not critical for our results, and we impose it here to simplify the exposition. The general dynamic model in Online Appendix A.2 features heterogeneity in wages, benefits, and disability severity θ, which can lead to situations where $\theta_i^A > \theta_i^R$ for some individuals i. Our empirical implementation follows the general model, and we thus do not rely on the assumption $\theta^A < \theta^R$.

[9]We assume that the DI assessment process is informative so that the award probability increases with the severity of the disability, or $\partial p(\theta; \theta^*)/\partial \theta \geq 0$. This assumption implies that, in an applicant pool with more severely disabled, a smaller fraction of DI assessments falls short of the cutoff θ^*.

individuals minus expenditures for the disability and other welfare programs:

$$G(\theta^*, b) = \tau\left[F(\theta^A) + \int_{\theta^A}^{\theta^R} 1 - p(\theta; \theta^*)\, dF(\theta)\right] - b\int_{\theta^A}^{\infty} p(\theta; \theta^*)\, dF(\theta) - z\int_{\theta^R}^{\infty} 1 - p(\theta; \theta^*)\, dF(\theta). \tag{4}$$

We are now ready to discuss the welfare effects of stricter DI eligibility rules and lower DI benefits. We frame the discussion around implementing a more restrictive DI system because most policy debates center around reducing the financial burden of the DI program. Of course, analogous arguments hold for reforms that increase the generosity of the DI system.

Stricter DI Eligibility Rules: Marginal Increase in θ^*

The utilitarian government sets DI eligibility rules θ^* to maximize social welfare W. In Online Appendix A.1, we show that the welfare effect of increasing θ^* is

$$\frac{\partial W(\theta^*, b)}{\partial \theta^*} = \underbrace{\int_{\theta^A}^{\infty} \frac{\partial p(\theta; \theta^*)}{\partial \theta^*}\left(v(c^b) - \max\{v(c^z), u(c^w) - \theta\}\right) dF(\theta)}_{\text{insurance losses}} + \lambda \underbrace{\left[B(\theta^*) + M(\theta^*)\right]}_{\text{fiscal cost reduction}}. \tag{5}$$

Condition (5) highlights the two opposing effects of stricter DI eligibility rules on social welfare. On the one hand, a higher θ^* reduces social welfare because fewer agents are awarded DI (insurance losses). On the other hand, a higher θ^* raises social welfare because it saves taxpayers money (fiscal cost reduction).

Panel (b) of Figure 1 graphically illustrates the effects of stricter eligibility criteria. Stricter criteria shift down the award probability curve (from $p(\theta)$ to $p(\theta)'$). How much the curve is shifted downwards at each θ depends on how the stricter rules affect the award probabilities at different disability levels. A higher θ^* has two distinct effects: a behavioral and a mechanical effect.

A higher θ^* reduces DI applications because the lower award probability reduces the expected value of an application, shifting the marginal applicant θ^A to the right to $\theta^{A'}$. The diagonally striped area in Panel (b) of Figure 1 captures the effect of this change in application behavior on DI entry. This change in behavior creates no direct insurance loss because of the envelope theorem. Changes in behavior thus do not show up in the insurance loss in condition (5).[10] However, the change in application behavior has a fiscal impact. Individuals who no longer apply return to work with fiscal consequences $[b + \tau]$ (save DI benefits plus collect additional tax revenue). The behavioral fiscal effect $B(\theta^*) \equiv (\partial\theta^A/\partial\theta^*)p(\theta^A; \theta^*)f(\theta^A)[b + \tau]$ in condition (5) therefore corresponds to the diagonally striped area in Panel (b) multiplied by $[b + \tau]$.

A higher θ^* also mechanically reduces DI entry by $-\int_{\theta^A}^{\infty}(\partial p(\theta; \theta^*)/\partial\theta^*)\, dF(\theta)$, the area between the old and new $p(\theta)$-curves in Panel (b) of Figure 1. This mechanical increase in

[10]For the formal argument, see Online Appendix A.1.

rejections creates the insurance loss in condition (5). For rejected applicants who return to work, the dotted area between θ^A and θ^R with mass $M_W \equiv -\int_{\theta^A}^{\theta^R} (\partial p(\theta;\theta^*)/\partial\theta^*)\,dF(\theta)$, the utility loss is $v(c^b) - (u(c^w) - \theta) > 0$. For rejected applicants who substitute to welfare benefits, the vertically dashed area above θ^R with mass $M_Z \equiv -\int_{\theta^R}^{\infty} (\partial p(\theta;\theta^*)/\partial\theta^*)\,dF(\theta)$, the utility loss is $v(c^b) - v(c^z) > 0$. Next to creating an insurance loss, the increase in rejections also has a fiscal impact. This mechanical fiscal effect is $M(\theta^*) \equiv M_W \cdot [b + \tau] + M_Z \cdot [b - z]$ because each rejected applicant resuming work saves the amount $[b + \tau]$ and each rejected applicant substituting DI for welfare benefits saves $[b - z]$ to the taxpayer. The total fiscal cost reduction of stricter eligibility criteria is then the sum of the behavioral and mechanical fiscal effect $B(\theta^*) + M(\theta^*)$ and is valued at λ—the value of public funds.

The optimal strictness of eligibility rules θ^* balances the trade-off between insurance losses and fiscal cost reductions, that is, (5) is set to zero. For later use, we rewrite this condition as

$$\frac{\partial W}{\partial \theta^*} \gtreqless 0$$

$$\Leftrightarrow \quad 1 + \frac{B(\theta^*)}{M(\theta^*)} \gtreqless \frac{\displaystyle\int_{\theta^A}^{\infty} -\frac{\partial p(\theta;\theta^*)}{\partial\theta^*}\big(v(c^b) - \max\{v(c^z), u(c^w) - \theta\}\big)\,dF(\theta)}{\lambda M(\theta^*)}. \quad (6)$$

The two sides of the inequality have an intuitive interpretation. The left-hand side is the fiscal multiplier, $1 + B(\theta^*)/M(\theta^*)$. It measures the reduction in the financial burden for the taxpayer per *mechanically* saved dollar (the hypothetical fiscal gain when application behavior remains unchanged). The right-hand side measures the reduction of the insurance value in monetary units per mechanically saved dollar.

Lower DI Benefits

The second key DI policy parameter is the level of DI benefits b. Reducing benefits has again two opposing effects on social welfare. On the one hand, a lower b reduces the insurance value because all DI recipients receive a lower transfer. The insurance value equals the benefit reduction times the marginal utility of consumption of DI recipients, $\int_{\theta^A}^{\infty} p(\theta;\theta^*)v'(c^b)\,dF(\theta)$. On the other hand, a lower b reduces DI program expenditures through a mechanical and a behavioral fiscal effect. The mechanical fiscal effect captures that lower benefits reduce DI expenditures for current beneficiaries, $M(b) \equiv \int_{\theta^A}^{\infty} p(\theta;\theta^*)\,dF(\theta)$. The dotted area in Panel (c) of Figure 1 illustrates this mechanical effect: All DI recipients are affected by the reduction in benefits $[b - b']$. Lower benefits also induce fewer agents to apply for DI benefits, which reduces DI entry (the diagonally striped area in Panel (c)). The behavioral fiscal effect is then the reduction in entry multiplied by its fiscal impact, $B(b) \equiv (\partial\theta^A/\partial b)p(\theta^A;\theta^*)f(\theta^A)[b + \tau]$, that is, the diagonally striped area in Panel (c) multiplied by $[b + \tau]$. The multiplier $1 + B(b)/M(b)$ captures the total fiscal gain for a one-dollar reduction in DI benefits.

The optimal DI benefit level balances the insurance loss and the fiscal effects of a benefit change (see Online Appendix A.1 for details):

$$\frac{\partial W}{\partial(-b)} \gtreqless 0 \quad \Leftrightarrow \quad 1 + \frac{B(b)}{M(b)} \gtreqless \frac{\displaystyle\int_{\theta^A}^{\infty} p(\theta;\theta^*)v'(c^b)\,dF(\theta)}{\lambda M(b)}. \quad (7)$$

Reducing DI benefits is welfare-improving if the total fiscal gain exceeds the insurance loss, which is the same logic as in the Baily–Chetty formula for optimal unemployment benefits.

Stricter Eligibility Rules versus Lower Benefits

How can we compare the two instruments, and which one should a policymaker use to curb DI program expenditures? From a welfare perspective, she should use the instrument with a higher ratio of the welfare effect to total cost savings. That is, she should choose stricter eligibility criteria θ^* over lower benefits b if and only if $\frac{W_{\theta^*}}{T_{\theta^*}} > \frac{W_b}{T_b}$.[11] We can rewrite this condition as the ratio of fiscal multipliers to the ratio of relative insurance losses:

$$\frac{1+B(\theta^*)/M(\theta^*)}{1+B(b)/M(b)} > \frac{V_{\theta^*}}{V_b}, \tag{8}$$

where V_{θ^*} and V_b denote the insurance losses associated with stricter eligibility and lower benefits, respectively. The relative insurance losses are given by

$$\frac{V_{\theta^*}}{V_b} = \frac{\dfrac{1}{\lambda M(\theta^*)}\displaystyle\int_{\theta^A}^{\infty} \frac{-\partial p(\theta;\theta^*)}{\partial\theta^*}\big(v(c^b)-\max\{v(c^z),u(c^w)-\theta\}\big)\,dF(\theta)}{\dfrac{1}{\lambda M(b)}\displaystyle\int_{\theta^A}^{\infty} p(\theta;\theta^*)v'(c^b)\,dF(\theta)}, \tag{9}$$

which nicely illustrates the trade-off between screening and benefits: Stricter screening affects only individuals who are screened out $\frac{-\partial p(\theta;\theta^*)}{\partial\theta^*}$ (dotted and dashed areas in Panel (b) of Figure 1) but they might experience a large insurance loss $v(c^b) - \max\{v(c^z), u(c^w) - \theta\}$. Less generous DI benefits affect all DI recipients (dotted area in Panel (c) of Figure 1), even the most deserving, but the insurance loss might be smaller. Hence, the relative insurance losses depend on the targeting effects of the two instruments: Who loses how much?

The key challenge to empirically estimate equation (9) is that the insurance loss associated with stricter eligibility rules, V_{θ^*}, consists of differences in utility levels and the true disability level θ is unobserved. The techniques to estimate insurance losses in other social insurances (mostly UI) identify differences in marginal utilities. To make progress, we proceed in two steps. First, we derive upper and lower bounds for the relative insurance losses that express equation (9) in marginal utilities of consumption. Second, we use results from Fadlon and Heien Nielsen (2019) that relate the relative insurance losses to spousal labor supply responses. Here, we derive the bounds, and Section 7 discusses the empirical implementation using spousal labor supply. We find that the empirical upper and lower bounds are tight enough to be informative.

For the upper bound of $\frac{V_{\theta^*}}{V_b}$, we use the following upper bound for V_{θ^*}:

$$V_{\theta^*} = \frac{1}{\lambda M(\theta^*)}\int_{\theta^A}^{\infty} \frac{-\partial p(\theta;\theta^*)}{\partial\theta^*}\big(v(c^b)-\max\{v(c^z),u(c^w)-\theta\}\big)\,dF(\theta) \tag{10}$$

[11] Online Appendix A.1 provides the formal proof of this condition.

$$\leq \frac{1}{\lambda M(\theta^*)} \int_{\theta^A}^{\infty} \frac{-\partial p(\theta;\theta^*)}{\partial \theta^*} \big(v(c^b) - v(c^z)\big)\, dF(\theta) \tag{11}$$

$$\approx \frac{1}{\lambda M(\theta^*)} \int_{\theta^A}^{\infty} \frac{-\partial p(\theta;\theta^*)}{\partial \theta^*} v'(c^b)(c^b - c^z)\, dF(\theta) \tag{12}$$

$$\leq \frac{1}{\lambda M(\theta^*)} \int_{\theta^A}^{\infty} \frac{-\partial p(\theta;\theta^*)}{\partial \theta^*} v'(c^b)(b - z)\, dF(\theta) \tag{13}$$

$$\leq \frac{1}{\lambda} \int_{\theta^A}^{\infty} v'(c^b) \frac{\Delta DI(b-z)}{M(\theta^*)}\, dF(\theta). \tag{14}$$

The first inequality (11) uses a revealed preference argument. Individuals who return to work would be worse off if they instead received other welfare benefits. Hence, the effective insurance loss must be weakly smaller than the insurance loss of moving from DI to other benefits. Equation (12) uses a first-order Taylor approximation to express the insurance loss in terms of marginal utilities. The quality of this approximation depends on the curvature of the utility function. In Online Appendix A.1, we derive a bound for the approximation error as a function of risk aversion and find that the error is small in our application for reasonable values of risk aversion.[12] Equation (13) uses that the consumption drop is smaller than the income drop due to self-insurance mechanisms. In (14), ΔDI is an indicator function for individuals who no longer make it into DI because of the stricter rules. Either they stop applying, or they do apply and are mechanically screened out. Thus, (14) simply adds the marginal utility of marginal applicants, $p(\theta^A;\theta^*)v'(c^b)(b-z)$, to the integral in (13). Bound (14) is convenient because it also holds for non-marginal changes in θ^*. Moreover, we can directly implement it empirically. Our upper bound therefore is[13]

$$\frac{V_{\theta^*}}{V_b} \leq \frac{\displaystyle\int_{\theta^A}^{\infty} v'(c^b) \frac{\Delta DI(b-z)}{M(\theta^*)}\, dF(\theta)}{E\big[v'(c^b)|\text{on DI}\big]}. \tag{15}$$

As lower bound, we use

$$\frac{V_{\theta^*}}{V_b} \geq \frac{\displaystyle\int_{\theta^R}^{\infty} v'(c^b) \frac{\frac{-\partial p(\theta;\theta^*)}{\partial \theta^*}(b-z)}{M(\theta^*)}\, dF(\theta)}{E\big[v'(c^b)|\text{on DI}\big]}, \tag{16}$$

which assumes that rejected individuals who return to work do not experience an insurance loss, and uses a first-order Taylor approximation to express the insurance losses in marginal utilities.

[12]In Online Appendix A.1, we show that the approximation error in our application is at most $\frac{1}{8}\cdot\gamma$, where γ denotes the coefficient of relative risk aversion. Chetty (2006b) argued that labor supply estimates from the literature imply $\gamma \approx 1$. In this case, our approximation error is at most 0.125. Hence, in Table IV, the upper bound would become 0.48 and 0.63 for the RSA-58 and RSA-59 change, respectively. Given our fiscal multiplier estimates, we would need extremely high values of risk aversion of $\gamma > 11$ to make our insurance value bounds inconclusive.

[13]It is straightforward to rewrite $V_b = \frac{1}{\lambda}\int_{\theta^A}^{\infty} v'(c^b)\frac{p(\theta)}{\int_{\theta^A}^{\infty} p(\theta)\,dF(\theta)}\, dF(\theta) = \frac{1}{\lambda}E[v'(c^b)|\text{on DI}]$.

Another challenge for the empirical implementation of equation (8) is the estimation of the mechanical fiscal effect of stricter eligibility criteria, $M(\theta^*)$. We address this empirical challenge in Section 6. Note that the mechanical fiscal effect simply rescales both sides of the incentive-insurance trade-off in equation (8). Hence, the mechanical fiscal effect estimate does not change any conclusions if we know the insurance losses and total fiscal impacts, which raises the question of why we rescale equation (8) by the mechanical fiscal effect in the first place. Rescaling by the mechanical fiscal effect allows us to independently evaluate the two sides of the incentive-insurance trade-off. It converts the insurance losses and fiscal impacts of stricter eligibility versus reduced benefits to the same units. Having meaningful estimates of both sides of the trade-off is valuable since the implementation of the two sides is based on different identifying assumptions.

Interestingly, type I and type II errors of the screening process—false rejections and false acceptances—do not directly appear in our welfare analysis. On the one hand, with a continuum of disability types, the definition of type I and type II errors is not obvious (who is truly deserving?). On the other hand, from a welfare perspective the classification errors must be weighted by their social costs and benefits, as the bounds in (15) and (16) show. This point relates to the findings of Deshpande and Lockwood (2022), who documented that the U.S. disability program plays a vital role in insuring nonhealth risks. Even for applicants with less severe health conditions, who might not meet the disability definition, the insurance value of DI benefits can exceed the social costs of providing these benefits. Their results, in line with our model, imply that classification errors based on health only are not sufficient for welfare analysis.[14]

Extensions

The static model highlights the key trade-offs but misses two critical features for evaluating DI reforms: heterogeneity across individuals beyond θ and intertemporal choices. As Online Appendix A.2 shows, the main insights from the static model generalize to a model that allows for multiple sources of heterogeneity and periods. Our empirical analysis exploits an increase in relaxed screening standards from age R to a higher age $R + \Delta$. The question "when should I apply?" becomes crucial in this context. The general model with multiple periods captures the intertemporal nature of the DI application choice and ensures a close connection between theory and empirical analysis.

The welfare analysis is derived for marginal changes in θ^* and b, but the empirical analysis exploits a discrete change in θ^* from a lenient to a strict standard. Online Appendices A.1 and A.2 show for the static and the dynamic model that the upper and lower bounds on the relative insurance values are robust to non-marginal changes in θ^* that are small.[15] For the mechanical effect, the distinction between always and marginal applicants becomes essential because the mechanical fiscal effect is driven by always applicants only. Empirically, we shed light on the responses of marginal and always applicants in Section 6 and estimate the fiscal multiplier for a discrete change in eligibility standards.

[14]In the Supplementary Material S.1, we discuss type I and type II errors formally.

[15]The non-marginal change in θ^* has to be small in the sense that it does not generate general equilibrium effects like changes in wages or aggregate saving behavior.

3. INSTITUTIONAL BACKGROUND AND DATA

3.1. *Institutional Background and Policy Variation*

Like many developed countries, Austria has three transfer programs providing income replacement in case of an economic or health shock: disability insurance (DI), sickness insurance (SI), and unemployment insurance (UI). The DI program is financed by a payroll tax. It provides partial earnings replacement to workers below the full retirement age with at least 5 insurance years in the last 10 years.[16] DI applicants must apply to the local DI office, where employees first check whether the applicant meets the formal requirements for DI receipt. Unlike in the United States, DI applicants are not required to stop working. Then a team of physicians and disability examiners assesses the severity of the medical impairment and the applicant's residual earnings capacity. An impairment is severe if it lasts at least six months and limits the applicant's mental or physical ability to engage in substantial gainful activity. Once benefits are awarded, DI beneficiaries receive monthly payments until their return to work, medical recovery, or death. DI benefits can be granted for a temporary period, but less than 4 percent of claimants ever leave the DI rolls.

DI Eligibility Rules

Applicants' residual earnings capacity depends on work experience and whether they are below or above a *relaxed screening age (RSA)* threshold, currently set at 60. Applicants below the RSA qualify for DI benefits only if the earnings capacity is less than 50% of the earnings capacity of a healthy person in *any reasonable* occupation the individual could be expected to carry out. Applicants above the RSA, who have worked for at least 10 years in the last 15 years, already qualify if the earnings capacity is less than 50% in a *similar* occupation.

The RSA was 57 until the end of 2012, but the 2. Stability Act (2. Stabilitätsgesetz) increased it in three one-year steps to age 60.[17] We exploit the reform-induced variation in the RSA to identify the labor market effects of stricter DI eligibility rules. The 2. Stability Act was announced in April 2012 and had two objectives: reduce expenditures in the public pension systems and foster employment among older workers. The only change to the DI program was the increase in the RSA to age 58 in January 2013, followed by further increases to age 59 in January 2015 and age 60 in January 2017. Individuals who worked in a similar occupation for less than 10 years in the last 15 years were unaffected by these increases. We focus on the RSA increases to ages 58 and 59 because the available data preclude the analysis of the RSA increase to age 60.

DI Benefits

DI benefits are subject to income and payroll taxes and replace approximately 70 percent of pre-disability net earnings up to a maximum of about 4500 Euros per month. The level of DI benefits is calculated by multiplying a pension coefficient, which varies by age

[16] Insurance years include contribution years (periods of employment, including maternity and sick leave) and non-contribution years (periods of unemployment, military service, or secondary education). The required insurance years increase by one month for every two months above age 50 up to a maximum of 15 insurance years.

[17] Staubli (2011) studied the labor market effects of an earlier increase in the RSA for men from 55 to 57 in 1996. However, he had no application data and could not study application behavior, which is essential to assess the welfare effects of DI reforms.

and insurance years, with an assessment basis, which is the average indexed capped earnings over a given period (e.g., the best 16 years in 2004). Applicants below age 56 with limited work experience qualify for a special increment to supplement their benefits. DI beneficiaries may continue work, but those earning more than an exempt threshold lose up to 50 percent of their benefits.

In January 2004, the Austrian government enacted a large pension reform (Pensionsreform 2003) that modified the DI benefit formula. The reform lowered potential DI benefits for most individuals by reducing the pension coefficient and increasing the length of the assessment basis from 16 to 40 years. These changes were phased in gradually between 2004 and 2017. Only individuals with limited work history experienced an increase in potential benefits, as the reform increased the special supplement's age limit from 56 to 60 between 2004 and 2010. The public heavily criticized the significant reduction in benefits. In response to the backlash, the Austrian government passed legislation in 2005, limiting the maximum benefit reduction to 5 percent of the projected pre-reform benefits. The maximum benefit reduction was then increased by 0.25 percent each year; in 2017, it was equal to 8.25 percent of pre-reform benefits. We use this variation to identify the labor market impacts of changes in DI benefit levels.[18]

SI and UI Benefits

In case of a temporary illness, employers continue to pay 100% of earnings for up to 12 weeks. Once the 12 weeks are exhausted, individuals may claim SI benefits which are taxed and replace approximately 65% of the last net wage up to the same maximum that applies to DI benefits. The potential SI benefit duration is 52 (26) weeks for individuals who have worked at least (less than) 6 months in the previous 12 months. UI benefits replace 55 percent of the last wage subject to a minimum and maximum. The potential UI benefit duration is 39 weeks for workers below 50 and 52 weeks for workers above 50 (provided they have paid UI contributions for at least 9 years in the last 15 years). Job losers who exhaust the UI benefits duration can apply for unemployment assistance. These means-tested transfers last indefinitely and are about 70 percent of regular UI benefits.

3.2. *Data*

We merge data from two administrative registers. First, the Austrian Social Security Database (ASSD) contains detailed longitudinal information for the universe of workers in Austria between 1972 and 2018. The ASSD records all employment, unemployment, disability, sick leave, and retirement spells, and some background characteristics (gender, month and year of birth, blue- or white-collar status). In addition, spells before 1972 are available for individuals who have claimed a public pension by the end of 2008. The ASSD also contains firm-specific information such as geographic region, industry affiliation, and firm identifiers that allow us to link individuals and firms. See Zweimüller et al. (2009) for a detailed data description. Second, we use data on all DI applications covering 2004 to 2017. They contain the application date, the decision date, the decision itself (i.e., reject or accept), the medical impairment of the applicant, and the stage of the application (i.e., first application, re-application, or appeal).

[18]Figure T.13 in the Supplementary Material illustrates the reform's impact on benefits by showing the changes in potential DI benefits between 2004 and 2017.

Starting from the population data set, we impose three restrictions. First, we exclude women. During our observation window, their eligibility age for an old-age pension gradually increased from age 56 to age 60, making it difficult to disentangle the impacts of the DI reforms and retirement age increases.[19] Second, we exclude the self-employed and civil servants because they are covered by a separate pension system. Third, we stop following people at age 62 when many become eligible for an old-age pension. Our sample covers over three-quarters of active labor market participants in Austria. Since we observe complete work histories, we can precisely calculate the amount of DI benefits individuals would get at any point in time and whether individuals have sufficient work experience to apply for DI benefits under the relaxed eligibility criteria above the RSA.

To study the effect of stricter DI eligibility, we focus on 54–61-year-old men born between 1954 and 1957. We split the sample into men with more and less than 10 employment years in the past 15 years (measured at age 56). Only men with more than 10 employment years in the past 15 years are eligible for relaxed DI eligibility.[20] We will use the sample of eligible men for our main effects and the sample of ineligible men for placebo tests. Each year, individuals are observed on the 1st of March, June, September, and December.

To study the effect of lower DI benefits, we focus on 30–60-year-old men from 2004 to 2017. Following Mullen and Staubli (2016), we define a reference date, January 1, and obtain all information to compute potential DI benefits and other relevant individual characteristics as of this date for each year an individual is not receiving DI benefits. Then, we estimate the effects separately for the age groups 30–56 and 57–60.[21]

4. THE EFFECT OF TIGHTER DI ELIGIBILITY RULES

4.1. *Estimation Strategy*

We exploit the policy-induced variation in the RSA across birth cohorts in a difference-in-differences (DID) design. The RSA is 57 for men who turn 57 before December 2012 (born before December 1955). We label this cohort the RSA-57 cohort.[22] Conversely, the RSA is 58 for men who turn 57 between December 2012 and November 2013 (born after November 1955 and before December 1956). We label this cohort the RSA-58 cohort. Finally, the RSA is 59 for men who turn 57 after November 2013 (born after November 1956). We label this cohort the RSA-59 cohort.[23]

The RSA increases imply that the strictness of DI eligibility rules varies at certain ages across birth cohorts. The RSA-57 birth cohort, our control cohort, faces lenient DI eligibility rules already at age 57. In contrast, the RSA-58 and RSA-59 birth cohorts, our

[19] Staubli and Zweimüller (2013) showed that this increase had sizeable employment and unemployment effects. Men in our sample were not affected by this increase; their eligibility age for an old-age pension was always age 62.

[20] Note that only individuals who worked in a *similar occupation* for 10 of the last 15 years are eligible for relaxed DI eligibility. In contrast, our definition is based on whether somebody has worked in *any occupation* for 10 years in the last 15 years because we can only observe industry affiliation and not occupation. Consequently, our sample will include some individuals who are not eligible for relaxed screening, but this number is likely small because what constitutes a similar occupation is broadly defined.

[21] Tables T.7 and T.9 in the Supplementary Material show summary statistics for the RSA and benefit generosity samples.

[22] Applications are assessed using the rules in the month after filing. Therefore, if someone turns 57 in December 2012 and applies to DI, his application is evaluated in January 2013, when the new RSA of 58 applies.

[23] Figure U.14 in the Supplementary Material provides descriptive evidence on the labor market effects of the RSA increases.

treated cohorts, face lenient eligibility rules only at age 58 and age 59, respectively. Thus, we can identify the effect of stricter DI eligibility rules by comparing the age profiles of treated and control birth cohorts. This comparison is implemented by estimating regressions of the following type:

$$
\begin{aligned}
y_{\mathrm{ict}} = \alpha + \sum_{k=54\setminus 56}^{61} \beta_k \cdot T_{ic} \cdot I[\mathrm{age}_i = k] \\
+ \sum_{k=54\setminus 56}^{61} \gamma_k \cdot I[\mathrm{age}_i = k] + \pi_c + \lambda_t + X'_{\mathrm{ict}}\delta + \varepsilon_{\mathrm{ict}},
\end{aligned} \tag{17}
$$

where i denotes individual, c denotes the year and month of birth, and t denotes the year and quarter of calendar time; y_{ict} is the outcome variable of interest (such as an indicator for receiving DI benefits). We include age-in-years fixed effects to control for age-specific levels in the outcome variable ($I[\mathrm{age}_i = k]$), year-month of birth fixed effects to capture time-constant differences across birth cohorts (π_c), year fixed effects to capture common time shocks (λ_t), and individual or region-specific characteristics to control for any observable differences that might confound the analysis (X_{ict}).[24] We cluster standard errors within birth year, birth month, and state of residence.

The key variables of interest are the indicators $T_{ic} \cdot I[\mathrm{age} = k]$, which are equal to 1 if an individual belongs to a treated cohort ($T_{ic} = 1$) and the age is equal to k, where k runs from 54 to 61 using $k = 56$ as the reference age. Each β_k-coefficient measures the average causal effect of an RSA increase at age k. To obtain the average impact of an RSA increase over a wider age interval, we can take the average of different β_k-coefficients. For example, $\sum_{k=57}^{61} \beta_k/5$ measures the average change in the outcome variable at each age in the age interval 57 to 61.

Since the RSA changes discontinuously by date of birth, we could evaluate its impact using a regression discontinuity design. Instead, we opt for a DID design because it allows us to assess the robustness of the welfare conclusions to a one-year (from 57 to 58) versus a two-year RSA increase (from 57 to 59). Specifically, we estimate the effects of the RSA-58 and RSA-59 change separately, always using the RSA-57 cohort as the control cohort. Moreover, the DID design has more statistical power, which helps us identify spousal labor supply responses and mechanical fiscal effects where sample sizes are smaller.

The identification assumption is that, absent the increase in the RSA, the change in y_{ict} across age would have been comparable between treated and control birth cohorts. A potential concern is that age-specific trends in the outcome variable could change across birth cohorts for reasons unrelated to the RSA increase. The estimated β_k-coefficients for $k < 57$ provide placebo checks for spurious trends. They should not be statistically significant if the identification assumption holds, although they could also pick up anticipation effects. As an additional placebo check, we estimate equation (17) for men with less than 10 employment years in the past 15 years. They are not eligible for the lenient eligibility rules and should not respond to the changes in the RSA. This check provides further robustness that the estimated effects are caused by the RSA increases and not any other cohort-specific policy changes that affect all men.

[24]The fixed effects are identified because we have quarterly reference dates, so that we have variation in the age-in-years (θ_a) conditional on the year-month of birth (π_c) and calendar year (λ_t). For example, men born in July 1955 are 56 in March and June 2012 and 57 in September and December 2012.

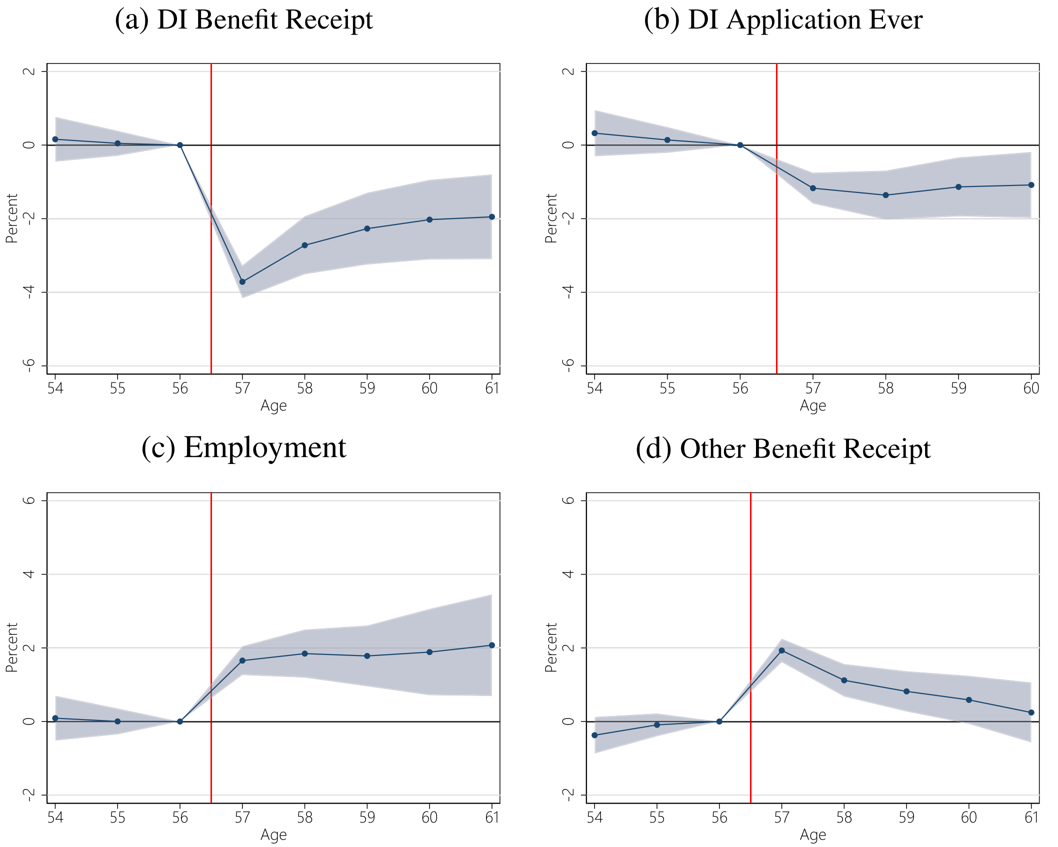

FIGURE 2.—Effects of on labor market states and DI application by age. Notes: The figure shows the estimated β_k-coefficients from the econometric specification in (17) for the -58 increases using the sample of eligible men. The shaded area denotes the 95 percent confidence interval.

4.2. *Empirical Results*

For brevity, we focus on the one-year increase to age 58 (see Online Appendix B.2 for the -59 results). Figure 2 shows the estimated β_k-coefficients from equation (17) for four key outcomes: a dummy for whether an individual is receiving DI benefits (DI benefit receipt), a dummy for whether an individual has ever applied for DI (DI application ever), a dummy for whether an individual is employed (employment), and a dummy for whether an individual is receiving UI or SI benefits (other benefit receipt). The shaded area denotes the 95 percent confidence interval.

The estimates before the pre-reform of 57 are close to zero and statistically insignificant, suggesting that differential trends do not confound the estimates. Panel (a) shows that DI benefit receipt drops by about 4 percentage points at age 57 and remains permanently lower, even though eligibility rules become more lenient at age 58. If applying for DI is costly, we expect fewer people to apply when eligibility criteria are strict. Indeed, Panel (b) shows that DI application rates drop at all ages above 56. Panels (c) and (d) show that stricter DI eligibility rules increase employment and other benefit receipt above age 56. The expansion in employment persists until age 61, while the rise in other benefit receipt is temporary.

TABLE I

AVERAGE EFFECT OF STRICTER DI ELIGIBILITY RULES, 58.

Labor market effects (%-points)			Fiscal effects (Euro)		
	Estimate	Mean		Estimate	Mean
DI benefit receipt	−2.54	18.56	DI benefits	−885	6756
	(0.44)		(A)	(161)	
DI application ever	−1.17	21.81	Payroll taxes	263	11,185
	(0.36)		(B)	(56)	
Employment	1.85	68.36	Other benefits	172	1217
	(0.39)		(C)	(46)	
Other benefit receipt	0.94	7.55	Total fiscal effect	−976	−3213
	(0.25)		(A − B + C)	(185)	
No. Observations	2,444,975			2,444,975	

Note: The table reports the average effect of the for the ages above age 56. The estimates are constructed by taking the average of the β_k-coefficients from equation (17) for $k \geq 57$. *Mean* denotes the mean above the for the −57 cohort. Fiscal effects are reported in 2018 Euros. Standard errors are reported in parentheses and are clustered within birth year, birth month, and state of residence.

The lasting impacts of a one-year increase on DI applications, DI receipt, and employment can occur for three reasons. One reason is higher application costs, but the reforms did not change the application or medical screening process. A second, more likely, reason is that a one-year increase lowers the lifetime value of DI benefits by a year because DI benefits are automatically replaced by an old-age pension at age 62. The lower lifetime value induces some marginal applicants to stop applying for DI permanently. A third reason is that health shocks could be temporary for some applicants. They would not apply at age 57 under the strict rules, but they would also not apply after age 57 because their health has improved.

One way to better understand the drivers of the persistent effects is to look at transitions from one labor market state to another. Specifically, we can estimate our main DID-equation using as outcome variables transitions from employment and other benefit receipt to DI benefit receipt, employment, and other benefit receipt. Online Appendix Figures B.2 and B.3 plot the corresponding β_k-coefficients for each increase. They show that the persistent impacts on DI receipt and employment are entirely driven by people who are already employed. Moreover, they also drive the permanent drop in DI applications (Online Appendix Figure B.4).

We can estimate the average effect of tighter DI eligibility rules between ages 57 and 61 by taking the average of the β_k-coefficients over these ages. Table I reports the average effects for the outcomes from Figure 2 and the corresponding fiscal impacts. We focus on four fiscal outcomes: DI benefits, payroll taxes, other benefits, and the total fiscal effect defined as the sum of DI benefits plus other benefits minus payroll taxes. We calculate the outcomes individually, multiplying at each age the number of days an individual spends in a given labor market state times the daily benefit received or taxes paid in that state.[25] Table I shows that tighter DI eligibility rules lessen spending on DI benefits (885 Euros per individual and year) and raise tax revenues from increased work activity (263 Euros) but also raise spending on other benefits because of benefit substitution (172 Euros). Overall, increasing the to 58 lowers fiscal costs at each age above 56 by 976 Euros per individual. The results for the −59 cohort are qualitatively similar but are about twice as large as for the −58 cohort (Online Appendix Figure B.5 and Online Appendix Table B.I).

[25]The figures in Supplementary Material U.2 show the estimated β_k-coefficients for the fiscal outcomes.

As a placebo test, Supplementary Material U.3 shows the estimates for men with too little work experience to be eligible for the lenient DI eligibility rules. For these placebo groups, we find that DI benefit receipt, DI application ever, employment, and other benefit receipt do not differ significantly across birth cohorts, even after age 56. The null effect among placebo groups strongly suggests that the estimated effects for the main sample are caused by the increases and not any other cohort-specific policy changes affecting all men.

5. IMPACT OF BENEFIT GENEROSITY

The ideal experiment to analyze the impact of a change in DI benefits would be to randomize the level of DI benefits across individuals. We emulate this ideal experiment with a quasi-experimental research design that exploits variation in DI benefits from a large pension reform. Our approach follows Mullen and Staubli (2016), who estimated the elasticity of DI claiming with respect to benefit generosity using variation in DI benefits in Austria from several reforms between 1987 and 2010. We differ from their study in two aspects. First, we update their estimates for a more recent period (2004 to 2017). This period is characterized by lower replacement rates and stricter disability screening compared to the 1980s and 1990s, which could affect the responsiveness of DI claiming and applications to benefit levels. Second, we study the effect on a novel set of outcomes, including employment, other benefit receipt, and fiscal costs. The budgetary effects of a change in benefit generosity are vital for assessing the welfare impact.

5.1. *Estimation Strategy*

Our estimation strategy exploits the variation in DI benefit levels stemming from the 2003 pension reform described in Section 3. We are interested in estimating the following regression:

$$y_{it} = \alpha + X_{it}'\beta + \gamma b_{it}(Z_{it}) + \lambda_t + \varepsilon_{it}, \tag{18}$$

where i denotes individual, t denotes year, y_{it} is the outcome variable of interest such as applying for DI, X_{it} is a vector of demographic and labor market characteristics, $b_t(Z_{it})$ are log potential DI benefits which are a function of labor market characteristics $Z_{it} \in X_{it}$ (age, insurance years, and the assessment basis), λ_t are year fixed effects, and ε_{it} are any unobserved factors affecting the outcome such as taste for work. The parameter of interest is γ, which measures the average effect of a change in benefit levels on the outcome variable.

Regression (18) has an endogeneity problem as labor market characteristics Z_{it} affect benefits and outcomes simultaneously. We can solve the problem by exploiting the 2003 reform because it creates variation in b independent from Z_{it}. Mullen and Staubli (2016) showed that the policy-induced variation in b can be isolated by including the individual-specific (log) hypothetical benefits under each policy regime as additional controls in equation (18). Because of the phased-in nature of the 2003 policy reform, we have 14 different hypothetical benefits for each year from 2004 to 2017:

$$y_{it} = \alpha + X_{it}'\beta + \gamma b_{it}(Z_{it}) + \sum_{r=2005}^{2017} \delta_r b_r(Z_{it}) + \lambda_t + \varepsilon_{it}, \tag{19}$$

where $b_r(Z_{it})$ denotes hypothetical DI benefits under the policy regime r. Controlling for hypothetical DI benefits ensures that actual potential benefits are uncorrelated with any unobservable factors affecting the outcome variable so that γ identifies the causal effect of DI benefits.[26] We cluster standard errors within birth year and birth month.

The identification assumption necessary for the consistency of our estimates is the standard common trends assumption, requiring that absent the 2003 reform, the outcome variable would have evolved similarly across groups with differential changes in benefit levels. To test the appropriateness of our identification strategy, we estimate 1000 placebo regressions in which we randomly assign individuals within each cell defined by year, insurance-year decile, and assessment decile potential benefits $b_r(Z_{it})$ from a different year. If our empirical strategy isolates the policy-induced variation in DI benefits, the placebo estimates should be clustered around zero. Figure V.20 in the Supplementary Material presents the results, showing that placebo increases in benefit generosity lead to estimates close to zero.

5.2. *Empirical Results*

Table II summarizes our main results providing estimates of equation (19) for labor market- and fiscal outcomes. As in Mullen and Staubli (2016), we define individuals' labor market outcomes as flows within a calendar year: a dummy for applying for DI (DI application ever), a dummy for entering DI (DI inflow), a dummy for exiting employment (employment outflow), and a dummy for leaving UI or SI benefits (other benefit outflow). Fiscal outcomes are calculated individually and correspond to the change in annual benefits received or taxes paid from entering or exiting a given labor market state.

We find that a one percent increase in DI benefits increases the propensity to apply for DI benefits by 0.17 percentage points for the age group 57–60, equivalent to a 0.64 percent increase in the baseline application rate. DI inflow increases by 0.093 percentage points. These estimates imply an award rate of 54 percent ($= 0.093/0.171$) for the marginal applicant in the age group 57–60. An increase in benefit levels does not affect employment

TABLE II

AVERAGE EFFECT OF BENEFIT GENEROSITY FOR 57–60-YEAR-OLD INDIVIDUALS.

Labor market effects (%-points)			Fiscal effects (Euro)		
	Estimate	Mean		Estimate	Mean
DI application ever	0.170	26.71	DI benefits	36.95	4516
	(0.019)		(A)	(3.16)	
DI inflow	0.093	18.68	Payroll taxes	−2.37	9915
	(0.015)		(B)	(1.12)	
Employment outflow	−0.004	71.43	Other benefits	−20.62	1944
	(0.011)		(C)	(2.33)	
Other benefit outflow	0.097	9.89	Behavioral fiscal effect	18.69	−3455
	(0.012)		(D = A-B+C)	(3.14)	
Observations	1,453,448			1,453,448	

Note: The table reports estimates for γ from the econometric specification in (19). Fiscal effects are reported in annual 2018 Euros. *Mean* denotes the mean in levels for the year 2004. Standard errors are reported in parentheses and are clustered within birth year and birth month.

[26]Figure V.18 in the Supplementary Material assesses the quality of our prediction of hypothetical DI benefits under different policy regimes. Actual benefits track our predicted benefits very closely.

but significantly increases outflow from other benefits, suggesting that marginal enrollees received other benefits before being awarded DI benefits.[27]

Concerning the fiscal effects, we observe that a one percent increase in DI benefits expands annual spending on DI benefits by 36.95 Euros for a 57–60-year-old individual, lowers tax revenue by 2.37 Euros, and lessens spending on other benefits by 20.62 Euros because of benefit substitution. Overall, the behavioral responses to a one percent increase in DI benefits raise annual fiscal spending by 18.69 Euros per 57–60-year-old individual.[28] We report the estimates for the age group 30–56 in Online Appendix Table C.III. The effects in this younger age group are much smaller.

6. ESTIMATING THE FISCAL MULTIPLIER OF DI REFORMS

The relative welfare effect of stricter eligibility rules versus lower benefits depends on the relative insurance values and the relative fiscal multipliers. This section discusses how we estimate fiscal multipliers for each policy instrument and presents the estimates.

Always-versus Marginal Applicants: A Complier Analysis

We start by examining characteristics of marginal applicants, who apply under lenient eligibility rules only, and always applicants, who apply under strict and lenient eligibility rules. We expect the two groups to respond differently to changes in stricter eligibility rules, and they also affect the fiscal multiplier differently: The mechanical fiscal effect is driven by always applicants, while the behavioral effect is driven by marginal applicants.

We cannot directly identify always and marginal applicants, but we can describe their observable characteristics using a complier analysis for DID settings (Imbens and Rubin (1997), Abadie (2003), De Chaisemartin and D'Haultfoeuille (2018)).[29] Online Appendix D.1 discusses the details of the analysis. We estimate a share of always applicants of $\pi^{AA} = 0.070$. The shares of marginal and never applicants are $\pi^{MA} = 0.014$ and $\pi^{NA} = 0.916$. Always applicants also differ systematically from marginal applicants in observable characteristics: They are more likely to have already applied for DI before age 57, more likely to be on sick leave at age 56 (a good proxy for health), less likely to be employed at age 56, more likely to work in blue-collar occupations, and more likely to suffer from musculoskeletal disorders. These patterns suggest that always applicants are less healthy and have lower employment potential than marginal applicants, consistent with existing evidence for other countries (e.g., Maestas, Mullen, and Strand (2013)).

[27]Note that this pattern is inconsistent with the simple model from Section 2. The simple model assumes that $\theta^A < \theta^R$, that is, the marginal applicants are at the margin between working and receiving DI benefits. Based on the simple model, we would thus expect the inflow effect to come from people who stop working. However, the inflow effect being driven by benefit substitution is consistent with the more general model in Online Appendix A.2. Heterogeneity in wages, benefits, and disability severity θ can lead to situations where $\theta_i^A > \theta_i^R$ for some individuals i.

[28]The fiscal estimates for a benefit change represent behavioral fiscal effects because they only encompass the behavioral responses of marginal applicants who no longer apply for DI when benefits decline. To obtain the total budgetary impact, we need to add the fiscal cost savings from paying lower DI benefits to current beneficiaries (mechanical fiscal effect). In contrast, the fiscal estimates for a change in eligibility rules represent total fiscal effects. The reason is that they capture both the behavioral fiscal impact from marginal applicants (who no longer apply under the stricter rules) and the mechanical fiscal impact from always applicants (who are screened out of DI under the strict rules).

[29]Individuals in the −58 cohort who apply at age 57 are always applicants because they face strict eligibility rules and apply. However, individuals in the −57 cohort who apply at 57 under the lenient rules are either always or marginal applicants.

Fiscal Multiplier Estimates for Stricter DI Eligibility

We calculate the fiscal multiplier by decomposing the total fiscal effect of stricter eligibility rules (Table I) into a mechanical and a behavioral component. To estimate the mechanical fiscal effect, we use the insights from the complier analysis to identify a subgroup that is observationally similar to always applicants. Our main subgroup consists of individuals who already applied for DI between 50 and 57 and were rejected, but we also explore robustness using other subgroups. We label this subpopulation *pre-57 applicants*, and the complier analysis shows that they predominantly consist of always applicants. However, they also tend to be healthier than always applicants; they have been rejected before, and healthier applicants are more likely to be rejected. As a result, they experience a larger drop in award probabilities from the increases than always applicants, which leads us to overstate the mechanical effect and understate the fiscal multiplier. Thus, this strategy is conservative and provides a lower bound for the fiscal multiplier.[30]

We then perform the same DID analysis for pre-57 applicants as for the main sample. Figure 3 plots the age profile of the DID estimates. Panel (a) shows that DI benefit receipt drops significantly at age 57 and then steadily catches up to the level of the control group by age 60. Always applicants experience a slight employment increase at age 57 that vanishes afterward (Panel b), while the rise in other benefit receipt is sizeable and persistent (Panel c). The temporary nature of the estimated effects implies that the persistent increases in employment and disability receipt in the main sample (Figure 2) are driven by the behavioral effects of marginal applicants and not by the mechanical effects of always applicants.[31]

We use the DID estimates from Figure 3 to estimate the mechanical fiscal effect and calculate the fiscal multiplier of stricter eligibility rules. The first two columns of Table III present the results. The mechanical fiscal effect for 58 is equal to $\mathbb{E}[M(\theta_R^*)] = 5585\text{Euro} * 0.070 = 391\text{Euro}$, where $\pi^{AA} = 0.070$ is the share of always applicants from the complier analysis and 5585Euro is the total fiscal effect among always applicants ($\mathbb{E}[\Delta G(\theta_R^*)|\text{pre-57 applicant}]$).[32] The behavioral fiscal effect is the difference between the total fiscal and the mechanical fiscal effect, $\mathbb{E}[B(\theta_R^*)] = 976 - 391 = 585\text{Euro}$. This decomposition implies a fiscal multiplier of 2.50 for 58 and a fiscal multiplier of 2.03 for 59. We test whether the fiscal multipliers are statistically different from 1, which would be the value of the multiplier if stricter eligibility rules had only mechanical and no behavioral effects. We can reject the null hypothesis of no behavioral responses (p-values: 0.046 for -58 and 0.001 for -59). Online Appendix Table D.VI shows that the mechanical fiscal effect and the fiscal multiplier are robust to alternative specifications and definitions of

[30]Online Appendix D.2 provides further evidence that pre-57 applicants are representative of always applicants. First, we show that pre-57 applicants continue to apply for DI benefits when eligibility is strict, as always applicants would do. Second, we focus on the -58 cohort and show that trends in DI benefit receipt and the total fiscal effect after applying for DI are similar for pre-57 applicants who reapply at age 57 and all always applicants (all individuals in the -58 who apply at age 57).

[31]Supplementary Material Figure W.21 shows that, for the -59 cohort, the mechanical effect persists for two years and then starts to disappear, as one would expect.

[32]$\mathbb{E}[\Delta G|\text{pre-57}]$ is the total fiscal effect among pre-57 applicants who reapply at age 57. It is calculated by first estimating the total fiscal effect among all pre-57 applicants: $\mathbb{E}[\Delta G|\text{pre-57}] * \Pr(\text{reapply at } 57) = 1167\text{Euro}$. We then divided by the probability to reapply at age 57, $P(\text{reapply at } 57) = 0.209$, to obtain $\mathbb{E}[\Delta G|\text{pre-57}] = 5585\text{Euro}$. If our identifying assumption holds, we have $\mathbb{E}[\Delta G(\theta_R^*)|\text{pre-57}] = \mathbb{E}[M(\theta_R^*)|\text{pre-57}] = \mathbb{E}[M(\theta_R^*)]/\pi^{AA}$. The first equality says that the total fiscal effect of pre-57 applicants is purely mechanical. The second equality says that pre-57 applicants are representative for always applicants in the whole population.

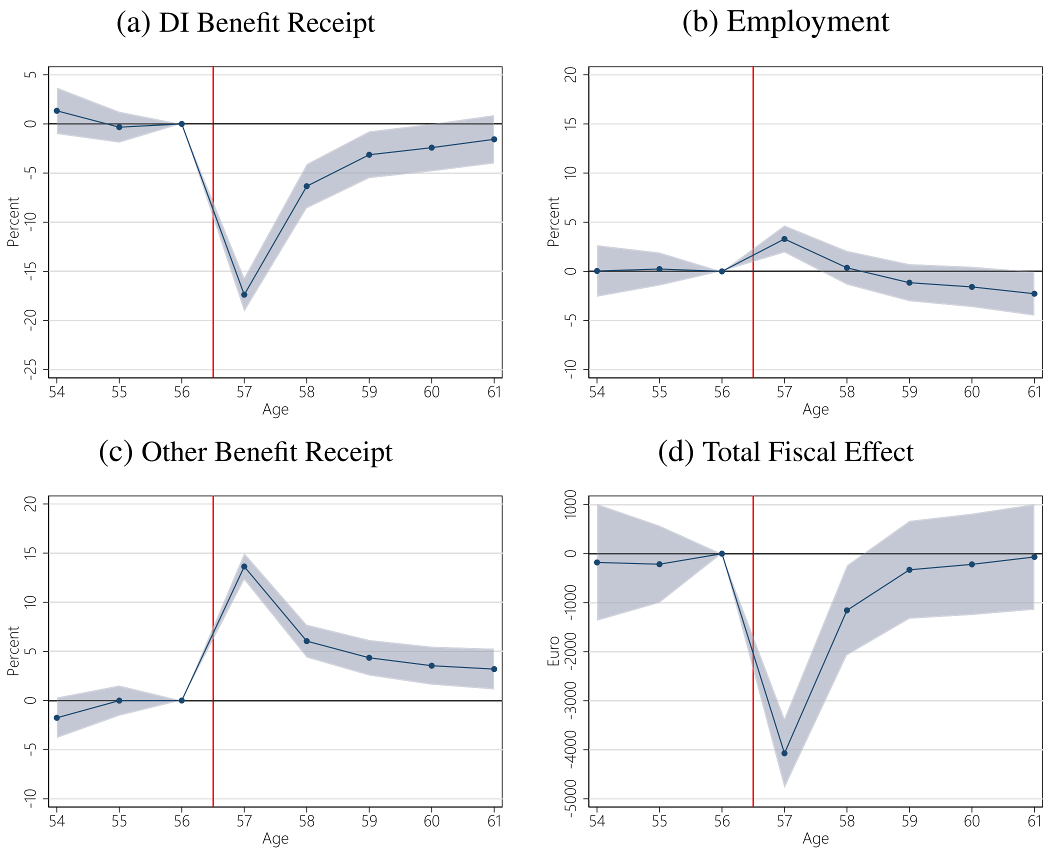

FIGURE 3.—Mechanical effects of −58 increase by age. Notes: The figure shows the estimated β_k-coefficients from the econometric specification in (17) for the −58 increase using the sample of always applicants. Always applicants comprise individuals who have applied for DI between age 54 and age 56. The shaded area denotes the 95 percent confidence interval.

always applicants. If at all, our main specification overstates the mechanical effect and understates the fiscal multiplier.

Fiscal Multiplier Estimates for Lower DI Benefits

The last two columns of Table III show the fiscal effects and fiscal multiplier of lower benefits for 57–60- and 30–56-year-olds. We have already estimated the behavioral labor and fiscal responses to a change in DI benefits in Section 5. This analysis reveals that a 1% cut in DI benefits induces a behavioral fiscal effect $\mathbb{E}[B(b_s)]$ of 18.69 Euros per year for 57–60-year-olds (and of 1.18 Euros per year for 30–56-year-olds).

To determine the fiscal multiplier, we additionally need to estimate the mechanical fiscal effect $\mathbb{E}[M(b_s)]$. It is equal to one percent of the pre-reform mean annual DI benefits. This mean is 4516 Euros for 57–60-year-olds (Table II), resulting in a yearly mechanical fiscal effect of 45.16 Euros (the same calculation for 30–56-year-olds yields a mechanical fiscal effect of 3.24 Euros). The total fiscal effect, the sum of behavioral and mechanical fiscal effects, is 63.85 Euros for 57–60-year-olds (and 4.42 Euros for 30–56-year-olds). To get the fiscal multiplier, we divide the total fiscal effect by the mechanical fiscal effect. This

TABLE III

FISCAL MULTIPLIER FOR ELIGIBILITY RULES AND BENEFIT GENEROSITY.

	Eligibility rules		Benefit generosity	
	58	59	Ages 57–60	Ages 30–56
Total fiscal effect	976	1770	63.85	4.42
	(83)	(111)	(2.65)	(0.18)
Mechanical fiscal effect (M)	391	871	45.16	3.24
	(100)	(140)	(0.33)	(0.02)
Behavioral fiscal effect (B)	585	899	18.69	1.18
	(94)	(144)	(2.63)	(0.18)
Fiscal multiplier (1+B/M)	2.50	2.03	1.41	1.36
	(0.75)	(0.32)	(0.06)	(0.05)
p-value: multiplier = 1	0.046	0.001	< 0.001	< 0.001

Note: Table presents estimates of the fiscal multiplier for stricter eligibility rules and more generous DI benefits. The fiscal multiplier of stricter eligibility is constructed as follows. The total fiscal effect is taken from Tables I and B.I. The mechanical fiscal effect is estimated using the sample of pre-57 applicants and then re-scaled by the population share of always applicants (see text for details). The behavioral fiscal effect is the total fiscal effect minus the mechanical fiscal effect. The fiscal multiplier of benefit generosity is constructed as follows. The behavioral fiscal effect is taken from Tables II and C.III. The mechanical fiscal effect captures a 1% increase in DI benefits for all DI beneficiaries in 2004. It is obtained by multiplying the mean DI benefits in Tables II and C.III with 0.01. The total fiscal effect is the sum of the mechanical and behavioral fiscal effects. Bootstrapped standard errors are reported in parentheses.

division yields fiscal multipliers of lower DI benefits of 1.41 for 57–60-year-olds and 1.36 for 30–56-year-olds. The fiscal multipliers for benefit generosity are precisely estimated, and we can reject the null hypothesis that they are equal to 1. We also test whether the fiscal multipliers for benefit generosity and stricter eligibility are statistically different. The probability that the stricter eligibility multiplier is not larger than the benefit generosity multiplier is 8% (3%) for −58 (−59). The statistical power is a bit limited because the pre-57 applicant sample is smaller and the mechanical effect for stricter eligibility thus less precisely estimated.

Family Insurance and Fiscal Multipliers

A growing literature studies the role of the family in sharing risks and smoothing consumption to adverse income shocks (see, e.g., Blundell, Pistaferri, and Saporta-Eksten (2016)). In our context, the DI reforms may affect spouses' labor supply to compensate for husbands' income losses. Such added worker effects could alter our welfare conclusions because they affect the size of the behavioral fiscal effect and, consequently, the fiscal multiplier.

To estimate spousal responses, we construct a sample of spouses who we can link to the men in the main sample (see Supplementary Material Tables T.8 and T.9 for summary statistics). We then estimate our main regression equations using the same outcome variables but this time for the spouse (e.g., spousal employment). As Online Appendix Table B.II and Figure B.6 show, stricter eligibility rules have no spousal labor market or fiscal impacts. The point estimates are quantitatively small and statistically insignificant.[33]

[33]Moreover, Online Appendix Figure B.7 shows that increases have the same labor market and financial impacts for all men, married men (men for whom we can match a spouse), and the household (married men and their spouses), implying that the fiscal effects of all men are representative for married men. This finding is important because we use spousal labor supply responses to estimate the insurance value of stricter disability eligibility criteria relative to lower benefits. This strategy identifies the insurance losses of married men, as discussed in the next section.

We also find no spousal responses to changes in benefit generosity (Online Appendix Table C.IV). Therefore, accounting for potential added worker effects leaves the fiscal multiplier almost unchanged. The multipliers for benefit generosity, −58, and −59 are 1.41, 2.5, and 2.03 without and 1.46, 2.7, and 2.09 with spousal responses.[34]

7. ESTIMATING THE RELATIVE INSURANCE VALUES

In this section, we document that stricter eligibility rules are associated with higher spousal earnings of DI entrants. Following Fadlon and Heien Nielsen (2019), we then show that we can identify the relative insurance values of stricter eligibility versus lower benefits (equations (15) and (16)) from the spousal labor supply patterns of DI recipients.

Panel (a) of Figure 4 plots average spousal earnings by their husband's quarter of birth for husbands who enter DI at age 57. We see a jump in spousal earnings at the reform cutoff. Spouses whose husband enters DI under the stricter rules (right of the cutoff) have on average 2500 Euros higher annual earnings when their husbands are 57–61 years old. Panel (b) provides an event-study of spousal earnings around the husband's DI entry separately for the control and treatment cohorts of the −58 increase. The event-study shows that spouses increase earnings when their husbands enter DI. Spouses whose husbands enter DI under the stricter eligibility rules (treatment cohorts) increase their earnings

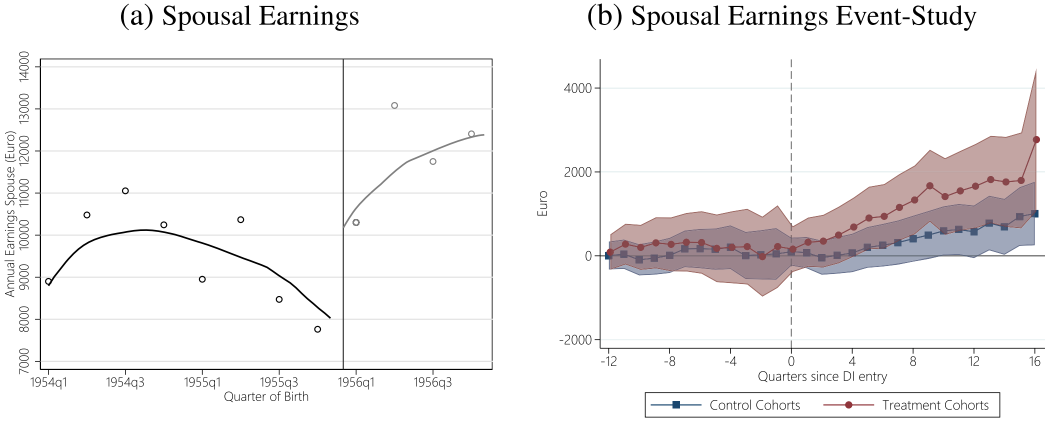

FIGURE 4.—Spousal labor supply patterns for −58 change. Notes: Panel (a) plots average spousal earnings by their husband's quarter of birth for husbands who enter DI at age 57 (the lines are local linear regressions with a bandwidth of 12 months). The jump in spousal earnings at the reform cutoff indicates that spouses whose husbands enter DI under the stricter rules (right of the cutoff) have on average around 2500 Euros higher earnings during husbands' age 57–61. Panel (b) provides an event-study of spousal earnings around the husband's DI entry separately for the control and treatment cohorts of the −58 increase. The event-study shows that spouses increase their earnings when their husbands enter DI. Spouses whose husbands enter DI under the stricter eligibility rules (treatment cohorts) increase their earnings more compared to spouses of husbands who enter DI under the lenient rules (control cohorts). While the event-study estimates are not statistically significant at each quarter, they are jointly significantly different from each other at the 10% percent level.

[34]The −58 fiscal multiplier with spousal responses is $1 + (976 + 80 - 391)/391$, where 976 is the total fiscal effect for men, 80 is the total fiscal effect for spouses, and 391 is the mechanical fiscal effect. The −59 fiscal multiplier with spousal responses is calculated in the same way. The benefit generosity fiscal multiplier with spousal responses is calculated as $1 + (18.69 + 2.07)/45.16$, where 18.69 is the behavioral fiscal effect of men, 2.07 is the behavioral fiscal effect of spouses, and 45.16 is the mechanical fiscal effect.

more than spouses of husbands who enter DI under the lenient rules (control cohorts).[35] Spouses' labor supply reacts mostly at the intensive margin and less at the extensive margin (Online Appendix Figure E.10, Panel (a)). Moreover, there is no detectable difference in spousal take-up of social insurance benefits at the cutoff (Online Appendix Figure E.10, Panels (b) and (c)).

We next relate these labor supply patterns to the relative insurance values. We follow the idea of Fadlon and Heien Nielsen (2019), who showed that the directly-affected spouse's marginal utility of consumption links to the indirectly-affected spouse's labor supply in a collective household model. Specifically, risk-sharing couples will choose consumption such that their marginal utility of consumption is the same, $v'(c^b) = u_2'(c_2^b)$, where $v'(c^b)$ and $u_2'(c_2^b)$ are the marginal utility of consumption of men on DI and their spouses. Moreover, spouses choose labor supply such that their marginal utility of consumption equals their marginal disutility of work, $u_2'(c_2^b) = \frac{1}{w_2(1-\tau_2)}\varphi_2'(l_2^b)$, where $\varphi_2'(l_2^b)$ is the marginal disutility of work and $w_2(1-\tau_2)$ is the net wage. Using the first-order condition $v'(c^b) = \frac{1}{w_2(1-\tau_2)}\varphi_2'(l_2^b)$ and a first-order Taylor approximation, we can express the upper bound of the relative insurance values from equation (15) as a function of observable spousal labor supply moments and the preference parameter ρ:[36]

$$\frac{V_{\theta^*}}{V_b} \leq \underbrace{\int_{\theta^A}^{\infty} \frac{\Delta DI \cdot (b-z)}{M(\theta^*)}\, dF(\theta)}_{\text{IP1}} + \rho \cdot \underbrace{\int_{\theta^A}^{\infty} \left[\frac{l_2^b - E[l_2^b|\text{on DI}]}{E[l_2^b|\text{on DI}]}\right] \frac{\Delta DI \cdot (b-z)}{M(\theta^*)}\, dF(\theta)}_{\text{IP2}}. \tag{20}$$

The bound depends on the two terms IP1 and IP2 and the preference parameter $\rho \equiv \frac{\varphi_2''(E[l_2^b|\text{on DI}])}{\varphi_2'(E[l_2^b|\text{on DI}])} E[l_2^b|\text{on DI}] \geq 0$. ρ measures the curvature of spousal disutility of labor $\varphi_2(\cdot)$. IP1 is the ratio of the hypothetical fiscal effect of stricter eligibility, $\Delta DI \cdot (b-z)$, if every marginal entrant (those with $\Delta DI = 1$) substituted to other benefits and the mechanical fiscal effect, $M(\theta^*)$. IP2 depends on the marginal entrants' spousal labor supply (l_2^b for individuals with $\Delta DI = 1$) relative to the spousal labor supply of all (pre-reform) DI recipients ($E[l_2^b|\text{on DI}]$). The upward jump in Panel (a) of Figure 4 indicates that marginal entrants' spousal labor supply is lower than that of always entrants. Hence, IP2 must be negative.[37] Interpreted through the lens of Fadlon and Heien Nielsen (2019)'s approach,

[35]At first sight, this finding seems contradictory to Section 6, where we show that spousal responses do not confound our fiscal effect estimates, and the multipliers remain almost unchanged when accounting for spousal responses. The key difference to Section 6 is that Figure 4 conditions on DI receipt. At the aggregate level, those responses are small. The change in DI receipt of stricter eligibility is −2.54 pp (Table I), and the RD estimate of spousal labor supply differences from Figure 4(a) is 2570 Euros. Together with a payroll tax rate of 0.28, this implies an aggregate fiscal effect of spousal labor supply of about 18 Euros ($0.0254 * 2570 * 0.28 = 18$ Euros), which is quantitatively not important for the fiscal effect estimates.

[36]Supplementary Material S.2 and S.4 present the full household model and all derivations. The model focuses on intensive-margin labor supply responses because Panel (a) of Figure 4 and Panel (a) of Online Appendix Figure E.10 show that spouses primarily react at the intensive margin and less at the extensive margin. We also refrain from modeling spouses' social insurance benefits as we see no response at this margin (Panels (b) and (c) of Online Appendix Figure E.10).

[37]Unless the weighting by $\frac{b-z}{M(\theta^*)}$ would significantly change the average.

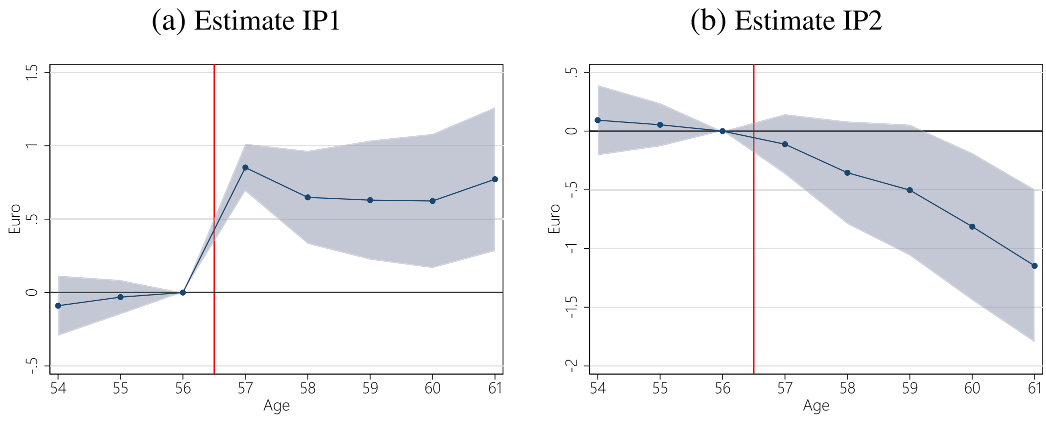

FIGURE 5.—Insurance value bounds estimation for −58 change. Notes: The figure shows the estimated β_k-coefficients from the econometric specification in (17) for IP1 and IP2 defined in equation (20). The shaded area denotes the 95 percent confidence interval.

the lower spousal labor supply of marginal entrants means that they value DI benefit receipt less than the average DI recipient. Consequently, the insurance losses of stricter eligibility are smaller than the loss of lower benefits, that is, $\frac{V_{\theta*}}{V_b} \leq 1$, for any value of the preference parameter ρ if IP1 ≤ 1.

Our difference-in-differences approach allows us to directly estimate IP1. To estimate IP1, we run equation (17) using as outcome variable $Y_i^{IP1} = \frac{1}{M(\theta^*)} DI_i(b_i - z_i)$, where $M(\theta^*)$ is the mechanical fiscal effect of stricter eligibility rules (Table III), DI_i is an indicator for whether individual i receives DI benefits, and $b_i - z_i$ is the difference between potential DI benefits b_i and other benefits z_i. Figure 5, Panel (a) presents the IP1-estimates for each age in the interval 54 to 61. At all ages, IP1 is below 1, and Table IV shows that for the −58 increase, the average IP1 over the age interval 56–61 is 0.71.

TABLE IV

RELATIVE INSURANCE VALUES FOR ELIGIBILITY RULES AND BENEFIT GENEROSITY.

	Bound for RSA-58		Bounds for RSA-59	
	Upper	Lower	Upper	Lower
IP1	0.71	0.22	0.72	0.19
	(0.25)	(0.13)	(0.13)	(0.07)
IP2	−0.59	−0.21	−0.37	0.06
	(0.24)	(0.21)	(0.10)	(0.14)
Relative insurance values ($IP1 + 0.6 \cdot IP2$)	0.36	0.09	0.50	0.23
	(0.17)	(0.19)	(0.11)	(0.12)
P-value: relative insurance values = 1	< 0.001	< 0.001	< 0.001	< 0.001
Relative multipliers ($\frac{1+B(\theta^*)/M(\theta^*)}{1+B(b)/M(b)}$)	1.77		1.44	
	(0.54)		(0.24)	
P-value: relative multipliers = 1	0.154		0.067	
Relative multipliers minus insurance values	1.41	1.68	0.94	1.21
	(0.44)	(0.53)	(0.18)	(0.23)

Note: Table presents estimates for the upper and lower bounds of the relative insurance values. We set $\rho = 0.60$. Bootstrapped standard errors are reported in parentheses.

We can also estimate IP2 with our difference-in-differences strategy using as outcome $Y_i^{IP2} = \frac{l_{i2}^b - E[l_2^b|\text{on DI}]}{E[l_2^b|\text{on DI}]}\frac{DI_i(b_i - z_i)}{M(\theta^*)}$, where l_{i2}^b measures annual spousal earnings of DI recipients and $E[l_2^b|\text{on DI}] = 9656$ measures the average annual spousal earnings of DI recipients in the -57 cohort when the husband is 57 to 61 years old. Consistent with labor supply patterns in Panels (a) and (b) of Figure 4, the IP2 estimates in Panel (b) of Figure 5 are negative at all ages above age 57 and become more negative with age. Table IV reports the average IP2 estimates to be -0.59 for the -58 increase. For the -59 change, we find the same pattern. Table IV reports the average estimates for the -59 change to be IP1 = 0.72 and IP2 = -0.37 and Online Appendix Figure E.11 provides the estimates by age.

Therefore, the relative insurance values are bounded by 1 for any value of ρ ($0.71 - 0.59\rho < 1$ for -58 and $0.72 - 0.37\rho < 1$ for RSA-59). Since the ratio of fiscal multipliers exceeds 1 (1.77 for RSA-58 and 1.44 for RSA-59), only the upper bound is decisive, and our conclusion holds for any choice of ρ. Our analysis suggests that stricter eligibility rules top lower DI benefits to curb DI program costs. They generate higher fiscal savings (larger multiplier) and a smaller insurance loss.

For completeness, we also estimate the lower bound for the relative insurance values (equation (16)). Following the same steps as for the upper bound, we can write the lower bound as $\frac{V_{\theta^*}}{V_b} \geq \int_{\theta^R}^{\infty} \frac{\Delta\text{DIB}\cdot(b-z)}{M(\theta^*)}\,dF(\theta) + \rho\int_{\theta^R}^{\infty}[\frac{l_2^b - E[l_2^b|\text{on DI}]}{E[l_2^b|\text{on DI}]}]\frac{\Delta\text{DIB}\cdot(b-z)}{M(\theta^*)}\,dF(\theta)$. The difference is that we assign an insurance value loss from stricter rules only for individuals who do not get into DI and end up on other benefits (those with $\Delta\text{DIB} = 1$). We also need an estimate of ρ to have a meaningful lower bound. Fadlon and Heien Nielsen (2019) showed that $\rho \equiv \frac{\varphi_2''(l_2^b)}{\varphi_2'(l_2^b)}l_2^b = \frac{1+\text{MPE}}{\varepsilon-\text{MPE}}$, where MPE is the marginal propensity to earn out of unearned income and ε is the uncompensated labor supply elasticity. Bargain, Orsini, and Peichl (2014) estimated uncompensated labor supply elasticities for several European countries and found $\varepsilon = 0.34$ for married women in Austria. The MPE's magnitude is debated in the literature with estimates ranging from -0.04 to -1; we choose a middle ground with MPE = -0.5.[38] Using these numbers, we estimate $\rho = \frac{1-0.5}{0.34+0.5} = 0.60$ and a lower bounds of 0.09 for RSA-58 and 0.23 for RSA-59 (columns 2 and 4 of Table IV). The corresponding upper bounds for RSA-58 and RSA-59 using $\rho = 0.60$ are 0.36 and 0.50. Overall, our upper and lower bounds are sufficiently tight to make meaningful welfare statements even if the fiscal multipliers were more similar. Moreover, we find that the difference between the relative fiscal multipliers and the relative insurance values is always positive and statistically significant at the 1% level, implying that from a welfare perspective, stricter eligibility rules dominate lower benefit levels to curb DI program expenditures.

One obvious caveat of our approach here is that it can only identify the relative insurance losses of married individuals. Since most men are married in the relevant age ranges (70% of 55–64-year-old men in Austria are married (Statistik Austria (2018))), our approach should provide a reasonable estimate of the relative insurance losses in the aggregate. However, it could well be that the relative insurance losses are different

[38]Exploiting lotteries, Cesarini, Lindqvist, Notowidigdo, and Östling (2017) reported an average MPE of around -0.15 in Sweden, Picchio, Suetens, and van Ours (2018) estimated an MPE of -0.046 in the Netherlands, Imbens, Rubin, and Sacerdote (2001) found an MPE of -0.11 for the U.S., and Golosov, Graber, Mogstad, and Novgorodsky (2021) found an MPE of -0.5 for U.S. lottery winners. Exploiting changes in survivor insurance generosity, Giupponi (2019) estimated an MPE of -1 in Italy, and Böheim and Topf (2021) estimated an MPE of -0.6 to -1.2 for Austrian men. We interpret these results that the MPE is rather large for our relevant group, women close to retirement, and set MPE = -0.5.

for single individuals.[39] To directly address this issue, one could estimate the insurance value bounds using the consumption-based approaches from the UI literature, for example, Gruber (1997) or Landais and Spinnewijn (2021). This approach, however, requires large-scale consumption data, which do not exist for Austria.

8. CONCLUSION

This paper provides a sufficient statistics framework to quantify the incentive-insurance trade-off associated with the two main DI policy parameters: DI eligibility rules and DI benefits. DI programs are characterized by imperfect screening, eligibility rules, and benefit generosity. Our analysis goes beyond the standard sufficient statistics framework by incorporating eligibility rules and by exploring the trade-off between eligibility and benefit generosity.

We empirically implement our new framework using two restrictive DI reforms in Austria. We find that stricter DI eligibility rules are more effective in reducing the DI system's fiscal burden and create smaller insurance losses than reducing benefit generosity. In the Austrian context, the conclusion is that stricter DI eligibility rules dominate lower DI benefits in scaling back the DI program.

While our empirical application is a local evaluation of the Austrian DI program, our framework is more general. The basic features of DI systems are similar across countries. For example, the U.S. DI program features age cutoffs where vocational factors are taken into account similar to the RSA in Austria. Whether the same conclusions would arise in the U.S. is an empirical question.[40] Carefully implementing our framework in other countries can enhance our understanding of when stricter eligibility criteria are more effective at reducing DI program costs relative to reducing DI benefits. Moreover, the same trade-off between eligibility and generosity that we emphasize in the context of DI may also arise in other social insurance contexts. Hence, our framework may provide guidance for studying the trade-off between eligibility and benefit generosity also in other social insurance programs.

It is also important to keep in mind that the two DI policy instruments studied here are not the only policy instruments that affect the costs and benefits of the DI system. In particular, the generosity of other social insurance programs, such as UI, social welfare, and (early) retirement programs, may well affect the DI inflow to the extent that benefit substitution is quantitatively important (as in the Austrian context). Moreover, a DI reform could target the precision of the imperfectly functioning disability-assessment system to minimize false acceptances and false rejections (type I and type II errors). Unlike changing DI eligibility criteria, improving the precision of DI screening requires resources, such as more extensive medical checks, better equipment, additional monitoring of DI applicants, and the like. The welfare calculations of such a policy need to consider society's

[39] Autor et al. (2019) documented differences in the insurance value of DI receipt for accepted and rejected appellants between married and unmarried individuals in Norway. Their evidence, however, does not speak to potential differences between married and unmarried regarding the *relative* insurance losses of stricter eligibility versus lower benefits. In Section 6, we showed that our fiscal estimates from the whole sample are representative for the fiscal effects of married men. Therefore, our analysis and welfare conclusions do hold for married individuals irrespective of the insurance losses of unmarried individuals.

[40] There are many differences between the U.S. and Austrian settings. For example, workers in Austria can apply to DI while still employed, which leads to a direct rejection of the application in the U.S. program. The U.S. social safety net is less comprehensive than the Austrian safety net. The U.S. labor market is more flexible than the Austrian labor market. All these differences can affect both the multipliers and insurance losses of DI reforms in non-obvious ways.

willingness to pay for improved DI screening. Studying this trade-off is an interesting direction for future research.

REFERENCES

ABADIE, ALBERTO (2003): "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113 (2), 231–263. [99]

AUTOR, DAVID, ANDREAS KOSTØL, MAGNE MOGSTAD, AND BRADLEY SETZLER (2019): "Disability Benefits, Consumption Insurance, and Household Labor Supply," *American Economic Review*, 109 (7), 2613–2654. [82,107]

AUTOR, DAVID H., AND MARK G. DUGGAN (2003): "The Rise in the Disability Rolls and the Decline in Unemployment," *Quarterly Journal of Economics*, 118 (1), 157–206. [82]

BAILY, MARTIN NEIL (1978): "Some Aspects of Optimal Unemployment Insurance," *Journal of Public Economics*, 10 (3), 379–402. [80]

BALL, STEFFAN, AND HAMISH LOW (2014): "Do Self-Insurance and Disability Insurance Prevent Consumption Loss on Disability?" *Economica*, 81 (323), 468–490. [80]

BARGAIN, OLIVIER, KRISTIAN ORSINI, AND ANDREAS PEICHL (2014): "Comparing Labor Supply Elasticities in Europe and the United States: New Results," *Journal of Human Resources*, 49 (3), 723–838. [106]

BENITEZ-SILVA, HUGO, MOSHE BUCHINSKY, AND JOHN RUST (2004): "How Large Are the Classification Errors in the Social Security Disability Award Process?" Working Paper 10219, National Bureau of Economic Research. [82]

BLUNDELL, RICHARD, LUIGI PISTAFERRI, AND ITAY SAPORTA-EKSTEN (2016): "Consumption Inequality and Family Labor Supply," *American Economic Review*, 106 (2), 387–435. [102]

BÖHEIM, RENÉ, AND MICHAEL TOPF (2021): "Unearned Income and Labor Supply: Evidence From Survivor Pensions in Austria," Available at SSRN 3773355. [106]

BORGHANS, LEX, ANNE C. GIELEN, AND ERZO F. P. LUTTMER (2014): "Social Support Substitution and the Earnings Rebound: Evidence From a Regression Discontinuity in Disability Insurance Reform," *American Economic Journal: Economic Policy*, 6 (4), 34–70. [82]

BOUND, JOHN (1989): "The Health and Earnings of Rejected Disability Insurance Applicants," *American Economic Review*, 79 (3), 482–503. [82]

BOUND, JOHN, AND RICHARD V. BURKHAUSER (1999): "Economic Analysis of Transfer Programs Targeted on People With Disabilities," in *Handbook of Labor Economics*, Vol. 3. Elsevier, 3417–3528. [82]

BOUND, JOHN, TODD STINEBRICKNER, AND TIMOTHY WAIDMANN (2010): "Health, Economic Resources and the Work Decisions of Older Men," *Journal of Econometrics*, 156 (1), 106–129. [82]

CAMPOLIETI, MICHELE (2004): "Disability Insurance Benefits and Labor Supply: Some Additional Evidence," *Journal of Labor Economics*, 22 (4), 863–889. [82]

CAMPOLIETI, MICHELE, AND CHRIS RIDDELL (2012): "Disability Policy and the Labor Market: Evidence From a Natural Experiment in Canada, 1998–2006," *Journal of Public Economics*, 96 (3–4), 306–316. [82]

CESARINI, DAVID, ERIK LINDQVIST, MATTHEW J. NOTOWIDIGDO, AND ROBERT ÖSTLING (2017): "The Effect of Wealth on Individual and Household Labor Supply: Evidence From Swedish Lotteries," *American Economic Review*, 107 (12), 3917–3946. [106]

CHEN, SUSAN, AND WILBERT VAN DER KLAAUW (2008): "The Work Disincentive Effects of the Disability Insurance Program in the 1990s," *Journal of Econometrics*, 142 (2), 757–784. [82]

CHETTY, RAJ (2006a): "A General Formula for the Optimal Level of Social Insurance," *Journal of Public Economics*, 90 (10–11), 1879–1901. [80]

——— (2006b): "A New Method of Estimating Risk Aversion," *American Economic Review*, 96 (5), 1821–1834. [89]

DE CHAISEMARTIN, CLÉMENT, AND XAVIER D'HAULTFOEUILLE (2018): "Fuzzy Differences-in-Differences," *The Review of Economic Studies*, 85 (2), 999–1028. [99]

DESHPANDE, MANASI, AND YUE LI (2019): "Who Is Screened out? Application Costs and the Targeting of Disability Programs," *American Economic Journal: Economic Policy*, 11 (4), 213–248. [82]

DESHPANDE, MANASI, AND LEE LOCKWOOD (2022): "Beyond Health: Non-Health Risk and the Value of Disability Insurance," *Econometrica*, 90 (4), 1781–1810. [90]

DESHPANDE, MANASI, TAL GROSS, AND YALUN SU (2021): "Disability and Distress: The Effect of Disability Programs on Financial Outcomes," American Economic Journal: Applied Economics. [80]

DIAMOND, PETER A., AND EYTAN SHESHINSKI (1995): "Economic Aspects of Optimal Disability Benefits," *Journal of Public Economics*, 57 (1), 1–23. [80,83,85]

FADLON, ITZIK, AND TORBEN HEIEN NIELSEN (2019): "Household Labor Supply and the Gains From Social Insurance," in *Journal of Public Economics*, Vol. 171, 18–28. trans-Atlantic Public Economics Seminar 2016. [80-82,88,103,104,106]

FRENCH, ERIC, AND JAE SONG (2014): "The Effect of Disability Insurance Receipt on Labor Supply," *American Economic Journal: Economic Policy*, 6 (2), 291–337. [82]

GARCIA-MANDICÓ, SÍLVIA, PILAR GARCÍA-GÓMEZ, ANNE C. GIELEN, AND OWEN O'DONNELL (2020): "Earnings Responses to Disability Insurance Stringency," *Labour Economics*, 66, 101880. [82]

GELBER, ALEXANDER, TIMOTHY J. MOORE, AND ALEXANDER STRAND (2017): "The Effect of Disability Insurance Payments on Beneficiaries' Earnings," *American Economic Journal: Economic Policy*, 9 (3), 229–261. [82]

GIUPPONI, GIULIA (2019): "When Income Effects Are Large: Labor Supply Responses and the Value of Welfare Transfers," CEP Discussion Paper No. 1651. [106]

GODARD, MATHILDE, PIERRE KONING, AND MAARTEN LINDEBOOM (2022): "Application and Award Responses to Stricter Screening in Disability Insurance," Journal of Human Resources, 1120–11323R1. [82]

GOLOSOV, MIKHAIL, MICHAEL GRABER, MAGNE MOGSTAD, AND DAVID NOVGORODSKY (2021): "How Americans Respond to Idiosyncratic and Exogenous Changes in Household Wealth and Unearned Income," Working Paper 29000, National Bureau of Economic Research. [106]

GRUBER, JONATHAN (1997): "The Consumption Smoothing Benefits of Unemployment Insurance," *American Economic Review*, 87 (1), 192–205. [107]

——— (2000): "Disability Insurance Benefits and Labor Supply," *Journal of Political Economy*, 108 (6), 1162–1183. [82]

GRUBER, JONATHAN, AND JEFFREY D. KUBIK (1997): "Disability Insurance Rejection Rates and the Labor Supply of Older Workers," *Journal of Public Economics*, 64, 1–23. [82]

HALLER, ANDREAS, STEFAN STAUBLI, AND JOSEF ZWEIMÜLLER (2024): "Supplement to 'Designing Disability Insurance Reforms: Tightening Eligibility Rules or Reducing Benefits?'," *Econometrica Supplemental Material*, 92, https://doi.org/10.3982/ECTA19021. [83]

——— (2023): "Designing Disability Insurance Reforms: Tightening Eligibility Rules or Reducing Benefits? Supplementary Material," *SSRN Working Paper*, https://ssrn.com/abstract=4597237. [83]

HENDREN, NATHANIEL, AND BEN SPRUNG-KEYSER (2020): "A Unified Welfare Analysis of Government Policies," *Quarterly Journal of Economics*, 135 (3), 1209–1318. [80]

IMBENS, GUIDO W., AND DONALD B. RUBIN (1997): "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *The Review of Economic Studies*, 64 (4), 555–574. [99]

IMBENS, GUIDO W., DONALD B. RUBIN, AND BRUCE I. SACERDOTE (2001): "Estimating the Effect of Unearned Income on Labor Earnings, Savings, and Consumption: Evidence From a Survey of Lottery Players," *American Economic Review*, 91 (4), 778–794. [106]

KOSTØL, ANDREAS, AND ANDREAS MYHRE (2021): "Labor Supply Responses to Learning the Tax and Benefit Schedule," *American Economic Review*, 111 (11), 3733–3766. [82]

KOSTOL, ANDREAS R., AND MAGNE MOGSTAD (2014): "How Financial Incentives Induce Disability Insurance Recipients to Return to Work," *American Economic Review*, 104 (2), 624–655. [82]

LANDAIS, CAMILLE, AND JOHANNES SPINNEWIJN (2021): "The Value of Unemployment Insurance," *The Review of Economic Studies*, 88 (6), 3041–3085. [107]

LEE, DAVID S., PAULINE LEUNG, CHRISTOPHER J. O'LEARY, ZHUAN PEI, AND SIMON QUACH (2021): "Are Sufficient Statistics Necessary? Nonparametric Measurement of Deadweight Loss in Unemployment Insurance," *Journal of Labor Economics*, 39 (S2), S455–S506. [80]

LOW, HAMISH, AND LUIGI PISTAFERRI (2015): "Disability Insurance and the Dynamics of the Incentive Insurance Trade-Off," *American Economic Review*, 105 (10), 2986–3029. [82,83]

——— (2020): "Disability Insurance: Theoretical Trade-Offs and Empirical Evidence," *Fiscal Studies*, 41 (1), 129–164. [82]

MAESTAS, NICOLE, KATHLEEN J. MULLEN, AND ALEXANDER STRAND (2013): "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt," *American Economic Review*, 103 (5), 1797–1829. [82,99]

MEYER, BRUCE D., AND WALLACE K. C. MOK (2019): "Disability, Earnings, Income and Consumption," *Journal of Public Economics*, 171, 51–69. [80]

MOORE, TIMOTHY J. (2015): "The Employment Effect of Terminating Disability Benefits," *Journal of Public Economics*, 124, 30–43. [82]

MULLEN, KATHLEEN J., AND STEFAN STAUBLI (2016): "Disability Benefit Generosity and Labor Force Withdrawl," *Journal of Public Economics*, 143, 49–63. [82,93,97,98]

PARSONS, DONALD O. (1991): "Self-Screening in Targeted Public Transfer Programs," *Journal of Political Economy*, 99 (4), 859–876. [82]

PICCHIO, MATTEO, SIGRID SUETENS, AND JAN C. VAN OURS (2018): "Labour Supply Effects of Winning a Lottery," *The Economic Journal*, 128 (611), 1700–1729. [106]

RUH, PHILIPPE, AND STEFAN STAUBLI (2019): "Financial Incentives and Earnings of Disability Insurance Recipients: Evidence From a Notch Design," *American Economic Journal: Economic Policy*, 11 (2), 269–300. [82]

STATISTIK AUSTRIA (2018): "Personen in Privathaushalten ab 15 Jahren nach Familienstand und Alter," Mikrozensus-Arbeitskräfteerhebung. [106]

STAUBLI, STEFAN (2011): "The Impact of Stricter Criteria for Disability Insurance on Labor Force Participation," *Journal of Public Economics*, 95 (9–10), 1223–1235. [91]

STAUBLI, STEFAN, AND JOSEF ZWEIMÜLLER (2013): "Does Raising the Early Retirement Age Increase Employment of Older Workers?" *Journal of Public Economics*, 108, 17–32. [93]

ZWEIMÜLLER, JOSEF, RUDOLF WINTER-EBMER, RAFAEL LALIVE, ANDREAS KUHN, JEAN-PHILIPE WUELLRICH, OLIVER RUF, AND SIMON BÜCHI (2009): "Austrian Social Security Database," Vol. 401, 1–80. Working Paper, University of Zurich. [92]

Co-editor Oriana Bandiera handled this manuscript.

Manuscript received 5 October, 2020; final version accepted 11 June, 2023; available online 18 September, 2023.