Mental Health and the Targeting of Social Assistance

Canishk Naik

UC Berkeley

January 2026

Latest version here

ABSTRACT. People living with mental disorders are vulnerable, yet often struggle to navigate complex barriers to accessing assistance. How this shapes the effectiveness of the social safety net remains unclear. Using administrative data covering the population of the Netherlands, I find that people with poor mental health are $3\times$ more likely to fall below the poverty line but only receive social assistance at similar rates to those with good mental health. Moreover, a policy that increases barriers disproportionately screens out people suffering from mental disorders. This is not because they value benefits less; individuals with poor mental health respond more to exogenous variation in the benefit level, suggesting a strong redistributive motive. To interpret these patterns, I develop a simple framework showing how take-up behaviour reveals whether the neediest are being screened out. Calibrating the model shows that those with mental disorders reveal a 57% higher marginal cost of overcoming barriers and a $2\times$ higher marginal value of benefits (need), controlling for income. Therefore, similar take-up rates mask poor targeting. In this context, reducing barriers is $2.3\times$ more cost-effective than increasing benefits.

JEL Codes: I38, I14, H53, D91

Contact: Haas School of Business, UC Berkeley (email: canishk@berkeley.edu).

I am incredibly thankful for all the advice and support given to me by my supervisors Nava Ashraf, Daniel Reck and Johannes Spinnewijn. This project also benefited greatly from insightful discussions with Doug Bernheim, Sonia Bhalotra, János Betkó, Jonneke Bolhaar, Katie Brown, Gharad Bryan, Francesco Caselli, Gaby Deschamps, Matthias Doepke, Jack Fisher, Nilmini Herath, Amen Jalal, Xavier Jaravel, Camille Landais, Gabriel Leite Mariante, Kristóf Madarász, Joana Naritomi, Will Parker, Charlie Rafkin, Gautam Rao, Michelle Rao, Patrick Schneider, Kate Smith, Caterina Soto Vieira, Krishna Srinivasan, Neil Thakral, Timo Verlaat, Heike Vethaak, Jenny Wang, Sarah Winton, Alwyn Young, Alice Zulkairnan and various seminar participants. I gratefully acknowledge financial support from STICERD (Grant #108968) and from the Economic and Social Research Council DTP (Grant #ES/P000622/1).

1. Introduction

Poor mental health is an urgent societal issue. Almost 1 billion people live with a mental disorder (WHO, 2022). In 2010, the economic cost of mental illness due to lost productivity and poor health was estimated to be $2.5 trillion, and is expected to double by 2030 (Bloom et al., 2012).

This paper examines how mental health influences the design of the social safety net. Mental disorders simultaneously impact many key inputs to the policy process. They themselves represent welfare-relevant hardship,[1] and also correlate strongly with economic vulnerability: people with poor mental health face up to three times the risk of poverty (Ridley et al., 2020). At the same time, mental illness may impair decision-making (Bierman et al., 2008; Hammar and Årdal, 2009; Gross and Muñoz, 1995) and make it difficult to navigate complex barriers to assistance (Herd and Moynihan, 2025). However, how mental disorders shape the effectiveness of traditional economic programs remains poorly understood. I provide the first comprehensive evidence on this question in the context of social assistance targeting in the Netherlands, using administrative data covering the full population.

Mental disorders pose important challenges in determining effective targeting of the social safety net. Eligibility for welfare benefits is determined by individuals having few resources, but income alone does not capture all dimensions of need. Theoretically, barriers screen for *general* unobservable vulnerabilities through self-targeting. The challenge is that poor mental health affects both the *cost* of overcoming barriers (decreasing take-up) and the *need*, or marginal value of benefits (increasing take-up). Therefore, take-up cannot distinguish between these channels: it reflects not only how willing people are to overcome barriers to receive benefits, but also how able. Separately identifying need from cost is essential: barriers are effective if the needy can afford the cost while the less needy cannot and are screened-out.

Empirically, measuring mental health at scale is also challenging. Survey data suffers from small samples and under-reporting due to stigma (Bharadwaj et al., 2017), while administrative data also face issues: objective outcomes are often extreme and people with poor mental health forgo care (Cronin et al., 2024).

I address these challenges in three steps. First, I develop a simple theoretical framework showing how to disentangle the need for benefits from the cost of overcoming barriers using take-up responses to changes in benefits and barriers. Second, I examine the empirical implications of mental health for the targeting of social assistance. I quantify average targeting and estimate how

[1] Common symptoms of mental disorders include feelings of worthlessness, confused thinking, withdrawal from support networks, fear, fatigue, guilt and, in the extreme case, suicidality (APA, 2013)

take-up responds to policy using Dutch administrative data ($N \approx 17$ million). The data contain rich socioeconomic and mental health information from administrative sources, a large ($N \approx 400$k) linked survey, and social assistance take-up. Finally, I combine theory and empirics to calculate how need and cost vary with mental health and examine the welfare consequences of targeting.

The key theoretical finding of this paper is that combining differences in average take-up levels across groups with take-up responses to changes in benefits and barriers is sufficient to evaluate people's marginal value of benefits (need) and marginal cost of having to overcome barriers (cost). I develop a framework allowing for heterogeneity in both need and cost across people with the same income.[2] The framework illustrates that three empirical moments relating to take-up reveal this underlying heterogeneity: (i) Differences in average take-up levels across groups reflect differences in *average* value net of cost. (ii) If one group responds more to a change in benefits, either they have higher need (*marginal* value) or they are more likely to lie at the margin take-up (i.e. *average* value net of cost closer to 0. This can be isolated by difference in take-up levels). (iii) Once need has been separately identified through (i) and (ii), this information can be combined with take-up responses to changes in barriers to identify cost.

Identifying how the need for benefits and cost of barriers depend on mental health provides key behavioural primitives for understanding targeting effectiveness. For instance, I derive the welfare effects of a budget-neutral increase in barriers, where money saved from lower take-up finances higher benefit levels. Translating these primitives to welfare effects requires modelling assumptions, most critically whether revealed costs reflect true welfare burdens or behavioural frictions. I characterise the robustness of welfare effects to the normative weight placed on ordeal costs following Naik and Reck (2025).

Empirically, I study social assistance take-up and mental health using Dutch administrative data. I examine the flagship social assistance program in the Netherlands, a cash transfer designed for people who do not have enough money to subsist. I combine detailed information on socio-economic demographics for the years 2011 - 2020 to construct an accurate measure of eligibility with low measurement-error.[3] Furthermore, the data contain rich mental health information, coming from three classes of outcomes: care usage, extreme outcomes and subjective mental health from a large linked survey. Each source is a potentially biased proxy on its own.

[2]Need encompasses economic vulnerability and perceptions about treatment effects of assistance. Cost comprises deadweight loss from compliance, information and psychological barriers. Both may contain bias, especially given the cognitive and emotional constraints of mental disorders. The framework thus provides a revealed preference approach to identify state-dependent utility (Finkelstein et al., 2013).

[3]Accurately calculating eligibility is a key challenge facing the take-up literature (Ko and Moffitt, 2024). I find that measurement-error is likely small: the estimated $\mathbb{P}[SA|\text{ Ineligible}] = 1\%$.

However, I find consistent patterns *across* measures throughout the empirical analysis, reflecting latent mental health status driving the results.

Three key findings arise from my empirical analysis. The first is descriptive. I find that people with poor mental health are substantially more likely to be eligible for social assistance than those with good mental health. One quarter of those eligible for social assistance are being treated for a mental disorder, more than double the rate for the general population. However, the average take-up levels (62.5%) hardly differ by mental health status conditional on eligibility, income and other covariates. Therefore, social assistance is not meaningfully targeted towards the more vulnerable group on average.

Second, increases in barriers to accessing social assistance disproportionately screen out people with poor mental health. I exploit the introduction of the Participation Act (Ministerie van SZW, 2015), a policy which increased access barriers by intensifying the obligations that recipients have to satisfy and incentivising municipalities to restrict inflow (SCP, 2019), a goal they pursued through (the threat of) sanctions (Ministerie van SZW, 2022). I use a difference-in-differences design to show that the reform disproportionately discourages people with poor mental health from entering the program, reducing their inflow by 10% compared to those with good mental health, conditional on income and eligibility. There is no differential outflow, indicating that psychological costs of anticipated obligations and difficult interactions with municipal workers, rather than actual compliance burdens, drive the screening effect.

Third, people with poor mental health respond twice as much to changes in benefits compared to those with good mental health, suggesting a revealed-preference rationale for redistributing resources toward this group. In the Netherlands, social assistance tops up income to an eligibility threshold, creating a kinked benefit schedule as a function of income (100% marginal tax rate below the threshold, 0% above). I leverage this kinked schedule as a novel instrument for benefit-level, made possible by my construction of the income concept used to determine eligibility. Using this source of variation in a regression kink design, I estimate elasticities of social assistance receipt with respect to benefits of 0.38 for those with poor mental health and 0.16 for those with good mental health.

Mental health emerges as the dimension of heterogeneity showing the strongest joint responsiveness to both barriers and benefits conditional on income, exceeding consumption, gender, migration status, and physical health.

Combining theory and empirical estimates yields the final key finding: people with poor mental health need benefits $2.2\times$ more than those with good mental health, conditional on income, but face 57% higher costs of overcoming barriers. Thus, equal take-up rates mask poor targeting. The neediest people, to whom the government has the strongest motive to redistribute money, find it most difficult to access assistance, undermining barriers as an effective policy tool to target social assistance.

I characterise welfare consequences of social assistance targeting by mental health using the marginal value of public funds (MVPF) framework (Hendren and Sprung-Keyser, 2020). Under the baseline assumption that revealed need and cost are welfare-relevant, reducing barriers yields an MVPF of 2.75 compared to 1.18 for increasing benefits, implying barrier reductions are 2.3× more cost-effective. If ordeal costs instead reflect behavioural biases, increasing benefits becomes relatively more attractive. This is because the welfare effects of benefit increases are determined by how well targeted the program is, regardless of whether targeting is driven by true ordeal costs or behavioural frictions.

Contribution to the Literature: There is a long literature in psychology and growing literature in economics studying mental disorders. Poor mental health not only imposes cognitive burden (Bierman et al., 2008; Hammar and Årdal, 2009) but also impairs emotion regulation (Gross and Muñoz, 1995), both of which hinder everyday functioning (Kessler et al., 2003; Evans et al., 2014). Economics research demonstrates that mental healthcare interventions such as psychotherapy improve self-confidence, patience, risk-tolerance and reduce decision costs (Bhat et al., 2022; Shreekumar and Vautrey, 2021; Angelucci and Bennett, 2024a,b). The literature also explores how mental health affects economic outcomes (Barker et al., 2021; Baranov et al., 2020; Ridley et al., 2020; Serena, 2024) and how income impacts mental health (Christian et al., 2019; Miller et al., 2024; Schmidt et al., 2021; Silver and Zhang, 2022).

I provide one of the first characterisations of the consequences of mental disorders for the design of traditional economic policies. Although the behavioural public policy literature has explored how mental health correlates with take-up (Arulsamy and Delaney, 2022; Bell et al., 2022; Martin et al., 2023a,b), I contribute by first developing a simple model to show that understanding effectiveness requires estimating heterogeneous take-up responses to both benefits and barriers, and then identifying these using quasi-experimental variation and rich administrative data.

Social safety net targeting is one domain where the unique features of mental disorders become particularly important for policy design. Ordeal-based screening is predicated on the assumption that need and barrier costs are negatively correlated (Nichols and Zeckhauser, 1982), precisely the relationship mental health threatens. That being said, there is an ongoing empirical debate examining whether ordeals screen out the "wrong" people in terms of income and other proxies for need (Alatas et al., 2016; Deshpande and Li, 2019; Giannella et al., 2023; Homonoff and Somerville, 2021; Wu and Meyer, 2023). Studies estimating welfare effects highlight how take-up frictions (Finkelstein and Notowidigdo, 2019), adverse selection (Shepard and Wagner, 2022) and implementation costs (Unrath, 2024) can undermine effectiveness, though Rafkin et al. (2025) argues self-targeting can be socially beneficial on average.

Accounting for mental disorders raises overlooked challenges. A full cost-benefit analysis requires quantifying the trade-off between ordeal costs and the need for benefits, potentially extending beyond poverty, that can be redistributed to the stayers when barriers screen people out. Neither take-up levels nor responses to ordeals alone are sufficient statistics for this trade-off, precisely because need can co-vary with cost conditional on income. I show that an additional moment, take-up responses to changes in benefits, is necessary to evaluate welfare implications.

Why don't governments simply target benefits based on mental health directly? Mental health is difficult to observe and contract upon (Godard et al., 2022), raising challenges for putting it into eligibility criteria directly. In any case, evaluating many alternative policy reforms, including direct targeting, requires benefit take-up elasticity estimates (Anders and Rafkin, 2022; Rafkin et al., 2025). Such estimates remain limited. I provide new evidence using a regression kink design that exploits quasi-experimental variation through the benefit schedule, revealing substantial heterogeneity by mental health status.

Outline: Section 2 develops a simple framework to guide the empirical analysis by showing how take-up behaviour reveals need and cost. In Section 3, I describe the context and data. I quantify descriptive patterns in social assistance take-up and targeting by mental health in Section 4. I estimate take-up responses to changes in barriers in Section 5 and benefits in Section 6. Section 7 calibrates the model and Section 8 concludes.

2. Theoretical Framework

I adapt the model from Finkelstein and Notowidigdo (2019), allowing for heterogeneous marginal value of €1 (need) even across individuals with the same income. This generalization reflects that vulnerability among people with poor mental health extends beyond their risk of poverty. Proofs and extensions are in Appendix A.

2.1. Positive Model of Social Assistance Take-up.

2.1.1. *Setup.* Individuals are indexed by mental health, θ. Social assistance is defined by two policy parameters. B is the (monetary) benefit, Λ is the barrier that individuals have to overcome to receive B. Each θ chooses whether to receive social assistance:

$$SA_\theta = \mathbb{1}\{\text{overcome barrier } \Lambda \text{ to receive benefit } B\} \tag{2.1}$$

Individuals derive value $v_\theta(B)$ from benefits B and face take-up cost $\kappa_\theta(\Lambda)$ from overcoming barrier Λ. This model takes a stylised reduced-form approach, where values and costs are catch-all quantities that could arise from various mechanisms and may reflect welfare-relevant primitives and/or contain psychological frictions.[4]

Take-up depends on an independent additive choice-shock $\varepsilon \sim F$ which captures decision-relevant unobservables unaffected by policy. Therefore take-up satisfies:

$$SA_\theta = 1 \iff v_\theta(B) > \kappa_\theta(\Lambda) + \varepsilon \tag{2.2}$$

Behaviour follows a threshold-rule: if $\varepsilon \leq \varepsilon^*_\theta = v_\theta(B) - \kappa_\theta(\Lambda)$, $SA = 1$ and if $\varepsilon > \varepsilon^*_\theta$, $SA = 0$. Rate of receipt is given by:

$$\mathbb{P}[SA]_\theta = F\big(v_\theta(B) - \kappa_\theta(\Lambda)\big) \tag{2.3}$$

2.2. **Identification.** Equation (2.3) shows that take-up levels do not distinguish between the key primitives **need** $:= v'_\theta(B)$ and **cost** $:= \kappa'_\theta(B)$ which determine targeting effectiveness. Remark 2.1 previews how take-up *responses* to changes in B and Λ can help tease them apart.

Remark 2.1. Barrier screening effects are characterised by Equation (2.4), and benefit take-up effects by Equation (2.5).

$$\underbrace{\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}}_{\text{Barrier screening effects}} = -\underbrace{\kappa'_\theta}_{\text{cost}} \cdot f_\varepsilon(v_\theta - \kappa_\theta) \tag{2.4}$$

$$\underbrace{\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}}_{\text{Benefit take-up effects}} = \underbrace{v'_\theta}_{\text{need}} \cdot f_\varepsilon(v_\theta - \kappa_\theta) \tag{2.5}$$

[4]Appendix A presents a micro-foundation of $v_\theta(B)$ for completeness. Value arises from extra consumption and recovered costs of work. Income depends on take-up but is fixed otherwise: $y_\theta^{SA=1}$ refers earned-income while receiving social assistance and $y_\theta^{SA=0}$ represents earned-income when not.

Intuitively, Λ is a price of taking up. Therefore, responsiveness to barriers is larger for a group when that group is more price-responsive (κ' large) or more likely to lie at the margin of take-up $\big(f_\varepsilon(\cdot)$ large$\big)$. Similarly, responsiveness to a change in benefit level is governed by need (v') and the likelihood of being marginal. Section 2.2.1 shows how the primitives are jointly identified.

2.2.1. *Three-step Identification.* The method takes as inputs: take-up levels $\mathbb{P}[SA]_\theta$, barrier screening effects $\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}$ and benefit take-up effects $\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}$ and uses these to identify need (v'_θ), cost (κ'_θ) and the likelihood of being marginal $\big(f_\varepsilon(v_\theta - \kappa_\theta)\big)$. I discuss the key assumptions underpinning the identification in Section 2.4. The intuition is as follows:

Suppose that the benefit take-up effect $\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}$ amongst one group is large relative to another. That reflects *either* higher need (v'_θ), by revealed-preference, *or* a higher likelihood to lie on the margin of take-up ($f_\varepsilon(v_\theta - \kappa_\theta)$). The latter occurs when average value net of cost $v_\theta - \kappa_\theta$ is closer to 0, i.e. θ closer to indifference, and can be examined separately using differences in average take-up levels $\mathbb{P}[SA]_\theta$. Once isolated, this identifies need (up to normalising $v'_{\theta_0} = 1$ for one type). Combining this with barrier screening effects $\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}$ identifies cost.

Step 1 (Average take-up levels): To aid intuition, suppose that we are in a special case of equalised take-up levels: $\mathbb{P}[SA]_\theta = \mathbb{P}[SA]_{\tilde{\theta}}$.[5] I.e. $F(v_\theta - \kappa_\theta) = F(v_{\tilde{\theta}} - \kappa_{\tilde{\theta}})$. Then, $f(v_\theta - \kappa_\theta) = f(v_{\tilde{\theta}} - \kappa_{\tilde{\theta}})$ because the cdf F is monotonic and $\varepsilon \perp \theta$. More generally if $\mathbb{P}[SA]_\theta \neq \mathbb{P}[SA]_{\tilde{\theta}}$, $f(v_\theta - \kappa_\theta)$ is identified in terms of $f(v_{\tilde{\theta}} - \kappa_{\tilde{\theta}})$ using a first-order Taylor expansion of difference in average take-up levels $\mathbb{P}[SA]_\theta - \mathbb{P}[SA]_{\tilde{\theta}}$. This requires additional structure, and is set out in Appendix A.2. At the end of **Step 1**, we know how $f(v_\theta - \kappa_\theta)$ compares across types.

Step 2 (Benefit take-up effects): If we know how $f(v_\theta - \kappa_\theta)$ compares across types, and estimate benefit take-up effects for each type - then we can quantify need. This done by dividing Equation (2.5) across types to give:

$$\frac{\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}}{\frac{\partial \mathbb{P}[SA]_{\tilde{\theta}}}{\partial B}} = \frac{v'_\theta}{v'_{\tilde{\theta}}} \cdot \underbrace{\frac{f_\varepsilon(v_\theta - \kappa_\theta)}{f_\varepsilon(v_{\tilde{\theta}} - \kappa_{\tilde{\theta}})}}_{\text{Estimated in } \textbf{Step 1}} \tag{2.6}$$

If we normalise $v'_{\theta_0} = 1$ for some θ_0 we can calculate v'_θ for all other θ using Equation (2.6). This normalization is without loss, and effectively scales all welfare effects in terms of θ_0's willingness-to-pay for €1. v'_θ is then understood as θ's need relative to θ_0.

[5]This is motivated by the empirical application, where I find no meaningful difference in average take-up levels by mental health.

Step 3 (Barrier screening effects): Finally, divide barrier screening effects from Equation (2.4) by benefit take-up effects from Equation (2.5) *within type* to identify κ'_θ for all θ as follows: [6]

$$\frac{-\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}}{\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}} = \kappa'_\theta \cdot \underbrace{\frac{1}{v'_\theta}}_{\text{Estimated in } \textbf{Step 2}} \tag{2.7}$$

2.3. **Welfare under Revealed Preference.** Why is it so important to identify need and cost? These primitives are key components of the welfare effects of changing B and Λ. The welfare consequences of marginal changes in policies affecting targeting serve as measures of targeting effectiveness as a whole. Translating the positive model to normative analysis requires making several important assumptions, most critically regarding whether choices reveal preferences. As a tractable benchmark, I first treat both $v_\theta(B)$ and $\kappa_\theta(\Lambda)$ as normatively relevant, then discuss robustness in Section 2.4.

2.3.1. *Individual Welfare.* Denote U_θ as θ's utility (which depends on take-up), and $\mathcal{U}_\theta$ expected utility. Following the setup in Section 2.1.1, I normalise utility to 0 if $SA = 0$. The normalisation is innocuous because in the succeeding analysis, I only consider *marginal* changes to B and Λ, and $\frac{d\mathbb{E}[\text{Utility}_\theta | \text{ SA=0}]}{dB} = \frac{d\mathbb{E}[\text{Utility}_\theta | \text{ SA=0}]}{d\Lambda} = 0$. Recall $\varepsilon^*_\theta = v_\theta(B) - \kappa_\theta(\Lambda)$. Then:

$$\begin{aligned}
\mathcal{U}_\theta = \mathbb{E}[U_\theta] &= \mathbb{P}[SA]_\theta \cdot \mathbb{E}[\text{Utility}_\theta | \text{ SA} = 1] + \big(1 - \mathbb{P}[SA]_\theta\big) \cdot \underbrace{\mathbb{E}[\text{Utility}_\theta | \text{ SA} = 0]}_{\text{Normalised to 0}} \\
&= \int_{-\infty}^{\varepsilon^*_\theta} [v_\theta(B) - \kappa_\theta(\Lambda) - \varepsilon] \; dF(\varepsilon)
\end{aligned}$$

2.3.2. *Social Welfare.* Let $\mu(\theta)$ be the distribution of types, and λ_θ social welfare weights. Of course, a major driver of redistributive motives is v'_θ, so λ_θ captures any deviations between the private and social marginal utility of transferring €1 to θ. All income (including benefits) is taxed at fixed marginal tax rate τ. The government's problem is:

[6]This within-type identification method is the same as the method used in Haller and Staubli (2024), and echoes the identification of the marginal rate of substitution in Russo (2023). The across-type identification is new. Here, the key novelty is that I can estimate the difference in average take-up levels conditional eligibility and use it to inform differences in average value net of cost across types, reminiscent of the "correlation test" from Rafkin et al. (2025).

$$W = \max_{\Lambda, B} \int \lambda_\theta \, \mathcal{U}_\theta \, d\mu(\theta)$$

$$\text{s.t.} \underbrace{\int \tau y_\theta^{SA=0} \cdot \big(1 - \mathbb{P}[SA]_\theta\big) + \tau(y^{SA=1} + B) \cdot \mathbb{P}[SA]_\theta \, d\mu(\theta)}_{\text{Tax Revenue}} = \underbrace{\int B \cdot \mathbb{P}[SA]_\theta \, d\mu(\theta)}_{\text{Program Costs}} \tag{2.8}$$

In this framework, I assume eligibility criteria for benefits are fixed (though not explicitly modelled).[7] The government does not observe individuals' private types (θ, ε), making targeted policy design challenging. Instead, it must rely on blunt instruments, benefit levels (B) and barriers to access (Λ), which do not vary by θ to indirectly target those most in need.

2.3.3. *Welfare Effects of a Budget-Neutral Increase in Barriers.* I consider a policy experiment capturing the idea behind using barriers to target social assistance: raise barriers, using the money saved due to lower take-up to finance an increase in benefits.[8] This is a budget neutral increase in Λ (B adjusts). This conceptual exercise highlights the importance of need and cost for policy design.

Proposition 2.1. *The marginal welfare effect of a budget-neutral increase in ordeals financing an increase in benefits is given by:*

$$\frac{dW}{d\Lambda} = \int \lambda_\theta \, \mathbb{P}[SA]_\theta \left[\underbrace{v_\theta'(B)}_{\textit{Need}} \cdot \frac{dB}{d\Lambda} - \underbrace{\kappa_\theta'(\Lambda)}_{\textit{Cost}} \right] d\mu \tag{2.9}$$

Budget Neutrality implies:

$$\frac{dB}{d\Lambda} = \frac{-\int FE_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda} d\mu}{(1-\tau) \cdot \int \mathbb{P}[SA]_\theta \, d\mu + \int FE_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial B} d\mu} \tag{2.10}$$

where:

$$FE_\theta = \tau \cdot (y_\theta^{SA=0} - y_\theta^{SA=1}) + (1-\tau) \cdot B \tag{2.11}$$

Equation (2.9) follows from the Envelope Theorem. The welfare effect $\frac{dW}{d\Lambda}$ is large when take-up ($\mathbb{P}[SA]_\theta$) is concentrated among those with highest need (v_θ' large) and lowest ordeal-costs (κ_θ' small): barriers target effectively when neediest receive assistance. Conversely, a positive correlation between need and cost can push $\frac{dW}{d\Lambda}$ to be negative.

---

[7] I discuss how to explicitly model eligibility in detail in Appendix A.

[8] The policy experiment differs from Rafkin et al. (2025) in what we compare total ordeal-costs $\kappa_\theta(\Lambda)$ to. They consider moving to automatic enrolment (comparing $\kappa_\theta(\Lambda)$ to 0) whereas I consider reducing barriers (comparing $\kappa_\theta(\Lambda)$ to $\kappa_\theta(\Lambda - \delta\Lambda)$). Hence, I require variation in barriers as well as benefits.

The intuition behind Equation (2.10) is as follows. The government can increase B more if *more* people are screened out by ordeals, if people exhibit *larger* benefit take-up effects, and if there are *fewer* beneficiaries at baseline. FE_θ is the fiscal externality of θ applying: there is a moral hazard fiscal externality due to labour supply response $y^{SA=0} \rightarrow y^{SA=1}$ which costs the government $\tau(y^{SA=0} - y^{SA=1})$, and a net direct cost $(1-\tau)B$ paid out to θ.

$\frac{dW}{d\Lambda}$ is my overall metric for the social welfare consequences of targeting using ordeal mechanisms. In the calibration, I follow Hendren and Sprung-Keyser (2020) by deriving the marginal value of public funds (MVPF) of increasing barriers vs increasing benefits. I derive "redistributive" MVPFs as social-welfare weighted willingness-to-pay divided by government cost, directly capturing the redistributive motives central to social assistance targeting. The formulae for the redistributive MVPFs of dB and $d\Lambda$ are in Appendix A.1.

2.4. **Discussion of assumptions.** Before presenting the empirical analysis, it is important to discuss the key assumptions underlying the identification of need and cost. In Section 7.2, I return to these assumptions and characterise how relaxing them impacts welfare effects.

2.4.1. *The Role of Bias.* The revealed-preference framework treats observed take-up as reflecting welfare-relevant preferences, and serves as a useful benchmark for characterising effectiveness. However, take-up decisions may reflect behavioural frictions, particularly for people with poor mental health facing cognitive and emotional constraints. Rather than make untestable paternalistic assumptions about the normative weights on choices in different psychological states, I characterise how welfare effects vary as a function of whether choices reflect true welfare (Naik and Reck, 2025). This allows governments to evaluate the robustness of policy conclusions to beliefs about the normative-relevance of decisions made with poor mental health.

Appendix A.3 presents a formal extension where perceived costs may diverge from the welfare-relevant truth. The three-step identification method from Section 2.2.1 continues to apply, but now reveals *perceived* need and cost rather than welfare-relevant primitives. The welfare effects of Section 2.3 remain similar in structure, but now include an additional term (the "behavioural welfare effect"). For example, if people overstate the cost $\kappa(\Lambda)$, they take-up too much relative to their private optimum. Reducing Λ then generates welfare gains not only through the direct cost reduction but also through an indirect internality correction from debiasing behaviour.

2.4.2. *Other key assumptions.* I assume ε is an additive independent shock to the take-up equation: $SA_\theta = 1 \iff v_\theta > \kappa_\theta + \varepsilon$. Independence enables **Step 1** in the identification but is threatened by differential selection of recipients on other characteristics by mental health. I explore relaxing independence in Appendix F.3.1 fitting a heteroskedastic probit with parametric form for $F_\varepsilon = N(\mu_\theta, \sigma_\theta^2)$. This model utilises any differential selection on other characteristics to identify the variance of unobservables affecting take-up.

Additivity separates need/cost from the scaling factor $f_\varepsilon(v_\theta - \kappa_\theta)$ in Equations (2.4) and (2.5), allowing me to extrapolate from take-up responses of marginals to infer preferences for infra-marginals. In practice, this extrapolation may not be well-approximated as linear. I explore relaxing additivity in Appendix F.3.2 following the bounding argument of Haller and Staubli (2024), though how exactly to reveal ordeal-costs among infra-marginals without additivity remains an active area of research.

Identifying marginal value v'_θ separately across types requires making interpersonal comparisons about the value of money (see e.g. Finkelstein et al., 2013). Standard practice avoids this by writing welfare effects in terms of each type's willingness-to-pay, leaving social welfare weights λ_θ as free normative judgement parameters. For sceptics of interpersonal comparisons, my welfare estimates are equivalent to setting $\frac{\lambda_{\text{Poor MH}}}{\lambda_{\text{Good MH}}} = \frac{\frac{\partial \mathbb{P}[SA]_{\text{Poor MH}}}{\partial B}}{\frac{\partial \mathbb{P}[SA]_{\text{Good MH}}}{\partial B}}$, a reduced-form, revealed preference signal of redistributive motive.

In my framework, the likelihood of being on the margin of take-up, $f(v_\theta - \kappa_\theta)$, is identical for benefit and barrier instruments. This follows from θ and ε being one-dimensional and allows for minimal structure on the take-up equation. However, the compliers may depend on the instruments themselves (Kline and Walters, 2019; Mogstad et al., 2024). Under additional parametric assumptions, $f(v_\theta - \kappa_\theta)$ can depend on the instrument as long as it scales proportionally across instruments for all mental health types, $\forall\theta, f^{d\Lambda}(v_\theta - \kappa_\theta) = \alpha f^{dB}(v_\theta - \kappa_\theta)$. Appendix F.3.3 shows welfare effects in this case.

I focus on mental health θ as the sole dimension of heterogeneity, such that other observables X do not independently affect (v, κ), conditional on θ. Under this conditional homogeneity assumption, the welfare results represent average effects holding X fixed, integrating over the distribution of θ conditional on X. Characterising unconditional welfare effects that account for heterogeneity along multiple dimensions would require solving a multidimensional screening problem which is beyond the scope of this paper. However, I show empirically that mental health is the dominant driver of responsiveness even when controlling for X.

3. Context and Data

I now turn to the empirical analysis of mental health and the targeting of social assistance in the Netherlands. In this section, I describe the institutional setting and administrative data used to measure mental health and take-up.

3.1. Dutch Institutional Context.

3.1.1. *Social Assistance.* In the Netherlands, social assistance, or *algemene bijstand*, is a non-contributory social safety net program. It is intended for individuals who do not have enough income or assets to subsist, and who do not qualify for any other benefit. Over the period of this study, around 450,000 people claim benefits each year. This translates to around 5% of the working-age population and is more than the number of people on disability and unemployment insurance.

Eligibility: Eligibility rules are determined at the national level. The benefits are means-tested: income and assets must be below a threshold in order to be entitled. The income threshold is 100% of the full-time national minimum wage for couples, and 70% for singles. The threshold depends on household composition. Income used to determine eligibility includes not just labour income, but from capital and other benefits. Additionally, eligibility requires being at least 18 years old and Dutch citizenship or residing lawfully in the Netherlands, not in prison or a detention centre. Mental health does not directly affect eligibility.

Application: Applicants must submit information to verify eligibility (e.g. residency proof, income / bank statements etc) as well as potentially go to the municipal office for an interview. The municipality legally must make a decision within 8 weeks of application.

Receipt: If accepted, income is topped-up to the eligibility threshold - i.e. there is a 100% marginal tax rate. The national minimum wage, and couples' threshold, is around €16.5k per year during the observation period. Often, people earn some income - on average, benefits paid equal around €12.7k per year. Conditional on receipt, people stay on social assistance for around 5 years - there is no time-limit to take-up. Municipalities can grant additional benefits, such as housing, health insurance and children subsidies. In this paper, I focus on the take-up of the general welfare benefit.[9]

[9] This is a reasonable simplification because the take-up of these additional benefits is uncorrelated with receipt of social assistance, after controlling for income and wealth (Berkhout et al., 2019). Furthermore, these subsidies are phased-out according to different schedules to social assistance.

Obligations: Social assistance is a workfare program: conditional on take-up, recipients must comply with several obligations. These include keeping all information up-to-date and work re-integration.[10] Single parents with young children and people with full and permanent incapacity to work can apply for an exemption from these obligations. In the event of non-compliance, municipalities can impose sanctions or (temporarily) reduce benefits. Exclusion from assistance is an uncommon, extreme outcome.

3.1.2. *Health Insurance.* The Netherlands has a mandated and subsidised private health insurance system. GPs are the first port-of-call for mental health issues, and can prescribe medications or refer to specialized care. In the general population, around 10% of people are dispensed psychopharmacological medications each year. Access to mental healthcare is roughly equalised by income (Naik et al., 2026), although quality of care may differ (Lopes et al., 2023).

3.1.3. *Disability Insurance.* Disability benefits count towards eligibility for social assistance. Insofar as people receive full disability benefits (e.g. people with severe mental disorders), they have income above the social minimum, are ineligible for social assistance and do not appear in my main analysis. Moreover, disability insurance is a contributory program replacing past earnings after work-limiting health shocks. Many people receiving social assistance do not have prior work history, so are ineligible for disability benefits.

3.2. **Data.** I use administrative data from the population of the Netherlands ($N \approx 17$million) accessed via CBS, the Statistics Agency of the Netherlands. The data contain information on socio-economic demographics determining eligibility for social assistance, rich characteristics on social assistance receipt and comprehensive information about mental health.

3.2.1. *Socio-economic information:* I create a new dataset containing eligibility for social assistance in the years 2011-2020 for all working-age individuals in the Netherlands. To do so, I extend the work of Inspectie SZW (2021) to calculate eligibility by merging detailed information on income, wealth, household composition and size, work status, education and other demographics, following the rules set out by law (Ministerie van SZW, 2015). These data are yearly, and so the measure reflects eligibility on average each year. I observe all eligibility determinants in the period 2011-2020 and construct each individual's "eligibility income", the income concept used to determine

[10]Full list: acceptance of work or voluntary activities (i.e. "participate"), wearing the correct clothing doing so, being prepared to travel a distance with a total travel time of 3 hours per day to find work, keeping all eligibility and benefit-level information up-to-date, complying with information requests and even home-visits, being willing to relocate municipality, achieving a good command of the Dutch language and acquiring and retaining knowledge and skills necessary for acquiring wealth (Ministerie van SZW, 2015).

eligibility. I focus on working-age individuals throughout the study.[11] Among this population, around 7% of people are eligible for social assistance each year. Table B.1 shows summary statistics about the socio-economic demographic variables, for the general population and for the eligible.

The administrative data show receipt of social assistance (among other benefits) for each person, as well as benefits received, which household-composition-dependent threshold has been applied, any income earned, exemptions and sanctions. I use these data to calculate the take-up rate of social assistance - defined as $\mathbb{P}[\text{Take-up SA}|\text{Eligible}]$. Over the study period, the take-up-rate is around 60%, in line with Inspectie SZW (2021). I find $\mathbb{P}[\text{Take-up SA}|\text{Ineligible}] = 1\%$, suggesting low measurement-error.

3.2.2. *Mental health information:* Finally, the data contain three classes of mental health measures: take-up of mental healthcare (mental healthcare expenditures and dispensations of psychotropic medications by ATC4-code), extreme outcomes (hospitalizations with ICD-10 codes corresponding to mental health disorders, and deaths by intentional self-harm–suicides), and psychological distress (Kessler's 10), loneliness and perceived control over one's life (in a linked representative survey for $\approx$ 400,000 people in 2012 and 2016 as a repeated cross-section). Table B.2 shows summary statistics about (mental) health.

Figure I shows the prevalence of poor mental health in the Netherlands across the general population, those eligible for social assistance, and recipients. The eligible are at least $2.5\times$ more likely to suffer from poor mental health than the general population. Social assistance recipients have similar mental health to the eligible population: self-targeting to the vulnerable appears muted, although for a full analysis, see Section 4.

3.3. **Key Analysis Variables.** In the rest of the paper, I empirically analyse the take-up of social assistance heterogeneously by mental health. Throughout, I define take-up as $SA_{it} = \mathbb{1}\{i \text{ receives SA in period } t\}$. For almost all results, this will refer to a stock. My main measure of poor mental health is $\text{Poor MH}_{it} = \mathbb{1}\{i \text{ dispensed psychotropic medications in year } t\}$.

(Naik et al., 2026) shows that this is an accurate proxy for poor mental health status. In the Netherlands, usage of mental healthcare is strongly positively correlated with subjective psychological distress, and the relationship between the two does *not* depend on income. Prescriptions are done by GPs, who are the first access point to healthcare. In general, access to healthcare in

[11] As in Inspectie SZW (2021), eligibility is noisier for students (age $\leq$ 27) and people above pension-age (65) and so these groups are excluded.

the Netherlands is excellent, and people often still receive care even if they default on their premiums (Roos et al., 2021). Indeed, 0.4% of *poor* households report unmet medical needs in the Netherlands, relative to 8.5% of *all* households in the US (Danesh et al., 2024).

Of course, even in the Netherlands there will be some non-take-up of mental healthcare by people with truly poor mental health. Therefore, throughout the empirical analysis, I validate that all findings remain broadly consistent across proxies (usage of care, survey, extreme outcomes).

4. Descriptive Patterns in Take-Up by Mental Health

Figure I shows that people with poor mental health are much more likely to be eligible for social assistance and hints at weak targeting. This section explores descriptive patterns in the take-up and targeting of social assistance with respect to mental health in more depth.

4.1. **Take-up Levels: Design.** First, I focus on take-up levels following **Step 1** in the three-step identification from Section 2.2.1. I employ regression specification 4.1 to ask whether people with poor mental health take-up social assistance more or less than people with good mental health, conditional on eligibility and income (and other observables).

$$SA_{it} = \beta \cdot \text{Poor MH}_{it-1} + X'_{it-1}\theta + \varepsilon_{it} \tag{4.1}$$

X_{it} are flexible controls of household disposable income,[12] wealth, education, household composition, work status, work sector and year, age, gender and municipality fixed effects. ε_{it} is an idiosyncratic error term. β measures average targeting of social assistance towards people with poor mental health and is the coefficient of interest.

Importantly, I estimate Equation (4.1) on the *eligible* population. Higher overall take-up rates by a group could come from higher probability of being eligible, or more frequent receipt conditional on eligibility. I focus on the latter in this paper because non-take-up by ineligible individuals is not attributable to the main forces of interest - need and ordeal costs.

4.2. **Take-up Levels: Results.** Table I shows the main results using Poor $\text{MH}_{it} = \mathbb{1}\{i$ dispensed psychopharma. in year $t\}$. I find that people with poor mental health take-up social assistance only slightly more than those with good mental health. In the context of mental disorders, social assistance fails to meaningfully target those facing greater hardship on average.

[12] I include household disposable income percentile (moving average $t-4 \rightarrow t-2$) fixed effects.

The table shows estimates of β, for seven specifications of increasing saturation. Throughout, people with poor mental health have similar take-up rates to those with good mental health. After conditioning on year, age, and gender fixed effects, the difference in take-up levels between groups ranges from around -1 to +1p.p, depending on inclusion of additional controls. Estimates are statistically significant but economically small.[13] In the full specification (Column 5), people with poor mental health take-up social assistance less than 1% more than those with good mental health, holding all else constant.

Figure II presents the estimates from Equation (4.1), varying the measure of poor mental health. $\hat{\beta}$ are plotted for each mental health status variable: $\mathbb{1}\{$dispensed of psychotropic meds$\}$, $\mathbb{1}\{$positive mental healthcare costs$\}$, $\mathbb{1}\{$Hospitalized for a mental health condition$\}$, and surveyed $\mathbb{1}\{$Severe psychological distress$\}$, $\mathbb{1}\{$Severe lack of control over own life$\}$, and $\mathbb{1}\{$Severe loneliness$\}$, relative to average take-up amongst those with good mental health. Qualitatively, these estimates are broadly consistent and show a small positive difference in rate of receipt by people with poor mental health vs people with good mental health.

Of course, there is a broad spectrum of mental disorders with different severities. Using psychotropic drug dispensations to measure poor mental health provides a practical approach to distinguish these differences. I find that social assistance targeting towards people with poor mental health is decreasing in severity on average. Figure C.2 shows coefficients from a regression of social assistance receipt on dummies for *type* of psychotropic drug dispensed, and all controls. Anti-depressant dispensation is associated with a higher rate-of-receipt, whereas anti-psychotic dispensation is associated with a *lower* rate-of-receipt. Figure C.3 shows individuals in poor mental health episodes take-up social assistance more than those suffering from chronic illness.

4.3. **Differential selection on other characteristics.** Take-up levels are similar across mental health groups, conditional on observables X. This average similarity is the focus of the stylised model. This section provides complementary evidence by examining differential selection on X, the characteristics controlled for in the take-up analysis, revealing which observable traits differ between recipients and eligible non-recipients across mental health groups.

[13]How does $\hat{\beta}$ compare to other covariates? Household disposable income percentile fixed effects range from 20 to -20, age from 0 to 10, and municipality from -15 to 5. Two controls drive most variation: lagged income increases R^2 from 0.05 to 0.16, and lagged work status increases R^2 to 0.64, with $\hat{\theta}_{SA_{t-2}=1} = 42.35$, showing strong autocorrelation in take-up. The positive $\hat{\beta}$ with individual fixed effects supports this, as does differential selection in social assistance history (see Table C.3).

First, Equation (4.2) examines differential selection on *lagged* observables X, controlling for background characteristics Z.

$$X_{it-1} = \delta \cdot \text{Poor MH}_{it} + \gamma \cdot SA_{it} + \phi \cdot \text{Poor MH}_{it} \cdot SA_{it} + Z'_{it-1}\theta + \nu_{it} \tag{4.2}$$

δ captures the baseline correlation between mental health and X, γ reflects targeting along the X dimension, and ϕ is the coefficient of interest. ϕ can be interpreted as a difference-in-difference. For example, if $X =$ consumption,[14] ϕ captures how much larger is the consumption gap between recipients and non-recipients with poor mental health compared to the baseline gap for those with good mental health. I only include household disposable income controls (and a constant) in Z, to hold economic status constant.

Figure IIIa plots $\hat{\phi}$. Generally, differential selection is small, with most estimates < 0.2SD. Recipients with poor mental health are more likely to be couples who have been working rather than receiving other benefits, have fewer kids, are more educated and are able to take on some household debt. Interestingly, recipients with poor mental health show no differential selection on lagged consumption relative to those with good mental health. This is consistent with $v' \perp \varepsilon$ because if consumption shapes the marginal value of income and v' were correlated with ε, we would expect differential patterns.

4.3.1. *What do they do with the money?* Consider the related specification, Equation (4.3). Here, ϕ measures differences in future outcomes between recipients and eligible non-recipients across mental health states, *holding fixed* selection on lagged outcomes. This provides descriptive evidence on recipients' usage of social assistance benefits, and how this depends on mental health.

$$X_{it+1} = \alpha \cdot X_{it-1} + \delta \cdot \text{Poor MH}_{it} + \gamma \cdot SA_{it} + \phi \cdot \text{Poor MH}_{it} \cdot SA_{it} + Z'_{it-1}\theta + \nu_{it} \tag{4.3}$$

Figure IIIb plots $\hat{\phi}$ for selected variable traits. Estimates are generally small (all z-scored outcomes $< 0.2SD$). The exception is future mental health: among those with poor mental health, recipients go on to have 3.85p.p. (6.4% of baseline) *fewer* psychopharma dispensations than non-recipients, compared to a baseline small positive gap for those with good mental health at baseline (recipients dispensed < 1p.p. *more*). The healthcare cost gap and labour supply gap are also smaller for those with poor mental health, while the consumption gap shows no differential. These descriptive patterns are consistent with social assistance improving health and alleviating labour

[14] I construct a rudimentary proxy for consumption using income and changes in wealth, see Appendix C.1.

supply costs and borrowing constraints more for people with poor mental health. Of course, the patterns could also reflect differential selection on health and economic trajectories.

5. Barrier Screening Effects

Section 4 shows that social assistance is hardly targeted towards people with poor mental health on average, despite their vulnerability. What gives rise to this worrying finding? This section shows that people with poor mental health are disproportionately screened-out by barriers.

I examine the effects of the Participation Act, a large reform to social assistance design in the Netherlands, which increased barriers to access. The policy was announced in 2014 (with significant public discourse - see Figure D.1) and implemented in January 2015. The reform was a response to rising caseloads following the Great Financial Crisis, and it cut municipal social assistance budgets from €1.4 billion in 2010 to around €500 million by 2018 (Heekelaar, 2021). Consequently, the Participation Act intensified all obligations and incentivized municipalities to restrict inflow through the (threat of) sanctions upon non-compliance (SCP, 2019; van der Veen, 2019). For more details, see Appendix D.1.

5.1. **Identification.** I exploit the Participation Act to estimate the heterogeneous take-up response to a change in barriers by baseline mental health. Practically, the specification, Equation (5.1), is a standard Difference-in-Difference design with people poor mental health as the treatment group. The interpretation of the treatment effects is the heterogeneous effect $\frac{\partial \mathbb{P}[SA]_L}{\partial \Lambda} - \frac{\partial \mathbb{P}[SA]_H}{\partial \Lambda}$. The identification assumption is that people with poor mental health's receipt would have evolved in parallel to those with good mental health.[15]

$$SA_{it} = \alpha + \eta_i + \gamma_t + \delta_t \times \text{Poor MH}_i + X_{it}'\theta + \varepsilon_{it} \tag{5.1}$$

η_i and γ_t are individual and year fixed-effects respectively. X is a vector of time-varying controls including household disposable income, education and muncipality, household composition and sector fixed effects. δ_t are the coefficients of interest and represent the heterogeneous treatment effects of the policy by baseline mental health. Poor $\text{MH}_i = \mathbb{1}\{i$ dispensed psychotropic drugs at some point in the pre-period (2011 - 2014)$\}$. Throughout, I cluster standard-errors at the level of municipality of residence in 2013.

[15]The formal parallel trends assumption is that the receipt of social assistance by those affected by the policy would have evolved in the same way as a (purely hypothetical) control group who did not experience the policy, for every level of baseline mental health (de Chaisemartin and D'Haultfœuille, 2023; Shahn, 2023).

I estimate Equation (5.1) for eligible middle-aged couples (ages 45-65), for whom the policy represents a clean exogenous increase in barriers only. I focus on couples because the eligibility threshold for single parents was cut in 2015, incentivising reclassification as single-person households. The take-up of social assistance pre/post 2015 by younger individuals is contaminated by inflow from a youth disability program (Wajong), where people with poor mental health are likely over-represented.[16] I focus on the eligible because the take-up responses for this group can be attributed to the change in barriers, not underlying changes in eligibility rates.[17] Sensitivity analyses confirm that the findings remain robust across various specifications and assumptions, as detailed in Section 5.5.

5.2. **Main Results.** Figure IV shows the estimates $\hat{\delta}_t$ according to Equation (5.1). The groups are on parallel trends before the policy announcement, giving confidence to the identification assumption.

The Participation Act disproportionately screens out people with poor mental health. The effect starts when the Act is announced, and then is especially pronounced in 2015. The overall difference-in-difference estimate of $\frac{\partial \mathbb{P}[SA]_L}{\partial \Lambda} - \frac{\partial \mathbb{P}[SA]_H}{\partial \Lambda} \approx -1$p.p. Although small in magnitude at face value, the disproportionate screening of people with poor mental health remains almost identical when controlling for the disproportionate screening of *other* characteristics correlated with mental health. Figure D.6 shows $\hat{\delta}$ for poor mental health when controlling for year fixed effects interacted with consumption, gender, $\mathbb{1}\{\text{foreign-born}\}$ and $\mathbb{1}\{\text{physical health condition}\}$. Comparing the interacted year fixed effects shows that mental health is the dimension of vulnerability most strongly screened-out by the increase in barriers induced by the Participation Act.

5.3. **Mechanism.** Figure V shows the effects on inflow and outflow. Outflow is not mechanically zero, the estimates are close to zero and have tight confidence intervals. This shows that the main results stem exclusively from a deterrence of inflow, aligning with Cook and East (2024) who argue work requirements can screen-out individuals at the extensive margin. The disproportionate reduction in inflow for people with poor mental health (1p.p.) is around 10% of the baseline control mean (see Figure D.2).

The reduction of inflow suggests people with poor mental health are deterred by increased obligations and unpleasant interactions with the municipality, given their incentives to reduce

[16]Figure D.7 shows the results when including adults aged 35-45. Similar results even though this group is more contaminated by Wajong entrants means that the main estimates are not driven by Wajong entrants.

[17]The effect holds also for those who are 'always-eligible' (eligible throughout 2011-2020), providing confidence that the main results are not driven by eligibility churn.

caseload. Qualitative evidence from (Ministerie van SZW, 2022) supports the latter mechanism. The authors state that beneficiaries experience a "feeling of shame" and highlight a "negative tone" from the municipality where "small event[s] can have major consequences". This creates "mutual distrust" and "fear" which creates a "barrier to applying for assistance, even when the need is great".[18] Differential screening at entry, without differential outflow, means (anticipatory) psychological costs dominate ongoing compliance difficulties. This pattern aligns with SCP (2019), who find no effect of the Participation Act on transition into paid work.

5.4. **Different Mental Health Measures.** How do the results compare when mental health is measured differently or across different disorders? Figure VIa shows that the Participation Act screens-out those using anti-psychotics twice as much as those using anti-depressants. Figure VIb shows that the barrier screening effects are more pronounced when poor mental health is measured by surveyed severe psychological distress. This shows that the main estimates are a lower-bound. The people screened out by barriers to social assistance likely also find it most difficult to take-up mental healthcare even if they have poor mental health. Taking these results together implies that ordeal-screening increases with severity and is driven by disorder symptoms themselves rather than treatment.

5.5. **Robustness.** The main results of Section 5.2 are robust to several threats to identification. First, the sample consists of couples eligible for social assistance each year, a population that changes over time due to eligibility churn from income fluctuations and the inflow of individuals with youth disabilities. This raises concerns that differential take-up rates upon entry and exit drive the main result. However, Figure D.3 shows consistent results for the always-eligible population; the 25% of couples who remain eligible throughout the sampling period. The smaller point estimates for the always-eligible are unsurprising: these people are less likely to be on the margin of take-up.

Appendix D presents a formal presentation of the sample-selection issue. Consequently Figure D.4, which shows that the estimates are virtually unchanged when removing all time-varying covariates, is further evidence that eligibility flows do not drive the results.

Secondly, there could be contemporaneous policy changes which affect the take-up of social assistance heterogeneously by mental health. One threat is a reform to the structuring of long-term care (WMO) (Kromhout et al., 2018). The remit of home support for people with mental health

[18] Translated from page 8 of Ministerie van SZW (2022). See Appendix D.1 for full quote.

issues was re-assigned to municipalities starting in 2015. Figure D.8 shows the WMO reform does not drive the results.

The main results are not driven by inflow from Wajong (an income-support program for those experiencing disability shocks before age 18), which merged into the Participation Act in 2015. My sample restricts to individuals above age 45 to conservatively exclude those who might have transitioned from Wajong to social assistance. The only way this group could contaminate the sample is if they experienced a disability shock at 18, did not take-up Wajong, survived without income support until age 45, and then opted for social assistance. Figure D.7 shows that the estimates are unchanged when including adults aged 35-45, confirming that the age restriction effectively controls for the potential contamination.

Thirdly, one might worry that the observed heterogeneous treatment effects reflect pre-existing differences in take-up levels rather than a causal effect of mental health on responses to barriers. However, this concern is inconsistent with the heterogeneity by disorder severity. If anything, individuals with good mental health have lower receipt than those on anti-depressants (moderate), but higher receipt than those on anti-psychotics (severe). Despite these opposing patterns in levels, both moderate and severe mental health groups are disproportionately screened out after 2015 relative to the good mental health group.

Fourthly, the groups are defined based on pre-period dispensations, so we might worry that the $\hat{\delta}$'s are capturing the effect of mental health treatment on social assistance receipt. However, Figure VIb shows that results when defining poor mental health based on self-reported symptoms, not prescriptions, are even stronger. In any case, Figure D.9 shows that poor mental health is reasonably well approximated by a fixed type for those eligible for social assistance.

A final related concern might be that the results could reflect positive effects on social assistance take-up following a mental health shock. However, Figure D.10 shows that even when mental health is defined after the policy, there is a noticeable drop in $\hat{\beta}$ in 2014 that persists over time. Additionally, Figure D.11 presents similar findings when poor mental health is defined as continuous psychopharma dispensations in all years from 2011 to 2020 (compared to none in any year). This group likely suffers from chronic mental illness, reinforcing confidence that the main results are not merely capturing the effects of a one-time mental health shock.

6. Benefit Take-up Effects

In the final empirical part of the paper, I estimate take-up responses to exogenous variation in benefits to provide revealed-preference evidence on the redistributive value of transferring money to those with poor mental health.

I leverage quasi-experimental variation in the benefit-level by exploiting the kinked benefits schedule as a function of income with a regression kink design (RKD) as in Card et al. (2015). The statutory benefits schedule is displayed in Figure VIIa. Before diving into the details of identification, Figure VIIb shows graphical evidence of the benefit take-up effects by poor vs good mental health. The figure indicates that people suffering from mental disorders take-up *more* in response to increasing benefit-levels than those with good mental health. I plot take-up within slices of *monthly* "eligibility income".[19] The take-up functions diverge starkly at the threshold for poor vs good mental health, with the former displaying almost twice the slope change.

People with income above the threshold take-up primarily due to measurement error: I do not observe the income concept used to determine eligibility (eligibility income). Therefore, I need to construct it. This Y_{calc} differs from Y_{true} because (a) some income information (e.g. from other benefits) is only recorded yearly, yet eligibility determined monthly. Unemployment spells are imputed. (b) Y is aggregated to the family level. Families are 1 or 2 adults (+ kids) who live together and share costs. The latter is unobservable.

To minimise attenuation from measurement error, I focus on single employees.[20] The data contain monthly income information for employees–minimising error due to (a) and singles are immune from issues in (b). Seeing as Y_{true} is (mostly) observed for the selected sample: recipients, Figure E.1 shows a histogram of $Y_{\text{true}} - Y_{\text{calc}}$ for the analysis population. Y_{true} is negatively selected for recipients, so we expect the distribution to be left-skewed. Measurement error has significant mass around 0. Both mean and median are small (-€51, -€13 respectively) and statistically indistinguishable from 0.

[19] Granular analysis is critical - hence the switch to monthly data. We can reconcile the findings in Figure VIIb with the small difference in average take-up levels estimated in Section 4 by recognizing that the overall results are largely driven by the 75% of eligible individuals who do not work. The RKD, however, is a LATE capture effects locally around the eligibility threshold.

[20] Details of the estimation are in Appendix E. Around the threshold, couples are significantly mismeasured because I cannot observe which adults live together as part of a family and which don't. This does not drive the barrier screening effects. Figure D.12 shows that the results remain the same when focusing on individuals away from the threshold. Unfortunately, this does mean that internal validity concerns restrict the samples differently for barrier screening and benefit take-up effects. This does not affect results about *relative* need and cost by mental health, as discussed in Section 7.2.3.

6.1. **Identification.**

6.1.1. *Theory.* Figure VIIb measures $\frac{d\mathbb{P}[SA]}{dy}$ for income y. In order to retrieve the take-up response $\frac{d\mathbb{P}[SA]}{dB}$, we need to re-scale by $1/\frac{dB}{dy}$. The statutory benefits schedule would imply $\frac{dB}{dy} = -1$ below threshold, and 0 above.

There is a challenge: municipalities can deviate from the policy formula through income exemptions - some or all of y is ignored when calculating B (Ministerie van SZW, 2015). Selection into take-up based on exemptions implies ex-post observed benefits conditional on receipt differ from the ex-ante schedule (i.e. the expected benefits a potential applicant is eligible for conditional on their income). Appendix E.5 sets out my theoretical approach to impute the ex-ante benefits schedule accounting for exemptions using Bayes Rule. Figure E.7 shows the results.

The imputation process is not perfect: it measures the ex-ante benefits schedule with error. Let B^* be the imputed (mis-measured) B: $B^* := B + U_B$. As discussed above, Y is also measured with error: $Y^* := Y + U_Y$. Therefore, I use a fuzzy RKD specification (Card et al., 2015). Proposition E.1 shows that a fuzzy RKD estimates a weighted average of marginal effects of B on $\mathbb{P}[SA]$.

6.1.2. *Estimation.* I estimate a standard fuzzy RKD specification, using local linear regression. I use a Calonico et al. (2014) robust bandwidth of €60, estimated separately for people who are (/not) dispensed psychotropic drugs in the year previously (poor/good mental health, respectively). The IV estimate $\frac{\hat{\beta}_1}{\hat{\delta}_1}$ measures $\frac{\partial \mathbb{P}[SA|Y=\bar{y}]}{\partial B}$. Standard-errors are clustered at the municipality level. See Appendix E.3 for more details on the estimation.

$$SA_{it} = \alpha + \beta_0 \cdot (y^*_{it} - \bar{y}_i) + \beta_1 \cdot \min\{y^*_{it} - \bar{y}_i, 0\} + \varepsilon_{it} \qquad \textbf{(Reduced Form)}$$

$$B^*_{it} = \gamma + \delta_0 \cdot (y^*_{it} - \bar{y}_i) + \delta_1 \cdot \min\{y^*_{it} - \bar{y}_i, 0\} + \varrho_{it} \qquad \textbf{(First Stage)}$$

Intuitively, the fact that the first-stage is also estimated on the mis-measured running variable y^*_{it} "accounts" for measurement-error as in Card et al. (2015).

Support for Identification Assumptions: The key identification assumption is that there is no manipulation of income around the threshold. Figure VIII shows no evidence for strategic income targeting around the eligibility threshold: McCrary (2008) tests with seventh-order polynomials show no statistically significant bunching. Although the threshold coincides with the full-time monthly minimum wage, the sample works much less than full-time on average (around 100 hours per month) and income used for eligibility does not only come from labour. Adjustment frictions are a likely reason for the lack of bunching (Kleven, 2016).

6.2. **Main Results.** First, I pool people with good and poor mental health together. Figure E.13 shows that people react significantly to the quasi-experimental variation in benefit-level. I estimate $\hat{\beta}_1 = -0.0258$ which translates to take-up increasing by $\approx 2.6p.p.$ for a €100 increase in the benefit level.

People with mental disorders have a two-times larger take-up response to a change in benefits than those with good mental health. The results are shown in Figure IX. I estimate $(\hat{\beta}_{1H}, \hat{\beta}_{1L}) = (-0.0218, -0.0508)$.[21] Measurement error is uncorrelated with mental health status - there is no statistically distinguishable difference in the slope above the threshold between good and poor mental health. Using the first stage in Figure E.14 to re-scale the above reduced-form and account for measurement error, we obtain IV estimates (and associated confidence intervals):

$$\text{Estimate of } \frac{\partial \mathbb{P}[SA|Y = \bar{y}]}{\partial B} = \frac{\hat{\beta}_1}{\hat{\delta}_1} = \begin{cases} \underset{[0.0080,0.0374]}{0.0227} \text{ p.p.} & \text{for Good MH} \\ \underset{[0.0164,0.0842]}{0.0503} \text{ p.p.} & \text{for Poor MH} \end{cases} \tag{6.1}$$

A higher benefit take-up elasticity demonstrates that the behaviour of people with poor mental health is more sensitive to benefit-levels, suggesting this money is more decision-relevant for them and pointing to social welfare gains from transferring €1 from those with good mental health to those with poor mental health. The RKD mechanically controls for income because it estimates a LATE at the eligibility threshold for both groups.

Translating these to take-up elasticities with respect to changes in benefits yields 0.16, 0.38 for good and poor mental health, respectively.[22] Although we have few recent estimates of benefit take-up elasticities of social insurance, it is worth mentioning that the elasticity for good mental health is at the lower-end of the range of previous estimates (Krueger and Meyer, 2002; McGarry, 1996), whereas the elasticity for poor mental health lies within-range. The full set of reduced-form and IV estimates (with and without controls) is contained in Table E.1.

[21] Whilst these estimates are somewhat noisy, recall that the Calonico et al. (2014) robust bandwidth does not take into account measurement-error, nor the efficiency of estimating *heterogeneous* treatment effects across groups. Indeed, optimal bandwidth selection in heterogeneous RDDs is an area of active research (Calonico et al., 2025). Here, the zoomed-out version in Figure VIIb gives us confidence that the take-up response is indeed twice as large for those with poor mental health. Moreover, Figures E.23 and E.24 confirm that for less extreme restrictions to the bandwidth, the estimates are similar, but more precise.

[22] $\mathbb{E}[B|Y = \bar{y}, \text{Good MH}] = €42.39$, $\mathbb{E}[B|Y = \bar{y}, \text{Poor MH}] = €48.98$, $\mathbb{P}[SA|Y = \bar{y}, \text{Good MH}] = 6.13\%$, $\mathbb{P}[SA|Y = \bar{y}, \text{Poor MH}] = 6.46\%$

6.2.1. *Robustness.* I assess the credibility of the design with standard robustness analyses whose results are described in Appendix E.7. Figures E.20 and E.21 show no statistical evidence of selection along observable characteristics around the kink. Indeed, Table E.1 shows that the addition of a rich set of covariates does not meaningfully affect the results. Figure E.22 displays a permutation test (Ganong and Jäger, 2018), and shows no evidence for worrying non-linearities. Figure E.23 and Figure E.24 explore sensitivity of the results to different bandwidths. Estimates are quite robust to lower bandwidths overall, and point estimates do not vary much in the heterogeneous case despite the confidence intervals overlapping with lower bandwidths.

6.2.2. *Different Mental Health Measures.* How do the results compare when mental health is measured differently or across different disorders? Figure E.15 and Figure E.16 show the take-up response to a change in benefits estimated for those dispensed anti-depressants and anti-psychotics, respectively, in each case relative to those with good mental health. The IV estimates are $\frac{\partial \mathbb{P}[SA]_{\text{Dep.}}}{\partial B} = 0.0704$, $\frac{\partial \mathbb{P}[SA]_{\text{Psycho.}}}{\partial B} = 0.120$, for anti-depressants and anti-psychotics respectively. Therefore, benefit take-up effect magnitudes increase with severity.

Moreover, Figure E.17 shows that the larger take-up response for those with poor mental health is present when poor mental health is alternatively measured by severe psychological distress reported in the survey. Therefore, responsiveness of take-up to changes in benefits seems to reflect general psychological distress, rather than treatment. Figure E.18 demonstrates that the main finding holds when examining inflows rather than receipt, confirming the elasticity reflects active take-up decisions in response to current benefit levels rather than persistence from past decisions.

6.2.3. *Mechanisms.* Why do people with poor mental health respond more strongly to benefit changes, conditional on income? Mental health correlates with other dimensions of heterogeneity which could independently affect benefit responsiveness. To isolate the role of mental health, I estimate whether the differential benefit take-up effect persists when controlling for interactions between the kink term and these other characteristics, mirroring Figure D.6 for barriers. Specifically, I estimate:

$$\begin{aligned} SA_{it} = &\alpha + \beta_0 \cdot (y^*_{it} - \bar{y}_i) + \beta_1 \cdot \min\{y^*_{it} - \bar{y}_i, 0\} \\ &+ \sum_k \delta_k \cdot Z_{kit} \\ &+ \sum_j \gamma_j \cdot X_{ji} \times \min\{y^*_{it} - \bar{y}_i, 0\} + \varepsilon_{it} \end{aligned} \tag{6.2}$$

where X_{ji} includes poor mental health, gender, foreign-born status, physical chronic conditions, and household consumption (all interacted with the kink term), and Z_{kit} includes non-interacted controls for household composition, education, age, year, month, sector, municipality, and wealth. The key parameter is $\gamma_{\text{Poor MH}}$, which captures the mental health effect conditional on other interacted characteristics. If the Calonico et al. (2014) robust bandwidth is not designed for heterogeneous RKDs, it is certainly not designed for comparing heterogeneous RKDs across multiple dimensions of heterogeneity. I therefore also present results using a wider (€180) bandwidth.

Table E.2 shows poor mental health is the dominant dimension of heterogeneity. At the narrow bandwidth, the coefficient barely changes from the specification without covariate interactions, and is $2.9\times$ larger than the gender effect, $2.2\times$ the foreign-born effect and $> 10\times$ the effect for physical conditions and consumption. At the wide bandwidth, poor mental health again generates the strongest response and the other correlated dimensions of heterogeneity explain only 12% of the disproportionate sensitivity to benefit-level of poor vs good mental health.

7. QUANTIFYING TARGETING EFFECTIVENESS

In this section, I calibrate the simple model from Section 2 using the empirical results about take-up behaviour from Sections 4 to 6. I quantify the key behavioural primitives need (v') and cost (κ') and describe how these estimates translate into welfare effects under different assumptions, with particular attention to the assumption of revealed-preference (Section 7.2.2). Section 7.2.3 and appendix F.3 contain other robustness analyses. Table II summarises all findings.

7.1. **Identifying Need and Cost.** For the calibration, I assume $\theta \in \{L, H\}$: mental health is either poor or good. The tax rate $\tau \approx 37\%$.

7.1.1. *Three-step Identification.* I employ the three-step identification method set out in Section 2.2.1. First, Section 4.2 shows no meaningful difference in average take-up levels conditional on eligibility between poor and good mental health. Therefore, I apply the special case of **Step 1**, where equalized take-up levels implies equalized likelihood of being at the margin of take-up.

$$f_\varepsilon(v_L - \kappa_L) = f_\varepsilon(v_H - \kappa_H)$$

This result reflects the fact that *average* value net of cost seems to be roughly the same across mental health states. However, this does not necessarily pin down *marginal* value (need) - nor does it separate need from cost.

First, I normalize $v'_H = 1$. As discussed in **Step 2**, this effectively scales need by the willingness-to-pay for €1 amongst people with good mental health. Moreover, it means that the benefit take-up response for people with good mental health measures $f_\varepsilon(v_H - \kappa_H)$. To match the theory, I re-scale the response estimated in Section 6.2 by $(1-\tau)$, the net-of-tax rate, because in the theory B is understood as gross benefits, whereas the regression kink design estimates responsiveness to net benefit level. I estimate $\frac{\partial \mathbb{P}[SA]_H}{\partial B} = f_\varepsilon(v_H - \kappa_H) = \underbrace{0.63}_{1-\tau} \times \underbrace{0.000227}_{\text{Estimate from RKD}}$.

Need: I apply **Step 2** and divide the benefit take-up response for people with poor mental health by the response for good mental health. The above implies $\frac{f_\varepsilon(v_L-\kappa_L)}{f_\varepsilon(v_H-\kappa_H)} = 1$. This, combined with $v'_H = 1$ shows that need for benefits for people with poor mental health is revealed as the *relative* benefit take-up response for this group. I estimate $\frac{\partial \mathbb{P}[SA]_L}{\partial B} = 0.63 \times 0.000503$, which therefore implies $v'_L = 2.22$.[23]

Cost: Finally, I use the difference-in-differences results of Section 5 to calibrate $\kappa'_\theta(\Lambda)$. Because barriers operate exclusively at the inflow margin (see Figure V), I calibrate barrier screening effects using inflow responses. For people with good mental health, I use the descriptive pre/post change in inflow: $\frac{\partial \mathbb{P}[SA]_H}{\partial \Lambda} \approx -0.016$ (Figure D.2). The DiD identifies the differential $\frac{\partial \mathbb{P}[SA]_L}{\partial \Lambda} - \frac{\partial \mathbb{P}[SA]_H}{\partial \Lambda} = -0.009$, yielding $\frac{\partial \mathbb{P}[SA]_L}{\partial \Lambda} \approx -0.025$.

Applying **Step 3** with $f_\varepsilon(v_L - \kappa_L) = f_\varepsilon(v_H - \kappa_H) = 0.63 \times 0.000227$ implies $\kappa'_H = 110$ and $\kappa'_L = 173$. The descriptive inflow anchor is likely an overestimate (due to secular trends and slower-moving stock), making the $\kappa'_L/\kappa'_H = 1.57$ ratio estimate *conservative*. Indeed, Table II confirms that the key finding (people with poor mental health are higher need and have higher ordeal costs) is robust to variation in $\left|\frac{\partial \mathbb{P}[SA]_H}{\partial \Lambda}\right|$, and is strengthened if this response is weaker.

7.1.2. *Discussion.* These estimates imply people with poor mental health have a $2.2\times$ higher marginal value of benefits (need) versus those with good mental health, conditional on income. This implies a strong redistributive motive towards people with poor mental health. Thus equal take-up rates mask poor targeting. Individuals suffering from mental disorders experience a 57% higher cost of overcoming ordeals. Therefore, the neediest people, to whom the government has the strongest motive to redistribute money, find it most difficult to access assistance, undermining barriers as an effective policy tool to target social assistance.

Mechanism: What is driving the higher need and cost among people with poor mental health? On the cost side, no differential outflow (Figure V) means that psychological costs such as fear,

[23] In this section, I use the regression kink design estimates of the take-up response to a change in benefits with the Calonico et al. (2014) robust bandwidth of €60. In Appendix F I explore how welfare consequences change when alternatively using the magnitudes estimated with a wider bandwidth as in Figure VIIb.

shame and stress from (anticipated) unpleasant interactions with municipal workers likely underpin a large share of the disproportionate difficulty overcoming barriers in this context. This accords with psychological work that shows mental disorders are associated with greater avoidance of potentially adverse experiences (Kashdan et al., 2006) and difficulty planning and navigating complex bureaucratic settings due to executive function deficits (Millan et al., 2012).

On the need side, the a priori relationship between the marginal value of €1 and mental health is ambiguous: anhedonia may reduce the value of consumption (Der-Avakian and Markou, 2012; Treadway and Zald, 2013), while functional impairments could make income support more important. My findings support the latter. This could reflect anticipated improvements in mental health from cash transfers (Haushofer et al., 2020). Additionally, functional impairments from poor mental health exacerbate everyday stressors (Kessler et al., 2003; Evans et al., 2014). Figure IIIb is consistent with social assistance improving mental health and compensating for impaired functioning by releasing labour supply and credit constraints. This aligns with Sen's "capabilities approach" (Sen, 1999, 2008): people with poor mental health need more resources to get by. If anything, perceived need likely underestimates true need given pessimism characterises depression (Alloy and Ahrens, 1987), one of the most common mental disorders.

On both sides, heterogeneous take-up responses scale with disorder severity (anti-psychotics vs. anti-depressants), are amplified when measured through surveyed psychological distress rather than treatment and are not driven by mental health proxying for other characteristics. Therefore, both higher need and higher cost appear to stem from a common component of psychological distress across disorders. Impairments in cognitive function and emotion regulation are central mechanisms underlying this common component (Bierman et al., 2008; Hammar and Årdal, 2009; Hyman et al., 2006; Gross and Muñoz, 1995). My results align with a model in which these deficits both exacerbate the challenges of navigating access barriers *and* overcoming everyday stressors, enough to raise the marginal value of income support.

7.2. **Quantifying Welfare Effects.** I calibrate welfare effects of marginal changes in benefits and barriers using the estimates from previous sections, taking the Section 2 model at face value. The conditional homogeneity assumption, that primitives are constant within θ, limits the analysis to welfare consequences of targeting with respect to mental health. However, mental health is the dimension of heterogeneity most jointly sensitive to both barriers (Figure D.6) and benefits (Table E.2).

In the data, $\mu(L) = 0.25$. I set $\mathbb{P}[SA]_L = \mathbb{P}[SA]_H = 0.625$. The heterogeneous monthly net fiscal externalities $FE_\theta = \tau(y^{SA=0} - y^{SA=1}) + (1-\tau)B$ are:

$$FE_L = 0.37 \times (€512.22 - €331.27) + (1 - 0.37) \times €972.22 = €679.45 \tag{7.1}$$

$$FE_H = 0.37 \times (€574.29 - €390.95) + (1 - 0.37) \times €916.29 = €645.09 \tag{7.2}$$

The fiscal externality is larger for poor mental health because they earn less when on social assistance. I estimate take-up levels and responses on the eligible population but extrapolate to the general population for welfare calculations (Proposition F.1). The intuition is that the government budget constraint reflects that ineligible individuals fund benefits for recipients. I rescale eligible take-up responses by "effective eligibility", a constant for each θ that adjusts for baseline incomplete take-up and the possibility that some ineligible individuals are on the margin of take-up (just indifferent between earning above vs. below the threshold to qualify for social assistance).

7.2.1. *MVPFs of Ordeals and Benefits under Revealed Preference.* I calculate welfare effects using using the redistributive marginal value of public funds (MVPF) formulae derived in Appendix A.1. I accommodate the direct redistributive motive towards people with poor mental health evidenced by the previous sections by setting *utilitarian* $\lambda_\theta = v'_\theta(B)$. This approach incorporates €1 being worth more to people with poor mental health and effectively writes all quantities in units of the people with good mental health's willingness-to-pay. Proposition A.1 derives the MVPF of an increase in barriers ($d\Lambda$) as:

$$\begin{aligned} MVPF_{d\Lambda} &= \frac{-\int \lambda_\theta \cdot \mathbb{P}[SA]_\theta \cdot \frac{\kappa'_\theta(\Lambda)}{v'_\theta(B)}\, d\mu}{\int FE_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda} d\mu} \\ &= \frac{-2.2 \times 0.625 \times \frac{173}{2.2} \times 0.25 - 1 \times 0.625 \times \frac{110}{1} \times 0.75}{679.45 \times 0.25 \times (-0.023) \times \frac{1}{1-0.625\times 0.907} + 645.09 \times 0.75 \times (-0.01568) \times \frac{1}{1-0.625\times 0.954}} \\ &= 2.75 \end{aligned}$$

An $MVPF_{d\Lambda}$ of 2.75 means ordeals impose a direct cost of €2.75 on infra-marginals (weighted by need) for every €1 saved by the government. $MVPF_{d\Lambda} \gg 1$ means that $d\Lambda$ is a costly way to raise government revenue. The MVPF for a change in benefits (dB) is:

$$
\begin{aligned}
MVPF_{dB} &= \frac{\int \lambda_\theta \cdot \mathbb{P}[SA]_\theta \, d\mu}{(1-\tau)\cdot \int \mathbb{P}[SA]_\theta \, d\mu + \int FE_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial B} d\mu} \\
&= \frac{2.2 \times 0.625 \times 0.25 + 1 \times 0.625 \times 0.75}{0.63 \times (0.625 \times 0.25 + 0.625 \times 0.75) + 679.45 \times 0.25 \times \frac{0.63 \times 0.000503}{1-0.625\times 0.907} + 645.09 \times 0.75 \times \frac{0.63 \times 0.000227}{1-0.625 \times 0.954}} \\
&= 1.18
\end{aligned}
$$

$MVPF_{dB} > 1$ reflects the redistributive motive: beneficiaries' welfare gain (weighted by need) exceeds government cost. Comparing $MVPF_{d\Lambda} = 2.75$ to $MVPF_{dB} = 1.19$ implying that reducing barriers is $2.3\times$ more effective than increasing benefits. In other words, the government would be willing to reduce benefit levels to finance lower barriers. Note that poor targeting through *barriers* limits the effectiveness of increasing *benefits*. Since take-up is similar across mental health states, dB redistributes to all infra-marginals regardless of need, limiting its cost-effectiveness.

Table II shows the behavioural primitives υ'_θ, κ'_θ written as ratios of poor to good mental health, the MVPFs, and the welfare effects of the composite policy $\frac{dW}{d\Lambda}$ overall and by mental health.[24] Increasing barriers to increase benefits is 19% worse for people with poor mental health than good. While the conditional homogeneity assumption means these magnitudes should be interpreted cautiously (targeting within mental health types is *not* captured), I include them to support the qualitative finding that mental disorders exacerbate the inefficiency of barriers due to being disproportionately screened out despite high need and similar fiscal costs.

7.2.2. *The Role of Bias.* As discussed in Section 7.1.2, it is possible that a substantial share of the revealed ordeal costs κ' is driven by behavioural bias, and therefore should be excluded when calculating welfare effects.[25] Therefore, I consider the case where individuals overstate barrier costs: a share ψ of revealed costs represent true welfare burdens, while $1-\psi$ reflect pure hassle costs that affect behaviour but are not normatively relevant (Naik and Reck, 2025). Appendix A.3 derives the MVPF formulas in this case, and Table II shows welfare effects varying ψ.

If only $\psi = 35\%$ (or less) of perceived ordeal costs represent true welfare costs, this reverses the ranking $MVPF_{d\Lambda} > MVPF_{dB}$. This arises because $MVPF_{d\Lambda}$ is unsurprisingly sensitive to ψ. If ψ is low, increasing barriers imposes little deadweight cost to infra-marginals and may help with

[24] $\frac{dW_\theta}{d\Lambda} = \lambda_\theta \, \mathbb{P}[SA]_\theta \left[\underbrace{\upsilon'_\theta(B)}_{Need} \cdot \frac{dB}{d\Lambda} - \underbrace{\kappa'_\theta(\Lambda)}_{Cost} \right]$ i.e. the welfare effect per-type θ. These estimates treat other observables as fixed.

[25] While it is possible that υ' also may be driven by bias, the most likely scenario seems to be that the true $\upsilon' > \hat{\upsilon}'$ for people with poor mental health, which only strengthens the policy conclusions.

targeting. $MVPF_{dB}$ is robust to ψ because its cost-effectiveness is driven by how well-targeted social assistance is at baseline (do high v' types have high $\mathbb{P}[SA]$?), regardless of whether targeting is driven by true ordeal costs or behavioural frictions.

7.2.3. *Relaxing other Assumptions.* Table II displays how relaxing other modelling assumptions affects welfare conclusions. Overall, $MVPF_{dB}$ is quite robust to modelling assumptions, $MVPF_{d\Lambda}$ is slightly sensitive, and absolute welfare effects $\frac{dW}{d\Lambda}$ are quite sensitive. Throughout, the ratio $MVPF_{d\Lambda}/MVPF_{dB}$ remains robustly above 1 (excluding cases where perceived costs substantially overstate true welfare costs, discussed in Section 7.2.2). I conclude by highlighting two particularly important assumptions: (i) ε independent from θ, and (ii) same compliers across instruments. Appendix F.3 provides the full details.

(i) To test whether the distribution of unobservables differs by mental health, I impose a parametric form: $\varepsilon_\theta \sim N(\mu_\theta, \sigma_\theta^2)$ and estimate a heteroskedastic probit with variance equation $\ln(\sigma_i) = \delta_0 + \delta_1 \cdot \mathbb{1}\{\text{Poor MH}\}_i$. The model identifies σ_L/σ_H by examining how observables predict take-up across groups: noisier predictions for poor mental health reveal higher variance in unobservables. My preferred specification yields $\hat{\sigma}_L/\hat{\sigma}_H = 1.039$. Greater dispersion in unobservables means fewer people are concentrated near the threshold of take-up. Since we observe larger take-up responses for poor vs good mental health despite having fewer marginals, the marginal individuals must have higher need and cost to generate the same aggregate response. Therefore, the baseline estimates slightly understate true differences: adjusted estimates are $v'_L/v'_H = 2.31$ and $\kappa'_L/\kappa'_H = 1.63$, strengthening the positive correlation between need and cost. Full diagnostics appear in Appendix F.3.1.

(ii) For internal validity, I focus on different subsamples in Section 5 and Section 6, calling into question the extent to which marginal take-up responses can be compared. However, **Step 2** of the 3-step identification Section 2.2.1 can be applied separately to the two policy designs. Therefore, people with poor mental health having a relatively $2\times$ higher need and a relatively 57% higher cost does not rely on assumption (ii). (ii) is relevant for the comparison between need and cost *within-individuals*. I show in Appendix F.3.3 that relaxing (ii) through additional structure on the take-up equation implies that $f_\theta^{d\Lambda} < f_\theta^{dB}$. This means ordeal costs are a *lower-bound* and again pushes further in favour of reducing barriers over increasing benefits.

8. Conclusion

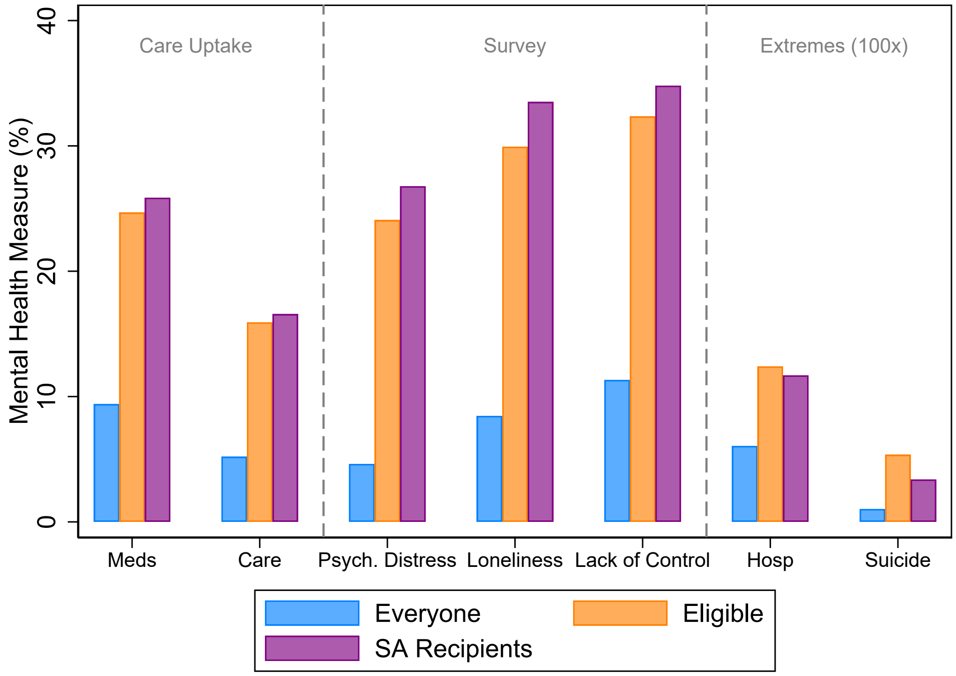
This paper shows that people with poor mental health are inefficiently excluded from social assistance despite high need. Mental disorders affect many key stages of the targeting process. They represent welfare-relevant hardship: controlling for income, people with poor mental health need benefits $2.2\times$ more than those with good mental health. They also correlate with economic vulnerability: those with poor mental health are $3\times$ overrepresented among the eligible. They impair decision-making: ordeal costs are 57% higher, driven primarily by psychological rather than compliance costs. And they undermine targeting: despite needing benefits twice as much, people with poor mental health take up assistance at the same rate as those without. Combining theory and empirics reveals that when evaluating targeting by mental health, reducing barriers would deliver twice the welfare gain as increasing benefits.

Throughout, I have assumed a static model where mental health does not respond endogenously to policy. However, barriers likely worsen mental health directly (Brewer et al., 2022), and descriptive evidence from Section 4 suggests social assistance may improve mental health. This raises the concern that barriers screen out those who would benefit most from the mental health improvements, creating a psychological poverty trap (Haushofer, 2019; Ridley et al., 2020). Work in progress calibrates a dynamic structural model of evolving mental health to quantify these feedback effects, which would amplify the welfare costs of barriers.

The mechanisms linking mental health and policy highlighted in this paper likely operate across many programs. A particularly troubling case is mental healthcare itself. Many people with mental health diagnoses forgo recommended psychotherapy, with non-monetary barriers playing a central role (Cronin et al., 2024). My framework can characterise the welfare consequences of those suffering from mental disorders being screened-out of the very treatments designed to improve their condition.

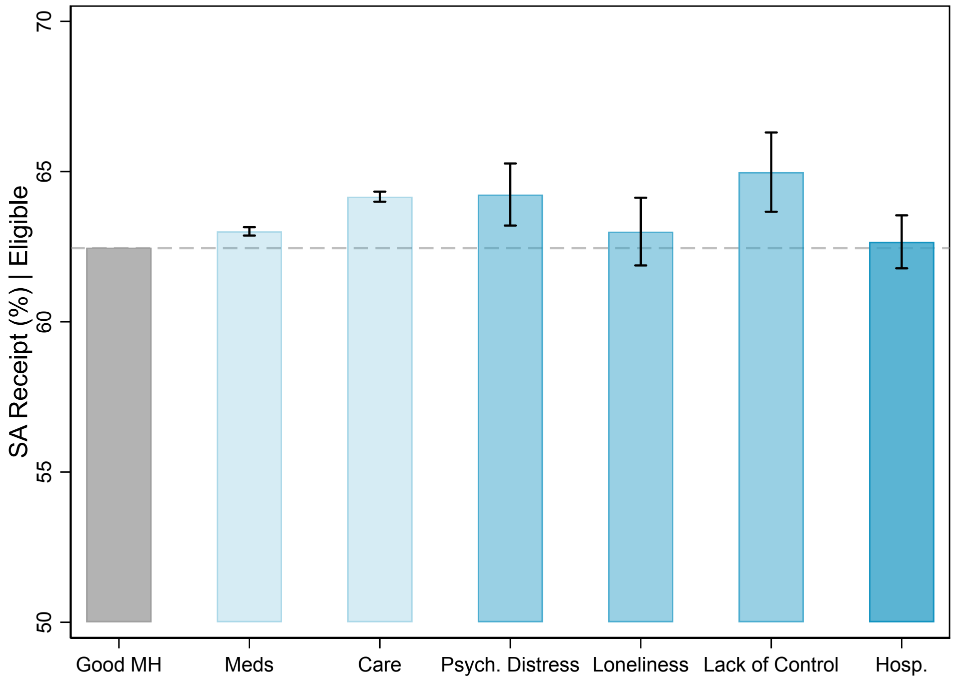
FIGURES

FIGURE I. Prevalence of poor mental health in the Netherlands.



Notes: This graph shows raw means of the seven mental health measures, across three different populations. All measures are in terms of percentages and are probabilities of the following: dispensed psychotropic medications, > 0 mental healthcare spending, surveyed severe psychological distress, surveyed severe loneliness, surveyed severe perceived lack of control over own life, hospitalisation due to a mental health condition, suicide. For the last two (extreme) outcomes, the probabilities are artificially inflated by $100\times$. The three populations are: everyone in the data, those eligible for social assistance, and the social assistance recipients, from 2011-2020 in each case.

FIGURE II. Limited targeting towards people with poor mental health (across measures) on average

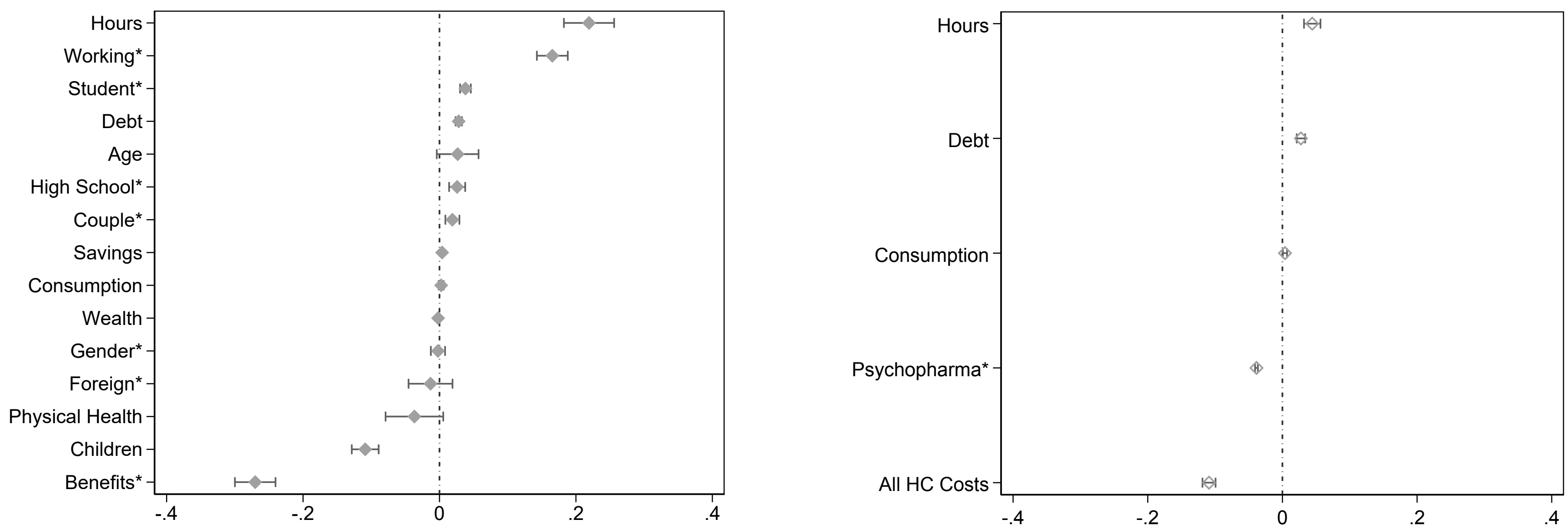


Notes: Coefficients of social assistance take-up regressed on mental health status–indicators of: psychotropic drugs, mental healthcare, severe surveyed psychological distress/loneliness/lack of control over own life, or mental health hospitalisation. Point estimates added to the control mean, with 95% confidence intervals. Lagged controls include household disposable income, wealth, education, work status, household composition, municipality, year, age, sector fixed effects, physical health, and benefits schedule. Eligible population from 2011 to 2020. Standard errors clustered at the municipality level.

FIGURE III. Differential selection on other observables by mental health

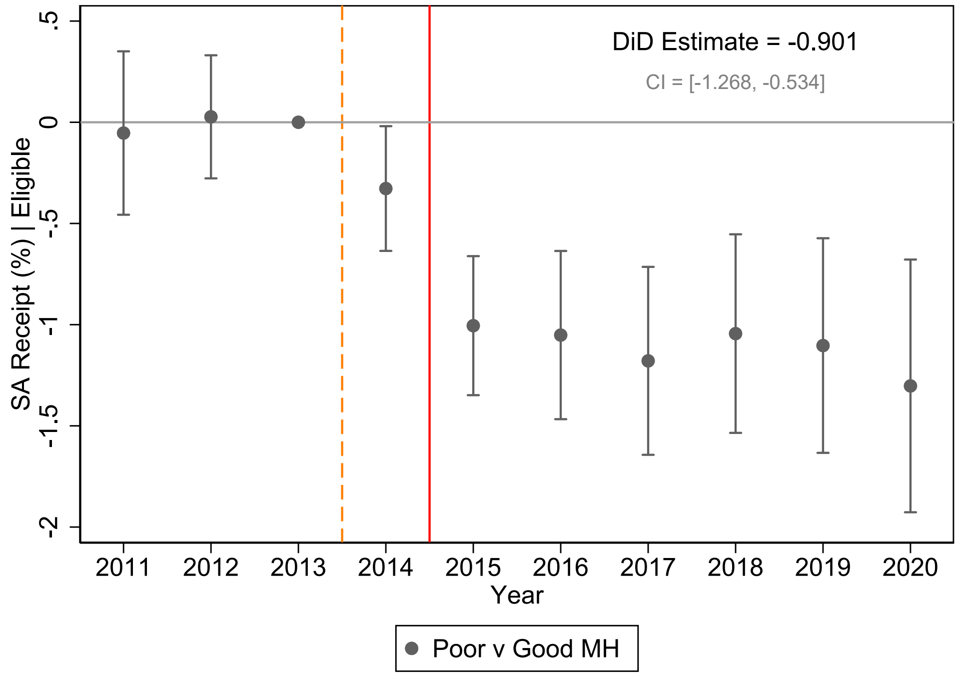
(A) Lagged observables

(B) Future observables | lagged observables



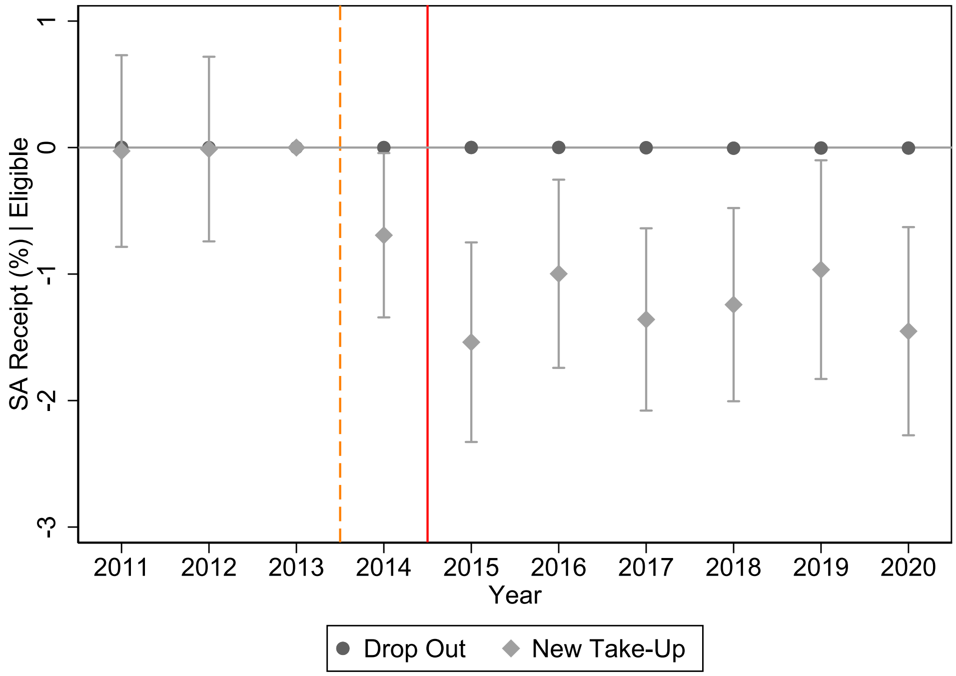
Notes: Figure IIIa shows estimates $\hat{\phi}$ from Specification 4.2 for different covariates X. Figure IIIb shows estimates $\hat{\phi}$ from Specification 4.3 for different covariates X. For each X, the outcome is z-scored apart from those X's labelled with a "*" which are discrete variables. Therefore, $\hat{\phi}$ is measured in percentage points and standard deviations for X's with/without "*"s, respectively. Point estimates are shown, alongside 95% confidence intervals with Bonferroni corrections for multiple hypothesis testing. Eligible population from 2011 to 2020. Standard errors clustered at the municipality level. See Table C.3 and Table C.4 for estimates and more information on outcomes.

FIGURE IV. Ordeals disproportionately screen out people with poor mental health



Notes: Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in pre-period. Controls include individual fixed effects, household disposable income, education and muncipality, hh composition and sector fixed effects. The TWFE estimate $\hat{\delta}$ in the regression $SA_{it} = \alpha + \eta_i + \gamma_t + \delta \cdot \mathbb{1}\{t \geq 2013\} \times \text{Poor MH}_i + X_{it}'\theta + \varepsilon_{it}$ is also shown. Standard-errors are clustered at the level of municipality of residence in 2013. Coefficient estimates are contained in Table D.2.

FIGURE V. Ordeals disproportionately deter inflow by people with poor mental health

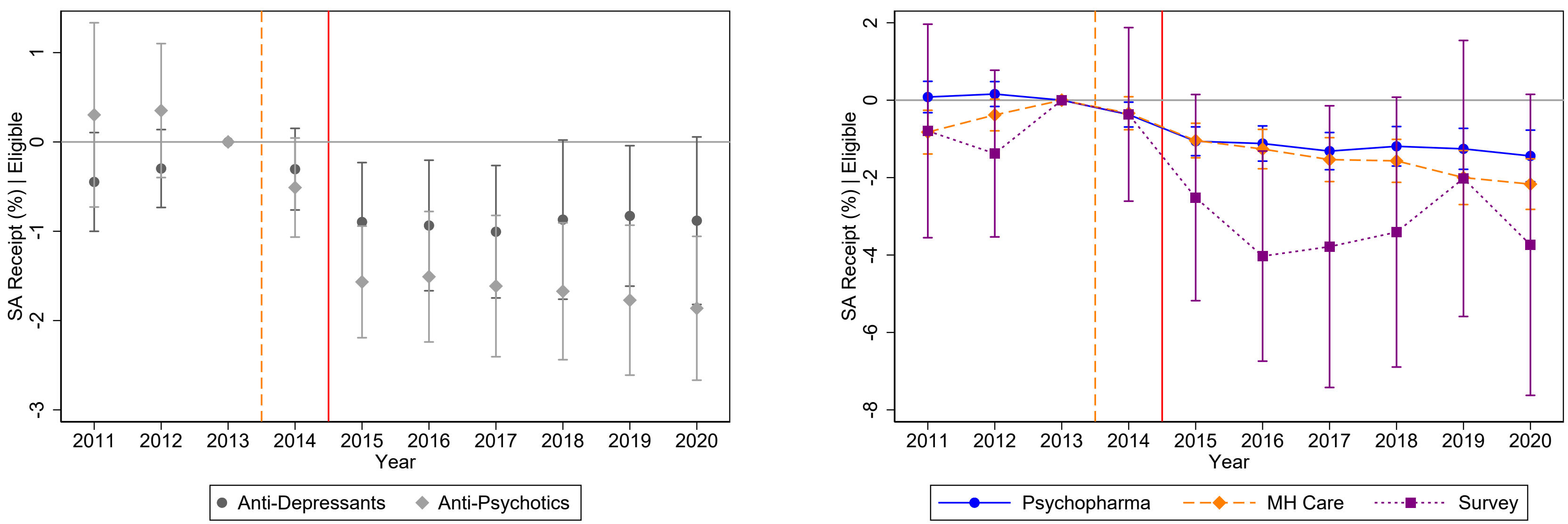


Notes: Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. Here, I split by inflow ($\mathbb{P}[SA_t = 1|\text{Eligible}_t = 1, SA_{t-1} = 0]$), and drop-out ($\mathbb{P}[SA_t = 0|\text{Eligible}_t = 1, SA_{t-1} = 1]$). The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in pre-period. Controls include individual fixed effects, household disposable income, education and muncipality, hh composition and sector fixed effects. Standard-errors are clustered at the level of municipality of residence in 2013. Coefficient estimates are contained in Table D.2.

FIGURE VI. Barrier screening effects across different mental health measures

(A) Effects intensify with severity

(B) Main results are lower bounds

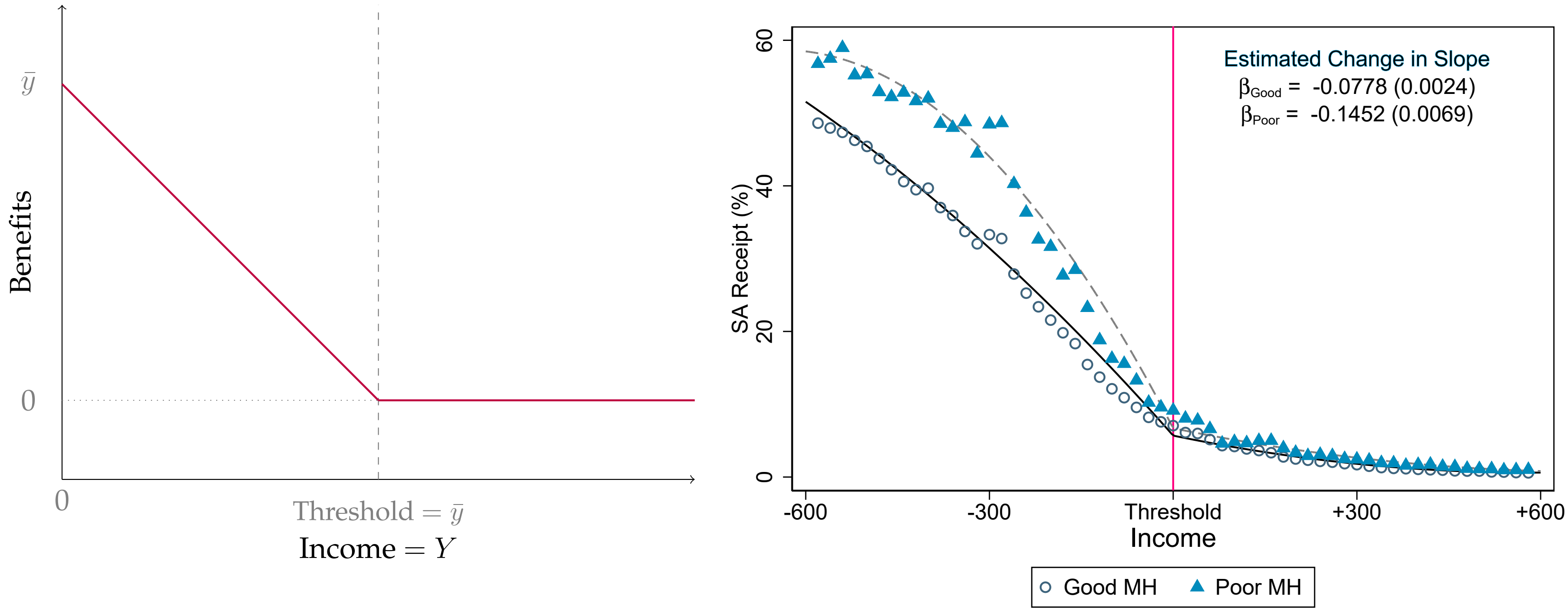


Notes: Estimates $\hat{\delta}_t$ from Equation (5.1) showing heterogeneous treatment effects of increased ordeals on rate-of-receipt by baseline mental health. Figure VIa: Poor MH_i takes 3 values: control, anti-depressants or anti-psychotics. Figure VIb: Poor MH_i is defined in 3 ways: prescription of psychotropic drugs, any mental healthcare costs, or surveyed severe psychological distress (2012). Prescription-based measures underestimate effects by missing individuals whose mental health prevents both treatment and social assistance take-up. Analysis population is eligible middle-age couples. Controls include individual fixed effects, household disposable income, education, municipality, household composition, and sector fixed effects. Standard errors clustered at municipality of residence in 2013. Coefficient estimates in Table D.2.

FIGURE VII. Benefits schedule and take-up effects by mental health

(A) Benefits schedule as a function of income

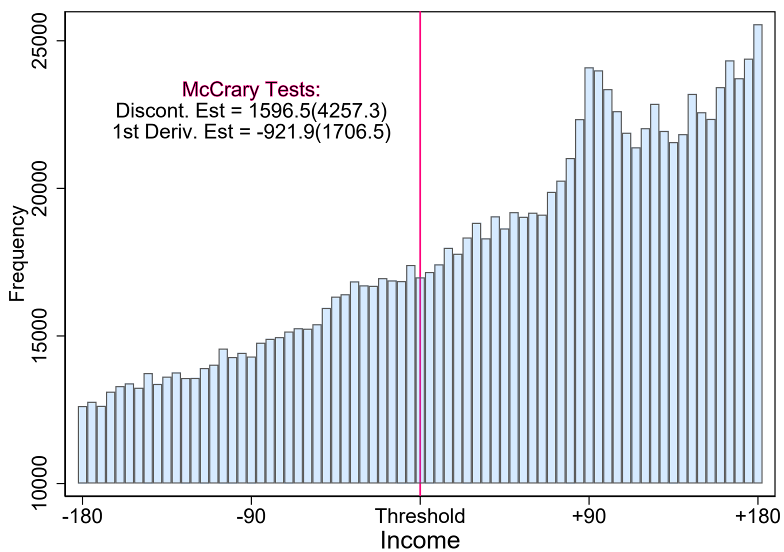
(B) Benefit take-up effects larger for people with poor mental health



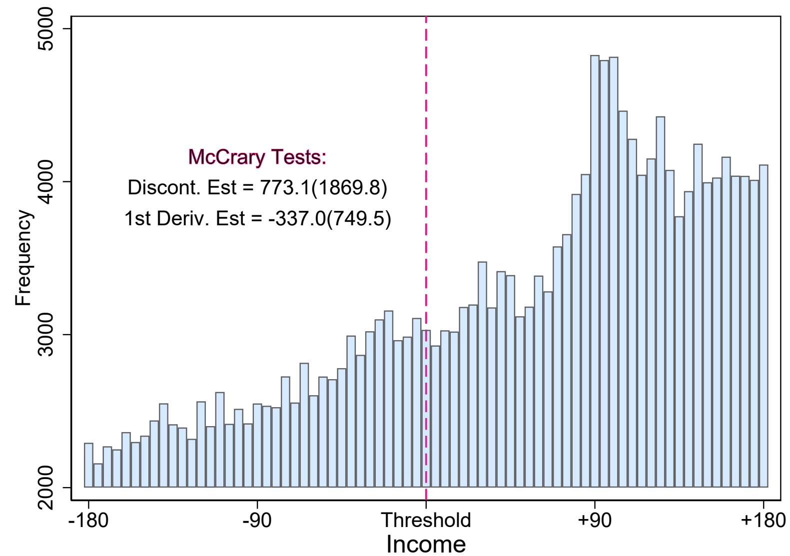
Notes: Figure VIIa shows the benefit schedule as a function of income. Figure VIIb shows average rate of receipt within income slice in a large window of income either side of the eligibility threshold. Income in Figure VIIb is monthly and refers to "eligibility income". Poor mental health is defined as receiving psychopharma in the year previously. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions.

Figure VIII. Density of income around the eligibility threshold

(A) Good Mental Health

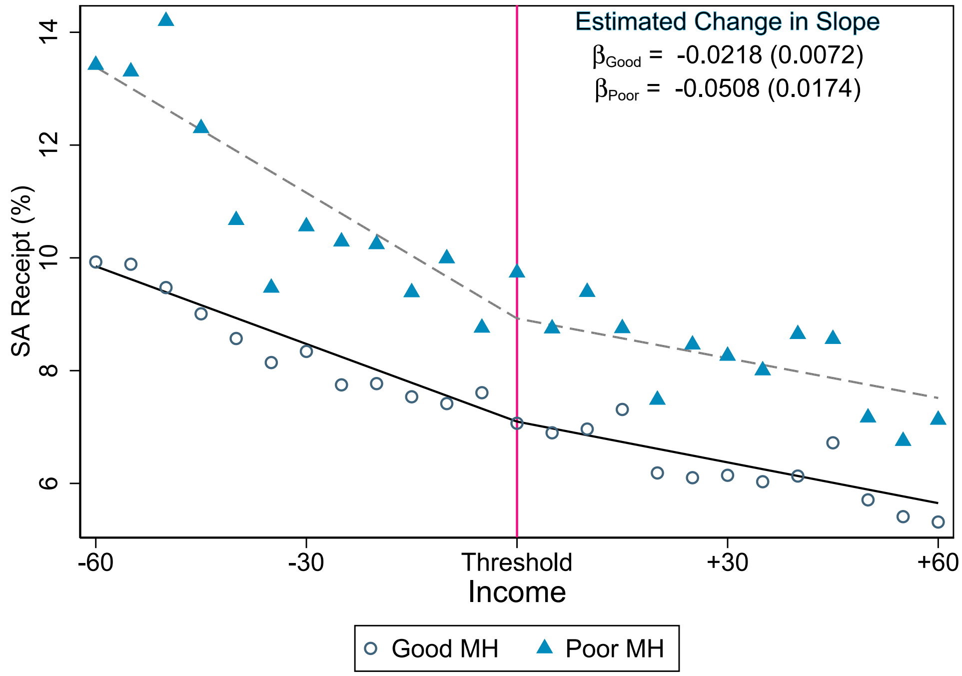


(B) Poor Mental Health



Notes: McCrary (2008) tests for discontinuity in levels and slopes around the threshold are shown. Income is monthly and refers to "eligibility income". Poor mental health is defined as receiving psychopharma in the year previously. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions.

FIGURE IX. Benefit take-up effects larger for people with poor mental health



Notes: Average rate of receipt within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly and refers to "eligibility income". Poor mental health is defined as receiving psychopharma in the year previously. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Regression lines are shown following Section 6.1.2, as well as estimated change in slopes from the regression kink design. Standard-errors are clustered at the municipality level.

TABLES

TABLE I. Limited targeting towards people with poor mental health on average

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\hat{\beta}$: Receipt of SA poor vs good MH (p.p.)	3.072***	0.491	-0.819*	1.429***	0.540***	1.984***	0.911***
	(0.810)	(0.699)	(0.362)	(0.095)	(0.071)	(0.065)	(0.0498)
Year, age and gender FEs		✓	✓	✓	✓		✓
Lagged hh disp income controls			✓	✓	✓		✓
Lagged work-status FEs				✓	✓		✓
Individual FEs						✓	✓
All other controls					✓		✓
Observations (people-years)	5,671,855	5,671,855	5,187,572	5,187,572	5,187,572	5,361,899	5,014,850
R^2	0.001	0.045	0.161	0.640	0.650	0.001	0.059
Baseline mean	59.97	59.97	62.45	62.45	62.45	60.07	62.00

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Results of a regression of receipt of social assistance on mental health status (measured by dispensation of psychotropic meds). First column shows the results with no controls. Second column shows results adding year, age and gender fixed effects. Third column shows results adding lagged household disposable income controls. Fourth column shows results adding lagged hh composition, education, municipality, wealth and work-status controls. Fifth column shows results adding sector, physical health and benefits schedule controls. Sixth column shows results with individual fixed effects only (no controls). Seventh column shows results with individual fixed effects and all controls. The sample contains the calculated eligible for SA in 2011-2020. Standard-errors are clustered at the municipality-level.

TABLE II. Welfare effects under different assumptions

	Primitives		**Welfare Effects**					
			MVPFs			Composite Policy ($d\Lambda \mid dB$)		
Specification	$\frac{v'_L}{v'_H}$	$\frac{\kappa'_L}{\kappa'_H}$	$MVPF_{d\Lambda}$	$MVPF_{dB}$	$\frac{MVPF_{d\Lambda}}{MVPF_{dB}}$	$\frac{dW}{d\Lambda}$	$\frac{dW_L}{d\Lambda}$	$\frac{dW_H}{d\Lambda}$
Baseline	2.22	1.57	2.75	1.18	2.33	-44.71	-50.67	-42.72
Role of Bias								
$\psi = 0.5$	2.22	1.57	1.66	1.31	1.27	-9.97	-3.27	-12.21
$\psi = 0.35$	2.22	1.57	1.33	1.34	0.99	0.45	10.95	-3.05
$\psi = 0$	2.22	1.57	0.56	1.43	0.39	24.76	44.13	18.30
Robustness to $\varepsilon \perp \theta$								
Preferred spec.	2.30	1.63	2.79	1.20	2.32	-45.19	-52.63	-42.72
Alternative spec.	2.12	1.50	2.71	1.16	2.34	-44.13	-48.37	-42.72
Different compliers	2.22	1.57	5.99	1.18	5.07	-136.89	-177.55	-123.33
Robustness to $d\Lambda$								
$d\Lambda = 5$	2.22	1.57	2.75	1.18	2.33	-71.63	-96.42	-63.36
$d\Lambda = 10$	2.22	1.57	2.75	1.18	2.33	-74.99	-102.14	-65.95
Robustness to $\frac{\partial P[SA]_H}{\partial \Lambda}$								
$\frac{\partial P[SA]_H}{\partial \Lambda} = -0.005$	2.22	2.80	2.76	1.18	2.33	-18.11	-38.12	-11.44
$\frac{\partial P[SA]_H}{\partial \Lambda} = -0.01$	2.22	1.90	2.75	1.18	2.33	-30.56	-43.99	-26.08
$\frac{\partial P[SA]_H}{\partial \Lambda} = -0.02$	2.22	1.45	2.75	1.18	2.33	-55.46	-55.75	-55.37
$\frac{\partial P[SA]_H}{\partial \Lambda} = -0.025$	2.22	1.36	2.75	1.18	2.33	-67.91	-61.62	-70.01
Alternative RKD specifications								
Zoomed out	1.86	1.57	0.80	0.57	1.42	-6.71	-6.71	-6.71
With controls	1.63	1.57	3.22	1.17	2.75	-58.36	-79.20	-51.42
Other robustness								
$P_L \neq P_H$	1.97	1.40	2.65	1.13	2.35	-43.32	-45.11	-42.72
Relaxing linearity			$\uparrow$	$\downarrow$	$\uparrow$	$\downarrow$		

Notes: This table presents welfare effects under various modeling assumptions. Columns show relative need (v'_L/v'_H), relative cost (κ'_L/κ'_H), MVPFs for barrier and benefit changes, and disaggregated welfare effects. Baseline uses revealed preference with $\psi = 1$ (all ordeal costs welfare-relevant). "Role of Bias" varies ψ from 0 (pure hassle costs) to 1 (see section 7.2.2). "Robustness to $\varepsilon \perp \theta$" tests sensitivity to independence assumption using heteroskedastic probit adjustments where $\varepsilon_\theta \sim N(\mu_\theta, \sigma^2_\theta)$ (see Appendix F.3.1). "Different compliers" relaxes the assumption that compliers are identical across benefit and barrier instruments (see Section 7.2.3). Section 7.2 implicitly treats the Participation Act as a 1 unit change in Λ. "Robustness to $d\Lambda$" tests sensitivity to the absolute change in barriers. "Robustness to $\frac{\partial P[SA]_H}{\partial \Lambda}$" varies the anchor for good mental health barrier screening effects. "Alternative RKD specifications" uses benefit elasticity estimates from the €600 bandwidth of Figure VIIb and the specification of Table E.1 (with controls). "$P_L \neq P_H$" allows for different baseline take-up rates between mental health groups. "Relaxing linearity" indicates directional effects when allowing non-linear take-up equations (see Appendix F.3.2). All monetary values in euros; welfare effects scaled by willingness-to-pay of individuals with good mental health.

References

Alatas, V., R. Purnamasari, M. Wai-Poi, A. Banerjee, B. A. Olken, and R. Hanna (2016): "Self-targeting: Evidence from a field experiment in Indonesia," *Journal of Political Economy*, 124, 371–427. [*Cited on page 5.*]

Alloy, L. B. and A. H. Ahrens (1987): "Depression and pessimism for the future: biased use of statistically relevant information in predictions for self versus others." *Journal of personality and social psychology*, 52, 366. [*Cited on page 28.*]

Anders, J. and C. Rafkin (2022): "The Welfare Effects of Eligibility Expansions: Theory and Evidence from SNAP." *SSRN*. [*Cited on page 5.*]

Angelucci, M. and D. Bennett (2024a): "Depression, Pharmacotherapy, and the Demand for a Preventive Health Product," *Available at SSRN 4808853*. [*Cited on page 4.*]

——— (2024b): "The economic impact of depression treatment in india: Evidence from community-based provision of pharmacotherapy," *American economic review*, 114, 169–198. [*Cited on page 4.*]

APA (2013): *Diagnostic and statistical manual of mental disorders: DSM-5*, vol. 5, American psychiatric association Washington, DC. [*Cited on page 1.*]

Arulsamy, K. and L. Delaney (2022): "The impact of automatic enrolment on the mental health gap in pension participation: Evidence from the UK," *Journal of Health Economics*, 102673. [*Cited on page 4.*]

Baranov, V., S. Bhalotra, P. Biroli, and J. Maselko (2020): "Maternal depression, women's empowerment, and parental investment: evidence from a randomized controlled trial," *American economic review*, 110, 824–59. [*Cited on page 4.*]

Bargain, O., K. Orsini, and A. Peichl (2014): "Comparing labor supply elasticities in Europe and the United States: New results," *Journal of Human Resources*, 49, 723–838. [*Cited on page 128.*]

Barker, N., G. T. Bryan, D. Karlan, A. Ofori-Atta, and C. R. Udry (2021): "Mental Health Therapy as a Core Strategy for Increasing Human Capital: Evidence from Ghana," *American Economic Review: Insights, Forthcoming*. [*Cited on page 4.*]

Bell, E., J. Christensen, P. Herd, and D. Moynihan (2022): "Health in Citizen-State Interactions: How Physical and Mental Health Problems Shape Experiences of Administrative Burden and Reduce Take-Up," *Public Administration Review*. [*Cited on page 4.*]

Berkhout, E., P. Koot, and N. Bosch (2019): "Gebruik (en niet-gebruik) van toeslagen in Nederland [Take-up (and non-take-up) of benefits in the Netherlands]," . [*Cited on page 12.*]

BHARADWAJ, P., M. M. PAI, AND A. SUZIEDELYTE (2017): "Mental health stigma," *Economics Letters*, 159, 57–60. [*Cited on page 1.*]

BHAT, B., J. DE QUIDT, J. HAUSHOFER, V. H. PATEL, G. RAO, F. SCHILBACH, AND P.-L. P. VAUTREY (2022): "The Long-Run Effects of Psychotherapy on Depression, Beliefs, and Economic Outcomes," Tech. rep., National Bureau of Economic Research. [*Cited on page 4.*]

BIERBAUM, M. AND F. GASSMANN (2016): "Limits of Freedom? Experiences with the Participation Act," *Policy Brief*. [*Cited on page 70.*]

BIERMAN, E. J., H. C. COMIJS, F. RIJMEN, C. JONKER, AND A. T. BEEKMAN (2008): "Anxiety symptoms and cognitive performance in later life: results from the longitudinal aging study Amsterdam," *Aging and Mental Health*, 12, 517–523. [*Cited on pages 1, 4, and 28.*]

BLOOM, D. E., E. CAFIERO, E. JANÉ-LLOPIS, S. ABRAHAMS-GESSEL, L. R. BLOOM, S. FATHIMA, A. B. FEIGL, T. GAZIANO, A. HAMANDI, M. MOWAFI, ET AL. (2012): "The global economic burden of noncommunicable diseases," Tech. rep., Program on the Global Demography of Aging. [*Cited on page 1.*]

BREWER, M., T. DANG, AND E. TOMINEY (2022): "Universal Credit: Welfare Reform and Mental Health," . [*Cited on page 32.*]

CALONICO, S., M. D. CATTANEO, M. H. FARRELL, F. PALOMBA, AND R. TITIUNIK (2025): "Treatment effect heterogeneity in regression discontinuity designs," *arXiv preprint arXiv:2503.13696*. [*Cited on page 24.*]

CALONICO, S., M. D. CATTANEO, AND R. TITIUNIK (2014): "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 82, 2295–2326. [*Cited on pages 23, 24, 26, 27, 89, and 119.*]

CARD, D., D. S. LEE, Z. PEI, AND A. WEBER (2015): "Inference on causal effects in a generalized regression kink design," *Econometrica*, 83, 2453–2483. [*Cited on pages 22, 23, 51, 87, 88, 91, and 93.*]

CHRISTIAN, C., L. HENSEL, AND C. ROTH (2019): "Income shocks and suicides: Causal evidence from Indonesia," *Review of Economics and Statistics*, 101, 905–920. [*Cited on page 4.*]

COOK, J. B. AND C. N. EAST (2024): "Work Requirements with No Teeth Still Bite: Disenrollment and Labor Supply Effects of SNAP General Work Requirements," Tech. rep., National Bureau of Economic Research. [*Cited on page 19.*]

CRONIN, C. J., M. P. FORSSTROM, AND N. W. PAPAGEORGE (2024): "What good are treatment effects without treatment? mental health and the reluctance to use talk therapy," *Review of Economic Studies*, rdae061. [*Cited on pages 1 and 32.*]

DANESH, K., J. T. KOLSTAD, J. SPINNEWIJN, AND W. D. PARKER (2024): "The Chronic Disease Index: Analyzing Health Inequalities Over the Lifecycle," Tech. rep., National Bureau of Economic Research. [*Cited on pages 15 and 66.*]

DE CHAISEMARTIN, C. AND X. D'HAULTFŒUILLE (2023): "Two-Way Fixed Effects and Difference-in-Differences Estimators with Heterogeneous Treatment Effects and Imperfect Parallel Trends," *Available at SSRN*. [*Cited on page 18.*]

DER-AVAKIAN, A. AND A. MARKOU (2012): "The neurobiology of anhedonia and other reward-related deficits," *Trends in neurosciences*, 35, 68–77. [*Cited on page 28.*]

DESHPANDE, M. AND Y. LI (2019): "Who is screened out? Application costs and the targeting of disability programs," *American Economic Journal: Economic Policy*, 11, 213–48. [*Cited on page 5.*]

EUROPEAN CENTRAL BANK (2024): "Fluctuation of the European Central Bank fixed interest rate from 2008 to 2024," Graph, Statista, accessed: January 31, 2026. [*Cited on page 67.*]

EVANS, V. C., G. L. IVERSON, L. N. YATHAM, AND R. W. LAM (2014): "The relationship between neurocognitive and psychosocial functioning in major depressive disorder: a systematic review," *The Journal of clinical psychiatry*, 75, 17306. [*Cited on pages 4 and 28.*]

FINKELSTEIN, A., E. F. LUTTMER, AND M. J. NOTOWIDIGDO (2013): "What good is wealth without health? The effect of health on the marginal utility of consumption," *Journal of the European Economic Association*, 11, 221–258. [*Cited on pages 2 and 11.*]

FINKELSTEIN, A. AND M. J. NOTOWIDIGDO (2019): "Take-up and targeting: Experimental evidence from SNAP," *The Quarterly Journal of Economics*, 134, 1505–1556. [*Cited on pages 5, 52, and 128.*]

GANONG, P. AND S. JÄGER (2018): "A permutation test for the regression kink design," *Journal of the American Statistical Association*, 113, 494–504. [*Cited on pages 25 and 114.*]

GELBER, A., D. JONES, D. W. SACKS, AND J. SONG (2020): "Using non-linear budget sets to estimate extensive margin responses: Evidence and method from the earnings test," *American Economic Journal: Applied Economics*. [*Cited on page 93.*]

GIANNELLA, E., T. HOMONOFF, G. RINO, AND J. SOMERVILLE (2023): "Administrative burden and procedural denials: Experimental evidence from SNAP," Tech. rep., National Bureau of Economic Research Cambridge, MA. [*Cited on page 5.*]

GODARD, M., P. KONING, AND M. LINDEBOOM (2022): "Application and award responses to stricter screening in disability insurance," *Journal of Human Resources*. [*Cited on page 5.*]

GROSS, J. J. AND R. F. MUÑOZ (1995): "Emotion regulation and mental health." *Clinical psychology: Science and practice*, 2, 151. [*Cited on pages 1, 4, and 28.*]

HALLER, A. AND S. STAUBLI (2024): "Measuring the Value of Disability Insurance from Take-Up Decisions," . [*Cited on pages 8, 11, and 124.*]

HAMMAR, Å. AND G. ÅRDAL (2009): "Cognitive functioning in major depression-a summary," *Frontiers in human neuroscience*, 3, 26. [*Cited on pages 1, 4, and 28.*]

HAUSHOFER, J. (2019): "Is there a Psychological Poverty Trap?" Tech. rep., Working Paper. [*Cited on page 32.*]

HAUSHOFER, J., R. MUDIDA, AND J. SHAPIRO (2020): "The comparative impact of cash transfers and psychotherapy on psychological and economic well-being," *NBER working paper*, 7. [*Cited on page 28.*]

HEEKELAAR, M. (2021): "Beschikbare en benodigde financiële middelen voor de Participatiewet: Analyse," . [*Cited on page 18.*]

HENDREN, N. (2020): "Measuring economic efficiency using inverse-optimum weights," *Journal of public Economics*, 187, 104198. [*Cited on page 53.*]

HENDREN, N. AND B. SPRUNG-KEYSER (2020): "A unified welfare analysis of government policies," *The Quarterly Journal of Economics*, 135, 1209–1318. [*Cited on pages 4, 10, and 53.*]

HERD, P. AND D. MOYNIHAN (2025): "Administrative Burdens in the Social Safety Net," *Journal of Economic Perspectives*, 39, 129–150. [*Cited on page 1.*]

HOMONOFF, T. AND J. SOMERVILLE (2021): "Program Recertification Costs: Evidence from SNAP," *American Economic Journal: Economic Policy*, 13, 271–98. [*Cited on page 5.*]

HYMAN, S., D. CHISHOLM, R. KESSLER, V. PATEL, AND H. WHITEFORD (2006): "Mental disorders," *Disease control priorities related to mental, neurological, developmental and substance abuse disorders*, 1–20. [*Cited on page 28.*]

INSPECTIE SZW (2021): "Niet-gebruik van de algemene bijstand: Een onderzoek naar de omvang, kenmerken, langdurigheid en aanpak," *Den Haag: Inspectie SZW*. [*Cited on pages 13, 14, and 70.*]

KASHDAN, T. B., V. BARRIOS, J. P. FORSYTH, AND M. F. STEGER (2006): "Experiential avoidance as a generalized psychological vulnerability: Comparisons with coping and emotion regulation strategies," *Behaviour Research and Therapy*, 44, 1301–1320. [*Cited on page 28.*]

KESSLER, R. C., P. BERGLUND, O. DEMLER, R. JIN, D. KORETZ, K. R. MERIKANGAS, A. J. RUSH, E. E. WALTERS, AND P. S. WANG (2003): "The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R)," *jama*, 289, 3095–3105. [*Cited on pages 4 and 28.*]

KLEVEN, H. J. (2016): "Bunching," *Annual Review of Economics*, 8, 435–464. [*Cited on pages 23 and 54.*]

KLINE, P. AND C. R. WALTERS (2019): "On Heckits, LATE, and numerical equivalence," *Econometrica*, 87, 677–696. [*Cited on page 11.*]

KNIJN, T. AND J. HIAH (2019): "Coping with the Participation Act: Welfare experiences in the Netherlands," *ETHOS country report for ETHOS report D*, 5. [*Cited on page 70.*]

KO, W. AND R. A. MOFFITT (2024): "Take-up of social benefits," *Handbook of Labor, Human Resources and Population Economics*, 1–42. [*Cited on page 2.*]

KROMHOUT, M., N. KORNALIJNSLIJPER, AND M. DE KLERK (2018): "Summary changing care and support for people with disabilities," . [*Cited on page 20.*]

KRUEGER, A. B. AND B. D. MEYER (2002): "Labor supply effects of social insurance," *Handbook of public economics*, 4, 2327–2392. [*Cited on page 24.*]

LANDAIS, C., A. NEKOEI, P. NILSSON, D. SEIM, AND J. SPINNEWIJN (2021): "Risk-based selection in unemployment insurance: Evidence and implications," *American Economic Review*, 111, 1315–1355. [*Cited on page 125.*]

LOPES, F. V., B. RAVESTEIJN, T. VAN OURTI, AND C. RIUMALLO-HERL (2023): "Income inequalities beyond access to mental health care: a Dutch nationwide record-linkage cohort study of baseline disease severity, treatment intensity, and mental health outcomes," *The Lancet Psychiatry*, 10, 588–597. [*Cited on page 13.*]

MARTIN, L., L. DELANEY, AND O. DOYLE (2023a): "Everyday administrative burdens and inequality," *Public Administration Review*. [*Cited on page 4.*]

MARTIN, L., L. DELANEY, O. DOYLE, ET AL. (2023b): "The Distributive Effects of Administrative Burdens on Decision-Making," *Journal of Behavioral Public Administration*, 6. [*Cited on page 4.*]

MCCRARY, J. (2008): "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of econometrics*, 142, 698–714. [*Cited on pages 23, 40, and 111.*]

MCGARRY, K. (1996): "Factors determining participation of the elderly in supplemental security income," *Journal of Human Resources*, 331–358. [*Cited on page 24.*]

MILLAN, M. J., Y. AGID, M. BRÜNE, E. T. BULLMORE, C. S. CARTER, N. S. CLAYTON, R. CONNOR, S. DAVIS, B. DEAKIN, R. J. DERUBEIS, ET AL. (2012): "Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy," *Nature Reviews Drug Discovery*, 11, 141–168. [*Cited on page 28.*]

MILLER, S., E. RHODES, A. W. BARTIK, D. E. BROOCKMAN, P. K. KRAUSE, AND E. VIVALT (2024): "Does Income Affect Health? Evidence from a Randomized Controlled Trial of a Guaranteed Income," Tech. rep., National Bureau of Economic Research. [*Cited on page 4.*]

MINISTERIE VAN SZW (2015): "Participatiewet," https://wetten.overheid.nl/BWBR0015703/2015-01-01/. [*Cited on pages 3, 13, 23, and 93.*]

——— (2022): "Participatiewet in Balans: uitkomsten beleidsanalyse," https://www.rijksoverheid.nl/documenten/kamerstukken/2022/06/21/bijlage-rapport-participatiewet-in-balans. [*Cited on pages 3, 20, 70, and 71.*]

MOGSTAD, M., A. TORGOVITSKY, AND C. R. WALTERS (2024): "Policy evaluation with multiple instrumental variables," *Journal of Econometrics*, 105718. [*Cited on page 11.*]

MULLAINATHAN, S., J. SCHWARTZSTEIN, AND W. J. CONGDON (2012): "A reduced-form approach to behavioral public finance," *Annu. Rev. Econ.*, 4, 511–540. [*Cited on page 56.*]

NAIK, C., W. PARKER, AND J. SPINNEWIJN (2026): "Social Determinants of Mental Health," . [*Cited on pages 13 and 14.*]

NAIK, C. AND D. RECK (2025): "Intrapersonal Utility Comparisons as Interpersonal Utility Comparisons: Welfare, Ambiguity, and Robustness in Behavioral Policy Problems," Tech. rep. [*Cited on pages 2, 10, and 30.*]

NICHOLS, A. L. AND R. J. ZECKHAUSER (1982): "Targeting transfers through restrictions on recipients," *The American Economic Review*, 72, 372–377. [*Cited on page 4.*]

PEI, Z., D. S. LEE, D. CARD, AND A. WEBER (2022): "Local polynomial order in regression discontinuity designs," *Journal of Business & Economic Statistics*, 40, 1259–1267. [*Cited on page 91.*]

RAFKIN, C., A. SOLOMON, AND E. SOLTAS (2025): "Self-Targeting in US Transfer Programs," *Available at SSRN 4495537*. [*Cited on pages 5, 8, and 9.*]

RIDLEY, M., G. RAO, F. SCHILBACH, AND V. PATEL (2020): "Poverty, depression, and anxiety: Causal evidence and mechanisms," *Science*, 370, eaay0214. [*Cited on pages 1, 4, and 32.*]

ROOS, A.-F., M. DIEPSTRATEN, R. DOUVEN, ET AL. (2021): "When Financials Get Tough, Life Gets Rough?: Problematic Debts and Ill Health," Tech. rep., CPB Netherlands Bureau for Economic Policy Analysis Hague, the Netherlands. [*Cited on page 15.*]

RUSSO, A. (2023): "Waiting or Paying for Healthcare: Evidence from the Veterans Health Administration," Tech. rep., Working Paper. [*Cited on page 8.*]

SCHMIDT, L., L. SHORE-SHEPPARD, AND T. WATSON (2021): "The Effect of Safety Net Generosity on Maternal Mental Health and Risky Health Behaviors," Tech. rep., National Bureau of Economic Research. [*Cited on page 4.*]

SCP (2019): "Eindevaluatie van de Participatiewet," . [*Cited on pages 3, 18, 20, 69, and 70.*]

SEN, A. (1999): *Development as Freedom*, New York: Oxford University Press. [*Cited on page 28.*]

——— (2008): "The idea of justice," *Journal of human development*, 9, 331–342. [*Cited on page 28.*]

SERENA, B. L. (2024): "The Causal Effect of Scaling up Access to Psychotherapy," . [*Cited on page 4.*]

SHAHN, Z. (2023): "Subgroup Difference in Differences to Identify Effect Modification Without a Control Group," *arXiv preprint arXiv:2306.11030*. [*Cited on page 18.*]

SHEPARD, M. AND M. WAGNER (2022): "Do ordeals work for selection markets? Evidence from health insurance auto-enrollment," Tech. rep., National Bureau of Economic Research. [*Cited on page 5.*]

SHREEKUMAR, A. AND P.-L. VAUTREY (2021): "Managing Emotions: The Effects of Online Mindfulness Meditation on Mental Health and Economic Behavior," . [*Cited on page 4.*]

SILVER, D. AND J. ZHANG (2022): "Impacts of Basic Income on Health and Economic Well-Being: Evidence from the VA's Disability Compensation Program," Tech. rep., National Bureau of Economic Research. [*Cited on page 4.*]

TREADWAY, M. T. AND D. H. ZALD (2013): "Parsing anhedonia: translational models of reward-processing deficits in psychopathology," *Current directions in psychological science*, 22, 244–249. [*Cited on page 28.*]

UNRATH, M. (2024): "Targeting, Screening, and Retention: Evidence from the Supplemental Nutrition Assistance Program in California," . [*Cited on page 5.*]

VAN DER VEEN, R. (2019): "Basic income experiments in the Netherlands?" *Basic Income Studies*, 14. [*Cited on pages 18 and 69.*]

WHO (2022): "World mental health report: transforming mental health for all," . [*Cited on page 1.*]

WOOLDRIDGE, J. M. (2019): "Correlated random effects models with unbalanced panels," *Journal of Econometrics*, 211, 137–150. [*Cited on page 75.*]

WU, D. AND B. D. MEYER (2023): "Certification and Recertification in Welfare Programs: What Happens When Automation Goes Wrong?" Tech. rep., National Bureau of Economic Research. [*Cited on page 5.*]

Supplementary Appendix

"Mental Health and the Targeting of Social Assistance" by Canishk Naik (2026)

Appendix A. Theory Appendix

Let θ have a type-specific indirect utility functions: $u_\theta(c, y)$ is increasing in consumption c and decreasing in earned income y. Income depends on take-up but is fixed otherwise:[26] let $y_\theta^{SA=1}$ refer to income earned if on social assistance and $y_\theta^{SA=0}$ if not. All income (including benefits) is taxed at marginal tax rate τ. Thus, $v_\theta(B)$ is given by:

$$v_\theta(B) \coloneqq u_\theta\left((1-\tau)\cdot\left[y_\theta^{SA=1}+B\right], y_\theta^{SA=1}\right) - u_\theta\left((1-\tau)\cdot y_\theta^{SA=0}, y_\theta^{SA=0}\right) \tag{A.1}$$

Thus, value is the net-utility gain from social assistance and comes from two main sources. First, if $y_\theta^{SA=0} \leq y_\theta^{SA=1} + B$, θ derives utility from the top-up in consumption $(1-\tau)y^{SA=0} \to (1-\tau)\cdot\left[y_\theta^{SA=1}+B\right]$. Second, if $y^{SA=1} < y^{SA=0}$, θ also derives value from a lowered cost of working when supported by social assistance. Importantly, heterogeneous value across types does not only come from different y_θ, the utility functions u_θ also differ.

Note that eligibility then is defined as $y \leq \bar{y}$ where $y = SA \cdot y^{SA=1} + (1-SA)\cdot y^{SA=0}$.

Proof of Proposition 2.1. Social welfare is defined as follows.

$$W = \int \lambda_\theta \mathcal{U}_\theta d\mu$$

Using the chain rule: $\frac{dW}{d\Lambda} = \frac{\partial W}{\partial \Lambda} + \frac{\partial W}{\partial B}\cdot\frac{\partial B}{\partial \Lambda}$, and using the Leibniz rule to differentiate under the integral gives Equation (2.9). Here, the Envelope Theorem implies the behavioural welfare effect is 0. For example,

$$\begin{aligned}
\frac{d\,\mathcal{U}_\theta}{d\Lambda} &= \frac{d}{d\Lambda}\int_{-\infty}^{\varepsilon_\theta^*}\left[v_\theta(B) - \kappa_\theta(\Lambda) - \varepsilon\right]\, dF(\varepsilon) \\
&= \frac{d\varepsilon_\theta^*}{d\Lambda}\cdot\underbrace{\left[v_\theta(B) - \kappa_\theta(\Lambda) - \varepsilon_\theta^*\right]}_{=0 \text{ by defn of } \varepsilon_\theta^*} + \int_{-\infty}^{\varepsilon_\theta^*}\left[-\kappa_\theta'(\Lambda)\right] dF(\varepsilon)
\end{aligned}$$

The above step is the Envelope Theorem at work.

$$= -\kappa_\theta'(\Lambda)\cdot F(\varepsilon_\theta^*)$$

[26]The assumption of no labour supply responses follows Finkelstein and Notowidigdo (2019) and simplifies the theoretical analysis. In the Netherlands, social assistance tops income up to a social minimum. Therefore, conditional on receipt, income ≈ 0 for many people. This means that the decision in practice can be reasonably approximated to take-up SA (and earn low/no income) vs do not take-up SA (and earn income).

Similarly, $\frac{d\,\mathcal{U}_\theta}{dB} = v'_\theta(B) \cdot F(\varepsilon^*_\theta)$. Therefore:

$$\frac{dW}{d\Lambda} = \int \lambda_\theta \mathbb{P}[SA]_\theta \left[v'_\theta(B) \cdot \frac{dB}{d\Lambda} - \kappa'_\theta(\Lambda) \right] d\mu$$

Let G be the government's budget. Budget neutrality implies $\frac{dG}{d\Lambda} = 0$. Using the chain and Leibniz rule again, and dropping θ subscripts:

$$\begin{aligned}\frac{dG}{d\Lambda} = \int & \left[\tau(y^{SA=0} - y^{SA=1}) + (1-\tau) \cdot B\right] \cdot \frac{\partial \mathbb{P}[SA]}{\partial \Lambda} + \left[\tau(y^{SA=0} - y^{SA=1}) + (1-\tau) \cdot B\right] \cdot \frac{\partial \mathbb{P}[SA]}{\partial B} \cdot \frac{dB}{d\Lambda} \\ & + (1-\tau) \cdot \mathbb{P}[SA] \cdot \frac{dB}{d\Lambda} d\mu = 0\end{aligned}$$

Rearranging gives Equation (2.10). □

A.1. **Redistributive MVPF Formulae.** The MVPF measures the ratio of the direct welfare effect to beneficiaries of a policy, divided by the cost to the government. Direct welfare effects are usually written in the units of each types' willingness-to-pay, with social marginal utilities of income usually appearing outside the MVPF formulae. The composite policy increasing Λ (B adjusts) is social-welfare improving, if the gains from increasing spending on dB exceed the losses from reducing spending through an increase $d\Lambda$. The proof is a simple re-writing of Hendren (2020) without separating out the social marginal utility of beneficiaries.

Proposition A.1. *(Hendren and Sprung-Keyser, 2020)*

The composite policy experiment of a budget-neutral increase in Λ financing an increase in B is good for welfare W iff:

$$MVPF_{d\Lambda} > MVPF_{dB} \tag{A.2}$$

*where the **redistributive** MVPF of an increase in ordeals is given by Equation (A.3).*

$$MVPF_{d\Lambda} = \frac{\overbrace{-\int \lambda_\theta \cdot \mathbb{P}[SA]_\theta \cdot \frac{\kappa'_\theta(\Lambda)}{v'_\theta(B)}\, d\mu}^{\textit{Direct Effect} <0}}{\underbrace{\int FE_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda} d\mu}_{\textit{behavioural Revenue Effect} <0}} \tag{A.3}$$

*and **redistributive** MVPF of an increase in benefit level is given by Equation (A.4).*

$$MVPF_{dB} = \frac{\overbrace{\int \lambda_\theta \cdot \mathbb{P}[SA]_\theta \, d\mu}^{\textit{Direct Effect} > 0}}{\underbrace{(1-\tau) \cdot \int \mathbb{P}[SA]_\theta \, d\mu}_{\textit{Mechanical Revenue Effect} > 0} + \underbrace{\int FE_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial B} d\mu}_{\textit{behavioural Revenue Effect} > 0}} \tag{A.4}$$

The direct effect of an increase in ordeals $d\Lambda$ is that it imposes dis-utility on infra-marginal individuals κ'_θ. Increasing barriers saves the government money through lower take-up, corresponding to the denominator. The direct effect of an increase in benefit level dB is that it transfers €1 of benefits to all infra-marginal individuals. The government has to pay for the mechanical extra program cost, as well as the new-entrants. See Appendix F.1 for how to calculate these formulas when sufficient statistics are estimated on the eligible population.

Proof of Proposition A.1. From the proof of Proposition 2.1,

$$\frac{\partial W}{\partial \Lambda} = -\int \lambda_\theta \mathbb{P}[SA]_\theta \kappa'_\theta d\mu \tag{A.5}$$

$$\frac{\partial W}{\partial B} = \int \lambda_\theta \mathbb{P}[SA]_\theta v'_\theta d\mu \tag{A.6}$$

$$\frac{\partial G}{\partial \Lambda} = \int FE_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda} d\mu \tag{A.7}$$

$$\frac{\partial G}{\partial B} = (1-\tau) \int \mathbb{P}[SA]_\theta \, d\mu + \int FE \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial B} d\mu \tag{A.8}$$

The first two equations follow by the Envelope theorem, as in the proof of Proposition 2.1. Dividing yields the MVPF formulas.

□

A.2. **Identification.** In this section, I set out how to identify the relationship between $f_\varepsilon(v_\theta - \kappa_\theta)$ across types using take-up levels and a first-order Taylor approximation. The key case is when $\mathbb{P}[SA]_\theta \neq \mathbb{P}[SA]_{\tilde{\theta}}$. For argument's sake - suppose that we are considering two types $\theta = L, H$.

This proposition requires some additional structure:

Let indirect utility $u_\theta(c, y) = v_\theta \cdot c - \frac{n_\theta}{1+1/e} \cdot \left(\frac{y}{n_\theta}\right)^{1+1/e}$: quasi-linear utility with scaling factor v– denoting the marginal value of income–and isoelastic disutility of labour, as in e.g. Kleven (2016). Individuals then differ based on their value of money, and their ability n_θ. For simplicity, Frisch elasticities are the same across types. In this case, $y^{SA=0} = \arg\max u\left((1-\tau)y, y\right) = n \cdot v \cdot (1-\tau)^e$. Suppose also that $\kappa(\Lambda) = \kappa_1 \cdot \Lambda + \kappa_0$. Therefore,

$$SA = 1 \iff u\left((1-\tau)\cdot\left(B + y^{SA=1}\right), y^{SA=1}\right) - \kappa\cdot\Lambda + \kappa_0 - \varepsilon \geq u\left((1-\tau)y^{SA=0}, y^{SA=0}\right) \quad \text{(A.9)}$$

Then:

Proposition A.2. *Identification of $f_L := f_\varepsilon(v_L - \kappa_L)$ in terms of $f_H := f_\varepsilon(v_H - \kappa_H)$ is given by:*

$$\mathbb{P}[SA]_L - \mathbb{P}[SA]_H \approx \left(\Psi\frac{\partial\mathbb{P}[SA]_L}{\partial B} + \Lambda\frac{\partial\mathbb{P}[SA]_L}{\partial\Lambda}\right)\cdot\left(\frac{f_H}{f_L} - 1\right) \quad \text{(A.10)}$$

where $\Psi = B + y^{SA=1} - \frac{y^{SA=0}}{1+e} - \frac{e}{1+e}\frac{\left(y^{SA=1}\right)^{1+1/e}}{\left(y^{SA=0}\right)^{1/e}}$.

Note that if the LHS = 0, the RHS will imply that $f_L = f_H$ as long as $\Psi\frac{\partial\mathbb{P}[SA]_L}{\partial B} + \Lambda\frac{\partial\mathbb{P}[SA]_L}{\partial\Lambda} \neq 0$.

Proof.

$$v(B) = u\left((1-\tau)\cdot(B + y^{SA=1}, y^{SA=1}\right) - u\left((1-\tau)y^{SA=0}, y^{SA=0}\right)$$

First, by Taylor's theorem:

$$\mathbb{P}[SA]_L - \mathbb{P}[SA]_H = F(v_L - \kappa_L) - F(v_H - \kappa_H) \approx [v_L - v_H - (\kappa_L - \kappa_H)]\cdot\underbrace{f(v_H - \kappa_H)}_{f_H}$$

Goal: approximate $v_L - v_H$ and $\kappa_L - \kappa_H$ using take-up responses to changes in B and Λ.

Given the structural assumptions, $v(B) = v\cdot(1-\tau)\{B + y^{SA=1} - y^{SA=0}\} - \frac{n}{1+1/e}\cdot\left(\frac{y^{SA=1}}{n}\right)^{1+1/e} + \frac{n}{1+1/e}\cdot\left(\frac{y^{SA=0}}{n}\right)^{1+1/e}$. But since $y^{SA=0} = n\cdot v\cdot(1-\tau)^e$, this means:

$$v(B) = v\cdot(1-\tau)\cdot\underbrace{\left\{B + y^{SA=1} - \frac{y^{SA=0}}{1+e} - \frac{e}{1+e}\frac{\left(y^{SA=1}\right)^{1+1/e}}{\left(y^{SA=0}\right)^{1/e}}\right\}}_{:=\Psi} \quad \text{(A.11)}$$

Note that: $v'(B) = v\cdot(1-\tau)$ in this setting. Finally, I assume $\kappa(\Lambda) = \kappa_1\cdot\Lambda + \kappa_0$ where $\kappa_1 = \kappa'(\Lambda)$. To match the empirical application, assume income is fixed across types.

$$\begin{aligned} F(v_L - \kappa_L) - F(v_H - \kappa_H) &\approx \left[(v'_L(B) - v'_H(B))\cdot\Psi - (\kappa'_L(\Lambda) - \kappa'_H(\Lambda))\cdot\Lambda - \Delta\kappa_0\right]\cdot f_H \\ &= \left(\frac{\partial\mathbb{P}[SA]_L}{\partial B}\cdot\frac{f_H}{f_L} - \frac{\partial\mathbb{P}[SA]_H}{\partial B}\right)\cdot\Psi \\ &\quad + \left(\frac{\partial\mathbb{P}[SA]_L}{\partial\Lambda}\cdot\frac{f_H}{f_L} - \frac{\partial\mathbb{P}[SA]_H}{\partial\Lambda}\right)\cdot\Lambda - \alpha \end{aligned}$$

by Equations (2.4) and (2.5) and where $\alpha = f_H \cdot \Delta\kappa_0$. Note that when the LHS = 0, we know that $f_L = f_H$. Therefore, $\alpha = \left(\frac{\partial \mathbb{P}[SA]_L}{\partial B} - \frac{\partial \mathbb{P}[SA]_H}{\partial B}\right) \cdot \Psi + \left(\frac{\partial \mathbb{P}[SA]_L}{\partial \Lambda} - \frac{\partial \mathbb{P}[SA]_H}{\partial \Lambda}\right) \cdot \Lambda$. Rearranging gives Equation (A.10).

□

A.3. **Robustness to Bias.** Suppose a share ψ of $\kappa_\theta(\Lambda)$ is a true cost, and $(1-\psi)$ is a hassle cost, which affects behaviour but not welfare. Then: $\mathbb{P}[SA]_\theta = F_\varepsilon\left[v_\theta(B) - \kappa_\theta(\Lambda)\right]$ still, but:

$$\mathcal{U}_\theta = \int_{-\infty}^{\varepsilon_\theta^*} \left[v_\theta(B) - \kappa_\theta(\Lambda) + MI_\theta - \varepsilon\right] dF(\varepsilon) \tag{A.12}$$

where $\varepsilon_\theta^* = v_\theta(B) - \kappa_\theta(\Lambda)$ and $MI_\theta = (1-\psi) \cdot \kappa_\theta(\Lambda)$ is the marginal internality (Mullainathan et al., 2012). Note that since the true cost $\psi \cdot \kappa \leq \kappa$, behaviour over-states the ordeal-cost, so take-up is too low relative to the private optimum. This means that a marginal increase in Λ has an extra negative behavioural welfare cost coming from people moving further away from the private optimum. A marginal increase in B has an extra positive behavioural welfare gain coming from the internality correction. This is shown in Proposition A.3.

Proposition A.3. *First order welfare effects when perceived cost differs from true cost.*

$$\frac{d\mathcal{U}_\theta}{d\Lambda} = -\psi \cdot \kappa_\theta'(\Lambda) \cdot \mathbb{P}[SA]_\theta + MI_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda} \tag{A.13}$$

$$\frac{d\mathcal{U}_\theta}{dB} = v_\theta'(B) \cdot \mathbb{P}[SA]_\theta + MI_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial B} \tag{A.14}$$

Proof.

$$\mathcal{U}_\theta = \int_{-\infty}^{\varepsilon_\theta^*} \left[v_\theta(B) - \kappa_\theta(\Lambda) - \varepsilon\right] dF(\varepsilon) + \int_{-\infty}^{\varepsilon_\theta^*} MI_\theta \, dF(\varepsilon)$$

which means that, by the Leibniz integral rule:

$$\frac{d\mathcal{U}_\theta}{d\Lambda} = -\kappa_\theta'(\Lambda) \cdot F(\varepsilon_\theta^*) + 0 + (1-\psi)\kappa_\theta(\Lambda)\frac{\partial F(\varepsilon_\theta^*)}{\partial \Lambda} + (1-\psi)\kappa_\theta'(\Lambda) \cdot F(\varepsilon_\theta^*)$$

where the 0 comes from $\varepsilon_\theta^* = v_\theta(B) - \kappa_\theta(\Lambda)$ - this is the Envelope Theorem at play. Rearranging gives Equation (A.13). Similarly,

$$\frac{d\mathcal{U}_\theta}{dB} = v_\theta'(B) \cdot F(\varepsilon_\theta^*) + 0 + (1-\psi)\kappa_\theta(\Lambda)\frac{\partial F(\varepsilon_\theta^*)}{\partial B}$$

and there is no final term because MI_θ is independent of B. □

These first order effects imply new MVPF formulas for the welfare effect of changing benefits and barriers. The fiscal externalities are unchanged - since they depend on behaviour only. However, the direct welfare effects reflect Equations (A.13) and (A.14).

Corollary A.1. *With bias:*

$$MVPF_{d\Lambda} = \frac{-\psi \cdot \int \lambda \cdot \frac{\kappa'(\Lambda)}{v'(B)} \mathbb{P}[SA] d\mu + (1-\psi) \cdot \int \lambda \frac{\kappa(\Lambda)}{v'(B)} \frac{\partial \mathbb{P}[SA]}{\partial \Lambda} d\mu}{\int FE \cdot \frac{\partial \mathbb{P}[SA]}{\partial \Lambda} d\mu} \tag{A.15}$$

$$MVPF_{dB} = \frac{\int \lambda \mathbb{P}[SA] d\mu + (1-\psi) \cdot \int \lambda \frac{\kappa(\Lambda)}{v'(B)} \frac{\partial \mathbb{P}[SA]}{\partial B} d\mu}{\int FE \cdot \frac{\partial \mathbb{P}[SA]}{\partial B} d\mu} \tag{A.16}$$

Appendix B. Context and Data

This section contains summary statistics about the data - comparing the general population to those eligible for social assistance. Pseudocode for my calculation of eligibility is presented in Algorithm 1

Algorithm 1 Eligibility Calculation

Procedure CalculateIncome(*calculation_type*)
  if (*calculation_type* == "Yearly")
    Income = income from work, assets & benefits.
    Deduct taxes & national insurance contributions
  else if (*calculation_type* == "Monthly")
    Gross Income = monthly employment income (spolis).
      Gross Income $\mapsto$ **Add** yearly income from business, assets, sickness/disability benefits $/12$
      Gross Income $\mapsto$ **Add** unemployment benefits over periods with no employment income
    Deductions = payroll taxes + national insurance contributions + employee insurance contributions
      Deductions $\mapsto$ **Add** other taxes (not on bijstand income)

Procedure DefineFamilies()
  Households = as in household income data (`rinpersoonkern`).
  Co-residents = people living at same address
  Families = ≤ 2 adult *Co-Residents* in same *Household*, plus children.

Procedure CostSharing()
  Cost-sharers = adults
  Remove students (age 21-30) *not* receiving student grants
  Threshold = `threshold` (# *Cost-sharers* in *Family*)
  Add norm-adjustment for all singles pre-2015.

Procedure CheckEligibility()
  Set Eligible = "Yes" if *Income* $\leq$ *Threshold*, *wealth* $\leq$ *wealth limit*, and *house value* $\leq$ *house limit*.
  Set Eligible = "No" if age < 21 `or` striking `or` living outside NL or in institutional hh `or` {age 21-27 student not receiving student grants}

Socio-economic Demographics	General Population	Eligible
Gender (%)		
Woman	49.9	53.8
Man	50.1	46.2
Education (%)		
Primary School	5.4	26.7
High School	31.8	46.8
Bachelor's	14.3	6.0
Masters-PhD	8.5	2.6
Unknown	40.1	17.9
Main source of Income (%)		
Employment or Civil Service Job	63.2	8.9
Director-shareholder	2.2	0.1
Self-employment	9.9	4.6
Other Job	0.2	0.0
Unemployment Insurance	2.0	2.5
Disability Insurance	5.5	6.5
Social Assistance	4.3	55.3
Other Benefits	1.9	12.9
Pension	3.8	1.3
Student Aid	0.6	3.3
Other (not active or without income)	6.1	4.7
Household Composition (%)		
Single person household	17.8	45.6
Couple without children	26.8	11.1
Couple with children	45.1	20.1
Single parent	6.4	19.6
Couples and parents with flatmates	2.1	1.9
Other shared households	1.0	1.6
Other Information		
Age	46.4 (11.0)	45.0 (11.3)
Foreign-born (%)	16.4 (37.0)	42.5 (49.4)
Household Std. Disposable Income (€)	66,949.4 (73,978.0)	13,125.2 (2,795.6)
Household Net Worth (€)	169,760.0 (4,227,453.1)	-5,497.5 (85,933.0)
Contracted Hours (per year)	1,509.7 (602.6)	471.1 (451.0)
Eligible (%)	6.6 (24.8)	100.0 (0.0)
Receipt of Social Assistance (%)	5.1 (21.9)	60.0 (49.0)

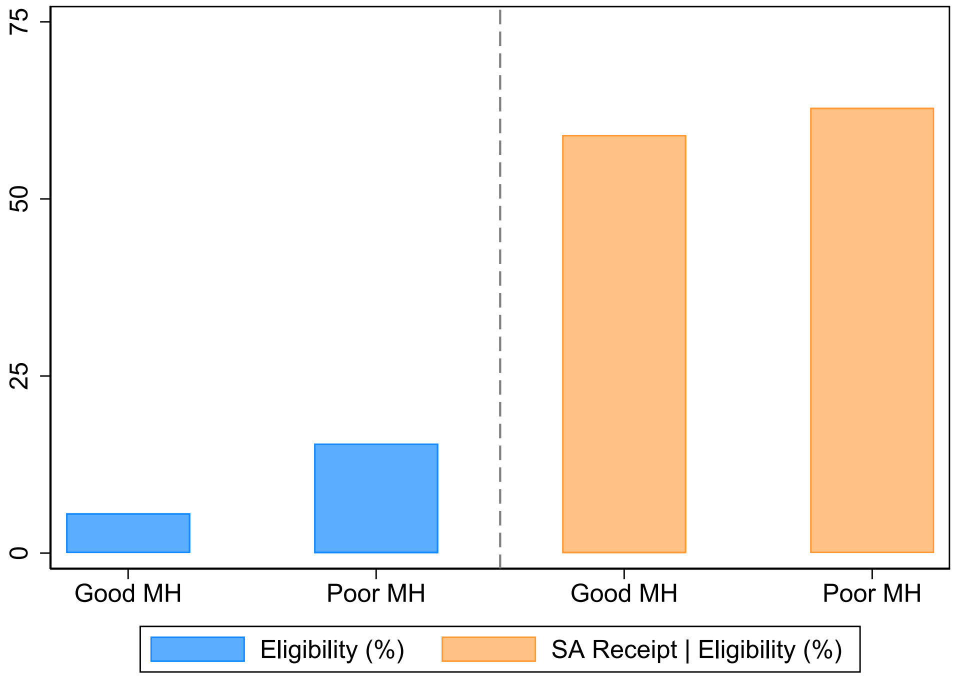
TABLE B.1. Summary Statistics for General and Eligible Populations

(Mental) Health Information	General Population Mean (SD)	Eligible Mean (SD)
General Health		
All Care Spending (€)	2,037.4 (7,181.0)	3,711.6 (11,015.0)
Physical Chronic Conditions (count)	0.67 (1.13)	1.03 (1.44)
Mental Health (admin)		
Mental Healthcare Spending (€)	274.3 (3,237.2)	1,055.9 (6,892.6)
Psychotropic Medication (%)	10.3 (30.3)	24.7 (43.1)
Anti-psychotics (%)	2.1 (14.4)	8.4 (27.7)
Anxiolytics (%)	2.2 (14.7)	8.0 (27.1)
Anti-depressants (%)	7.6 (26.6)	16.1 (36.7)
Hypnotics and Sedatives (%)	1.2 (11.1)	4.5 (20.7)
ADHD Medication (%)	0.7 (8.5)	1.7 (12.8)
Mental Health Hospitalizations (%)	0.05 (2.1)	0.12 (3.5)
Deaths by Suicide (%)	0.01 (1.2)	0.05 (2.3)
Mental Health (survey)		
Loneliness (0-11)	2.64 (3.14)	5.51 (3.82)
Life Control (7-35)	27.13 (5.09)	22.36 (5.72)
Kessler-10 Psychological Distress (10-50)	15.69 (6.43)	22.24 (9.82)

TABLE B.2. Summary Statistics for General and Eligible Populations

Appendix C. Descriptive Patterns in Take-Up by Mental Health: Additional Material

Figure C.1. Eligibility and receipt of social assistance by mental health



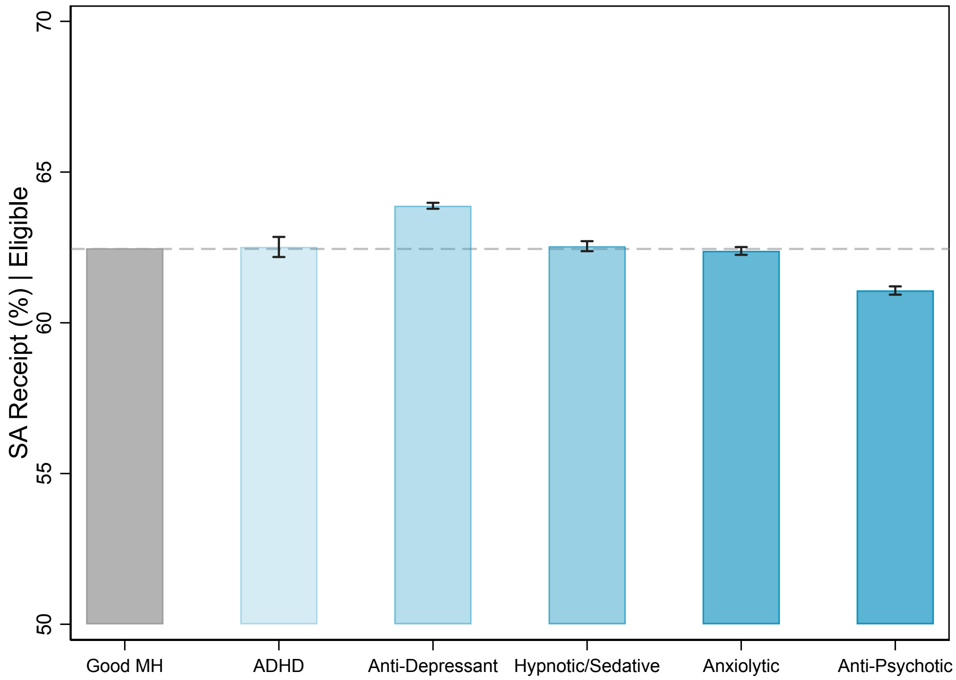
Notes: $\mathbb{P}$[Eligible] and $\mathbb{P}[SA|\text{Eligible}]$, compared for people with poor mental health (dispensed psychopharma in year previously) vs good mental health (not). Underlying population: 2011-2020 in each case.

	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\beta}$: Receipt of SA poor vs good MH (p.p.)	0.540***	1.695***	1.767***	0.533	2.513***	0.191
	(0.071)	(0.086)	(0.527)	(0.575)	(0.673)	(0.450)
Year, age and gender FEs	✓	✓	✓	✓	✓	✓
Lagged household disposable income controls	✓	✓	✓	✓	✓	✓
Lagged work-status FEs	✓	✓	✓	✓	✓	✓
Individual FEs						
All other controls	✓	✓	✓	✓	✓	✓
Observations (people-years)	5,187,572	5,162,351	14,402	12,718	6,514	3,690,830
R^2	0.650	0.650	0.690	0.695	0.746	0.639
Baseline mean	62.45	62.66	64.34	63.89	64.71	62.78

Standard errors in parentheses
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

TABLE C.1. Coefficients of social assistance take-up regressed on mental health status–indicators of: psychotropic drugs (1), mental healthcare (2), severe surveyed psychological distress (3)/loneliness (4)/lack of control over own life (5), or mental health hospitalisation (6). Point estimates and standard errors shown. Lagged controls include household disposable income, wealth, education, work status, household composition, municipality, year, age, sector fixed effects, physical health, and benefits schedule. Eligible population from 2011 to 2020 (2011-2017 for hospitalisations). Around 2% of the general population are surveyed. Standard errors clustered at the municipality level.

FIGURE C.2. Targeting of social assistance with respect to different mental health conditions



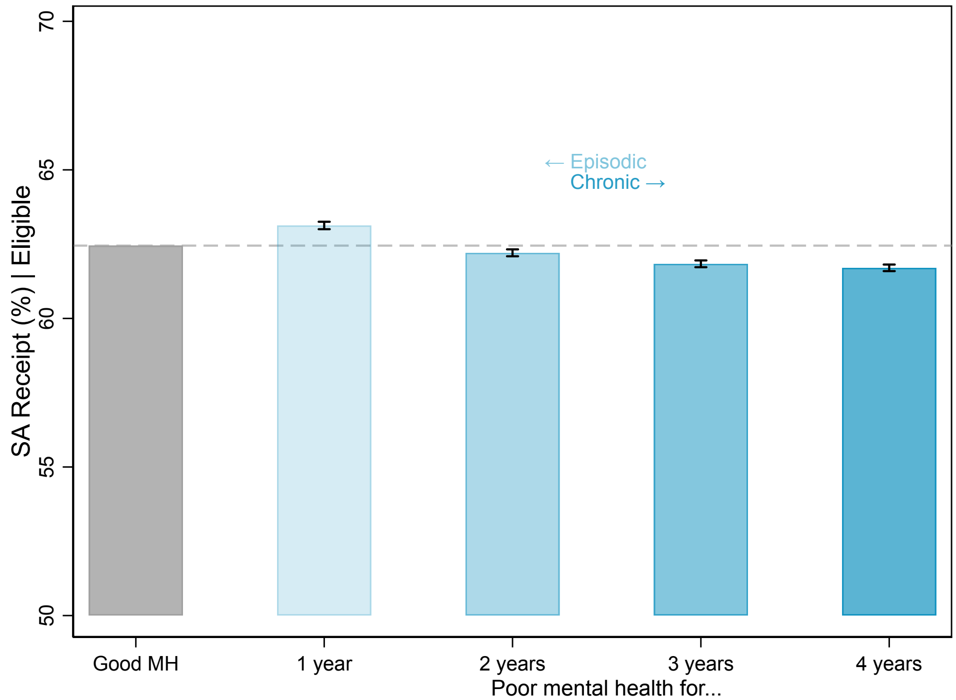
Notes: Coefficients of social assistance take-up regressed on psychopharmacology dispensation fixed effects (by type: ADHD medications, anti-depressants, hypnotics/sedatives, anti-anxiety medications and anti-psychotics). Point estimates added to the control mean, with 95% confidence intervals. Lagged controls include household disposable income, wealth, education, work status, household composition, municipality, year, age, sector fixed effects, physical health, and benefits schedule. Eligible population from 2011 to 2020. Standard errors clustered at the municipality level.

$\hat{\beta}$: SA receipt regressed on $\mathbb{1}\{$Dispensed psychotropic drug$\}$, coefficients relative to good mental health (no dispensation) (p.p.)	(1)
ADHD	0.0459
	(0.170)
Anti-Depressant	1.412***
	(0.0506)
Hypnotic/Sedative	0.0719
	(0.0845)
Anxiolytic	-0.0859
	(0.066)
Anti-Psychotic	-1.399***
	(0.0701)
Year, age and gender FEs	✓
Lagged household disposable income controls	✓
Lagged work-status FEs	✓
Individual FEs	
All other controls	✓
Observations (people-years)	5,187,572
R^2	0.650
Baseline mean	62.45

Standard errors in parentheses
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

TABLE C.2. Coefficients of social assistance take-up regressed on psychopharmacology dispensation fixed effects (by type: ADHD medications, anti-depressants, hypnotics/sedatives, anti-anxiety medications and anti-psychotics). Point estimates added to the control mean, with 95% confidence intervals. Lagged controls include household disposable income, wealth, education, work status, household composition, municipality, year, age, sector fixed effects, physical health, and benefits schedule. Eligible population from 2011 to 2020. Standard errors clustered at the municipality level.

FIGURE C.3. Targeting of social assistance to different durations of poor mental health



Notes: Coefficients of social assistance take-up regressed on psychopharmacology dispensation fixed effects (by duration). Point estimates added to the control mean, with 95% confidence intervals. Lagged controls include household disposable income, wealth, education, work status, household composition, municipality, year, age, sector fixed effects, physical health, and benefits schedule. Eligible population from 2011 to 2020. Standard errors clustered at the municipality level.

TABLE C.3. Differential targeting on lagged observables

	Hours-worked	Working*	Student*	Debt	Age	High School*	Couple*
Poor MH	-0.327***	-0.189***	-0.0424***	-0.0292***	0.0425**	-0.0310***	-0.0838***
	(0.0131)	(0.00776)	(0.00286)	(0.00156)	(0.0130)	(0.00380)	(0.00342)
SA	-0.310***	-0.213***	-0.0576***	-0.0308***	0.109***	-0.121***	-0.0471***
	(0.0206)	(0.0104)	(0.00462)	(0.00132)	(0.0104)	(0.00717)	(0.00384)
Poor MH × SA	0.219***	0.165***	0.0381***	0.0282***	0.0266*	0.0259***	0.0189***
	(0.0125)	(0.00772)	(0.00268)	(0.00160)	(0.0104)	(0.00402)	(0.00351)
Observations	5,206,819	5,206,819	5,206,819	5,160,527	5,206,819	4,184,533	5,206,325
R^2	0.075	0.131	0.044	0.051	0.051	0.024	0.039

	Savings	Consumption	Wealth	Man*	Foreign*	Physical Health	Children	Benefits*
Poor MH	-0.00293***	-0.0117***	0.00338***	-0.0437***	-0.103***	0.625***	-0.0682***	0.300***
	(0.000395)	(0.000761)	(0.000429)	(0.00649)	(0.00747)	(0.0202)	(0.00418)	(0.0105)
SA	-0.0109***	0.00260***	0.00139***	-0.0924***	0.172***	0.120***	0.184***	0.382***
	(0.000438)	(0.000778)	(0.000275)	(0.00376)	(0.00716)	(0.0139)	(0.00648)	(0.0142)
Poor MH × SA	0.00392***	0.00250	-0.00200***	-0.00224	-0.0131	-0.0368*	-0.109***	-0.270***
	(0.000371)	(0.00129)	(0.000426)	(0.00354)	(0.0110)	(0.0144)	(0.00671)	(0.0101)
Observations	5,160,527	5,008,768	5,160,527	5,206,819	5,206,819	5,206,819	4,720,094	5,206,819
R^2	0.019	0.007	0.007	0.024	0.052	0.058	0.069	0.258

Notes: Standard errors in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. This table shows estimates from Equation (4.2) for different lagged outcome variables. Consumption is proxied in Appendix C.1, physical health = number of dispensed medicines for physical chronic conditions (Danesh et al., 2024) and benefits = 1 if receive any benefits (social assistance, disability, sickness or unemployment). All outcomes are z-scored except for Man, Foreign, Couple, High School, Working, Benefits, and Student which are discrete variables. The interaction term (Poor MH × SA) captures differential selection on observable characteristics for recipients with poor mental health relative to the baseline difference among eligible non-recipients and is plotted in Figure IIIa. Lagged household disposable income controls. Eligible population from 2011 to 2020. Standard errors clustered at the municipality level.

TABLE C.4. Differential targeting on future observables | lagged observables

	Consumption	Hours-worked	Debt	All HC Costs	Psychopharma (p.p.)
Poor MH	-0.0139***	-0.156***	-0.0313***	0.333***	0.598***
	(0.00104)	(0.0046)	(0.00369)	(0.0071)	(0.00254)
SA	-0.00714***	-0.0248*	-0.0370***	0.00824*	0.00697***
	(0.000803)	(0.0105)	(0.0034)	(0.00356)	(0.000734)
Poor MH $\times$ SA	0.00378*	0.0443***	0.0275***	-0.109***	-0.0385***
	(0.00159)	(0.00626)	(0.0032)	(0.00496)	(0.00108)
Observations	3,865,248	4,598,094	4,008,905	4,056,491	4,598,094
R^2	0.003	0.325	0.113	0.17	0.613

Notes: Standard errors in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. This table shows estimates $\hat{\phi}$ from Equation (4.3) for different future outcome variables. Consumption is proxied in Appendix C.1, "all hc costs" means all healthcare expenditures under basic insurance. All outcomes except Psychopharma (p.p.) are z-scored. The interaction term (Poor MH $\times$ SA) captures differential outcomes for recipients with poor mental health relative to the baseline difference among eligible non-recipients and is plotted in Figure IIIb. Lagged household disposable income controls. Eligible population from 2011 to 2020. Standard errors clustered at the municipality level.

C.1. **Consumption.** To what extent is poor mental health proxying for lower consumption conditional on household disposable income, and thus driving take-up? In this section, I propose a test of this hypothesis. First, I construct a rudimentary measure of consumption as follows. Since financial wealth (savings minus debt) amount to 93% of household assets for this population ($\approx 3\%$ of those eligible for social assistance own property), then I can reasonably proxy:

$$C_{it} = Y_{it} - [F_{it} - F_{it-1}] \tag{C.1}$$

where C is consumption, Y income and F financial wealth. Then:

$$\Delta F_{it} = \Delta S_{it} - \Delta D_{it} + I^E_{it} - I^O_{it} \tag{C.2}$$

where S is savings, D debt, I^E interest earned and I^O interest owed. The ECB fixed interest rate was $< 1\%$ in this time period (European Central Bank, 2024), therefore I approximate $I^E = 0$ for this low-income population. 65% of those eligible for social assistance have non-0 debt. The majority of this (76%) is credit card debt, and minority student (18%) and mortgage (6%) debt. I assume a 1% student debt interest rate,[27] 3% mortgage rate,[28] and an 8% credit interest rate.[29]

How does consumption vary by mental health? People using psychotropic drugs have on average €1,136 lower consumption than those who are not. This difference is mostly driven by economic status: conditional on household composition, education, municipality FEs and wealth and work-status controls, people with poor mental health have €416 lower consumption each year. This difference is fully driven by people with poor mental health taking on lower debt (possibly due to borrowing constraints) than those with good mental health (with the same controls, €371 lower debt).

However, this does not explain the take-up of social assistance by mental health. Table C.5 shows that the addition of consumption as a control does not meaningfully affect the average targeting of social assistance with respect to mental health.

	(1)	(2)	(3)	(4)
$\hat{\beta}$: Receipt of SA poor vs good MH (p.p.)	0.540***	0.679***	-0.222	-0.382
	(0.071)	(0.0641)	(0.397)	(0.390)
Year, age and gender FEs	✓	✓	✓	✓
Lagged household disposable income controls	✓	✓	✓	✓
Consumption controls FEs		✓		✓
All other controls	✓	✓		
Observations (people-years)	5,187,572	4,990,349	5,206,819	5,008,768
R^2	0.650	0.656	0.142	0.143

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE C.5. Results of a regression of receipt of social assistance on mental health status (measured by dispensation of psychotropic meds). (1) vs (2) shows the results with/without controlling for consumption, while maintaining the full set of controls from Section 4. (3) vs (4) shows the same comparison, but only additionally controlling for year, age and gender FEs as well as lagged household disposable income. The sample contains the calculated eligible for SA in 2011-2020. Standard-errors are clustered at the municipality-level.

[27]This is a conservative estimate. The following article states that interest rates were at 0% for years before 2023: https://www.utoday.nl/news/74664/interest-rate-on-student-debt-remains-more-or-less-the-same

[28]https://www.dnb.nl/en/statistics/dashboards/residential-mortgages/bank-mortgage-lending-rates/

[29]https://www.statista.com/statistics/597483/fixed-interest-rate-less-than-1-year-new-consumer-credit-netherlands/

Appendix D. Barrier Screening Effects: Additional Material

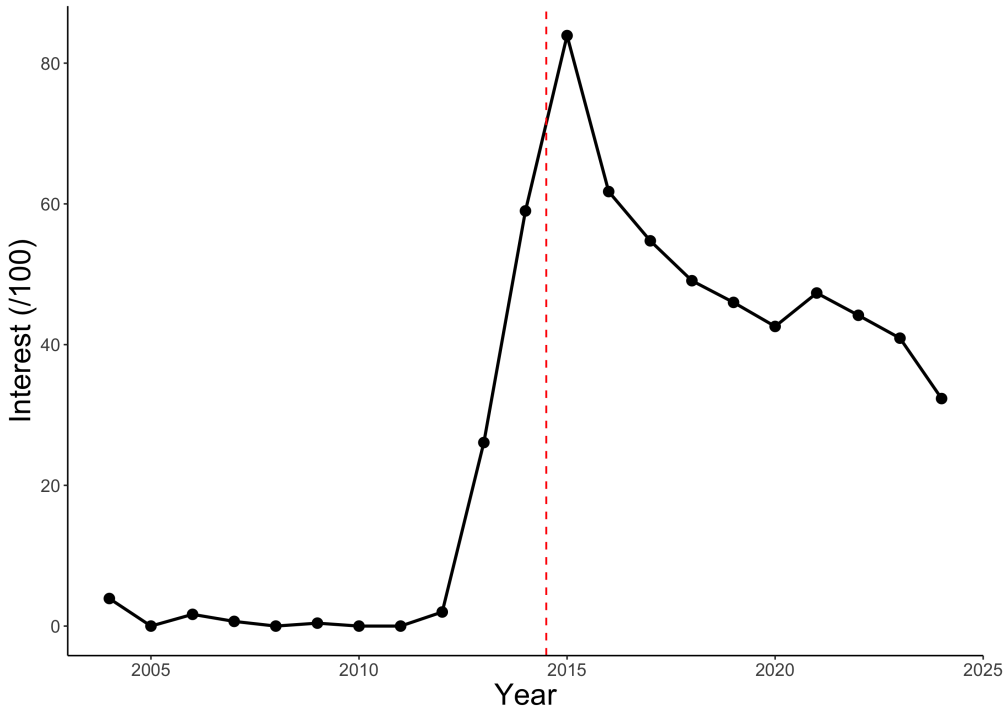


FIGURE D.1. Google Trends for *Participatiewet*, the Dutch translation of the "Participation Act" over time in the Netherlands.

D.1. **Detail on the Participation Act.** The Participation Act increased barriers to accessing social assistance. All of the major obligations were intensified. The full list is: achieving a good command of the Dutch language (made necessary in 2016), work re-integration, register as a job seeker, being willing to relocate municipality, being prepared to travel a distance with a total travel time of 3 hours per day to find work, acquiring and retaining knowledge and skills necessary for acquiring wealth, wearing the correct clothing in work/volunteering, and quid-pro-quo unpaid voluntary work. The policy also incentivised municipalities to reduce caseload which they did via (threat of) sanctions. This is supported by qualitative evidence:

- SCP (2019) state "Since the introduction of the Participation Act, extra attention has been paid to better compliance with the obligations associated with benefit entitlement, such as looking for and accepting work, the language requirement and the quid pro quo."

- van der Veen (2019) argues "the Participation Act introduced a much stricter regime of entitlement conditions, involving mandatory participation in 're-integration' activities [... and] introduced an important element of workfare, the so-called 'quid-pro-quo'".
- Additionally, "the Participation Act is complex and the fraud policy is strict for the time being. The focus on legitimacy could deter people, encourage non-use and thereby potentially undermine the effectiveness of general assistance" (Inspectie SZW, 2021).
- In their report entitled "Coping with the Participation Act: Welfare experiences in the Netherlands", Knijn and Hiah (2019) conclude that "boundary drawing as well as some exclusionary mechanisms hamper redistributive and recognitive justice as experienced by social assistance clients".
- In "Limits of Freedom? Experiences with the Participation Act" (Bierbaum and Gassmann, 2016), state that "obligation [can be unrealistic given] individual and the labour market situation. [...] obligations and related controls could conflict with feelings of privacy and freedom. [...] stress or panicking could be the result of obligations."

Quantitatively, SCP (2019), a report evaluating the Participation act, contains the results from a survey of 80 municipalities which asked representatives how often they impose obligations, and for each type of obligation how many impose these more after the introduction of the Participation Act. An overview of the results are shown in Table D.1.

Table D.1 shows that many municipalities say they intensified the various obligations. No surveyed municipality said that they imposed obligations less often.

There are plans in the Netherlands to repeal the Participation Act. Ministerie van SZW (2022) makes the case for a "Participation Act in Balance". The authors work with (former) social assistance recipients, municipalities and other experts to suggest that the obligations associated with the 2015 are too strict. They state:

> "Applying for social assistance is experienced by various experts as complex, tedious and too long. A negative tone [by the municipality] is also mentioned, threatening action from the outset and a creating a sense of mutual distrust. At the same time, citizens experience a high degree of dependence on the government. A feeling of shame prevails that they have to make use of social assistance, even though in situations they simply cannot (temporarily) do otherwise. People definitely understand the need for monitoring and enforcement, but the way in which this is done now is drastic. A small event can have major consequences. People do not

Obligation	Percent of Impose	Percent More Since PA15
Language	76.5	69.4
Work	93.8	26.3
Accept Jobs	95.1	19.5
Register	48.1	20.5
Move	13.6	54.5
Commute 3 hours	29.6	50.0
Acquire skills	75.3	24.6
Clothes	63.0	49.3
Quid-pro-quo	87.7	56.8

TABLE D.1. Percentage of municipalities surveyed who impose the full list of obligations (Column 2) and who impose obligations more often since the Participation Act (Column 3). The different obligations in-full are: achieving a good command of the Dutch language, work re-integration, register as a job seeker, being willing to relocate municipality, being prepared to travel a distance with a total travel time of 3 hours per day to find work, acquiring and retaining knowledge and skills necessary for acquiring wealth, wearing the correct clothing in work/volunteering, and quid-pro-quo unpaid voluntary work.

always feel heard or treated as an equal person. Fear also arises. This can create a barrier to applying for assistance, even when the need is great."[30]

D.2. **Results.**

[30] Translated from page 8 of Ministerie van SZW (2022)

$\hat{\delta}_t$ (p.p.) for different specifications	Figure IV	Figure V		Figure VIa		Figure VIb			
	FE + Ctrls	Inflow	Outflow	Anti-dep.	Anti-psych.	Pharma	Survey	Care	Hosp.
2011	-0.210	-0.0274	0.00182	0.312	0.303	0.0818	-0.795	-0.825**	-1.446
	(0.208)	(0.385)	(0.00170)	(0.309)	(0.525)	(0.206)	(1.402)	(0.287)	(1.851)
2012	-0.0451	-0.0118	-0.000374	0.249	0.351	0.159	-1.379	-0.382	0.272
	(0.154)	(0.371)	(0.00161)	(0.283)	(0.381)	(0.163)	(1.094)	(0.208)	(2.096)
2014	-0.292	-0.693*	0.000169	-0.212	-0.511	-0.372*	-0.367	-0.337	-0.557
	(0.150)	(0.330)	(0.00164)	(0.244)	(0.282)	(0.163)	(1.140)	(0.216)	(1.445)
2015	-1.021***	-1.539***	0.000696	-0.854*	-1.567***	-1.062***	-2.517	-1.042***	-3.109
	(0.171)	(0.401)	(0.00141)	(0.369)	(0.318)	(0.189)	(1.354)	(0.227)	(1.961)
2016	-0.984***	-0.998**	0.000818	-0.777*	-1.509***	-1.120***	-4.028**	-1.263***	-1.382
	(0.195)	(0.378)	(0.00154)	(0.304)	(0.371)	(0.231)	(1.380)	(0.259)	(2.488)
2017	-1.094***	-1.359***	-0.00122	-1.028**	-1.614***	-1.316***	-3.784*	-1.535***	-0.726
	(0.232)	(0.366)	(0.00164)	(0.319)	(0.402)	(0.244)	(1.850)	(0.288)	(2.418)
2018	-0.954***	-1.242**	-0.00496**	-0.773*	-1.672***	-1.191***	-3.408	-1.568***	-1.045
	(0.239)	(0.388)	(0.00152)	(0.392)	(0.390)	(0.260)	(1.772)	(0.282)	(2.666)
2019	-1.004***	-0.965*	-0.00344*	-1.078**	-1.772***	-1.257***	-2.022	-1.999***	-3.291
	(0.260)	(0.439)	(0.00166)	(0.350)	(0.427)	(0.269)	(1.812)	(0.355)	(2.874)
2020	-1.191***	-1.452***	-0.00370*	-1.430***	-1.863***	-1.440***	-3.738	-2.168***	-3.885
	(0.301)	(0.418)	(0.00174)	(0.398)	(0.409)	(0.339)	(1.977)	(0.332)	(2.908)
Time-varying controls	✓	✓	✓	✓	✓				
Observations (people-years)	982,749	393,415	589,334	982,749	982,749	1,036,353	15,007	1,036,353	1,036,353
R^2	0.022	0.025	0.020	0.018	0.018	0.004	0.008	0.004	0.003

Standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE D.2. Main results for Section 5.
Notes: Coefficient estimates and standard errors plotted in Figures IV through VIb. For more details see those graphs.

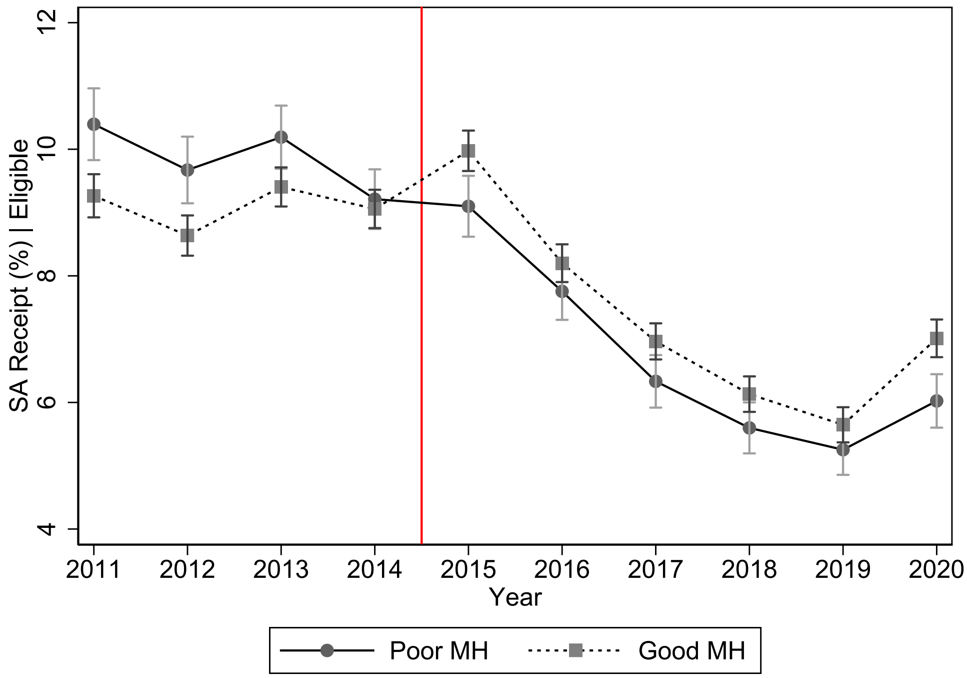


FIGURE D.2. Evolution of inflow of social assistance over time, split by people with poor mental health in the pre-period vs those with good mental health in this period. Raw means and respective 95% confidence intervals are shown. The introduction of the Participation Act in 2015 is shown by the red vertical line. Standard-errors are clustered at the level of municipality of residence in 2013.

$\hat{\delta}_t$ (p.p.) for robustness checks	Figure D.4		Figure D.5	Figure D.3	Figure D.7	Figure D.8	Figure D.11	Figure D.12
	FE	FE + Ctrls	Interacted Ctrls	Always Elig.	Younger	+ WMO	MH Always	Buffer
2011	0.0818	-0.210	-0.031	-0.398	-0.292	-0.0343	0.0226	0.0370
	(0.206)	(0.208)	(0.208)	(0.295)	(0.190)	(0.206)	(0.384)	(0.213)
2012	0.159	-0.0451	0.061	-0.141	0.0471	0.0383	0.128	0.135
	(0.163)	(0.154)	(0.152)	(0.209)	(0.149)	(0.155)	(0.271)	(0.167)
2014	-0.372*	-0.292	-0.286	-0.203	-0.387**	-0.330*	-0.586*	-0.307
	(0.163)	(0.150)	(0.161)	(0.157)	(0.146)	(0.157)	(0.231)	(0.173)
2015	-1.062***	-1.021***	-0.987***	-0.514**	-1.016***	-1.036***	-1.049**	-0.983***
	(0.189)	(0.171)	(0.176)	(0.189)	(0.172)	(0.175)	(0.327)	(0.196)
2016	-1.120***	-0.984***	-1.017***	-0.698**	-1.117***	-1.085***	-1.327***	-0.992***
	(0.231)	(0.195)	(0.218)	(0.234)	(0.175)	(0.210)	(0.369)	(0.243)
2017	-1.316***	-1.094***	-1.170***	-0.753***	-1.059***	-1.211***	-0.829*	-1.204***
	(0.244)	(0.232)	(0.244)	(0.220)	(0.219)	(0.236)	(0.395)	(0.218)
2018	-1.191***	-0.954***	-1.036***	-0.712**	-0.906***	-1.079***	-0.654	-0.989***
	(0.260)	(0.239)	(0.258)	(0.249)	(0.231)	(0.247)	(0.442)	(0.240)
2019	-1.257***	-1.004***	-1.061***	-0.702**	-0.944***	-1.136***	-0.658	-1.026***
	(0.269)	(0.260)	(0.278)	(0.256)	(0.216)	(0.266)	(0.487)	(0.242)
2020	-1.440***	-1.191***	-1.273***	-0.764**	-0.991***	-1.336***	-0.588	-1.184***
	(0.339)	(0.301)	(0.333)	(0.280)	(0.279)	(0.312)	(0.514)	(0.307)
Time-varying controls		✓	✓	✓	✓	✓	✓	✓
Observations (people-years)	1,036,353	982,749	982,749	321,793	1,340,950	982,749	518,967	937,016
R^2	0.004	0.022	0.006	0.014	0.022	0.018	0.019	0.003

Standard errors in parentheses

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE D.3. Robustness for Section 5.

Notes: Coefficient estimates and standard errors plotted in Figures D.3 through D.12. For more details see those graphs.

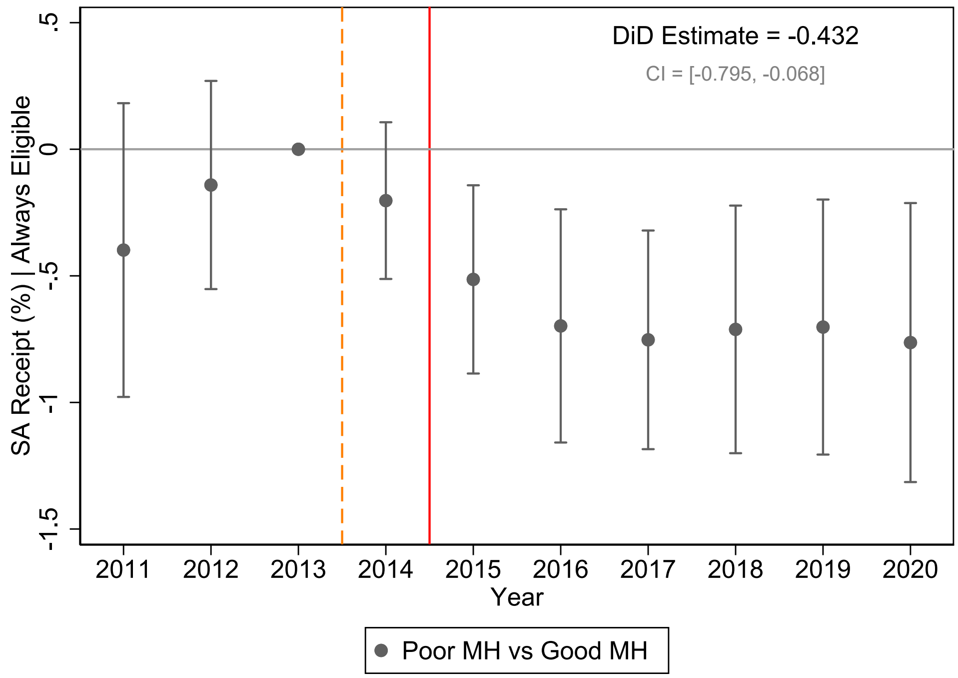


FIGURE D.3. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is always-eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the pre-period. Controls include individual fixed effects, household disposable income, education and muncipality, hh composition and sector fixed effects. The TWFE estimate $\hat{\delta}$ in the regression $SA_{it} = \alpha + \eta_i + \gamma_t + \delta \cdot \mathbb{1}\{t \geq 2013\} \times \text{Poor MH}_i + X_{it}'\theta + \varepsilon_{it}$ is also shown. Standard-errors are clustered at the level of municipality of residence in 2013.

Formally, the sample-selection issue can be framed as follows. Let $\mathbf{e}_i = (e_{i1}, ...e_{iT})$ where $e_{it} \in \{0, 1\}$ denotes eligibility. Let $\mathbb{X}_{it}$ be all explanatory variables (and $\mathbf{X}_i$ similarly). Essentially, we only "observe" ($\mathbf{X}_{it}$, SA_{it}) for i, t such that $e_{it} = 1$ - i.e. only these observations are included in the regression. Wooldridge (2019) shows that the necessary identification assumption in this setting is given by Equation (D.1).

$$\mathbb{E}[\varepsilon_{it} \big| \mathbf{X}_i, \eta_i, \mathbf{e}_i] = 0 \tag{D.1}$$

However, note that eligibility is a (non-linear) function of observables: $e_{it} := \phi(y_{it}, \bar{y}_i, ...)$. Therefore, controlling for $y_{it}, \bar{y}_i$ etc implies that selection is fully determined by observables. I.e. the standard assumption $\mathbb{E}[\varepsilon_{it} \big| \mathbf{X}_i, \eta_i] = 0$ is sufficient. In this case, it is particularly important to check that the time-varying controls are not driving the results.

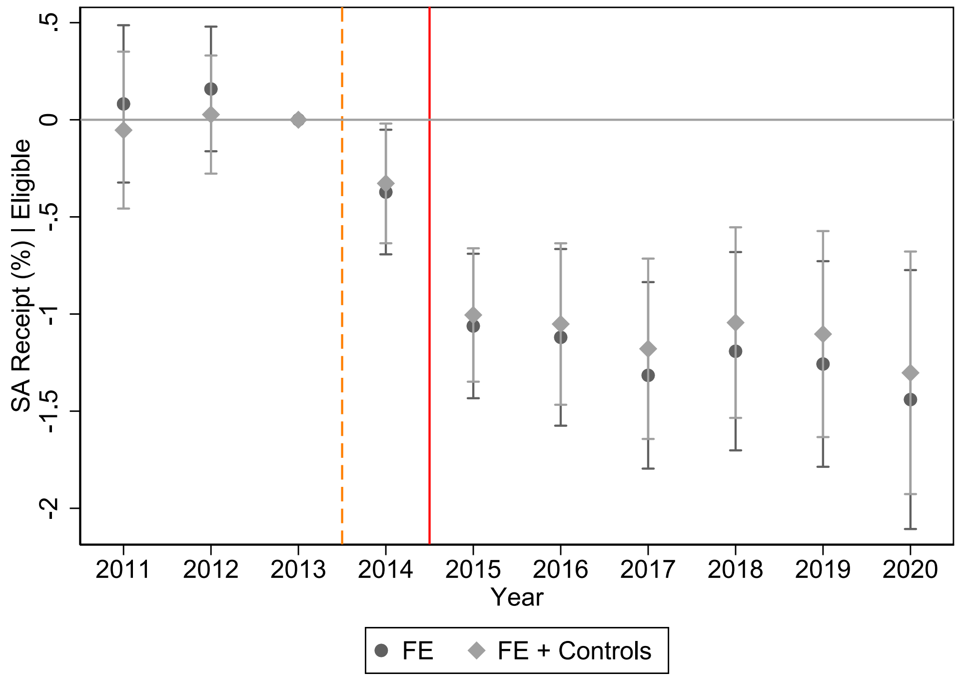


FIGURE D.4. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the pre-period. $\hat{\delta}_t$ are shown for two specifications - one with no time-varying controls (only individual FEs), and one with all time-varying controls - individual fixed effects, household disposable income, education and muncipality, hh composition and sector fixed effects. Standard-errors are clustered at the level of municipality of residence in 2013.

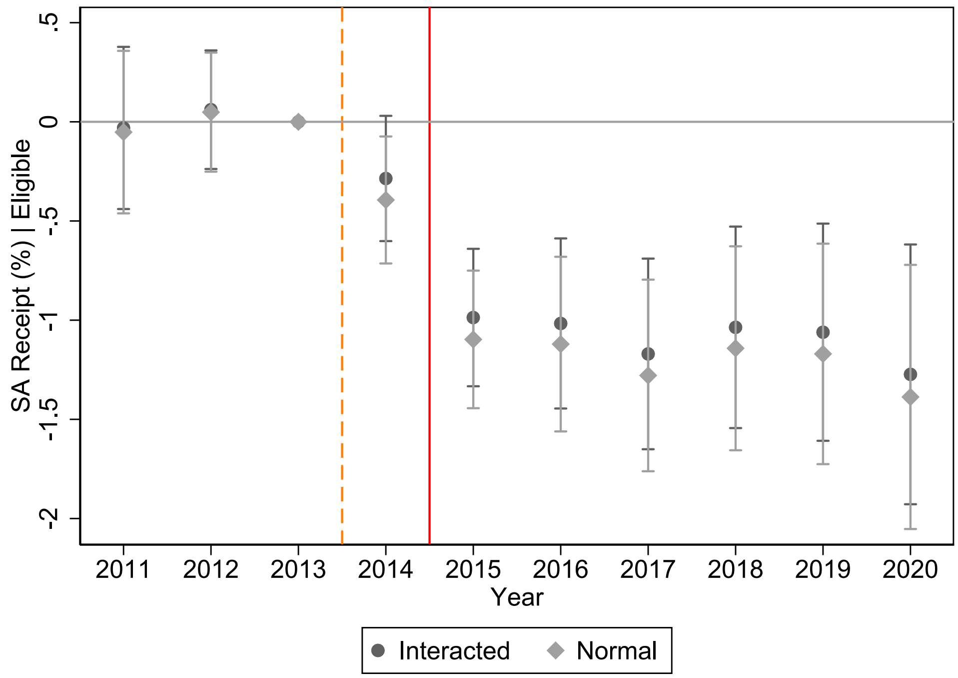


FIGURE D.5. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the pre-period. $\hat{\delta}_t$ are shown for two specifications - one with all time-varying controls - individual fixed effects, household disposable income, education and muncipality, hh composition and sector fixed effects and one where all of these controls have been interacted with a post-policy dummy. Standard-errors are clustered at the level of municipality of residence in 2013.

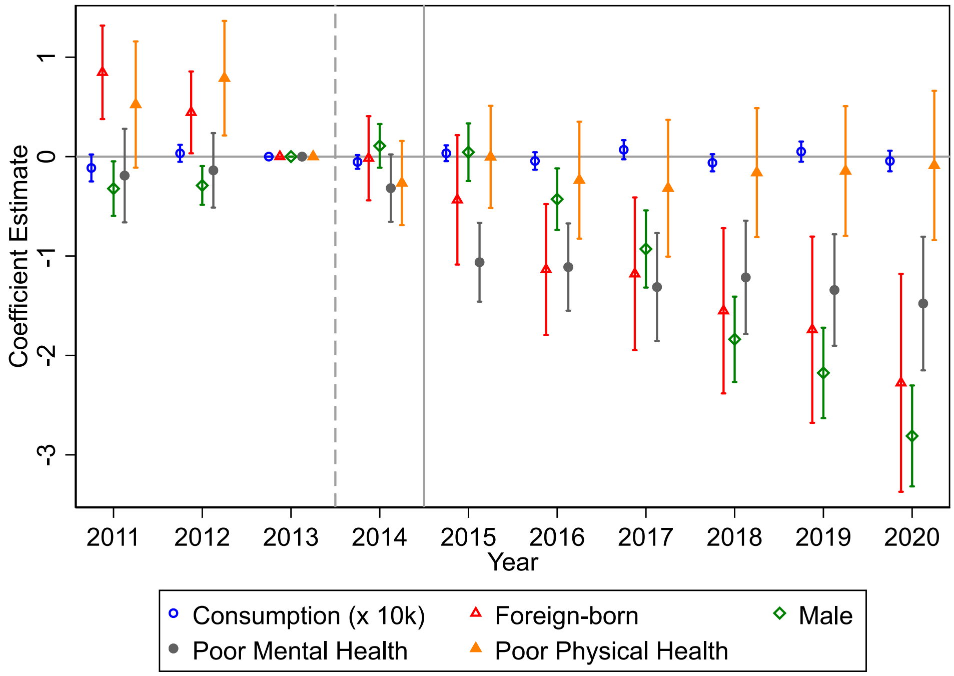


FIGURE D.6. Estimates $\hat{\delta}_{jt}$ from $SA_{it} = \alpha + \eta_i + \gamma_t + \sum_j \delta_{jt} \times X_{ji} + \varepsilon_{it}$ for X_{ji} = (consumption, gender, $\mathbb{1}$\{foreign-born\}, Poor MH$_i$ $\mathbb{1}$\{physical health condition\}). Other time-varying controls are omitted to avoid over-saturation. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the pre-period. Standard-errors are clustered at the level of municipality of residence in 2013.

Poor mental health is the dimension of vulnerability most strongly screened-out by the increase in barriers induced by the Participation Act. While men have larger effects by the end of the study period, these only emerge from 2016 onwards. Foreign-born also shows large effects but exhibits strong pre-trends, undermining causal interpretation.

Year	Consumption (×10,000)	Foreign	Male	Poor MH	Poor PH
2011	0.0114	0.848***	-0.322*	-0.191	-0.120
	(0.0069)	(0.241)	(0.143)	(0.240)	(0.274)
2012	0.0033	0.445*	-0.290**	-0.138	-0.073
	(0.0044)	(0.181)	(0.113)	(0.191)	(0.221)
2014	-0.0054	-0.017	0.108	-0.317	0.021
	(0.0035)	(0.166)	(0.105)	(0.173)	(0.207)
2015	-0.0034	-0.435	0.044	-1.063***	-0.063
	(0.0041)	(0.228)	(0.121)	(0.202)	(0.240)
2016	-0.0044	-1.136***	-0.428**	-1.111***	-0.120
	(0.0045)	(0.329)	(0.147)	(0.224)	(0.272)
2017	-0.0069	-1.179**	-0.929***	-1.312***	-0.094
	(0.0049)	(0.377)	(0.162)	(0.250)	(0.299)
2018	-0.0062	-1.551***	-1.838***	-1.258***	-0.124
	(0.0044)	(0.356)	(0.179)	(0.253)	(0.309)
2019	-0.0051	-1.741***	-2.176***	-1.115***	-0.145
	(0.0052)	(0.405)	(0.197)	(0.272)	(0.333)
2020	-0.0044	-2.276***	-2.810***	-1.287***	-0.089
	(0.0053)	(0.467)	(0.254)	(0.299)	(0.383)
Observations	857,206	857,206	857,206	857,206	857,206
R^2	0.006	0.006	0.006	0.006	0.006

Standard errors in parentheses
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

TABLE D.4. Estimates from Figure D.6

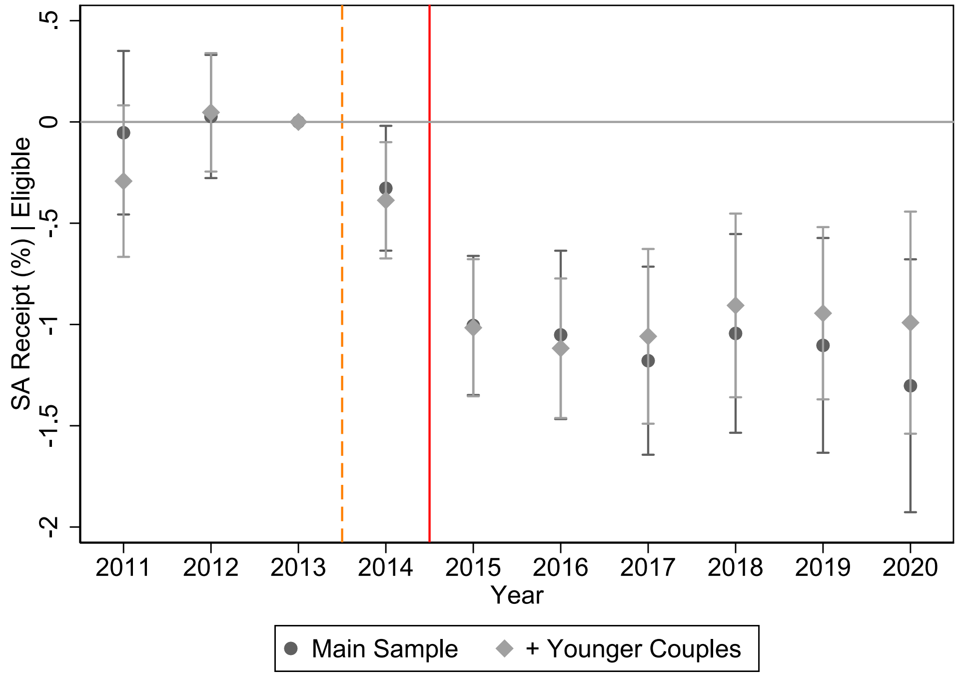


FIGURE D.7. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the pre-period. $\hat{\delta}_t$ are shown for two specifications - one with the standard analysis population, and the other with additionally including younger couples. Controls include individual fixed effects, household disposable income, education and muncipality, hh composition and sector fixed effects. Standard-errors are clustered at the level of municipality of residence in 2013.

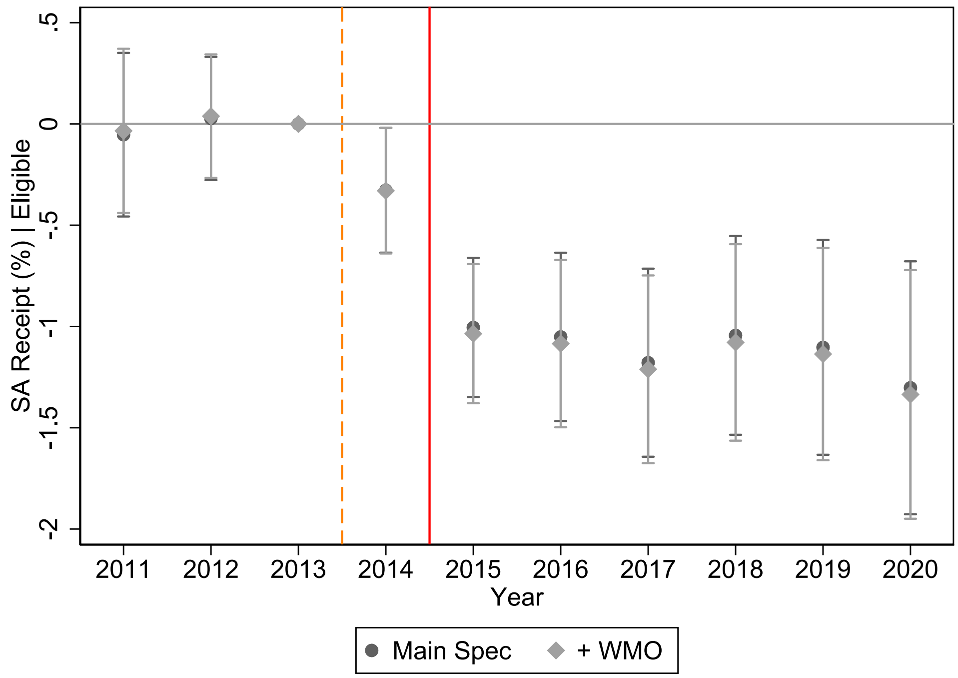


FIGURE D.8. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the pre-period. $\hat{\delta}_t$ are shown for two specifications - one with the main time-varying controls–individual fixed effects, household disposable income, education and muncipality, hh composition and sector fixed effects–and one adding controls for usage of home support via the WMO. Standard-errors are clustered at the level of municipality of residence in 2013.

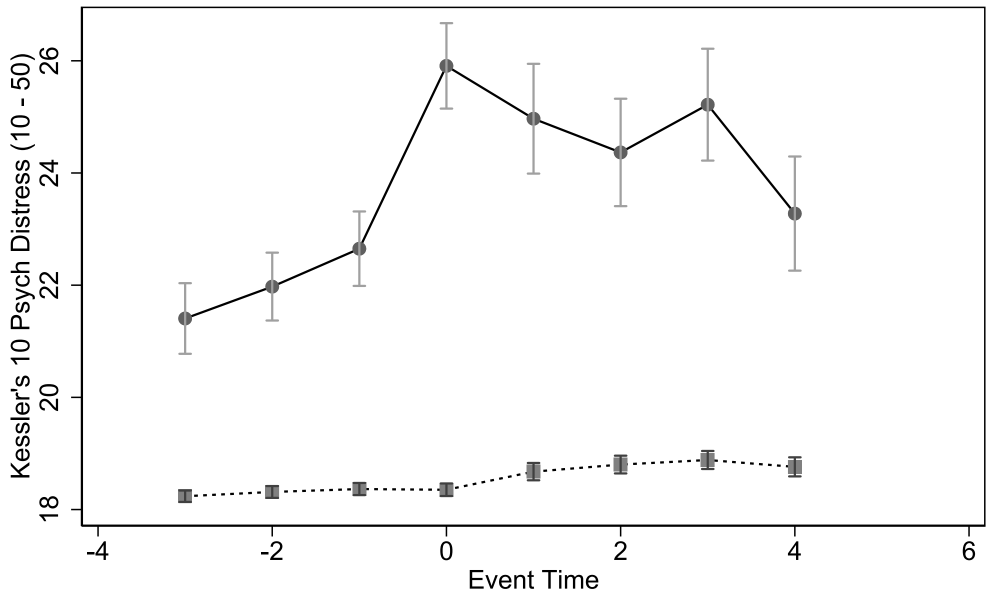


FIGURE D.9. This plot shows the mean subjective mental health (measured by Kessler's 10 Psychological Distress) for two groups: one group is prescribed psychopharma for the first time in Event Time 0, the other group has no prescriptions for all event times $t \leq 0$. Standard-errors are clustered at the level of municipality of residence in 2013.

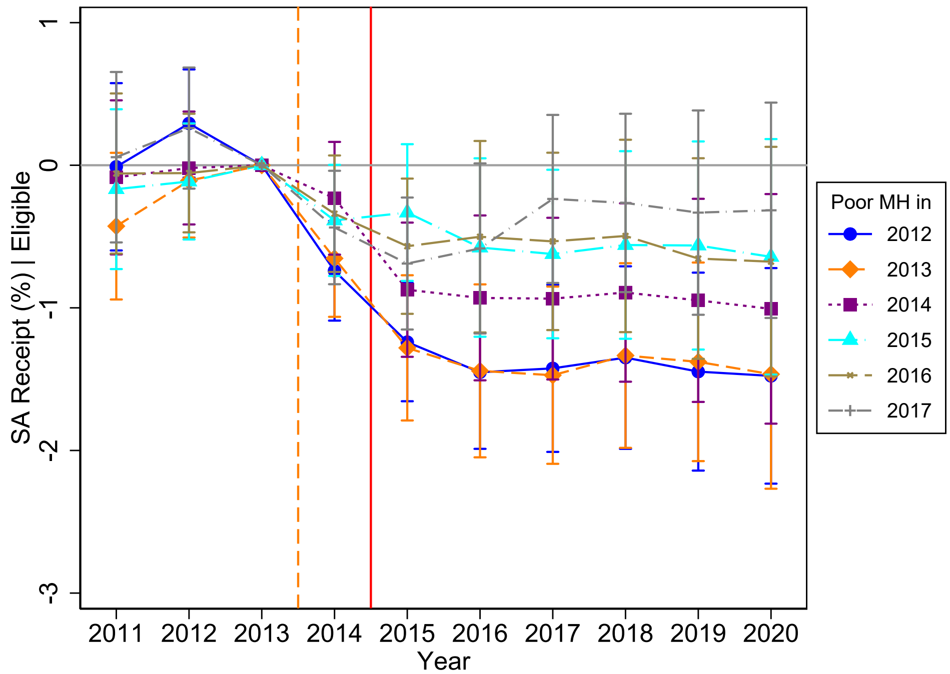


FIGURE D.10. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the pre-period. $\hat{\delta}_t$ are shown changing the definition of Poor MH_i. The different definitions are Poor $\text{MH}_i = \mathbb{1}\{\text{Prescribed Psychopharma in year } y\}$, showing estimates for $y \in \{2012, ..., 2017\}$. Controls include individual fixed effects, household disposable income, education and muncipality, hh composition and sector fixed effects. Standard-errors are clustered at the level of municipality of residence in 2013. Coefficient Estimates are contained in Table D.5

$\hat{\delta}_t$ (p.p.) for different timing of MH diagnosis	Figure D.10: Poor MH$_i$ $= \mathbb{1}\{\text{Psychopharma}\}_i$ in year:					
	2012	2013	2014	2015	2016	2017
2011	-0.0108	-0.427	-0.0850	-0.168	-0.0581	0.0564
	(0.298)	(0.261)	(0.275)	(0.285)	(0.286)	(0.304)
2012	0.293	-0.108	-0.0193	-0.114	-0.0547	0.261
	(0.193)	(0.203)	(0.201)	(0.207)	(0.211)	(0.216)
2014	-0.740***	-0.653**	-0.232	-0.387	-0.340	-0.437*
	(0.178)	(0.208)	(0.201)	(0.198)	(0.208)	(0.202)
2015	-1.240***	-1.281***	-0.873***	-0.332	-0.568*	-0.689**
	(0.211)	(0.259)	(0.239)	(0.244)	(0.241)	(0.235)
2016	-1.452***	-1.442***	-0.930**	-0.577	-0.501	-0.583
	(0.273)	(0.308)	(0.294)	(0.318)	(0.341)	(0.303)
2017	-1.424***	-1.472***	-0.935**	-0.623*	-0.534	-0.236
	(0.298)	(0.316)	(0.288)	(0.300)	(0.316)	(0.300)
2018	-1.349***	-1.334***	-0.892**	-0.559	-0.496	-0.264
	(0.325)	(0.329)	(0.318)	(0.334)	(0.343)	(0.318)
2019	-1.447***	-1.378***	-0.948**	-0.563	-0.654	-0.332
	(0.353)	(0.354)	(0.362)	(0.371)	(0.358)	(0.364)
2020	-1.477***	-1.465***	-1.007*	-0.642	-0.676	-0.316
	(0.384)	(0.408)	(0.409)	(0.420)	(0.409)	(0.384)
Time-varying controls	✓	✓	✓	✓	✓	✓
Observations (people-years)	656,049	661,794	659,773	658,504	654,347	648,909
R^2	0.018	0.018	0.018	0.019	0.019	0.019

Standard errors in parentheses
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

TABLE D.5. Coefficient Estimates of Figure D.10

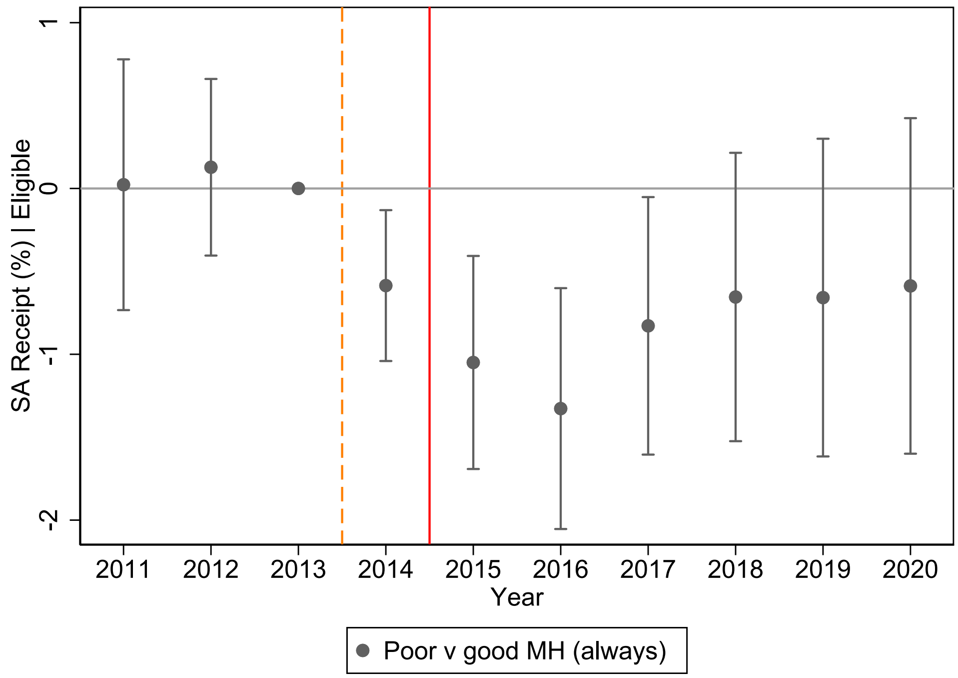


FIGURE D.11. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by Poor MH_i: dispensed psychopharma in *every* year 2011-2020, vs good mental health throughout. Controls include individual fixed effects, household disposable income, education and muncipality, hh composition and sector fixed effects. Standard-errors are clustered at the level of municipality of residence in 2013.

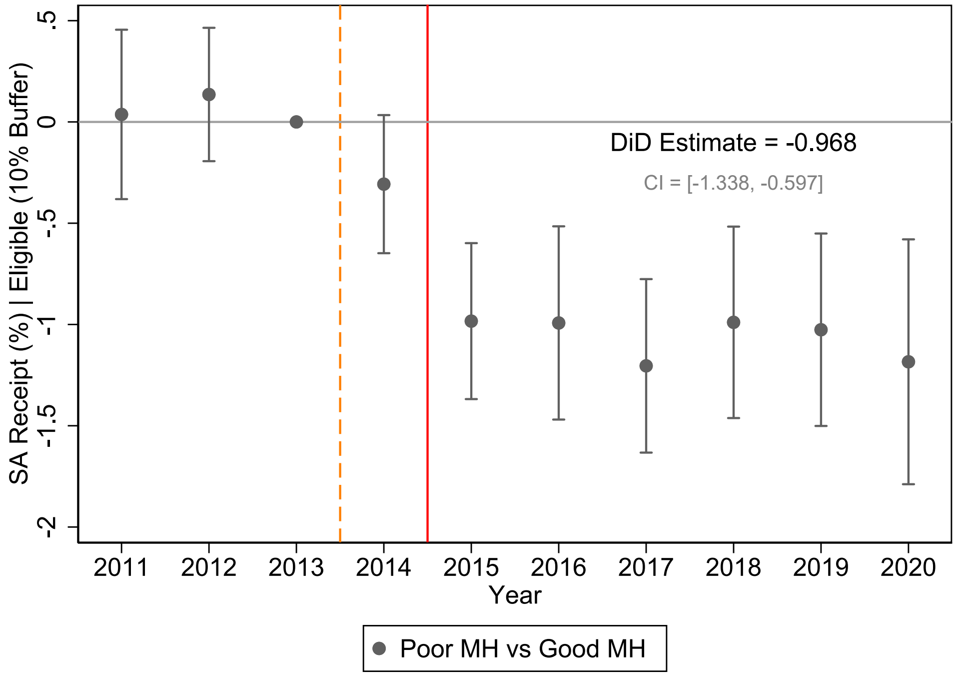


FIGURE D.12. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the pre-period. $\hat{\delta}$ for the main specification for individuals with eligibility income below 90% of the eligibility threshold. Poor $\text{MH}_i = \mathbb{1}\{$Prescribed Psychopharma in pre-period. Controls include individual fixed effects, household disposable income, education and muncipality, hh composition and sector fixed effects. Standard-errors are clustered at the level of municipality of residence in 2013.

Appendix E. Benefit Take-up Effects: Additional Material

This section contains additional material relating to the RKD estimation of the effect of changes in benefit level on SA receipt (heterogeneously by mental health).

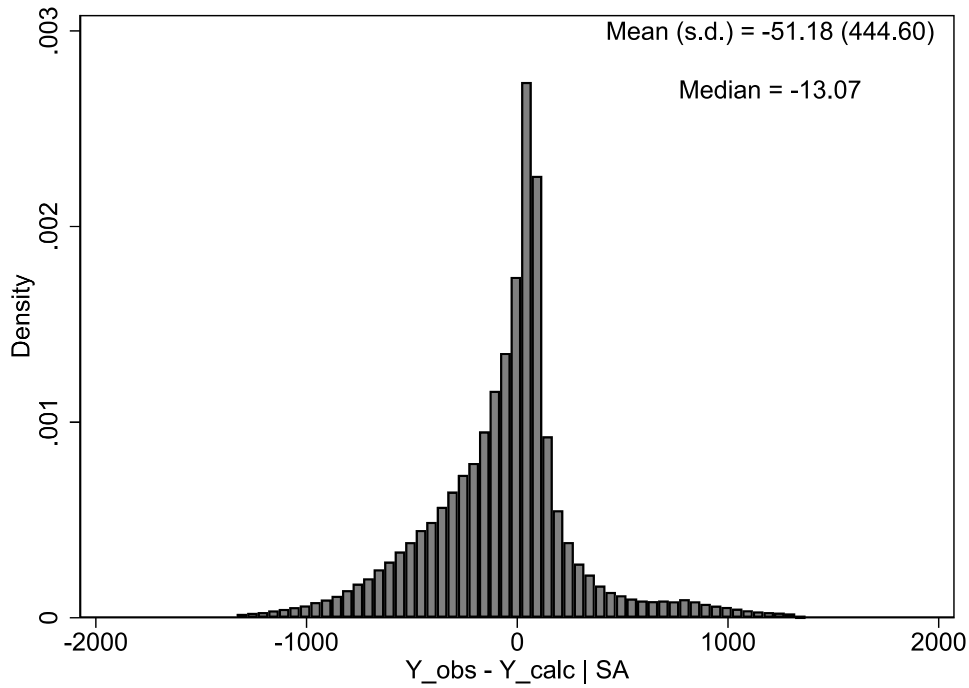


FIGURE E.1. Histogram of $Y_{\text{true}} - Y_{\text{calc}}$ for the analysis population of the RKD.

E.1. **Theory.** I use the generalized non-separable model of Card et al. (2015): receipt of SA is a function of benefit level B, income Y and error term ε: $\mathbb{P}[SA] = p(B, Y, \varepsilon)$. Let I_X be the support of random variable X which is potentially multi-dimensional, in which case represents a product space. Benefit level $B^* = B + U_B$ and income $Y^* = Y + U_Y$ are mis-measured.

Proposition E.1. *(Card et al. (2015)) Under regularity, smooth effect of income, y, first stage and non-negligible population at the kink, smooth density, smooth probability of no measurement error and monotonicity:*

(a) $\mathbb{P}[\varepsilon \leq e, \nu \leq v | Y = y]$ *continuously differentiable in* y^* *at* $y^* = \bar{y}$ $\forall (e, v) \in I_{\varepsilon,\nu}$.

(b)

$$\frac{\lim_{\xi\to\bar{y}^+}\left.\frac{d\mathbb{P}[SA|Y^*=y^*]}{dy^*}\right|_{y^*=\xi}-\lim_{\xi\to\bar{y}^-}\left.\frac{d\mathbb{P}[SA|Y^*=y^*]}{dy^*}\right|_{y^*=\xi}}{\lim_{\xi\to\bar{y}^+}\left.\frac{d\mathbb{E}[B^*|Y^*=y^*]}{dy^*}\right|_{y^*=\xi}-\lim_{\xi\to\bar{y}^-}\left.\frac{d\mathbb{E}[B^*|Y^*=y^*]}{dy^*}\right|_{y^*=\xi}} \tag{E.1}$$
$$=\int\frac{\partial\mathbb{P}[SA\mid B=b(\bar{y},v),Y=\bar{y},\varepsilon=e]}{\partial B}\cdot\varphi(e,v)\,dF_{\varepsilon,\nu}(e,v)$$

where weighting function

$$\varphi(e,v)=\frac{\mathbb{P}[U_Y=0|Y=\bar{y},\varepsilon=e,\nu=v]\big(b_1^+(v)-b_1^-(v)\big)\frac{f_{Y|\varepsilon=e,\nu=v}(\bar{y})}{f_Y(\bar{y})}}{\int\mathbb{P}[U_Y=0|Y=\bar{y},\varepsilon=e,\nu=\omega]\big(b_1^+(v)-b_1^-(v)\big)\frac{f_{Y|\varepsilon=e,\nu=\omega}(\bar{y})}{f_Y(\bar{y})}dF_\nu(\omega)} \tag{E.2}$$

The fuzzy RKD estimates a weighted average of marginal effects of B on $\mathbb{P}[SA]$ with weights $\varphi(e,v)$. The intuition is as follows. $\varphi(e,v)$ has three main components. $\frac{f_{Y|\varepsilon=e,\nu=v}(\bar{y})}{f_Y(\bar{y})}$ is the weight in a sharp RKD and reflects the relative likelihood an individual is located at the kink. $b_1^+(v)-b_1^-(v)$ reflects size of the kink: the fuzzy RKD upweights people with larger kinks. $\mathbb{P}[U_Y=0|Y=\bar{y},\varepsilon=e,\nu=v]$ reflects the probability that the assignment variable is correctly measured at threshold.

The Card et al. (2015) identification assumptions are stated in full in Appendix E.2. Two are key to my setting. (a) the density of Y^* is continuously differentiable at the threshold $\bar{y}$, (b) the benefits-schedule is continuous. This $\implies$ $\mathbb{P}[\text{Exemption}|Y=y]$ continuous at $\bar{y}$, where exemption refers to an 'income exemption'. For more detail, see Appendix E.5.

Figure E.19 and Figure VIII show no evidence for non-smoothness of the distribution of income. Discontinuous $\mathbb{P}[\text{Exemption}|Y=y]$ would imply discontinuous $\mathbb{E}[B|SA,Y=y]$ at the threshold. However, Figure E.5 exhibits no such discontinuity. Moreover, there are no conditions in the law which state income below/above the threshold should be exempted differently.

E.2. Card et al. (2015) assumptions for validity of fuzzy RKD.

(1) **Regularity:** (ε,ν) has bounded support. $p(\cdot,\cdot,\cdot)$ is **continuous** and partially differentiable w.r.t. first and second arguments. $p_1(b,y,e)$ continuous.
(2) **Smooth effect of** Y**:** $p_2(b,y,e)$ is continuous.
(3) **First Stage and Nonnegligible Population at Kink:** $b(y,v)$ continuous and $b_1(y,v)$ continuous apart from at $y=\bar{y}$. Positive mass at kink.
(4) **Smooth Density:** Density of Y is continuously differentiable
(5) **Smooth Probability of No Measurement Error:** $\mathbb{P}[U_Y=0,U_B=0|Y=y,\varepsilon,\nu]$ and partial derivative w.r.t. y is continuous.
(6) **Monotonicity:** Either $b_1^+(v)\geq b_1^-(v)$ for all v or $b_1^+(v)\leq b_1^-(v)$ for all v.

There are two conditions for identification specific to my context worth highlighting: Assumption 1 and Assumption 2

Assumption 1 (No 0-censoring)**.**

(a) Take-up is not censored to 0 below threshold:

$$\forall \mathbb{P}[SA|B = b, Y \leq \bar{y}] > 0 \tag{E.3}$$

(b) Take-up is not censored to 0 above threshold:

$$\exists \Delta > 0 \text{ s.t. } \mathbb{P}[SA|Y = y] > 0 \; \forall y \in [\bar{y}, \bar{y} + \Delta] \tag{E.4}$$

Assumption 2 (Continuous probability of exemption)**.**

$$\mathbb{P}[\text{Exemption}|Y = y] \text{ continuous at } \bar{y} \tag{E.5}$$

Without both parts of Assumption 1, the numerator of the estimand in Equation (E.1) will be 0, while without one part only, regularity is violated. In my sample, around 8% of people receiving social assistance have $Y_{\text{true}} > \bar{y}$. Figure E.2 provides support that $\lim_{B \to 0} \mathbb{P}[SA|B] > 0$. Assumption 2 is a corollary of $b(y, v)$ being continuous.

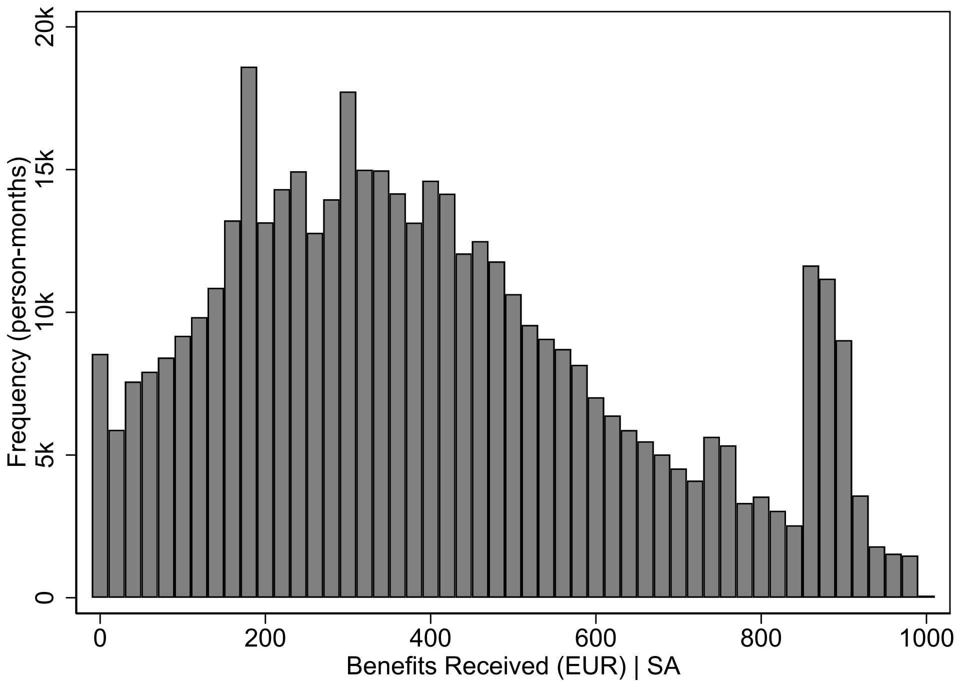
E.3. **Estimation.** I use monthly data for the regression kink design because eligibility is based on the previous month's income, making granular analysis crucial. While the data provide detailed monthly information on labor income and contracted hours, income from other benefits is only available yearly, which motivates my sample restrictions:

Sample Restrictions: I restrict the sample to employees (for whom their monthly income data is higher quality) who are singles before 2015, as misclassification near the threshold is more common for couples, and limit the period after the Participation Act to ensure the analysis is unaffected by changes in ordeals.

Specification: I estimate a standard fuzzy RKD specification, using local linear regression. I use a Calonico et al. (2014) (hereafter, CCT) robust bandwidth of approximately €60. For the CCT bandwidth selection algorithm, I do not use regularization. This is because the CCT framework is not designed to efficiently identify heterogeneous RKDs nor account for measurement error. Both would suggest the use of a larger bandwidth.[31] The non-regularized CCT bandwidth delivers a

[31] Indeed, the CCT robust bandwidth without regularization performs poorly in simulations (see Appendix E.4).

FIGURE E.2. Histogram of true benefits received, near 0



Notes: Plot shows a histogram of benefits received, conditional on receipt, from the administrative records. The non-random censored data issue illustrated in Figure E.5 should be borne in mind. Benefits are monthly. The sample contains single employees, years 2011-2014. Note that the bunching at 0 is driven by individuals staying on the recipient logs for one month as they transition off benefits.

larger bandwidth and has the same asymptotic properties as with regularization. The specification is as follows, where the IV estimate $\frac{\hat{\beta}_1}{\hat{\delta}_1}$ measures $\frac{\partial \mathbb{P}[SA|Y=\bar{y}]}{\partial B}$. I cluster standard-errors at the municipality level.

E.4. **Estimation Choices.** To assess the performance of the CCT robust bandwidth in my context, I perform simulation analyses on a simplified version of the model set out in Section 2. The motivation for these analyses is that the frameworks are not designed for (i) measurement error and (ii) efficiently detecting heterogeneous RKD effects.

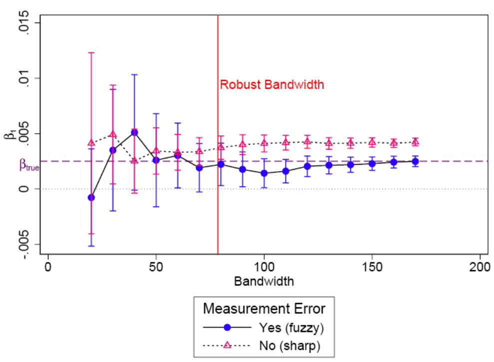
E.4.1. *Setup.* I simulate a million individuals which are characterised by ability $Y \sim U[500, 1500]$. This corresponds to their income. I set a fixed cost to be $\kappa = 150$ for everyone. Choice error $\varepsilon = \frac{U_1+U_2}{2}$ where $U_j \overset{\text{i.i.d.}}{\sim} U[-200, 200]$. I.e. ε follows a symmetric triangular distribution centered around 0. The threshold $\bar{y} = 1000$ for everyone. Benefits schedule $B(y)$ is programmed as $B(y) = \max\{\bar{y} - \nu \cdot y, 0\}$ where exemption $\nu \in \{0, 1\}$ and $\mathbb{P}[\nu = 1|Y = y] \equiv 0.1$. Individual y takes up iff:

$$B(y) \geq \kappa + \varepsilon \tag{E.6}$$

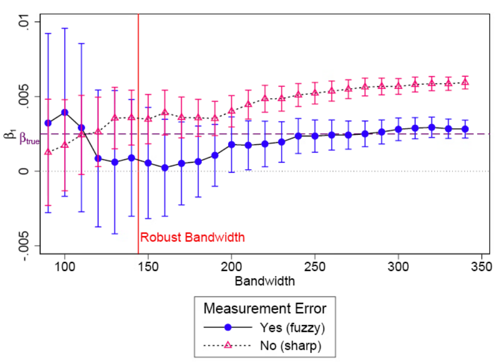
In the case of measurement error, I let $Y^* = Y + U_Y$ where $U_Y \sim N(0, 100)$. I then run the CCT robust bandwidth and RKD analyses exactly as in the main analysis. Specifically, I impute the benefits schedule as in Proposition E.2.

E.4.2. *Results.*

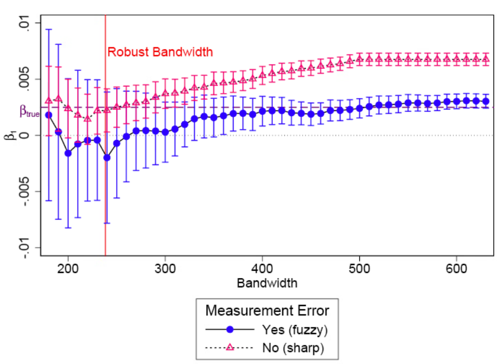
Polynomial order: applying rules-of-thumb from Pei et al. (2022) suggests a linear estimator. Furthermore, simulations show that with measurement-error - linear estimators out-perform higher order polynomials at the CCT robust optimal bandwidth. This result echoes Card et al. (2015) who suggests that the CCT bandwidths can be too small for RKDs.



(A) Local polynomial order $p = 1$



(B) Local polynomial order $p = 2$



(C) Local polynomial order $p = 3$

FIGURE E.3. Results of simulations showing estimates from RKDs using different bandwidths and different local polynomial orders. In each, the CCT robust bandwidth is shown.

Bandwidth: for linear estimation, CCT bandwidths seem to perform well, but estimates become noisy for lower values with measurement error. For the identification of heterogeneous effects under measurement error, CCT performs poorly: I now assume that half of my simulated individuals have value $\alpha = 1$, and half $\alpha = 2$. Individuals take-up iff:

$$\alpha \cdot B(y) \geq \kappa + \varepsilon \tag{E.7}$$

and rate of receipt $\mathbb{P}[SA|Y = y] = F_\varepsilon\big(\alpha \cdot B(y) - \kappa\big)$. I estimate the RKDs separately for $\alpha = 1, 2$ and test for a difference in the RKD estimates at different bandwidths. The estimates are shown in Figure E.4. The plot shows that the CCT bandwidth performs poorly (noisy and biased estimate of the heterogeneous RKD), whereas the estimators converge to the true effect for larger bandwidths.

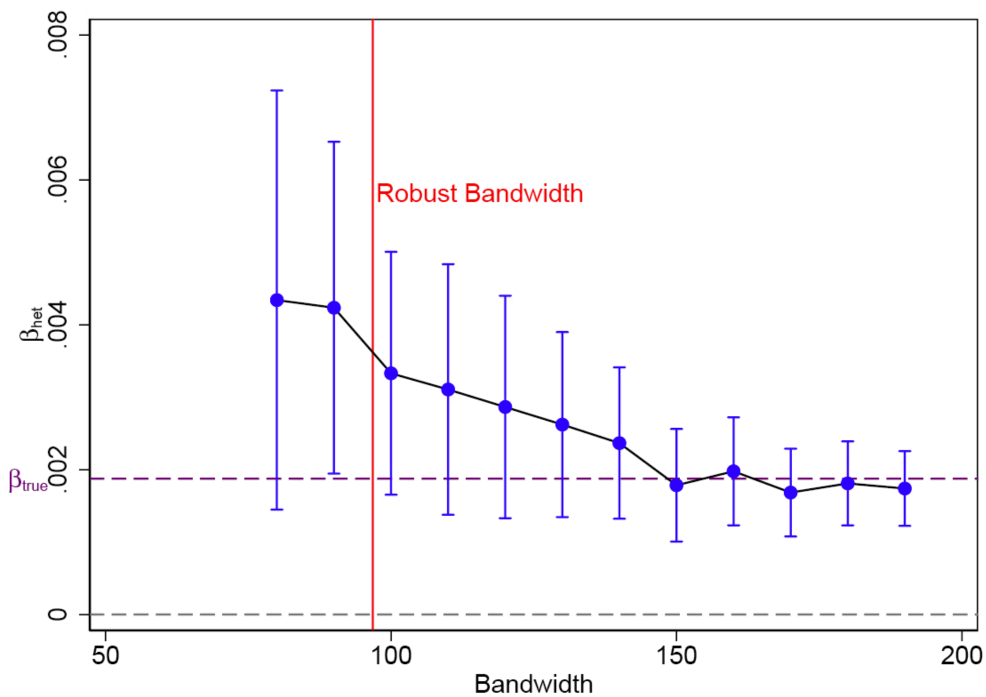


FIGURE E.4. Results of simulations showing estimates from heterogeneous RKD ($\alpha = 1$ vs 2) using different bandwidths. CCT robust bandwidth is shown.

Other: use standard triangular kernel.

E.5. **Income Exemptions.** Municipalities have the option to deviate from the statutory benefits schedule (Figure VIIa) in the form of income exemptions. In this case, income is exempted "insofar as, in the judgment of the [municipality], it contributes to [their] employment opportunities" (Ministerie van SZW, 2015) when calculating benefits that the person is eligible for. This complicates matters because now, B is no longer deterministic (it depends on case-worker leniency) and $\frac{dB}{dy} \neq -1$ necessarily. Let the true benefits schedule be denoted $B = b(Y, \nu)$ where ν captures noncompliance with policy formula due to exemptions.

To properly re-scale the reduced-form estimates, we need to know how B depends on Y ex-ante. However, there is selection into social assistance with respect to exemptions. This makes sense because applicants receive more money with an exemption vs without, holding income fixed. Figure E.5 shows the observed average benefits received conditional on income slice, as well as the probability that this information is missing from the data. The ex-post schedule departs from the statutory one, particularly at and above the threshold. In this region, applicants typically take-up social assistance if they receive an exemption. Selection on exemptions implies ex-post benefits received $\mathbb{E}[B|SA, Y = y]$ is not a good proxy for the ex-ante schedule $\mathbb{E}[B|Y = y]$. Moreover, there is a non-random censored data issue. Case workers appear to be non-randomly censoring their allocated benefits data entry. Indeed, Figure E.6 shows that the likelihood that administrative income data is missing increases substantially around the time when the administrative record of income relative to the threshold starts deviating from my own calculation. Therefore, I need to impute the ex-ante benefits schedule by imputing non-random missing data, as well as imputing the ex-ante benefits schedule from the ex-post.

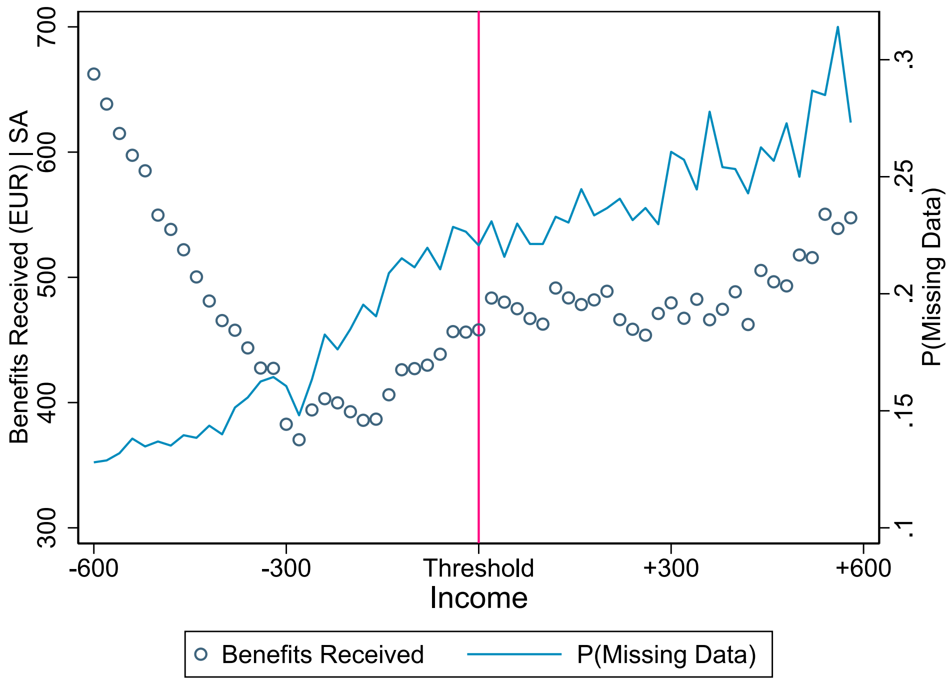
I first focus on imputing the ex-ante benefits schedule from the ex-post one, and leave the issue of missing data to one side by assuming $\mathbb{E}[B|SA, Y = y, \text{Missing}] = \mathbb{E}[B|SA, Y = y, \text{Not Missing}]$.[32] I impute the benefits schedule using a theoretical approach.[33] I recover the ex-ante schedule from the ex-post schedule using Bayes-rule and average receipt. This re-scaling exercise explained in-depth in Appendix E.5. While we may be worried about the endogeneity of using receipt in this calculation, I obtain similar results when I assume a less-flexible form for the probability of exemption - i.e. that it is constant w.r.t. y. In this case, the imputation does not depend on the full take-up function by income.

I model the unobserved benefits schedule as Equation (E.8).

[32] I explore the robustness of the following exercise to different assumptions in Appendix E.5.1.

[33] As in Gelber et al. (2020), the imputation generates measurement error in the first-stage as well, which the Card et al. (2015) framework accounts for.

FIGURE E.5. Ex-post benefits schedule for selected sample



Notes: Plot of benefits received conditional on receipt (and data non-missing), averaged within income slice (€10 bins). A window of €600 either side of the threshold is shown. $\mathbb{P}$[benefits data is missing] is overlaid. Income in this plot is monthly and refers to "eligibility income". The sample contains single employees, years 2011-2014.

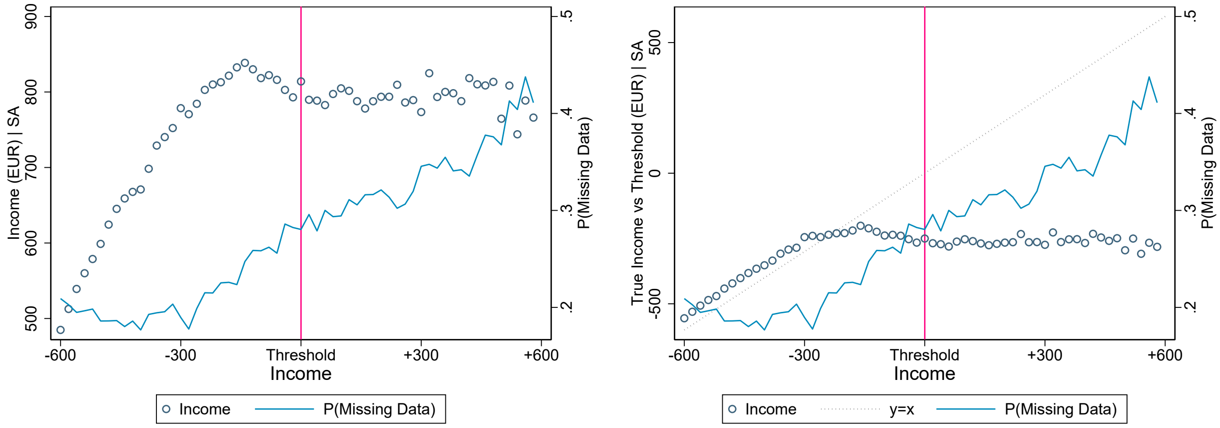
$$B = b(y,\nu) = \begin{cases} \bar{y} & \text{if exemption, } \nu = 1 \\ \max\{\bar{y} - y, 0\} & \text{if exemption, } \nu = 0 \end{cases} \tag{E.8}$$

where $\nu = 1$ with probability $p(y)$. This approach is motivated by the fact that $\mathbb{E}[B|SA, Y = y] \approx \bar{y}$ for $y \geq \bar{y}$. People with income above the threshold are not eligible for any benefits unless they receive an exemption, therefore $\mathbb{E}[B|SA, Y = y]$ is a good measure of benefits received conditional on exemption when $y \geq \bar{y}$. I allow for the possibility that exemptions can vary in reduced-form likelihood throughout the income distribution.

Proposition E.2 (Benefits-Schedule Imputation)**.** *Suppose that the benefits-formula is given by Equation (E.8). Then,* $\mathbb{E}[B|Y = y] = p(y) \cdot \bar{y} + \big(1 - p(y)\big) \cdot \max\{\bar{y} - y, 0\}$ *where:*

$$p(y) = \begin{cases} \frac{\left(\mathbb{E}[B|SA,Y=y] - (\bar{y}-y)\right) \cdot \mathbb{P}[SA|Y=y]}{y \cdot \mathbb{P}[SA|Y=y,\nu=1]} & \textit{if } y \leq \bar{y} \\ \frac{\mathbb{E}[B|SA,Y=y] \cdot \mathbb{P}[SA|Y=y]}{\bar{y} \cdot \mathbb{P}[SA|Y=y,\nu=1]} & \textit{if } y \geq \bar{y} \end{cases} \tag{E.9}$$

FIGURE E.6. Ex-post income (absolute and relative) for selected sample



(A) Income

(B) Income (relative to threshold)

Notes: Plot of administrative records of income (absolute, and relative to threshold) conditional on receipt (and data non-missing), averaged within income (own computation) slice (€10 bins). A window of €600 either side of the threshold is shown. $\mathbb{P}$[benefits data is missing] is overlaid. Income in this plot is monthly and refers to "eligibility income". The sample contains single employees, years 2011-2014.

The proof is a simple application of Bayes-rule.

Proof of Proposition E.2. Let $\mathbb{E}_y := \mathbb{E}[\cdot|Y = y]$ and $\mathbb{P}_y := \mathbb{P}(\cdot|Y = y)$

$$\begin{aligned}
\mathbb{E}[B|SA, Y = y] &= \mathbb{E}_y[B|SA] \\
&= \frac{\mathbb{E}_y[B \cdot \mathbb{1}\{SA\}]}{\mathbb{P}_y[SA]} \\
\mathbb{E}_y[B \cdot \mathbb{1}\{SA\}] &= \mathbb{E}_y[B \cdot \mathbb{1}\{SA\} \cdot \mathbb{1}\{\nu = 1\}] + \mathbb{E}_y[B \cdot \mathbb{1}\{SA\} \cdot \mathbb{1}\{\nu = 0\}] \\
&= \bar{y} \cdot \mathbb{P}_y[SA \cap \nu = 1] + \max\{\bar{y} - y, 0\} \cdot \mathbb{P}_y[SA \cap \nu = 0] \\
&= \bar{y} \cdot \mathbb{P}_y[SA \cap \nu = 1] + \max\{\bar{y} - y, 0\} \cdot \big[\mathbb{P}_y[SA] - \mathbb{P}_y[SA \cap \nu = 1]\big]
\end{aligned}$$

Note that $\mathbb{P}_y[\nu = 1] = p(y)$.

$$= \bar{y} \cdot p(y) \cdot \mathbb{P}_y[SA|\nu = 1] + \max\{\bar{y} - y, 0\} \cdot \left[\mathbb{P}_y[SA] - p(y) \cdot \mathbb{P}_y[SA|\nu = 1]\right]$$

$$= \begin{cases} [\bar{y} - y] \cdot \mathbb{P}_y[SA] + y \cdot p(y) \cdot \mathbb{P}_y[SA|\nu = 1] & \text{if } y \leq \bar{y} \\ \bar{y} \cdot p(y) \cdot \mathbb{P}_y[SA|\nu = 1] & \text{if } y \geq \bar{y} \end{cases}$$

Therefore, $\mathbb{E}_y[B|SA] = \begin{cases} \frac{[\bar{y}-y]\cdot\mathbb{P}_y[SA]+y\cdot p(y)\cdot\mathbb{P}_y[SA|\nu=1]}{\mathbb{P}_y[SA]} & \text{if } y \leq \bar{y} \\ \frac{\bar{y}\cdot p(y)\cdot\mathbb{P}_y[SA|\nu=1]}{\mathbb{P}_y[SA]} & \text{if } y \geq \bar{y} \end{cases}$

Rearranging for $p(y)$ gives the expression in Equation (E.9).

$\square$

Figure E.7 shows the results of the process to impute the ex-ante benefits schedule, heterogeneously by baseline mental health (measured by lagged psychopharma dispensations). People with poor mental health receive more exemptions that those without - presumably because they have larger costs of working and this incentivises the municipality promote re-integration more.

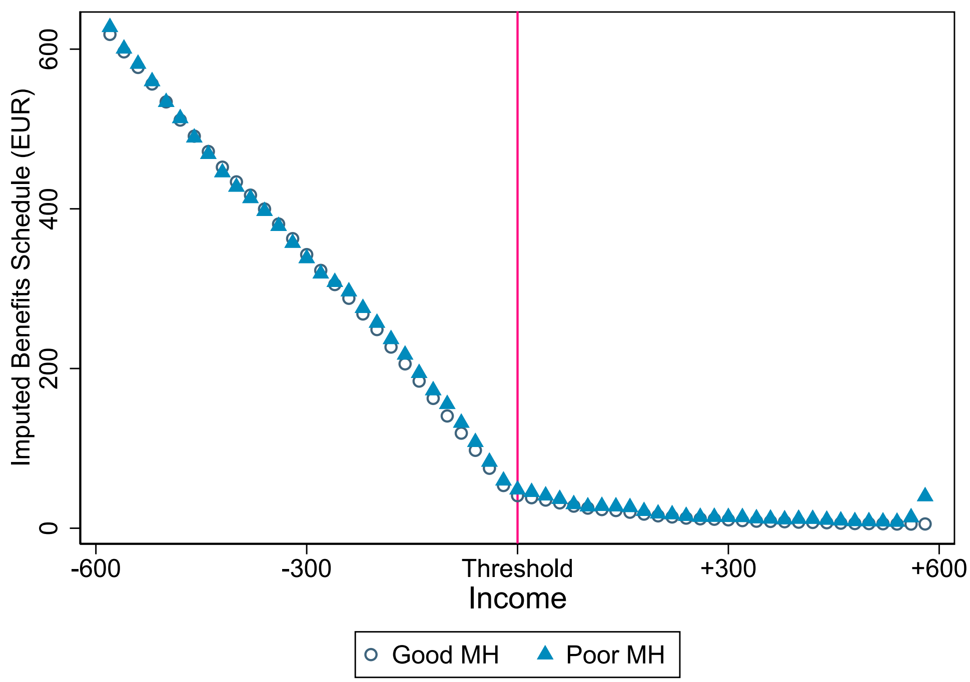


FIGURE E.7. Results of Imputation from Proposition E.2, calculated separately for people dispensed anti-depressants in the year previously (poor mental health) versus those who were not (good mental health).

E.5.1. *Robustness.* The process of imputing the ex-ante benefits schedule due to income exemptions involves several assumptions that could affect the results in theory. The first is that income-exemptions are the cause of the somewhat strange looking ex-post benefits schedule conditional on receipt. The ex-post schedule appears to be kinked around €300 below the threshold. What if this were the true threshold and income exemptions are not part of the explanation? This scenario is inconsistent with various other pieces of evidence.

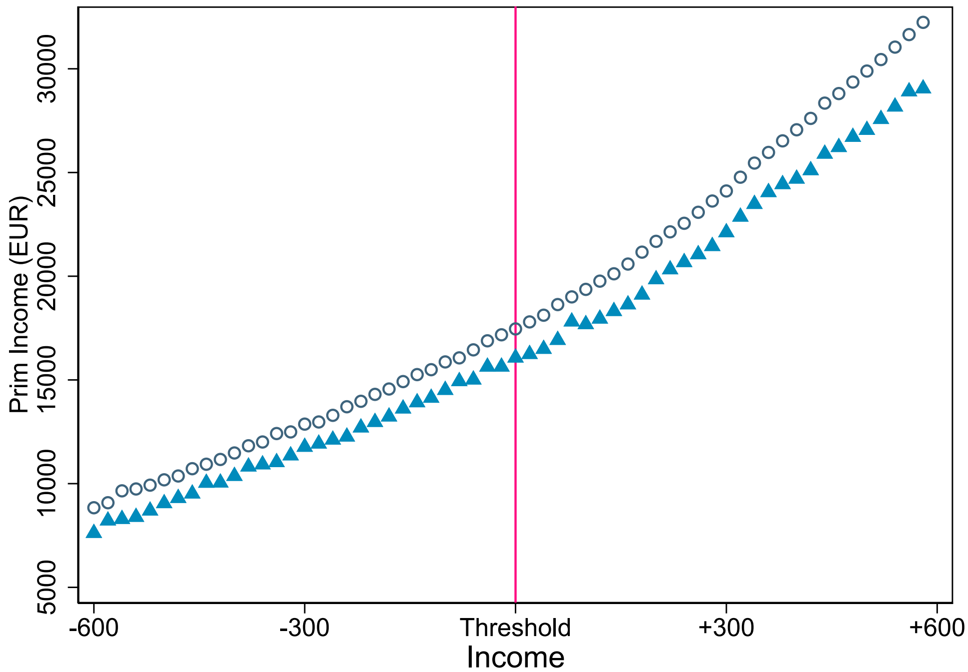
First, average take-up is around 40% at this point, this seems extremely unlikely to be the case if this were the eligibility threshold. Second, the ex-post schedule is increasing above €300 below the threshold. If income-exemptions are not the story, then an increasing ex-post schedule must come from substantial measurement error resulting in calculated income being negatively related to observed income. Note that we cannot test this using reported income conditional on receipt due to the extreme selection issue (data is only present for recipients, and even among recipients there is a non-random censored data issue). Therefore, Figure E.8 plots administrative yearly primary income (observed for everyone without error) by eligibility income concept and shows no evidence for non-monotonicities. Third, at this alternative kink €300 below the threshold, the reduced form RKD results are positive, implying a concave response to a convex kink. This is inconsistent with reasonable behaviour. Finally, note that the imputed ex-ante benefits schedule is not kinked €300 below the threshold. This is not mechanical, any form for the observed ex-post schedule is taken into account in the imputation.

To verify that income-exemptions are indeed sufficient to explain Figure E.5, I plot the ex-post benefits schedule in the simulated data described Appendix E.4, which hard-codes the probability of exemption = 0.1 uniformly, in Figure E.9. This figure confirms that income-exemptions can explain shape of the ex-post schedule.

I also explore the sensitivity to the failure of other assumptions relating to the imputation process. To test sensitivity, I redo the benefits imputation under alternative assumptions and plot below the new ex-ante schedule, as well as comment on how these change the IV estimates for $\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}$. Overall, the process is quite robust.

First, I use the observed (kinked) take-up function $\mathbb{P}[SA|Y = y]$ in the schedule, which could mechanically create a kink in the imputed schedule at the threshold. Figure E.10 shows the results of the imputation assuming that the probability of exemption $p(y)$ is constant in y, $p(y) \equiv \varrho$, allowing me only to use the overall take-up rate $\mathbb{P}[SA]$ (absent any kinks). These first stage kinks do not overturn the finding that $\frac{\partial \mathbb{P}[SA]_L}{\partial B} \gg \frac{\partial \mathbb{P}[SA]_H}{\partial B}$.

FIGURE E.8. Yearly administrative primary income, by eligibility income



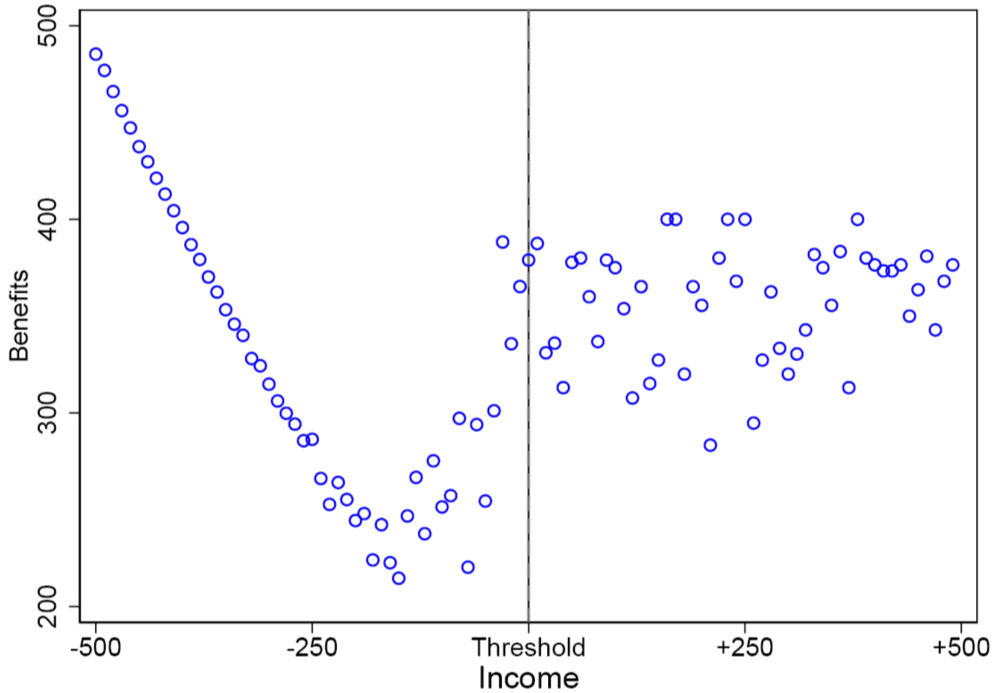
Notes: Average yearly administrative primary income within income (concept used to determine eligibility) slice in a small window of income either side of the eligibility threshold. Benefits are imputed assuming that the probability of exemption is constant. Income in this plot is monthly and refers to "eligibility income". Poor mental health is defined as receiving anti-depressants in the year previously. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Regression lines are shown following Section 6.1.2, as well as the estimated change in slopes following the regression kink design. Standard-errors are clustered at the municipality level.

Second, assuming that $\mathbb{E}[B|SA, Y = y, \text{Missing}] = \mathbb{E}[B|SA, Y = y, \text{Not Missing}]$ is *conservative*. Figure E.11 shows the results of the imputation when making the most extreme alternative assumption in the other direction: $\mathbb{E}[B|SA, Y = y, \text{Missing}] = \bar{y}$, i.e. any missing values conceal full benefits. This just exacerbates the main results.

Finally, empirically I assume $\mathbb{P}[SA|\nu = 1] = 1$ in the main results. However, the results are robust to a more reasonable value taken from $\mathbb{P}[SA|\nu = 1] = \mathbb{P}[SA|Y = 0]$, which may differ by mental health.

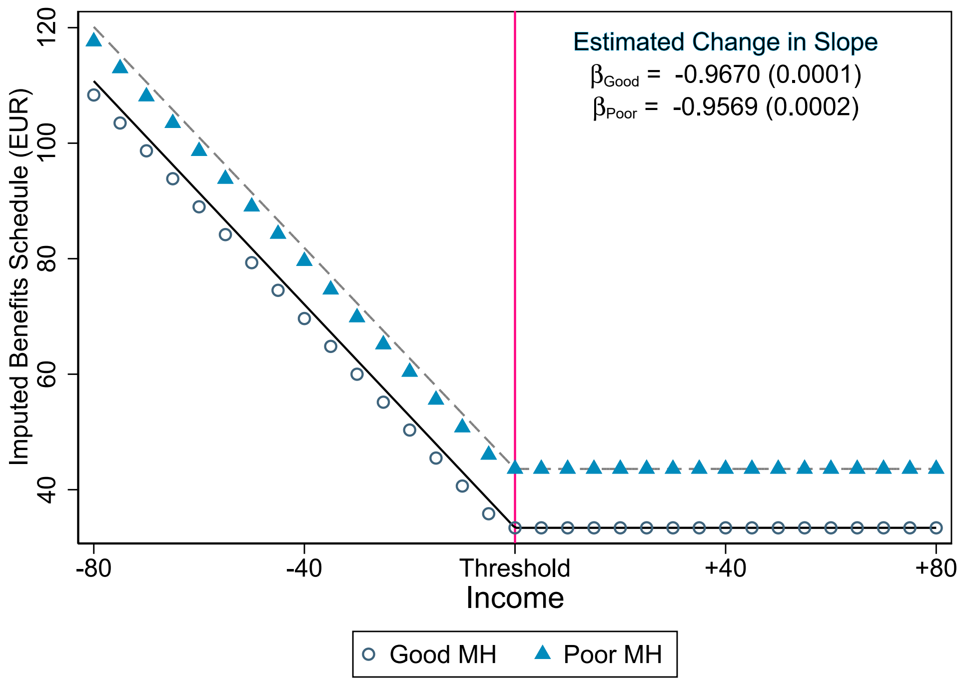
E.6. **Results.**

FIGURE E.9. Simulated ex-post schedule with exemptions



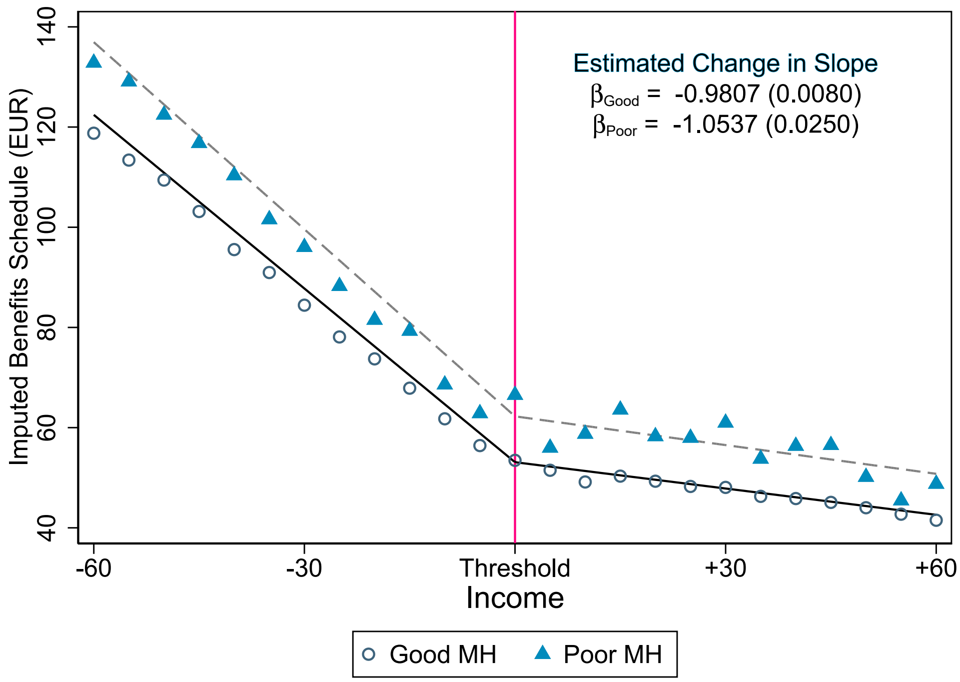
Notes: Plot of benefits received conditional on receipt in simulated dataset.

FIGURE E.10. Imputed benefits schedule (constant $p(y)$)



Notes: Average imputed benefits schedule within income slice in a small window of income either side of the eligibility threshold. Benefits are imputed assuming that the probability of exemption is constant. Income in this plot is monthly and refers to "eligibility income". Poor mental health is defined as receiving anti-depressants in the year previously. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Regression lines are shown following Section 6.1.2, as well as the estimated change in slopes following the regression kink design. Standard-errors are clustered at the municipality level.

FIGURE E.11. Imputed benefits schedule (censored data adjustment)



Notes: Average imputed benefits schedule within income slice in a small window of income either side of the eligibility threshold. Benefits are imputed assuming that $\mathbb{E}[B|SA, Y = y, \text{Missing}] = \bar{y}$. Income in this plot is monthly and refers to "eligibility income". Poor mental health is defined as receiving anti-depressants in the year previously. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Regression lines are shown following Section 6.1.2, as well as the estimated change in slopes following the regression kink design. Standard-errors are clustered at the municipality level.

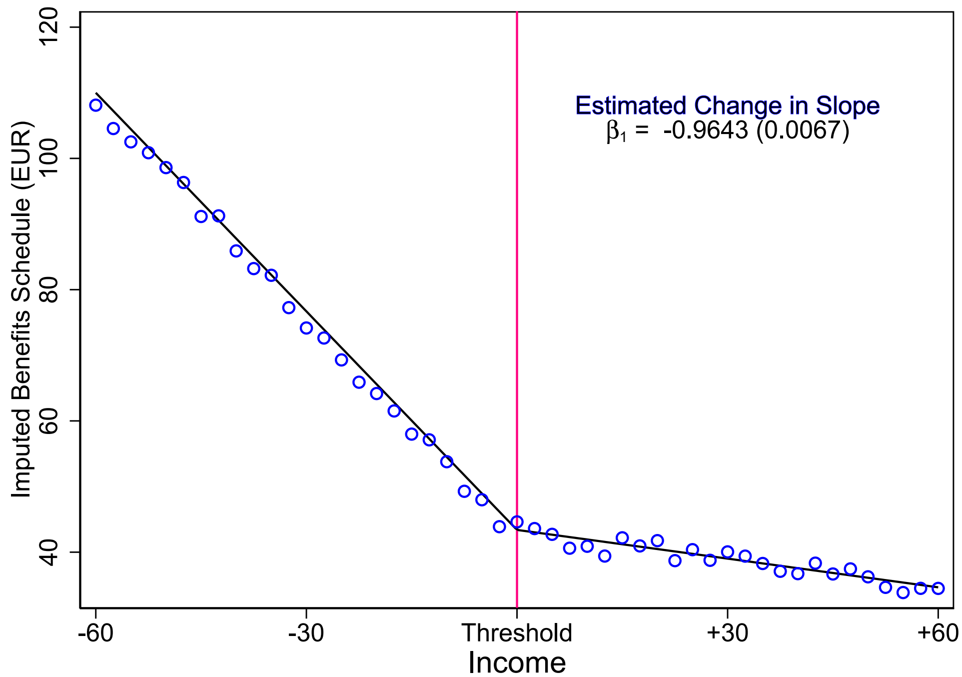


FIGURE E.12. Average imputed benefits schedule within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly and refers to "eligibility income". The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Regression lines are shown following Section 6.1.2, as well as the estimated change in slopes following the regression kink design. Standard-errors are clustered at the municipality level.

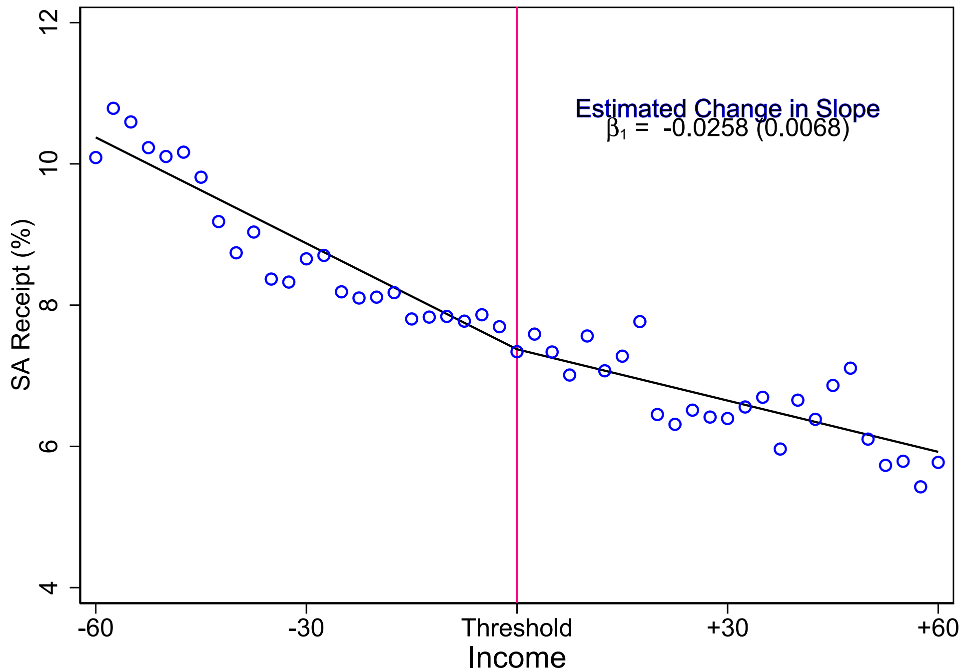


FIGURE E.13. Average rate of receipt within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly and refers to "eligibility income". The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Regression lines are shown following Section 6.1.2, as well as the estimated change in slopes following the regression kink design. Standard-errors are clustered at the municipality level.

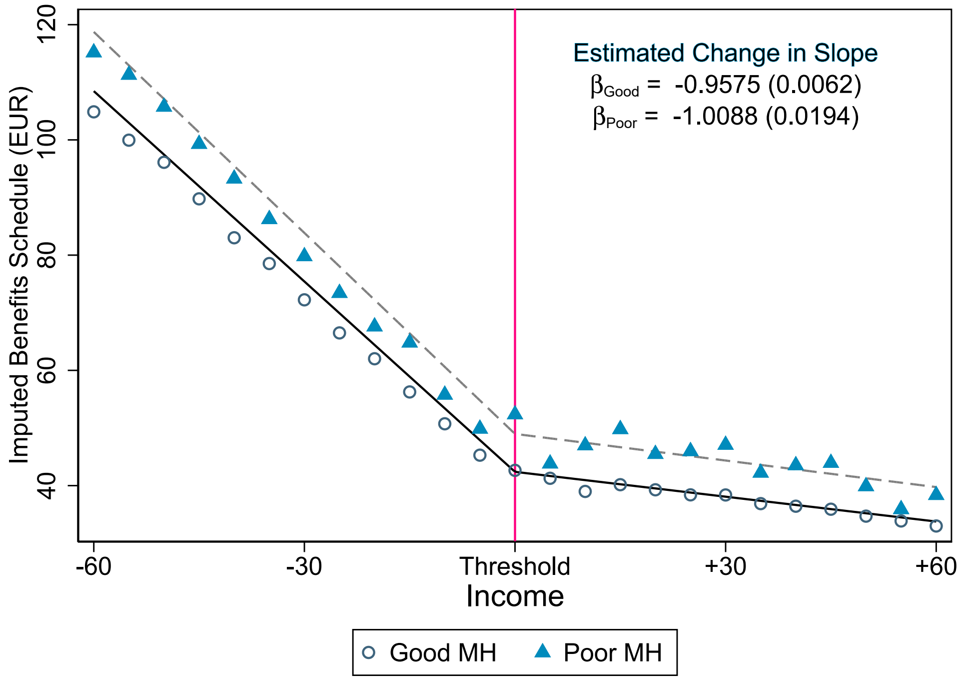


FIGURE E.14. Average imputed benefits schedule within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly and refers to "eligibility income". Poor mental health is defined as receiving anti-depressants in the year previously. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Regression lines are shown following Section 6.1.2, as well as the estimated change in slopes following the regression kink design. Standard-errors are clustered at the municipality level.

Effects on $\mathbb{P}[SA]$	**Reduced Form**				**IV**			
	Pooled		By Mental Health		Pooled		By Mental Health	
	Raw	+ Controls	Raw	+ Controls	Raw	+ Controls	Raw	+ Controls
Linear term (Income - Threshold):								
Everyone / Good MH	-0.0242***	-0.0186***	-0.0241***	-0.0174***	-0.0203***	-0.0153***	-0.0209***	-0.0144**
	(0.00386)	(0.00304)	(0.00394)	(0.00329)	(0.00481)	(0.00385)	(0.00494)	(0.00421)
Poor MH (**relative**)			0.00058	-0.00703			0.00506	-0.00513
			(0.00938)	(0.00826)			(0.0119)	(0.0104)
Kink term ($\min\{$Income - Threshold$, 0\}$):								
Everyone / Good MH	-0.0258***	-0.0207***	-0.0218**	-0.0187**				
	(0.00684)	(0.00585)	(0.00720)	(0.00653)				
Poor MH (**relative**)			-0.0290	-0.0136				
			(0.0182)	(0.0164)				
IV: Benefits effect								
Everyone / Good MH					0.0213***	0.0267***	0.0227**	0.0194**
					(0.00600)	(0.00707)	(0.00751)	(0.00677)
Poor MH (**absolute**)							0.0503**	0.0317*
							(0.0173)	(0.0145)
Observations (people-months)	487,475	448,307	487,475	448,307	487,475	448,307	487,475	448,307
R^2	0.002	0.225	0.003	0.226	0.001	0.226	0.001	0.226
Regressors	2	354	5	339	2	548	5	474

Standard errors (clustered at municipality level) in parentheses

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

TABLE E.1. Regression kink design estimates using a bandwidth of €60. “Pooled” columns estimate effects for the entire sample. “By Mental Health” columns estimate effects separately by mental health status. In the reduced form (kink term), coefficients show effects for good MH and the differential effect for poor MH (relative). In the IV stage, coefficients show absolute effects for each group. “Raw” specifications include only the running variable; “+ Controls” add fixed effects for month, year, age, gender, wealth, education, municipality, household composition, and sector. IV columns use the imputed benefits schedule as the first stage. Standard errors clustered at municipality level. Sample: single employees, 2011-2014. Income is monthly and refers to “eligibility income”.

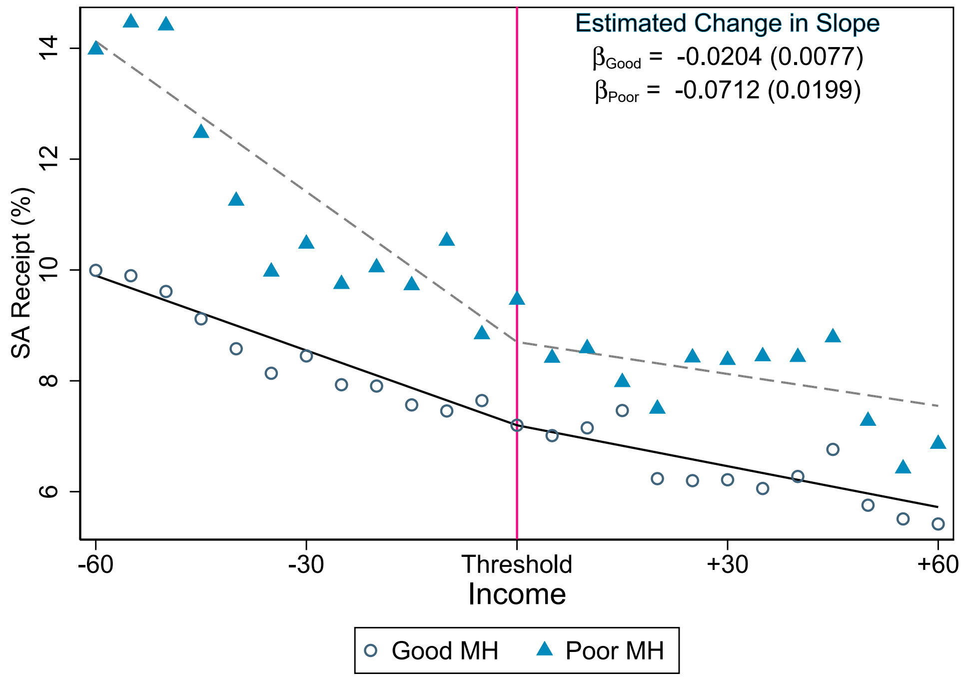


FIGURE E.15. Average rate of receipt within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly and refers to "eligibility income". Poor mental health is defined as receiving anti-depressants in the year previously. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Regression lines are shown following Section 6.1.2, as well as estimated change in slopes from the regression kink design. Standard-errors are clustered at the municipality level.

E.6.1. *Mechanisms.*

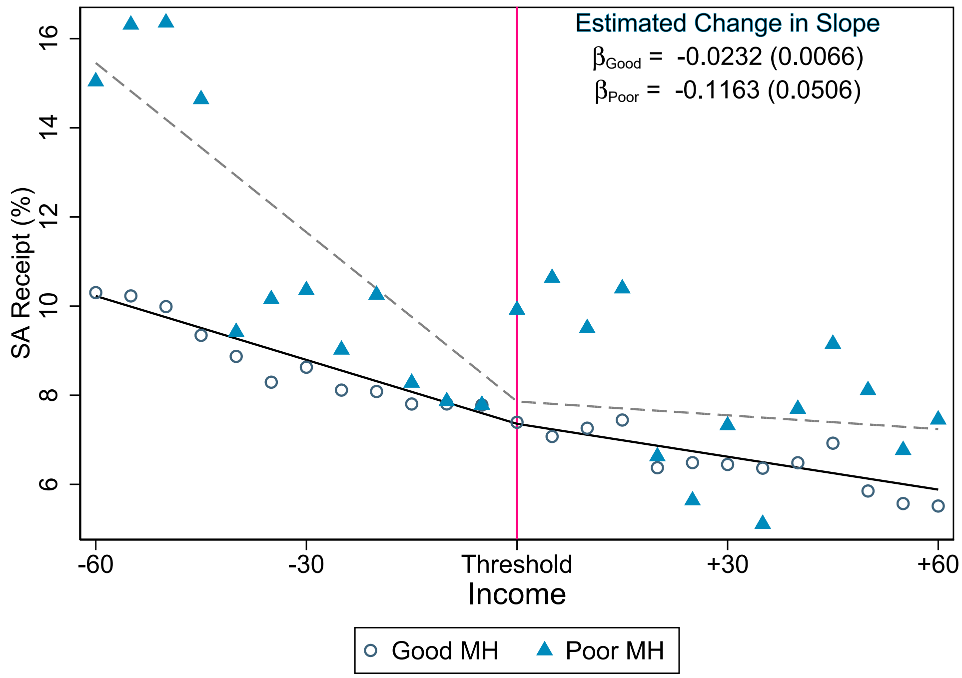


FIGURE E.16. Average rate of receipt within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly and refers to "eligibility income". Poor mental health is defined as receiving anti-psychotics in the year previously. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Regression lines are shown following Section 6.1.2, as well as estimated change in slopes from the regression kink design. Standard-errors are clustered at the municipality level.

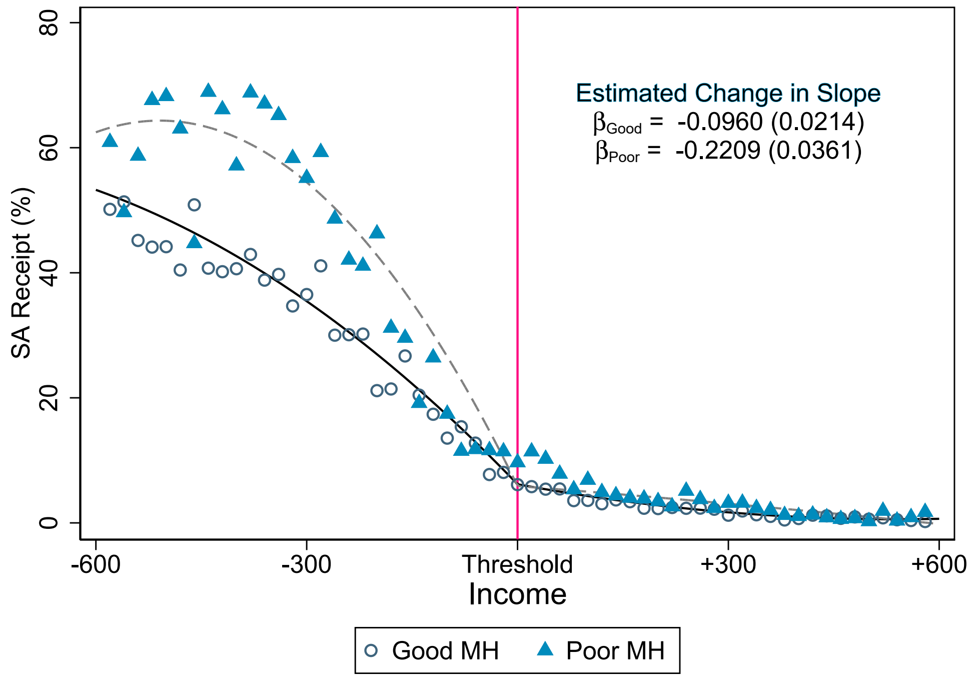
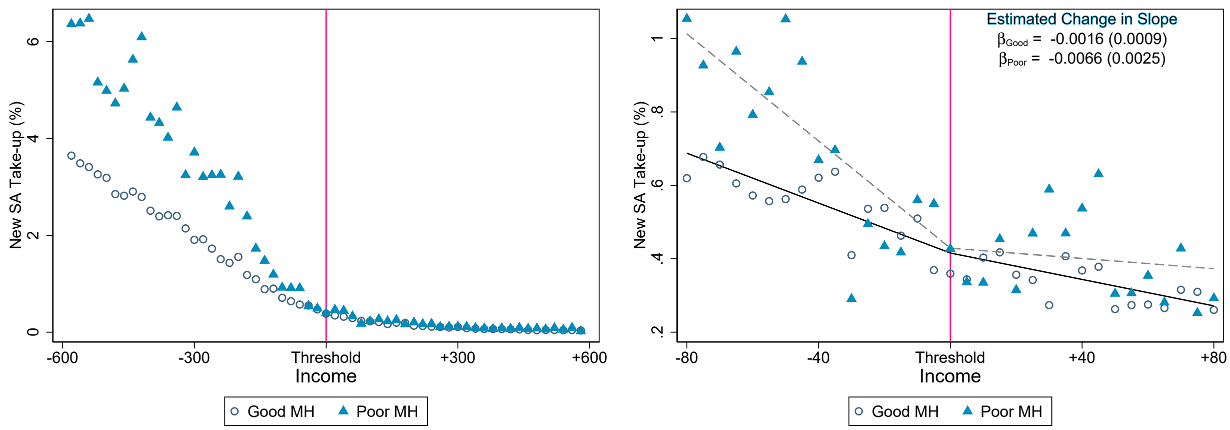


FIGURE E.17. Average rate of receipt within income slice in a large window of income either side of the eligibility threshold. Income in this plot is monthly and refers to "eligibility income". Poor mental health is defined as reporting severe psychological distress in the survey in 2012. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions.



(A) Overview (B) Zoomed

FIGURE E.18. Average inflow within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly and refers to "eligibility income". Poor mental health is defined as using psychotropic medications in the year previously. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Regression lines are shown following Section 6.1.2, as well as estimated change in slopes from the regression kink design. Standard-errors are clustered at the municipality level.

TABLE E.2. Regression Kink Design Estimates: Heterogeneous Effects by Mental Health and Bandwidth

	Narrow Bandwidth (€60)			**Wide Bandwidth (€180)**		
	Raw	+ Controls	+ Interactions	Raw	+ Controls	+ Interactions
Linear term:						
Income - Threshold	-0.0241***	-0.0243***	-0.0232*	-0.0189***	-0.0179***	-0.0127***
	(0.00394)	(0.00647)	(0.0113)	(0.000912)	(0.00144)	(0.00310)
Kink term:						
$\min\{$Income - Threshold$, 0\}$	-0.0218**	-0.0190	-0.0272	-0.0540***	-0.0549***	-0.0547***
	(0.00720)	(0.0123)	(0.0223)	(0.00207)	(0.00288)	(0.00559)
Interaction effects with kink term:						
Poor MH	-0.0290	-0.0334	-0.0326	-0.0414***	-0.0370***	-0.0326**
	(0.0182)	(0.0275)	(0.0280)	(0.00685)	(0.0101)	(0.0100)
Man			-0.0111			-0.0121*
			(0.0172)			(0.00526)
Foreign-born			0.0148			0.0130*
			(0.0206)			(0.00592)
Physical chronic condition			-0.000171			-0.0176***
			(0.0192)			(0.00493)
Household consumption $(\times 10,000)$			0.00425			0.00363*
			(0.00644)			(0.00154)
Observations (people-months)	487,475	211,388	211,388	1,506,599	636,043	636,043
Adjusted R^2	0.003	0.247	0.247	0.036	0.267	0.267
Number of clusters	355	355	355	355	355	355

Standard errors (clustered at municipality level) in parentheses

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Notes: This table presents regression kink design estimates of the heterogeneous effects of benefit changes on social assistance receipt. The narrow bandwidth uses observations within €60 of the eligibility threshold, while the wide bandwidth uses €180. "Raw" specifications include only the running variable and interactions; "Controls" add controls for hh consumption and fixed effects for month, year, age, gender, wealth, education, municipality, household composition, and sector; "Interacted" additionally includes interactions between the kink term and demographic characteristics. The sample consists of single employees from 2011-2014. Poor MH indicates dispensation of psychotropic medication in the previous year. Note that the discrepancy between controls column here vs Table E.1 is the addition of hh consumption as a control. Income is monthly and refers to "eligibility income".

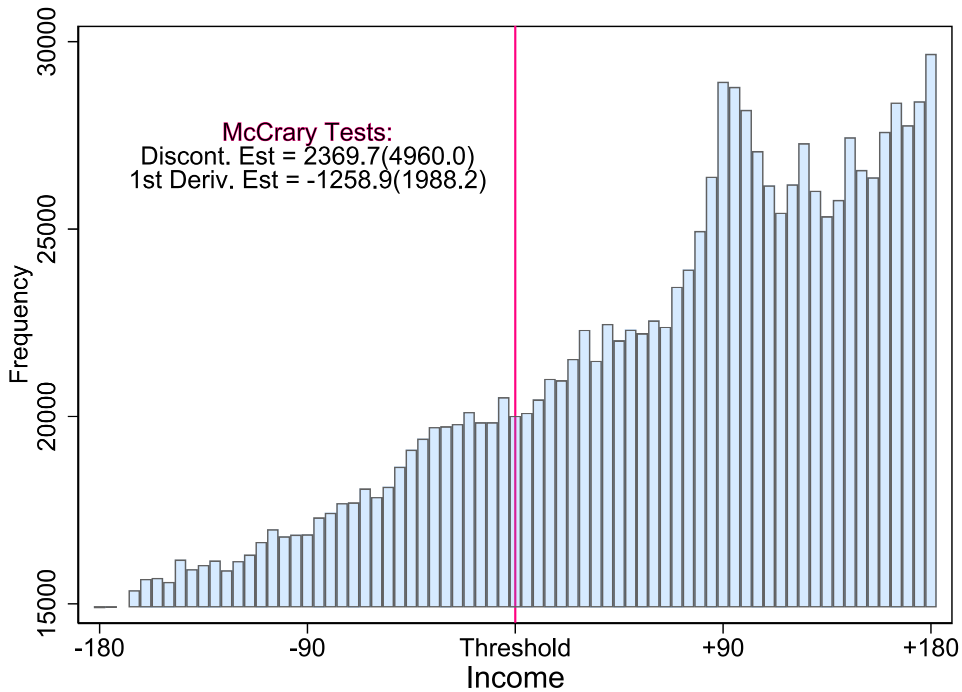


FIGURE E.19. Density of income around the eligibility threshold. McCrary (2008) tests for discontinuity in levels and slopes around the threshold are shown. Income in this plot is monthly and refers to "eligibility income". The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions.

E.7. **Validity of RKD.**

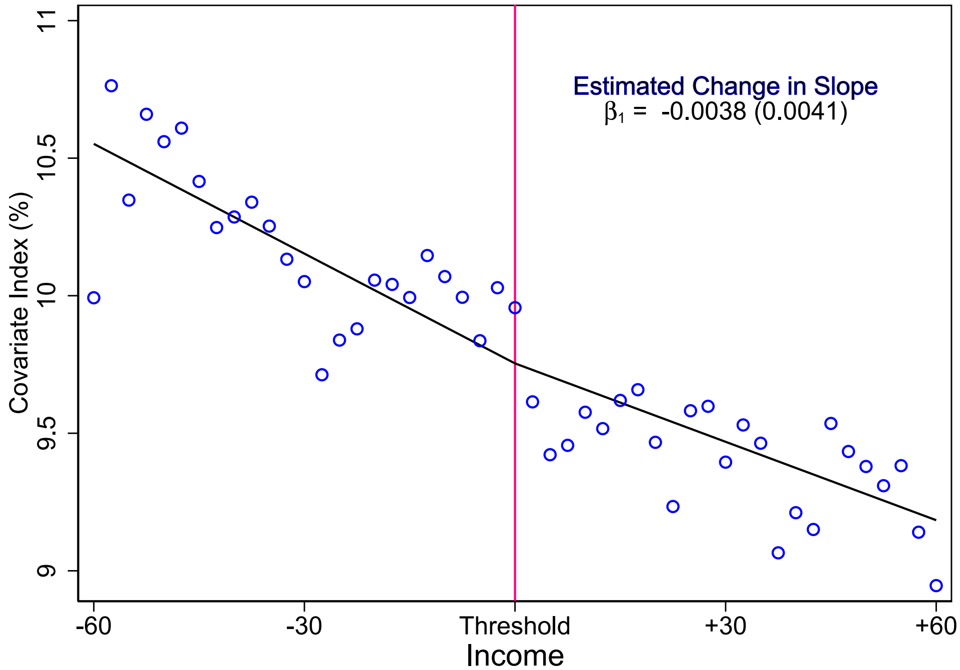


FIGURE E.20. Covariate Test: plot shows fitted values of a regression of social assistance take-up on all pre-determined controls used throughout this paper including income (disposable, rather than the concept used to determine eligibility), education, hh composition, municipality FEs. These fitted values form a "Covariate Index" which is binned. An RKD estimate with income as the running variable is also shown. Income in this plot is monthly and refers to "eligibility income". The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Standard-errors are clustered at the municipality level.

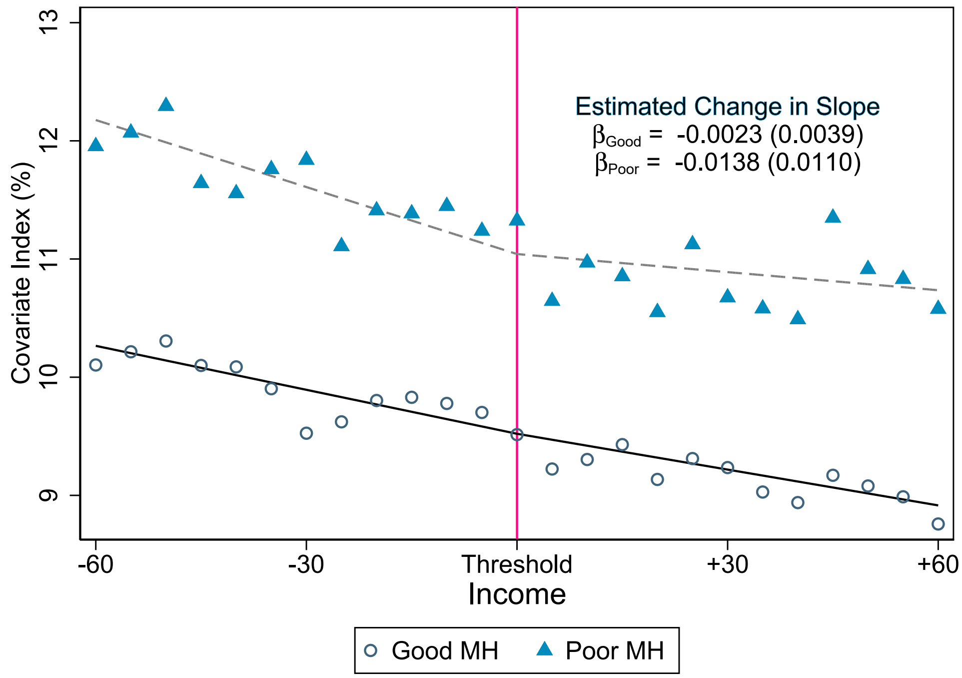


FIGURE E.21. Covariate Test: plot shows fitted values of a regression of social assistance take-up on all pre-determined controls used throughout this paper including income, education, hh composition, municipality FEs. These fitted values form a "Covariate Index" which is binned. An RKD estimate with income as the running variable is also shown. Separated by mental health. Income in this plot is monthly and refers to "eligibility income". Poor mental health is defined as receiving psychopharma in the year previously. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Regression lines are shown following Section 6.1.2, as well as the estimated change in slopes following the regression kink design. Standard-errors are clustered at the municipality level.

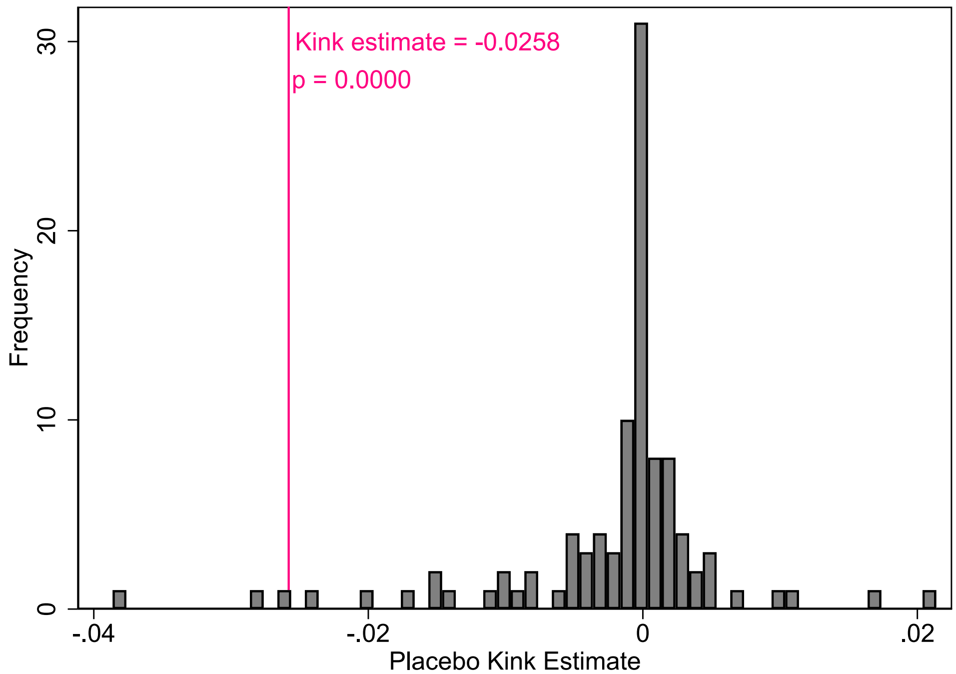


FIGURE E.22. Results of permutation test à la Ganong and Jäger (2018). I estimate RKDs on 100 placebo kinks in the range $[\bar{y} - 600, \bar{y} + 600]$ and plot a histogram of the estimates. A binomial test is used to check whether the true estimate is an outlier. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Standard-errors are clustered at the municipality level.

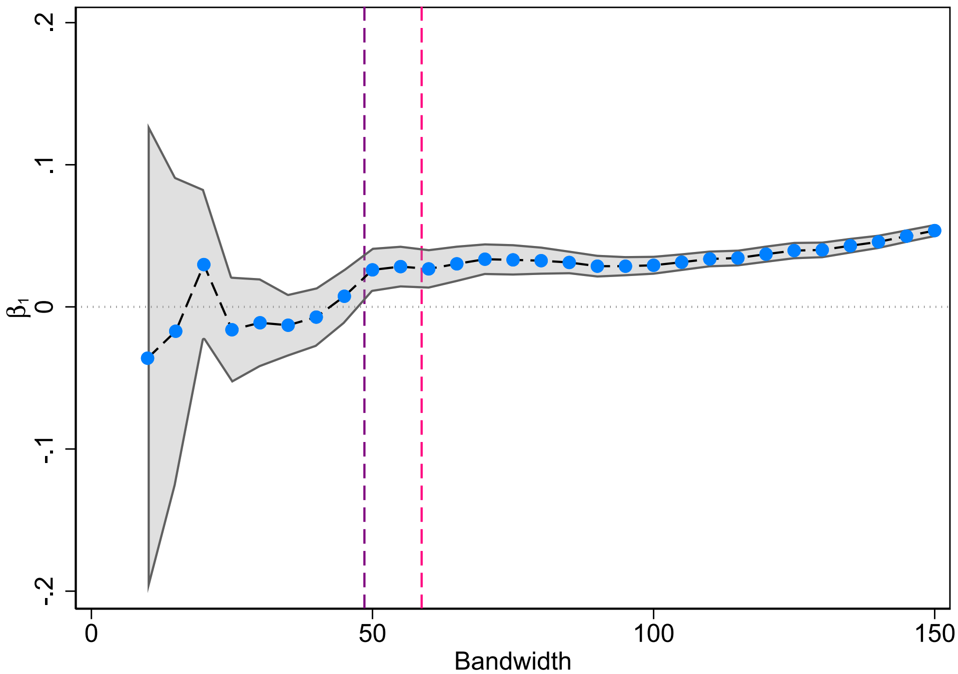


FIGURE E.23. Results of test of sensitivity to changes in bandwidth. I estimate RKDs changing the bandwidth, with the CCT robust bandwidth displayed. The lower purple dashed line indicates the CCT robust bandwidth with regularization, and the upper pink dashed line indicates the CCT robust bandwidth without regularization. This plot shows the estimates and confidence intervals. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Standard-errors are clustered at the municipality level.

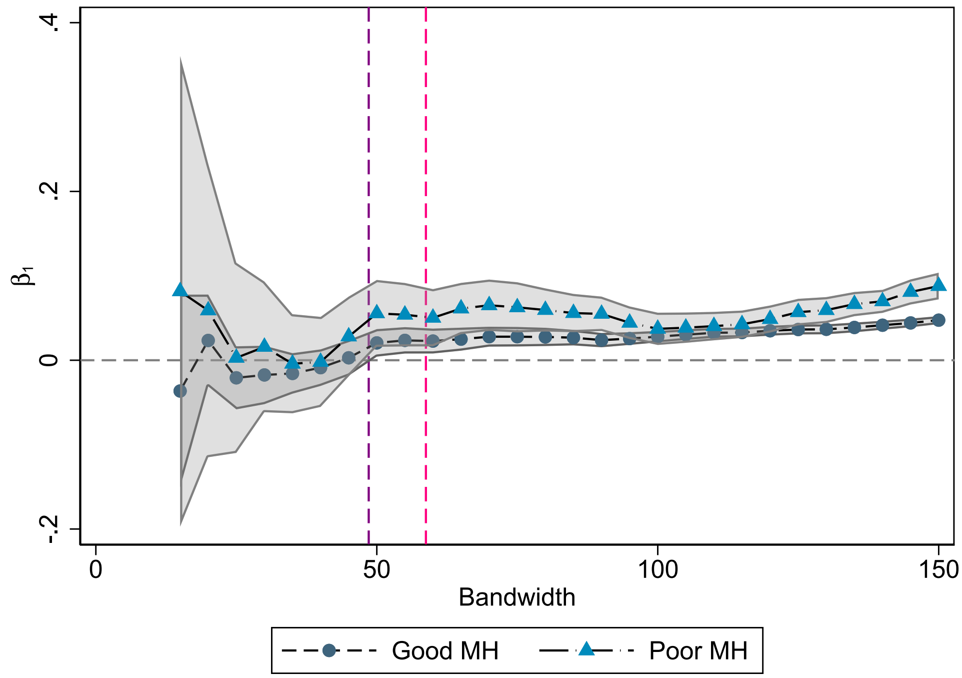


FIGURE E.24. Results of test of sensitivity to changes in bandwidth. I estimate heterogeneous RKDs changing the bandwidth, with the CCT robust bandwidth displayed. The lower purple dashed line indicates the CCT robust bandwidth with regularization, and the upper pink dashed line indicates the CCT robust bandwidth without regularization. This plot shows the estimates and confidence intervals. Poor mental health is defined as receiving psychopharma in the year previously. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Standard-errors are clustered at the municipality level.

Appendix F. Calibration of Welfare Effects: Additional Material

F.1. **Eligibility.** Throughout the empirical analysis, I focus on take-up levels and responses among the *eligible* population. This is because I am interested in take-up *behaviour* across types, and not underlying eligibility. However, the theoretical framework above does not model eligibility directly. Indeed, the government budget constraint as defined in Equation (2.8) makes much more sense if it holds for θ in the general population, and not the eligible. In reality, the ineligible fund benefits for the recipients, and not the eligible non-takers.

Proposition F.1 shows that identifying take-up levels and responses for the eligible population is sufficient for the general population as long as $\mathbb{P}[SA|\text{Ineligible}] = 0$.

Proposition F.1. *Assume* $\mathbb{P}[SA|\textit{Ineligible}] = 0$. *Then:*

$$\mathbb{P}[SA]_\theta = \mathbb{P}[SA \mid \textit{Eligible}]_\theta \cdot \mathbb{P}[\textit{Eligible}]_\theta \tag{F.1}$$

and take-up responses to policy X are given by:

$$\frac{\partial \mathbb{P}[SA]_\theta}{\partial X} = \frac{\partial \mathbb{P}[SA \mid \textit{Eligible}]_\theta}{\partial X} \cdot \underbrace{\left(\frac{\mathbb{P}[\textit{Eligible}]_\theta}{1 - \mathbb{P}[SA \mid \textit{Eligible}]_\theta \cdot \mathbb{P}[\textit{Ineligible} \mid \textit{No } SA]_\theta} \right)}_{EE_\theta\text{: Effective Eligibility}_\theta} \tag{F.2}$$

Proof of Proposition F.1. We make the following assumptions:

(1) $\mathbb{P}[SA|\text{Ineligible}] = 0$ - no take-up among the ineligible population.
(2) No labour supply responses to policy changes dB or $d\Lambda$: $y_\theta^{SA=0}$ and $y_\theta^{SA=1}$ are fixed with respect to B and Λ.
(3) Eligibility is determined by the threshold rule: a person is eligible if and only if $y \leq \bar{y}$ where $y = SA \cdot y^{SA=1} + (1 - SA) \cdot y^{SA=0}$.

Part 1: Take-up levels.

By the law of total probability:

$$\mathbb{P}[SA]_\theta = \mathbb{P}[SA|y \leq \bar{y}]_\theta \cdot \mathbb{P}[y \leq \bar{y}]_\theta + \mathbb{P}[SA|y > \bar{y}]_\theta \cdot \mathbb{P}[y > \bar{y}]_\theta$$

By Assumption 1, $\mathbb{P}[SA|y > \bar{y}]_\theta = 0$. Therefore:

$$\begin{aligned}\mathbb{P}[SA]_\theta &= \mathbb{P}[SA|y \leq \bar{y}]_\theta \cdot \mathbb{P}[y \leq \bar{y}]_\theta \\ &= \mathbb{P}[SA|\text{Eligible}]_\theta \cdot \mathbb{P}[\text{Eligible}]_\theta\end{aligned}$$

Part 2: Take-up responses.

Differentiating the result from Part 1 with respect to policy parameter X:

$$\frac{\partial \mathbb{P}[SA]_\theta}{\partial X} = \frac{\partial}{\partial X}\left[\mathbb{P}[SA|\text{Eligible}]_\theta \cdot \mathbb{P}[\text{Eligible}]_\theta\right]$$
$$= \frac{\partial \mathbb{P}[SA|\text{Eligible}]_\theta}{\partial X} \cdot \mathbb{P}[\text{Eligible}]_\theta + \mathbb{P}[SA|\text{Eligible}]_\theta \cdot \frac{\partial \mathbb{P}[\text{Eligible}]_\theta}{\partial X}$$

We now evaluate $\frac{\partial \mathbb{P}[\text{Eligible}]_\theta}{\partial X}$. By the law of total probability:

$$\mathbb{P}[\text{Eligible}]_\theta = \mathbb{P}[y \leq \bar{y}]_\theta$$
$$= \mathbb{P}[y \leq \bar{y}|SA = 1]_\theta \cdot \mathbb{P}[SA = 1]_\theta + \mathbb{P}[y \leq \bar{y}|SA = 0]_\theta \cdot \mathbb{P}[SA = 0]_\theta$$

By Assumption 1, $\mathbb{P}[SA = 1|\text{Ineligible}] = 0$, which by Bayes' rule implies $\mathbb{P}[\text{Ineligible}|SA = 1] = 0$. Therefore, $\mathbb{P}[y \leq \bar{y}|SA = 1]_\theta = 1$.

$$\mathbb{P}[\text{Eligible}]_\theta = \mathbb{P}[SA = 1]_\theta + \mathbb{P}[y \leq \bar{y}|SA = 0]_\theta \cdot (1 - \mathbb{P}[SA = 1]_\theta)$$
$$= \mathbb{P}[SA = 1]_\theta + (1 - \mathbb{P}[\text{Ineligible}|SA = 0]_\theta) \cdot (1 - \mathbb{P}[SA]_\theta)$$

where $\mathbb{P}[\text{Ineligible}|SA = 0]_\theta = \mathbb{P}[y > \bar{y}|SA = 0]_\theta$. Simplifying:

$$\mathbb{P}[\text{Eligible}]_\theta = 1 - \mathbb{P}[\text{Ineligible}|SA = 0]_\theta \cdot (1 - \mathbb{P}[SA = 1]_\theta)$$

By Assumption 2, labour supply conditional on take-up is fixed, so $\frac{\partial \mathbb{P}[\text{Ineligible}|SA=0]_\theta}{\partial X} = 0$. Differentiating:

$$\frac{\partial \mathbb{P}[\text{Eligible}]_\theta}{\partial X} = \mathbb{P}[\text{Ineligible}|SA = 0]_\theta \cdot \frac{\partial \mathbb{P}[SA = 1]_\theta}{\partial X}$$

Substituting back:

$$\frac{\partial \mathbb{P}[SA]_\theta}{\partial X} = \frac{\partial \mathbb{P}[SA = 1|\text{Eligible}]_\theta}{\partial X} \cdot \mathbb{P}[\text{Eligible}]_\theta + \mathbb{P}[SA = 1|\text{Eligible}]_\theta \cdot \mathbb{P}[\text{Ineligible}|SA = 0]_\theta \cdot \frac{\partial \mathbb{P}[SA = 1]_\theta}{\partial X}$$

Rearranging:

$$\frac{\partial \mathbb{P}[SA]_\theta}{\partial X}\left(1 - \mathbb{P}[SA = 1|\text{Eligible}]_\theta \cdot \mathbb{P}[\text{Ineligible}|SA = 0]_\theta\right) = \frac{\partial \mathbb{P}[SA = 1|\text{Eligible}]_\theta}{\partial X} \cdot \mathbb{P}[\text{Eligible}]_\theta$$

Solving for $\frac{\partial \mathbb{P}[SA]_\theta}{\partial X}$:

$$\frac{\partial \mathbb{P}[SA]_\theta}{\partial X} = \frac{\partial \mathbb{P}[SA|\text{Eligible}]_\theta}{\partial X} \cdot \frac{\mathbb{P}[\text{Eligible}]_\theta}{1 - \mathbb{P}[SA = 1|\text{Eligible}]_\theta \cdot \mathbb{P}[\text{Ineligible}|SA = 0]_\theta}$$

This completes the proof. $\square$

Proposition F.1 follows from Bayes Rule, the fact that eligibility is determined by $y \leq \bar{y}$ where $y = SA \cdot y^{SA=1} + (1-SA) \cdot y^{SA=0}$ and from the fact we have assumed no potential income responses to dB or $d\Lambda$ (although of course individuals can switch income when switching from $SA = 1$ to $SA = 0$ or vice versa). The intuition is as follows: we need to adjust for baseline incomplete take-up and the fact that ineligible people can still be on the margin of take-up (if they were just indifferent between earning income above the threshold and switching to earning income below the threshold and receiving social assistance) when mapping conditional take-up responses to the general population.

How should we implement Proposition F.1 when calculating welfare effects? When integrating against average take-up levels, Bayes Rule $\rightarrow \int \mathbb{P}[SA]_\theta \cdot H_\theta d\mu = \mathbb{P}[\text{Eligible}] \cdot \int \mathbb{P}[SA \mid \text{Eligible}]_\theta \cdot H_\theta d\mu_{\text{Eligible}}$. Where μ_{Eligible} is the conditional density of types θ. Similarly, Bayes Rule $\rightarrow \int \frac{\partial \mathbb{P}[SA]_\theta}{\partial X} \cdot H_\theta d\mu = \mathbb{P}[\text{Eligible}] \cdot \int \frac{\partial \mathbb{P}[SA|\text{Eligible}]_\theta}{\partial X} \cdot \frac{1}{1-\mathbb{P}[SA \mid \text{Eligible}]_\theta \cdot \mathbb{P}[\text{Ineligible} \mid \text{No } SA]_\theta} \cdot H_\theta d\mu_{\text{Eligible}}$.

Note that in the regression kink design (Section 6), income exemptions by municipalities mean that some individuals with $y > \bar{y}$ can effectively receive SA. In principle, the analysis above could be extended to account for these exemptions by redefining eligibility to include exempted income. However, I do not pursue this extension to simplify the analysis, and it would not materially affect the results given that $\mathbb{P}[SA|y > \bar{y}]$ is negligible.

F.2. **Identification.** For the identification, I use the point-estimates of the take-up response to a change in benefit level using the regression kink design with Calonico et al. (2014) robust bandwidth. These point estimates, $(0.000227, 0.000503)$ for good and poor mental health, respectively are smaller than those estimated on the larger bandwidth of €600 either side of the threshold as in Figure VIIb, $(0.000778, 0.00145)$ for good and poor mental health, respectively. If I alternatively use these larger estimates for the calibration, I find that $v'_L = 1.86$, $\kappa'_H = 20.2$ and $\kappa'_L = 31.7$. These estimates imply:

$$MVPF_{dB} = 0.57$$

$$MVPF_{d\Lambda} = 0.81$$

and increasing barriers is concluded as 42% more effective than increasing benefits, although both $MVPF$s are below 1. Since regression kink design estimates are intended to be local to the kink, the preferred estimate is the one using the smaller, robust bandwidth because the shape of the take-up function away from the threshold is affected by unobservables.

F.3. **Relaxing Modelling Assumptions.**

F.3.1. *Relaxing Independence.* Relaxing independence of ε from θ requires examining how the distribution of unobservables varies by mental health status. I explore this by allowing for a specific parametric form for how the distribution of σ varies with θ which maintains identification but allows me to test whether the role of unobservables affecting take-up differs by mental health: $\varepsilon_\theta \sim N(\mu_\theta, \sigma_\theta^2)$ where $\sigma_L \neq \sigma_H$ nor $\mu_L \neq \mu_H$ necessarily.

Heteroskedastic Probit Analysis: To test whether $\sigma_L \neq \sigma_H$, I estimate a heteroskedastic probit model:

$$\mathbb{P}[SA_{it} = 1 | \mathbb{1}\{\text{Poor MH}\}_i, X_{it}] = \Phi\left(\frac{\alpha_0 + \alpha_1 \mathbb{1}\{\text{Poor MH}\}_i + X_{it}'\beta}{\sigma_i}\right) \tag{F.3}$$

where $\ln(\sigma_i) = \delta_0 + \delta_1 \mathbb{1}\{\text{Poor MH}\}_i$. The heteroskedastic probit directly estimates the ratio σ_L/σ_H through the coefficient on poor mental health in the variance equation.

Table F.1 shows the results of this regression. I estimate $\ln\hat{\sigma} = 0.0379$, implying $\sigma_L/\sigma_H = \exp(0.0379) = 1.039$. This suggests individuals with poor mental health face approximately 4% more dispersion in unobservable factors affecting take-up. With more spread in unobservables, ε's probability mass is distributed more widely, resulting in a lower density at any given point $f_L < f_H$. Indeed, a standard F-test rejects $\sigma_L = \sigma_H$, which is consistent with Figure IIIa: my analysis of differential targeting suggests recipients with poor mental health have small but distinguishable differences in other characteristics versus versus those with good mental health, relative to the gap among eligible non-recipients.

Goodness of fit and selection of controls: the parametric model imposes more stringent assumptions about the shape of F_ε than a linear probability model. Below, I discuss the validity of these assumptions depending on the choice of controls. I focus on specifications (4) and (5) from Table F.1, examining whether lagged work-status controls should be included when fitting the probit model. For both specifications, I plot 3 objects: the heteroskedastic probit generalized residuals $r_i^{\text{gen}} = y_i \cdot \frac{\phi(X_i'\beta/\sigma_i)}{\Phi(X_i'\beta/\sigma_i)} - (1-y_i) \cdot \frac{\phi(X_i'\beta/\sigma_i)}{1-\Phi(X_i'\beta/\sigma_i)}$, standardized residuals $r_i^{\text{std}} = \frac{y_i - \Phi(X_i'\beta/\sigma_i)}{\sqrt{\Phi(X_i'\beta/\sigma_i)\cdot[1-\Phi(X_i'\beta/\sigma_i)]}}$ and fitted values $\hat{p}_i = \Phi(X_i'\beta/\sigma_i)$, all heterogeneously by mental health.

A well-specified model should exhibit: (i) smooth, unimodal distribution of predicted probabilities $\hat{p}_i$ across $[0,1]$, (ii) standardized residuals with a spike at zero indicating accurate predictions, and (iii) generalized residuals approximately normally distributed without systematic holes or gaps. The two rows of Figure F.1 show how well specifications (6) and (7) fit the data, respectively.

TABLE F.1. Heteroskedastic Probit: Social Assistance Take-Up by Mental Health

	(1) YAG	(2) + HH Disp Income	(3) + Consumption	(4) + HH Composition	(5) + Education	(6) + Municipality	(7) + B
Mean Equation							
Poor Mental Health	0.0037	0.0170	0.0095	0.0069	-0.0048	0.0002	0.0368***
	(0.0208)	(0.0117)	(0.0126)	(0.0126)	(0.0119)	(0.0105)	(0.0050)
Variance Equation							
$\log(\sigma)$: Poor MH	-0.0336	0.0267**	0.0616***	0.0766***	0.0361**	0.0379***	-0.0464***
	(0.0267)	(0.0098)	(0.0105)	(0.0109)	(0.0111)	(0.0115)	(0.0040)
Observations	5,671,855	5,073,831	5,008,768	4,990,323	4,990,323	4,990,323	4,990,320
Degrees of Freedom	49	149	150	203	217	358	359

Notes: Heteroskedastic probit estimates. Dependent variable: social assistance receipt. Mean equation includes Poor Mental Health and controls. Variance equation: $\log(\sigma^2) = \delta_0 + \delta_1 \cdot$ Poor MH. Standard errors clustered at municipality level. YAG = Year, Age, Gender FE. Sample: Eligible population 2011-2020. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

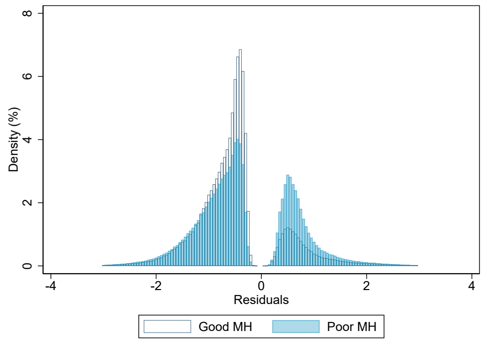
Both models suffer from different problems. Specification (6) is poorly calibrated, with few accurate predictions, however its smooth fitted values means the latent variable structure seems reasonable. Specification (7) has accurate predictions but violates the latent variable structure due to extreme bi-modality in the fitted-values distribution, likely suggesting quasi-separation due to overfitting. Therefore, I maintain a linear probability model for the main specification, but include these results for robustness. In particular, I show how each specification changes the estimates for need and cost below.

Implication for need and cost: Starting with specification (6), the fact that $\sigma_L > \sigma_H$ implies that the larger sensitivity to $d\Lambda$ and dB among people with poor mental health *understates* the true differences in need and cost. Greater dispersion in unobservables means fewer people are concentrated near the threshold of take-up. Since we observe larger take-up responses despite having fewer marginals, the marginal individuals must have substantially higher need and cost to generate the same aggregate response. Concretely, under the model in Equation (F.3),

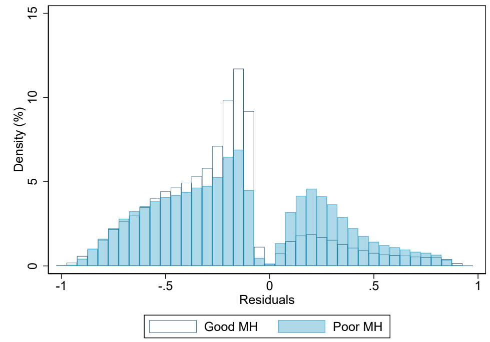
$$\begin{aligned}
\mathbb{P}[SA]_L = \mathbb{P}[SA]_H \implies & \frac{v_L - \kappa_L - \mu_L}{\sigma_L} = \frac{v_H - \kappa_H - \mu_H}{\sigma_H} \\
\implies \frac{f_L}{f_H} &= \frac{\sigma_H}{\sigma_L} \cdot \underbrace{e^{-\frac{(v_L-\kappa_L-\mu_L)^2}{2\sigma_L^2}+\frac{(v_H-\kappa_H-\mu_H)^2}{2\sigma_H^2}}}_{=1} \\
\implies \underbrace{\frac{\frac{\partial \mathbb{P}[SA]_L}{\partial \Lambda}}{\frac{\partial \mathbb{P}[SA]_H}{\partial \Lambda}}}_{1.57} &= -\frac{\kappa_L'}{\kappa_H'} \cdot \frac{f_L}{f_H} = -\frac{\kappa_L'}{\kappa_H'} \cdot \underbrace{\frac{\sigma_H}{\sigma_L}}_{=1/1.039} \\
\implies \frac{\kappa_L'}{\kappa_H'} &= 1.57 \times 1.039 = 1.63
\end{aligned}$$

This implies that $\frac{\kappa_L'}{\kappa_H'} = 1.63$ and $\frac{v_L'}{v_H'} = 2.30$. These adjustments only accentuate the main finding of positive correlation between need and cost across mental health. If I were to adopt specification (7), it would suggest $\sigma_L < \sigma_H$. This slightly attenuates, but does not overturn, the main results, implying $\frac{\kappa_L'}{\kappa_H'} = 1.50$ and $\frac{v_L'}{v_H'} = 2.12$.

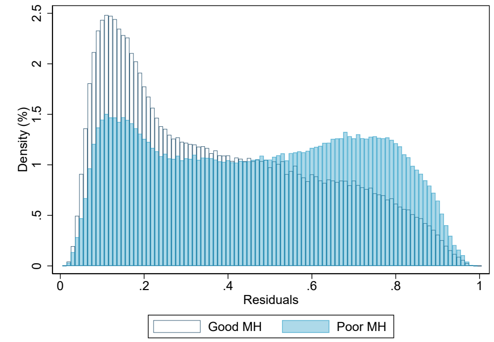
FIGURE F.1. Goodness-of-fit heteroskedastic probit



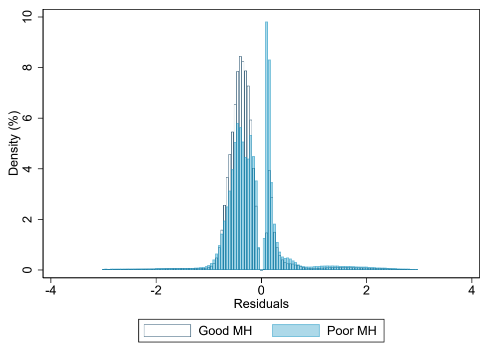
(A) Generalized residuals, spec (6)



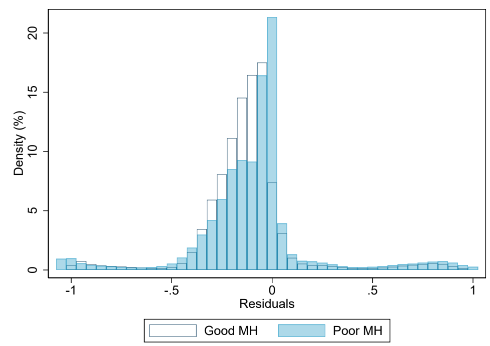
(B) Standardized residuals, spec (6)



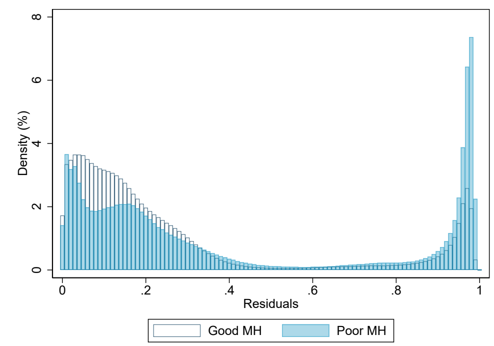
(C) Fitted values, spec (6)



(D) Generalized residuals, spec (7)



(E) Standardized residuals, spec (7)



(F) Fitted values, spec (7)

Notes: Goodness-of-fit plots for heteroskedastic probit model described in Equation (F.3). The sample contains people eligible for social assistance, years 2011-2020.

F.3.2. *Relaxing Additivity.* Additivity separates need/cost from the scaling factor $f_\varepsilon(v_\theta - \kappa_\theta)$ in Equations (2.4) and (2.5), allowing me to extrapolate from marginals' take-up responses to infer primitives for infra-marginals. Without additivity, a bounding argument can be made about $\frac{dW}{d\Lambda}$ following Haller and Staubli (2024). In this case, $SA = 1 \iff v_\theta(B,\varepsilon) > \kappa_\theta(\Lambda,\varepsilon)$. Given monotonicity of $v-\kappa$ with respect to ε, behaviour will follow a threshold-rule: $SA = 1 \iff \varepsilon \leq \varepsilon^*$ where ε^*_θ satisfies the implicit equation $v_\theta(B,\varepsilon^*_\theta) = \kappa_\theta(\Lambda,\varepsilon^*_\theta)$. Then:

$$\frac{\partial \mathbb{P}[SA]}{\partial \Lambda} = \frac{-\frac{\partial \kappa}{\partial \Lambda}}{\frac{\partial \kappa}{\partial \varepsilon} - \frac{\partial v}{\partial \varepsilon}} \cdot f(\varepsilon^*)$$
$$\frac{\partial \mathbb{P}[SA]}{\partial B} = \frac{\frac{\partial v}{\partial B}}{\frac{\partial \kappa}{\partial \varepsilon} - \frac{\partial v}{\partial \varepsilon}} \cdot f(\varepsilon^*)$$

Therefore, Equation (2.7) becomes

$$\frac{-\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}}{\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}} = \frac{\mathbb{E}\left[\kappa'_\theta(\Lambda,\varepsilon)|\varepsilon = \varepsilon^*_\theta\right]}{\mathbb{E}\left[v'_\theta(B,\varepsilon)|\varepsilon = \varepsilon^*_\theta\right]} \tag{F.4}$$

or, in words - that the ratio of the barrier-screening effect to the benefit take-up effect identifies the ratio of the expected cost for the *marginals* (to $d\Lambda$) to the expected need for the *marginals* (to dB).

Clearly in this model, welfare effects of $d\Lambda$ and dB depend on $\mathbb{E}\left[\kappa'_\theta(\Lambda,\varepsilon)|\varepsilon \leq \varepsilon^*_\theta\right]$ and $\mathbb{E}\left[v'_\theta(B,\varepsilon)|\varepsilon \leq \varepsilon^*_\theta\right]$, i.e. need and cost to the *inframarginals*. In Section 5 and Section 6, I estimate:

$$\frac{-\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}}{\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}} = \frac{\mathbb{E}\left[\kappa'_\theta(\Lambda,\varepsilon)|\text{Marginal to } d\Lambda\right]}{\mathbb{E}\left[v'_\theta(B,\varepsilon)|\text{Marginal to } dB\right]}$$

The most likely contributor (apart from mental health) to v' is consumption. Note, however, that those on the margin of the dB instrument (i.e. at the eligibility threshold) have the same disposable income as the infra-marginals (recipients of SA) because social assistance *tops income up to the threshold*. Moreover, Table C.3 showed that *conditional on disposable income*, there was no selection on consumption between good and poor mental health. That being said, the RKD is estimated on singles. Therefore if anything, approximating $\mathbb{E}\left[v'_\theta(B,\varepsilon)|\text{Inframarginal}\right] \leq \mathbb{E}\left[v'_\theta(B,\varepsilon)|\text{Marginal to } dB\right]$ seems reasonable.

On the other hand, those people screened out by the change of ordeals are likely less needy on average than the inframarginals. The analysis is estimated on older couples (already lower need), and Figure D.6 is consistent with marginals being more likely to be men and no more likely to have poor physical health or lower consumption. Given the positive correlation between need and cost, this suggests $\mathbb{E}\left[\kappa_\theta'(\Lambda, \varepsilon)|\text{Inframarginal}\right] \geq \mathbb{E}\left[\kappa_\theta'(\Lambda, \varepsilon)|\text{Marginal to } dB\right]$. These inequalities push in the direction of $MVPF_{d\Lambda} > MVPF_{dB}$, however are highly suggestive.

Understanding which empirical estimates help tighten these bounds is a fruitful avenue for future research. More broadly, the core empirical findings do not depend on additivity or the welfare framework. While it is debatable how to translate the fact that barriers disproportionately screen-out the vulnerable into welfare effects, the conclusion that barriers target poorly will most likely hold under most plausible modelling assumptions.

F.3.3. *Relaxing $f(v_\theta - \kappa_\theta)$ same across instruments.* Assumption (ii) follows in the case that types are one-dimensional (Landais et al., 2021). However, note that to maximise internal validity of the quasi-experimental design, sample restrictions are made both in Section 5 and Section 6. In Section 5, I focus on couples, as for them, the Participation Act was a change in ordeals only, and not also a change in benefit level. Note that the majority of individuals in this sample have income much below the threshold. In Section 6, I focus on singles, as I mis-classify couples more than singles, and in the RKD analysis, measurement error is much more consequential, because I zoom into a small window around the threshold. Moreover, I restrict to employees as I observe monthly income for this group. The samples for the two instruments are quite different, as confirmed by Tables F.2 and F.3, and the within-sample compliers may be even more different across instruments (as in Landais et al. (2021)). This is an important caveat.

However, this assumption can be relaxed if I impose additional structure. Note that **Step 2** of the identification method can be applied separately. Therefore, the result that people with poor mental health have a $2\times$ higher need and 57% higher cost than those with good mental health (relatively speaking) still holds.

Below, I employ some additional structure in order to use the correlation test to identify net value $-$ cost, which then allows for the quantification of all sufficient statistics without maintaining Assumption (ii). Instead, I maintain Assumption 3.

Assumption 3. Consider four groups defined the 2x2 combination of mental health $\theta \in \{L, H\}$ and sample $s \in$ {older couples (oc), single employees (se)}, i.e the samples for the RKD and DiD analyses respectively. First assume that $\varepsilon \sim F_s$: the choice error depends on the sample, but v_θ and κ_θ are constant across samples. Then, I assume

$$\forall \theta, \quad f_{oc}\left(v_\theta(B) - \kappa_\theta(\Lambda)\right) = \alpha \cdot f_{se}\left(v_\theta(B) - \kappa_\theta(\Lambda)\right)$$

Socio-economic Demographics	Middle-Age Couples	Single Employees
Gender (%)		
Woman	51.9	57.4
Man	48.1	42.6
Education (%)		
Primary School	3.5	12.3
High School	31.6	46.9
Bachelor's	3.5	5.0
Masters-PhD	1.9	1.9
Unknown	23.7	33.8
Main Source of Income (%)		
Employment or Civil Service Job	6.5	88.6
Director-shareholder	0.1	0.2
Self-employment	5.2	0.8
Other Job	0.1	0.0
Unemployment Insurance	2.4	2.5
Disability Insurance	8.9	1.7
Social Assistance	55.3	4.6
Other Benefits	9.9	0.4
Pension	2.2	1.0
Student Aid	0.2	0.2
Other (not active or without income)	9.2	0.0
Household Composition (%)		
Single Person Household	13.6	65.5
Couple Without Children	40.8	1.7
Couple With Children	42.1	3.8
Single Parent	13.6	26.6
Couples and Parents with Flatmates	2.1	1.3
Other Shared Households	1.4	1.1
Other Information		
Age	54.5 (5.9)	46.5 (7.8)
Foreign-born (%)	56.5 (49.6)	30.2 (45.9)
Household Std. Disposable Income (€)	5,731 (12,767)	21,739 (11,350)
Household Net Worth (€)	-5,309 (103,775)	-4,627 (51,309)
Contracted Hours (per year)	553 (498)	1,416 (552)
Eligible (%)	100.0 (0.0)	27.2 (44.5)
Receipt of Social Assistance (%)	60.4 (48.9)	9.4 (29.2)

TABLE F.2. Summary Statistics for Middle-Age Couples and Single Employees

(Mental) Health Information	**Middle-Age Couples** Mean (SD)	**Single Employees** Mean (SD)
General		
All Care Spending (€)	3,502 (9,850)	1,997 (6,818)
Physical Chronic Conditions (count)	1.61 (1.66)	0.73 (1.14)
Mental Health (admin)		
Mental Healthcare Spending (€)	381 (3,675)	14 (35)
Psychotropic Medication (%)	22.4 (41.7)	13.9 (34.6)
Anti-psychotics (%)	5.7 (23.1)	2.37 (15.2)
Anxiolytics (%)	7.1 (25.6)	2.57 (15.8)
Anti-depressants (%)	16.2 (36.8)	9.42 (29.2)
Hypnotics and Sedatives (%)	4.0 (19.5)	1.23 (11.0)
ADHD Medication (%)	0.5 (6.8)	0.64 (8.0)
Mental Health Hospitalizations (%)	0.08 (2.8)	0.06 (2.4)
Deaths by Suicide (%)	0.02 (1.4)	0.01 (1.1)
Mental Health (survey)		
Loneliness (0-11)	5.01 (3.77)	4.44 (3.75)
Life Control (7-35)	22.06 (5.73)	–
Kessler-10 Psychological Distress (10-50)	22.02 (9.81)	18.90 (8.11)

TABLE F.3. Summary Statistics for Middle-Age Couples and Single Employees (Mental Health)

I.e. that F_s implies the likelihood that θ lies on the margin of take-up scales with the *same factor* when moving from the RKD to DiD sample, for each θ.

Also assume linearity $\kappa_\theta(\Lambda) = \kappa'_\theta \cdot \Lambda$ ($\Delta\kappa_0 = 0$ without any information on heterogeneous fixed costs) and:

$$v(B) = v' \cdot (1-\tau) \cdot \underbrace{\left\{ B + y^{SA=1} - \frac{y^{SA=0}}{1+e} - \frac{e}{1+e} \frac{\left(y^{SA=1}\right)^{1+1/e}}{\left(y^{SA=0}\right)^{1/e}} \right\}}_{:=\Psi}$$

In this case, $\hat{\mathbb{P}}[SA]_L \approx \hat{\mathbb{P}}[SA]_H$, which holds for $s = se$ and $s = oc$, implies that $v'_L \cdot \Psi_L - \kappa'_L \cdot \Lambda = v'_H \cdot \Psi_H - \kappa'_H \cdot \Lambda$. Now, I estimate $\frac{\partial \hat{\mathbb{P}}[SA]_L - \hat{\mathbb{P}}[SA]_H}{\partial \Lambda}$ for $s = oc$ and $\frac{\partial \hat{\mathbb{P}}[SA]_\theta}{\partial B}$ for $s = se$. $\frac{\partial \hat{\mathbb{P}}[SA]_H}{\partial B} = 0.000227$, $\frac{\partial \hat{\mathbb{P}}[SA]_L}{\partial B} = 0.000503$ as before, implying $v'_L = 2.2$, normalising $v'_H = 1$. Now, $\frac{\partial \hat{\mathbb{P}}[SA]_H}{\partial \Lambda} = f_{oc}(v_H - \kappa_H) \cdot \kappa'_H$, as this is the group for whom I estimate the response. Therefore,

$$\begin{aligned}\frac{\partial \hat{\mathbb{P}}[SA]_L - \hat{\mathbb{P}}[SA]_H}{\partial \Lambda} &= f_{oc}(v_H - \kappa_H) \cdot \kappa'_H - f_{oc}(v_L - \kappa_L) \cdot \kappa'_L \\ &= f_{oc}(v_H - \kappa_H) \cdot \left(\kappa'_H - \kappa'_L\right)\end{aligned}$$

because $\hat{\mathbb{P}}[SA]_L \approx \hat{\mathbb{P}}[SA]_H$ for $s = oc$

$$= \alpha \cdot \underbrace{f_{se}(v_H - \kappa_H)}_{\frac{\partial \mathbb{P}[SA]}{\partial B}} \cdot \left(\kappa'_H - \kappa'_L\right)$$

$$\implies \Lambda \cdot \frac{\partial \hat{\mathbb{P}}[SA]_L - \hat{\mathbb{P}}[SA]_H}{\partial \Lambda} = \alpha \cdot (1-\tau) \times 0.000227 \cdot \left(\kappa'_H - \kappa'_L\right) \cdot \Lambda$$

I use Table D.1 to calibrate the percent change in ordeals $\frac{\partial \Lambda}{\Lambda} = 22.1\%$ - which comes from treating the final column as percent changes in each of the scores (second column) where the score cannot exceed 100%.

$$\implies \frac{-0.009}{0.221} = \alpha \cdot (1-\tau) \times 0.000227 \cdot \left(\kappa'_H - \kappa'_L\right) \cdot \Lambda$$

But $v'_L \cdot \Psi_L - \kappa'_L \cdot \Lambda = v'_H \cdot \Psi_H - \kappa'_H \cdot \Lambda$

$$\implies \frac{-0.009}{0.221} = \alpha \cdot (1-\tau) \times 0.000227 \cdot \left(v'_H \Psi_H - v'_L \Psi_L\right)$$

I also calibrate $\Psi_L = 506, \Psi_H = 501$ using $e = 0.2$ (Bargain et al., 2014).

$$\implies \frac{0.009}{0.221} = \alpha \cdot (1-\tau) \times 0.000227 \cdot (2.2 \times 506 - 501)$$

Rearranging gives $\alpha = 0.46$. Therefore, $\kappa'_H = -\frac{\partial \hat{\mathbb{P}}[SA]_H}{\partial \Lambda} \cdot \frac{1}{f_{oc}(v_H - \kappa_H)} = -\frac{\partial \hat{\mathbb{P}}[SA]_H}{\partial \Lambda} \cdot \frac{1}{\alpha \cdot f_{se}(v_H - \kappa_H)} = \frac{0.01568}{0.46 \times (1-\tau) \times 0.000227} = 238$, and $\kappa'_L = 376$. This only pushes in the direction of $MVPF_{d\Lambda} > MVPF_{dB}$. For estimates, see Table II ("Different Marginals").

F.4. **Welfare Effects.**

F.4.1. *Calibration with bias.* How does bias affect the quantification of welfare effects? This requires us to evaluate the size of MI_θ, the marginal internality for each type. According to the theory,

$$MI_\theta = (1 - \psi) \cdot \kappa_\theta(\Lambda) \tag{F.5}$$

Note that the marginal internality depends on *average* ordeal-costs, rather than marginal ordeal-costs. In order to evaluate this term, I make the linearisation $\kappa_\theta(\Lambda) = \kappa_\theta \cdot \Lambda$. Linearity is a natural approximation for $\kappa_\theta(\Lambda) = \kappa_\theta \cdot \Lambda$, where Λ represents the amount of ordeals and κ_θ is the per-unit cost and is assumed by Finkelstein and Notowidigdo (2019).

Therefore, evaluating the new $MVPF$ formulas requires taking a stance on what Λ is. As discussed in Appendix F.3, qualitative evidence from municipalities suggests the percent change in Λ due to the Participation Act is an increase of 22.1%. Further, I assume that the Participation Act represented an absolute change in Λ of 1 unit. Therefore, $\Lambda = 1/0.221 = 4.52$. For example, Λ could represent number of hours spent on obligations, and κ_θ is the welfare cost per hour spent. When $\kappa_\theta(\Lambda) = \kappa_\theta \cdot \Lambda$, $\kappa_\theta = \kappa'_\theta(\Lambda)$. Therefore, given the estimates from Section 7:

$$MI_L = (1 - \psi) \cdot 4.52 \cdot 173 = (1 - \psi) \cdot 782$$
$$MI_H = (1 - \psi) \cdot 4.52 \cdot 110 = (1 - \psi) \cdot 497$$

These estimates mean that we can quantify how large the $MVPF$ formulas are for different values of ψ. For $\psi = 1$ - the $MVPF$ are as Section 7. What if ordeal-costs were a pure bias which affects behaviour only but not welfare? Then:

$$MVPF_{d\Lambda}^{\psi=0} = 0.54$$
$$MVPF_{dB}^{\psi=0} = 1.43$$

$MVPF_{d\Lambda}^{\psi=0} < MVPF_{d\Lambda}^{\psi=1}$ as there is no direct welfare effect of the increase in barriers. $MVPF_{d\Lambda}^{\psi=0} \neq 0$, however, because of the negative behavioural welfare effect. $MVPF_{dB}^{\psi=0} > MVPF_{dB}^{\psi=1}$ because of the internality correction that an increase in benefits provides. Finally, we can quantify the level of bias ψ^* required to reverse the welfare ordering $MVPF_{d\Lambda} > MVPF_{dB}$. This turns out to be $\psi^* \approx 35\%$. That is to say, the government needs to be confident that at least 65% of the as-if ordeal-costs are purely a bias in order to reverse the welfare conclusions. Alternatively, as long as people don't over-estimate the size of the cost by a factor of 3, then the welfare conclusions remain robust. Finally, note that $d\Lambda$ is unsurprisingly more sensitive to bias than dB.