# Mental Health and the Targeting of Social Assistance

Canishk Naik
UC Berkeley

October 2025

Latest version here

Abstract. People living with mental disorders are at a higher risk of needing income-support programs but face greater difficulty overcoming barriers to access. This paper investigates whether social assistance effectively reaches people with poor mental health. I measure mental health and social assistance take-up using Dutch administrative data and develop a theoretical framework to show how take-up responses can identify the marginal value of benefits (need) and the cost of barriers. These are key components for evaluating targeting effectiveness. I find that a policy increasing barriers disproportionately screens-out those with poor mental health, indicating a 64% higher cost of these barriers. Despite their higher cost, people with poor mental health have the same average take-up levels as those with good mental health, conditional on eligibility, which suggests greater need. To assess this, I show that individuals with poor mental health are more responsive to plausibly exogenous variation in benefits than those with good mental health, demonstrating that their need is twice as high. These estimates imply that people with poor mental health are inefficiently excluded from low-income welfare assistance by barriers. Consequently, reducing barriers to take-up would be twice as effective as increasing benefits.

**JEL Codes:** D91, H51, H53, I14, I38

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, 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 bad health was estimated to be $2.5 trillion, and is expected to more than double by 2030 (Bloom et al., 2012). Symptoms of mental disorders include worthlessness, confused thinking, withdrawal from support networks, fear, fatigue, guilt and, in the extreme case, suicidality (APA, 2013). Additionally, people with poor mental health face up to three times the risk of poverty (Ridley et al., 2020). Therefore, people struggling with mental disorders are especially vulnerable.

Modern welfare states are rooted in the principle that society should protect its most vulnerable members. Ensuring that safety net programs effectively reach those in need is essential to upholding the social contract. However, administrative and psychological costs often make it difficult to access social support, which leads to widespread non-take-up (Ko and Moffitt, 2024). In theory, application barriers could help filter out people with lower need (Nichols and Zeckhauser, 1982), but in practice people suffering from mental illness find it more challenging to overcome take-up barriers than those with good mental health (Bell et al., 2022).

This paper investigates whether social assistance effectively reaches people with poor mental health. A key concern is that the very source of vulnerability is also what makes it difficult to overcome barriers to help. Nevertheless, the inefficiency arising from excluding individuals with mental disorders from assistance has remained undocumented.

Mental disorders pose important theoretical and empirical challenges in determining effective targeting. I focus on low-income welfare benefits in this study. Theoretically, eligibility for these programs is determined by people having few resources. The goal of barriers is then to target for *general* unobservable need. The challenge is that poor mental health affects the *cost* of overcoming barriers, which decreases take-up, as well as in principle the *need* for support, which increases take-up. Therefore, take-up does not distinguish between these channels, but separating them is essential for assessing effectiveness. This is because barriers target well if the needy can afford the cost and the less needy cannot.[1] Empirically, measuring mental health at scale is challenging with survey data, due to small samples and under-reporting due to stigma (Bharadwaj et al., 2017), but also with administrative data: objective outcomes are often extreme and people with poor mental health forgo care (Cronin et al., 2024).

[1]This theoretical challenge applies to the wide range of social programs where eligibility does not directly depend on mental health but the eligible population contains many people with mental disorders. The exception is disability insurance. Here, Godard et al. (2022) emphasize that the key policy challenge with mental disorders is that they are hard for case-workers to observe.

I address these challenges in three steps. First, I develop a theoretical framework to disentangle need for benefits from the cost of overcoming barriers using take-up responses to changes in benefits and barriers. Second, I empirically estimate take-up levels and responses of low-income benefits, heterogeneously by mental health, using Dutch administrative data. The data contain rich information on mental health from administrative sources and a large ($N \approx 400$k) linked survey, as well as on social assistance eligibility and take-up. Finally, I combine theory and empirics to calculate how need and cost vary with mental health and evaluate welfare consequences of the targeting of social assistance.

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 the marginal value of benefits (need) and the cost of barriers. To show this, I develop a theoretical framework allowing for heterogeneity in both need and cost.[2] There are three components to identification. (i) Differences in average take-up levels reflect how *average* value net of cost compares across the population. (ii) If an individual responds more to a change in benefits, either they have high need (*marginal* value) or they were at the margin of taking-up versus not (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 the cost of overcoming barriers depend on mental health is crucial for policy-making. The former is the social welfare gain from transferring €1 from someone with good mental health to someone with poor mental health. The latter reflects the welfare costs that barriers impose on individuals. Therefore, these key primitives characterise the benefits and costs of targeting social assistance using barriers. For example, need, cost and take-up responses to benefits and barriers are sufficient to calculate the welfare effects of a budget-neutral increase in barriers, where the money saved due to lower take-up is used to finance an increase in benefit level.[3]

My theoretical framework adopts a reduced-form approach to identify need and cost across groups without making strong assumptions about underlying mechanisms. The baseline analysis treats need and cost as welfare-relevant, but a natural question arises: should governments give

[2]In the framework, need and cost can vary across people with the same income. Thus, need cannot be controlled for by holding income constant, creating a new identification challenge relative to past work on targeting (Finkelstein and Notowidigdo, 2019; Rafkin et al., 2023).

[3]This is an example of a policy experiment which captures the essence of how effective it is to use barriers to target (Zeckhauser, 2021; Ko and Moffitt, 2024).

full normative weight to choices made under the cognitive and emotional constraints of mental disorders? Following Naik and Reck (2025), I characterize how policy conclusions change depending on the extent to which choices reveal true preferences. This approach delivers robust policy guidance without requiring strong prior commitments about the welfare-relevance of decision-making with poor mental health.

Empirically, I study social assistance take-up and mental health using administrative data for the population of the Netherlands (17 million people). I examine the flagship Dutch social assistance program, the *algemene bijstand*,[4] a cash transfer designed for people who don't have enough money to subsist. I combine detailed information on socio-economic demographics for the years 2011 - 2020 to newly construct an accurate measure of eligibility with low measurement-error.[5] 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 survey which is linked to the administrative data.[6] I combine these outcomes to reliably proxy for mental health status: this is not possible with survey or admin data alone (Kolstad et al., 2024).

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, however, conditional on eligibility, they take-up at the same rate. I find that one quarter of people eligible for social assistance have been diagnosed with a mental disorder, more than double the rate for the general population. However, the average take-up levels ($60\%$) do not meaningfully differ by mental health status conditional on eligibility, income and other covariates.

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 (threat of) sanctions (Ministerie van SZW, 2022). I use a difference-in-differences

[4]Literal translation: general assistance. Information about the benefit can be found on the Dutch Government website. I will refer to this program as social assistance (SA). Social assistance is more prevalent than unemployment or disability benefits, with around 400,000 recipients every year. Eligibility is primarily determined by income being below 100% of the full-time net minimum monthly wage for couples (70% for singles).

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

[6]The outcomes are: care usage (mental healthcare spending, dispensations of psychotropic drugs), extreme outcomes (hospitalisations for a mental health condition, deaths by suicide) and subjective mental health from a large survey (psychological distress, loneliness and perceived control over own life).

design to show that the reform disproportionately discourages people with poor mental health from flowing into the program, reducing their inflow by 10% compared to those with good mental health. There is no differential outflow, indicating that psychological costs of anticipated obligations—rather than actual compliance burdens—drive the screening effect.

Third, people with poor mental health respond twice as much to a change in benefits compared to those with good mental health. 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 benefits, 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.

Combining theory and empirical estimates yields the final key finding of the paper: people with poor mental health need benefits twice as much as those with good mental health, conditional on income, but also have a $64\%$ higher cost of overcoming barriers. These primitives suggest that governments have an incentive to redistribute money to people with poor mental health, but that barriers are not an efficient way to do so. I estimate the marginal value of public funds (Hendren and Sprung-Keyser, 2020), capturing the direct welfare effect of the policy divided by net government cost, of a reduction in barriers as 2.34 and of an increase in benefits as 0.91. This implies reducing barriers is an effective use of government funds, 2.4$\times$ more so than increasing benefits. This conclusion is robust to whether ordeal costs reflect true welfare costs or behavioural frictions: regardless, the welfare gain of increasing benefits is limited by poor targeting.

**Contribution to the Literature:** This paper contributes to the public economics literature on the targeting of government programs. There is an ongoing empirical debate on who gets screened out of assistance by barriers 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) or adverse selection (Shepard and Wagner, 2022) can undermine effectiveness. The classic view from Nichols and Zeckhauser (1982) suggests that ordeal mechanisms are effective when need and the cost of ordeals are weakly negatively correlated and Rafkin et al. (2023) argues that self-targeting can be socially beneficial on average. However, a full cost-benefit analysis requires quantifying the trade-off between the costs of ordeals and the need for benefits—potentially extending beyond poverty—that can be redistributed to infra-marginals.

My theoretical framework shows that neither take-up levels nor responses to ordeals are sufficient statistics for characterising this trade-off, precisely because need can co-vary with cost across the population. An additional moment—take-up responses to changes in benefits—is necessary to evaluate welfare implications. Importantly, I show the cost-benefit analysis of ordeals is robust to whether ordeal costs reflect true welfare costs or behavioural misperceptions: poor targeting limits the upside from benefit-level increases either way. While this framework is broadly applicable, it also brings attention to an especially understudied dimension: mental health.

I provide one of the first quantifications of the welfare consequences of excluding people with poor mental health from assistance. 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), understanding welfare effects requires assessing need. I show that individuals with mental disorders need benefits more than those with good mental health, even when conditioning on income. This highlights that vulnerability is often multi-dimensional and extends beyond poverty.

The idea that mental disorders not only increase need but also make it harder to navigate barriers mirrors the dual effects highlighted in the scarcity literature (Mullainathan et al., 2012). Financial strain can impair cognition (Mani et al., 2013; Kaur et al., 2021), yet it can also sharpen focus and lead to better decisions (Shah et al., 2012; Fehr et al., 2022). This paper proposes a theoretical framework to discipline these opposing forces and implements it empirically using rich data and policy variation in benefits and barriers.

Lastly, there is a growing literature in psychology and 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 healthcare affects economic outcomes (Barker et al., 2021; Baranov et al., 2020; Serena, 2024), how income impacts mental health (Christian et al., 2019; Miller et al., 2024; Schmidt et al., 2021; Silver and Zhang, 2022) and the drivers of psychotherapy demand (Abramson et al., 2024; Cronin et al., 2024; Roth et al., 2024).

I quantify the policy relevance of the cognitive and emotional burdens that mental disorders impose on individuals by empirically estimating the costs of ordeals for these people. Moreover, I estimate take-up responses to show that people with poor mental health have a higher *perceived* need for welfare benefits than those without mental disorders. This new finding shows that non-take-up of assistance among people with poor mental health does not stem from under-valuation, but rather the challenges of accessing benefits.

These results support Sen's "capabilities approach" (Sen, 1999, 2008); those facing greater daily challenges, such as disabilities, require more resources to satisfy basic needs. My analysis indicates that the same cognitive bandwidth and emotion regulation constraints that heighten the costs of overcoming barriers also appear to exacerbate everyday stressors enough to significantly raise the marginal value of additional income. These constraints plausibly affect everyone to varying degrees and have significant implications for inclusive program design, indicating that ordeals may not be effective.

**Outline:** Section 2 sets out my theoretical framework to characterize the social welfare consequences of targeting. In Section 3, I describe the context and data. I quantify social assistance take-up levels by mental health status in Section 4. I estimate take-up responses to changes in barriers in Section 5 and benefits in Section 6. Section 7 combines the theory and empirical estimates to calculate welfare effects. 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. I develop a method to separately identify need from the cost of overcoming barriers using take-up responses to changes in benefits and barriers.[7] The framework is agnostic about whether behaviour reflects welfare-relevant preferences or contains psychological frictions. I first derive empirically-implementable formulas for welfare effects of policies affecting targing under revealed preference as a baseline, then discuss in Section 2.4 how conclusions change depending on the extent of bias. Proofs and extensions are in Appendix A.

### 2.1. Model of Social Assistance Take-up.

[7] This distinction relates to Shepard and Wagner (2022), who show that adverse selection can undermine ordeal mechanisms due to correlation between value and cost. Importantly, in their setting cost refers to the cost of insurance (borne by the government), whereas I focus on the cost of ordeals (borne by the individual).

2.1.1. *Setup.* Individuals are indexed by $\theta$.[8] Social assistance is defined by two policy parameters. $B$ is the (monetary) benefit, $\Lambda$ is the barrier that individuals have to overcome to receive $B$. Each $\theta$ makes one key choice: whether to receive social assistance:

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

Preferences are defined as follows. Individuals derive value $v_\theta(B)$ from benefits $B$. There is an take-up cost $\kappa_\theta(\Lambda)$, which represents the individual-specific dis-utility from overcoming barrier $\Lambda$. Importantly, these quantities can be interpreted either as welfare-relevant primitives (under revealed preference) or as perceived values and costs (under psychological frictions). I also model take-up to depend on an independent additive choice-shock $\varepsilon \sim F$ which can be thought of as decision-relevant unobservables which are unaffected by policy. Therefore, the take-up equation for each $\theta$ is:

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

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

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

This model takes a stylised reduced-form revealed preference approach, where individual values and costs are modelled as catch-all quantities that could arise from various underlying mechanisms. Rather than imposing structure on the psychological processes generating these preferences, I focus on identifying the key sufficient statistics—need and cost—that determine targeting effectiveness. This approach prioritizes parsimony and empirical tractability, given limited evidence on welfare effects for individuals with poor mental health.[9]

Nevertheless, 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. All income (including benefits) is taxed at marginal tax rate $\tau$.

[8] In my empirical setting, $\theta$ will represent mental health status, but the following model applies to any other dimension of heterogeneity which could influence the marginal value of €1 as well as the take-up cost.
[9] See Section 2.3.2 for a discussion of all key identification assumptions.

2.2. **Welfare under Revealed Preference.** As a tractable benchmark, I first characterize welfare assuming choices reveal preferences. This treats both $v_\theta(B)$ and $\kappa_\theta(\Lambda)$ as normatively relevant.

2.2.1. *Individual Welfare.* Denote $U_\theta$ as $\theta$'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 $SA = 0$. Let $\varepsilon^*_\theta = v_\theta(B) - \kappa_\theta(\Lambda)$. Then:

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

2.2.2. *Social Welfare.* Let $\mu(\theta)$ be the distribution of types, and $\lambda_\theta$ social welfare weights. The government's problem is given by:

$$\begin{aligned} W = \max_{\Lambda, B} \int \lambda_\theta\, \mathcal{U}_\theta\, d\mu(\theta) \\ \text{s.t. } \underbrace{\int \tau y_\theta^{SA=0} \cdot \big(1 - \mathbb{P}[SA]_\theta\big) + \tau(y^{SA=1} + B) \cdot \mathbb{P}[SA]_\theta\, d\mu(\theta)}_{\text{Tax Revenue}} = \underbrace{\int B \cdot \mathbb{P}[SA]_\theta\, d\mu(\theta)}_{\text{Program Costs}} \end{aligned} \tag{2.4}$$

In this framework, I assume eligibility criteria for benefits are fixed (though not explicitly modelled).[10] $\tau$ is also fixed. The government does not observe individuals' private types $(\theta, \varepsilon)$, making targeted policy design challenging. Instead, it must rely on blunt instruments—benefit levels ($B$) and barriers to access ($\Lambda$)—which do not vary by $\theta$ to indirectly target those most in need. The policy-maker's goal is to allocate benefits to individuals with a high, unobservable marginal value of benefits ($v'_\theta(B)$, i.e., need). Barriers ($\Lambda$) effectively target when neediest receive assistance, while those with lower need do not. Section 2.2.3 derives formulas for the welfare effect of an example policy experiment capturing this mechanism.[11] This is one way of characterising the effectiveness of targeting using barriers.

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

[11] See, e.g. Ko and Moffitt (2024) who say that the "presence of costs induces the less needy to not apply, which saves government funds that can then be used to pay higher benefits to those in greater need, who have a higher probability of ending up as recipients."

2.2.3. *Welfare Effects of a Budget-Neutral Increase in Barriers.* I consider a policy experiment capturing the essence of using barriers to target social assistance: increase barriers, saving government funds due to lower take-up, in order to finance an increase in benefit level. This is a budget neutral increase in $\Lambda$ ($B$ adjusts).

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

$$\frac{dW}{d\Lambda} = \int \lambda_\theta \, \mathbb{P}[SA]_\theta \left[ \underbrace{v_\theta'(B)}_{\textit{Need}} \cdot \frac{dB}{d\Lambda} - \underbrace{\kappa_\theta'(\Lambda)}_{\textit{Cost}} \right] d\mu \tag{2.5}$$

*Budget Neutrality implies:*

$$\frac{dB}{d\Lambda} = \frac{-\int FE_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda} d\mu}{(1-\tau) \cdot \int \mathbb{P}[SA]_\theta \, d\mu + \int FE_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial B} d\mu} \tag{2.6}$$

*where:*

$$FE_\theta = \tau \cdot (y_\theta^{SA=0} - y_\theta^{SA=1}) + (1-\tau) \cdot B \tag{2.7}$$

Equation (2.5) follows from an application of the Envelope Theorem. The expression shows that the overall welfare effect is large whenever take-up is high ($\mathbb{P}[SA]_\theta$ large) among the $\theta$'s with the highest need ($v_\theta'$ large) and *lowest* ordeal-costs ($\kappa_\theta'$ small). Analogously, $\frac{dW}{d\Lambda}$ will be negative when need and cost are strongly positively correlated.

The intuition behind Equation (2.6) is as follows. Budget-neutral policy changes depend on aggregate responses only. The government can increase $B$ more if *more* people are screened out by ordeals, if people take-up *less* in response to changes in benefit level and if there are *fewer* beneficiaries at baseline. $FE_\theta$ is the fiscal externality of $\theta$ applying: there is a moral hazard fiscal externality due to labour supply response $y^{SA=0} \to y^{SA=1}$ which costs the government $\tau(y^{SA=0} - y^{SA=1})$, and a direct cost $(1-\tau)B$ paid out to $\theta$.

The welfare effects depend on four key sufficient statistics. Increasing barriers imposes a direct cost on infra-marginal individuals: $\kappa_\theta'(\Lambda)$. However, the government saves money due to lower take-up. This depends on the strength of *barrier screening effects*, $\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}$. Increasing benefits has redistributive value for infra-marginal individuals: $v_\theta'(B)$. However, it costs the government money. This depends on the strength of *benefit take-up effects*, $\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}$.[12]

[12]The policy experiment differs from Rafkin et al. (2023) in what we compare 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.

$\frac{dW}{d\Lambda}$ is my overall metric for the social welfare consequences of targeting using ordeal mechanisms. However, the units are hard to interpret. In the calibration, I use the framework of Hendren and Sprung-Keyser (2020) to aid intuition by also deriving the marginal value of public funds (MVPF) of a decrease in barriers vs an increase in benefit level. The MVPF is defined as willingness-to-pay divided by government cost (both money-metric). The formulae for the MVPF of $dB$ and $d\Lambda$ are derived in Appendix A.1. While MVPFs have interpretable units, they do not capture $\theta$'s having different marginal values of income unless social welfare weights are included. Therefore, I also calculate MVPFs with utilitarian (rather than money-metric) social welfare functions. The comparison of utilitarian $MVPF_{dB}$ and $MVPF_{d\Lambda}$ is isomorphic to $\frac{dW}{d\Lambda}$.

2.3. **Identification.** How should we empirically characterise the welfare consequences of targeting using barriers? Proposition 2.1 is an example showing that in order to know whether barriers target effectively, we must estimate four key "sufficient statistics": need ($v'_\theta$), cost ($\kappa'_\theta$), benefit take-up effects $\left(\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}\right)$ and barrier screening effects $\left(\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}\right)$.[13]

My goal is to quantify these statistics empirically. Therefore, it is helpful if there are as few as possible. First, I use theory to reduce the number of sufficient-statistics from 4 to 3. The key idea is that benefit take-up effects $\left(\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}\right)$ depend on the marginal value of benefits, i.e. need ($v'_\theta$). Similarly, barrier screening effects $\left(\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}\right)$ depend on the cost of barriers ($\kappa'_\theta$).

*Remark* 2.1. Barrier screening effects are characterised by Equation (2.8), and benefit take-up effects by Equation (2.9).

$$\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda} = -\kappa'_\theta \cdot f_\varepsilon(v_\theta - \kappa_\theta) \tag{2.8}$$

$$\frac{\partial \mathbb{P}[SA]_\theta}{\partial B} = v'_\theta \cdot f_\varepsilon(v_\theta - \kappa_\theta) \tag{2.9}$$

Intuitively, $\Lambda$ is a price of taking up. Therefore, responsiveness to take-up is large when consumers are price-responsive ($\kappa'$ large) or just 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 probability of being marginal. This means that there are only three key primitives which determine welfare effects: need, cost and $f_\varepsilon(v_\theta - \kappa_\theta)$, the likelihood of being on the margin of take-up. The latter is an scaling factor allowing for the inference of infra-marginal costs/benefits through marginal take-up responses.

[13]Therefore, my model aligns with the sufficient-statistics approach to public economics (Einav and Finkelstein, 2011; Baily, 1978; Chetty, 2008).

2.3.1. *Three-step Identification.* In this section, I present a three-step method to identify the three key statistics sufficient for evaluating welfare effects. 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 ($v'_\theta$), cost ($\kappa'_\theta$) and the likelihood of being marginal ($f_\varepsilon(v_\theta - \kappa_\theta)$). The intuition is as follows:

Difference in take-up levels $\mathbb{P}[SA]_\theta$ across types cannot distinguish between average value ($v_\theta$) and cost. However, they reflect how average value *net* of cost compares across types. This, in turn, influences how $f_\varepsilon(v_\theta - \kappa_\theta)$ compares across types. Using this information, cost can be inferred from barrier screening effects $\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}$. The idea is that $\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}$ being large reflects either large $\kappa'$ or average value net of cost being close to zero. The latter can be isolated using difference in take-up levels. Similarly, the contribution of $f_\varepsilon(v_\theta - \kappa_\theta)$ to benefit take-up effects $\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}$ can be isolated from need.

**Step 1 (Average take-up levels):** To aid intuition, suppose that we are in a special case of equalised take-up levels: $\mathbb{P}[SA]_\theta = \mathbb{P}[SA]_{\tilde{\theta}}$.[14] 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. 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 how need varies across types. This done by dividing Equation (2.9) *across types* to give:

$$\frac{\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}}{\frac{\partial \mathbb{P}[SA]_{\tilde{\theta}}}{\partial B}} = \frac{v'_\theta}{v'_{\tilde{\theta}}} \cdot \underbrace{\frac{f_\varepsilon(v_\theta - \kappa_\theta)}{f_\varepsilon(v_{\tilde{\theta}} - \kappa_{\tilde{\theta}})}}_{\text{Estimated in } \textbf{Step 1}} \tag{2.10}$$

Then, if we normalise $v'_{\theta_0} = 1$ for some $\theta_0$ we calculate $v'_\theta$ for all other $\theta$ using Equation (2.10). This normalization is without loss, and effectively scales all welfare effects in terms of $\theta_0$'s WTP for €1.[15]

**Step 3 (Barrier screening effects):** Finally, divide barrier screening effects from Equation (2.8) by benefit take-up effects from Equation (2.9) *within type* to identify $\kappa'_\theta$ for all $\theta$ as follows:[16]

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

[15] $v'_\theta$ is then understood as $\theta$'s need relative to $\theta_0$.

[16] 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

$$\frac{-\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}}{\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}} = \kappa'_\theta \cdot \underbrace{\frac{1}{v'_\theta}}_{\text{Estimated in } \textbf{Step 2}} \tag{2.11}$$

2.3.2. *Discussion of Key Assumptions.* Before presenting the empirical analysis, it is important to discuss the key assumptions underlying the identification of need and cost. In Section 7, I return to these assumptions and characterise how relaxing them impacts welfare effects.

I assume that $\varepsilon$ is an additive independent shock to the take-up equation: $SA = 1 \iff v_\theta > \kappa_\theta + \varepsilon$. Independence, as assumed in random utility models (McFadden, 1981; Woodford, 2020), enables **Step 1** in the identification. The assumption can be probed by examining how including additional covariates changes the three-step identification.[17] Without independence, the model is not identified and either $v_\theta$ constant across $\theta$ or $\kappa_\theta$ constant across $\theta$ must be assumed. Seeing as the purpose of the framework was to separate need and cost when both could depend on mental health, neither of these cases is useful.[18] Additivity allows me to separate need/cost from the scaling factor $f_\varepsilon(v_\theta - \kappa_\theta)$ in Equations (2.8) and (2.9).

The framework assumes that the likelihood of being on the margin of take-up, $f(v_\theta - \kappa_\theta)$, is the same for benefit and barrier instruments. This comes from $\theta$ and $\varepsilon$ being one-dimensional. The assumption is called into question by recent work arguing that the compliers to an instrument depend on the instrument itself (Kline and Walters, 2019; Mogstad et al., 2024). This assumption allows for minimal structure on the take-up equation. Relaxing $f(v_\theta - \kappa_\theta)$ to depend on the instrument is possible under additional parametric assumptions as long as $f(v_\theta - \kappa_\theta) = f(v_{\tilde{\theta}} - \kappa_{\tilde{\theta}})$ for all $\theta, \tilde{\theta}$, i.e. as long as the difference in complier characteristics across instruments is orthogonal to mental health. Appendix F shows welfare effects in this case.

In the theory, $\theta$ is treated as an immutable type, but in practice mental health evolves over time and in response to stimuli. This assumption is made in order to set out a tractable static framework. In Section 6, I show that social assistance does not appear to have a strong dynamic positive effect on mental health. However, ordeals likely worsen mental health, a dynamic I cannot quantify in this paper. This effect would imply that the welfare costs of increasing ordeals are a lower-bound.

is new. Here, the key novelty is that I can estimate take-up levels with information on eligibility and use these to inform differences in likelihood of being on the margin of take-up across types.

[17]Throughout my empirical analysis, including additional covariates does not meaningfully change the comparison between good and poor mental health, providing support for $\varepsilon \perp \theta$.

[18]In settings where it seems reasonable that $v_\theta \perp \theta$ or $\kappa_\theta \perp \theta$, models from Finkelstein and Notowidigdo (2019); Rafkin et al. (2023) should be used.

2.4. **The Role of Bias.** The revealed-preference framework above treats observed take-up responses as reflecting welfare-relevant preferences. This allows clean identification of need and cost and straightforward application of the Envelope Theorem when deriving welfare effects. However, take-up decisions may reflect behavioural frictions that create a wedge between revealed and true preferences. This is particularly salient for people with poor mental health, who face cognitive and emotional constraints that could distort decision-making.

A natural question is whether governments should give normative weight to choices made under such constraints. Adopting a purely paternalistic stance—where only decisions made with good mental health receive normative weight—requires the government to make untestable normative assumptions about which psychological states produce welfare-relevant preferences. My approach takes an intermediate position: I characterize welfare under revealed preference as a baseline, then assess how conclusions change depending on the extent to which behaviour deviates from true welfare (as in Naik and Reck, 2025). This allows policymakers to evaluate robustness without requiring strong prior commitments about the welfare-relevance of choices made with poor mental health.

Appendix A.3 presents a formal extension where perceived values and costs may diverge from true welfare-relevant quantities. The three-step identification method from Section 2.3.1 continues to apply, but now reveals *perceived* need and cost rather than true welfare primitives. The welfare effects of Section 2.2 remain similar in structure, but now include an additional behavioural welfare effect. For example, if people overstate the cost $\kappa(\Lambda)$, then reducing $\Lambda$ generates welfare gains not only through the direct cost reduction but also through a debiasing effect that brings behaviour closer to the intrinsic optimum.

Quantifying this behavioural welfare effect requires estimating average values and costs, not just their marginal counterparts. In the empirical application, I make linearity assumptions to achieve this identification. Linearity is a natural approximation for $\kappa_\theta(\Lambda) = \kappa_\theta \cdot \Lambda$, where $\Lambda$ represents the amount of ordeals and $\kappa_\theta$ is the per-unit cost.[19] Linearity is less attractive for $v_\theta(B)$, but the empirical evidence in Section 5 and Section 6 suggests behavioural frictions are more plausible on the cost side.

The rest of the paper evaluates the key sufficient statistics and discusses implications for welfare. In Section 4, I estimate average take-up levels and show that they do not meaningfully differ by mental health. I estimate barrier screening effects in Section 5 and benefit take-up effects in Section 6.

[19]Indeed, linearity is assumed by Finkelstein and Notowidigdo (2019).

# 3. Context and Data

## 3.1. **Institutional Context.**

3.1.1. *Social Assistance in the Netherlands.* 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 4.5% of the general population and is more than the number of people on disability and unemployment insurance. Figure B.1 shows the evolution of caseload from 2005 to 2021.

**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 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 center. 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.[20] 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.[21]

[20] Basic income experiments have been trialled in some municipalities, where some treatment arms reduce/remove obligations and other treatment arms reduce the 100% claw-back of benefits (Verlaat and Zulkarnain, 2022). Strict obligations are rationalised by wanting to incentivise activation and eventual transitioning out to paid work in the face of the 100% marginal tax rate.

[21] 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.[22] 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. *Healthcare in the Netherlands.* 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 with psychopharmacalogical medications each year. Access to mental healthcare appears to be roughly equalised by income (Kolstad et al., 2024), although quality of care may differ (Lopes et al., 2023).

3.1.3. *Disability Insurance in the Netherlands.* One potential concern about my analysis is that perhaps it's not social assistance people with poor mental health should be receiving, but disability insurance. However, 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. In sum, those with moderately poor mental health are in the target population.

3.2. **Data.** In order to quantify the nature of selection of SA recipients with respect to mental health, I use administrative data from the population of the Netherlands ($\approx$ 17 million people) 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.

[22]Full list of obligations can be found in Ministerie van SZW (2015). They include 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.

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 socio-economic information, including 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.[23] The main analysis sample is from 2011 to 2020 because I observe all eligibility determinants in this period. I focus on working-age individuals throughout the study - age 27-65.[24] Among this population, around 10% 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 in each month, 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 surveyed psychological distress (Kessler's 10), loneliness and perceived control over own life (in a linked survey for 400k people in 2012 and 2016). Table B.2 shows summary statistics about (mental) health.

Figure 1 shows the prevalence of poor mental health in the Netherlands, and how this varies when focusing on the general population, those eligible for social assistance and recipients. The figure shows that the eligible are at least $2.5\times$ more likely to suffer with poor mental health than the general population. Whereas, social assistance recipients seem to have similar mental health to the eligible population.

[23] I also calculate eligibility monthly for people who work - as the data contain monthly income information for employees. I use this for the regression kink design in Section 6.

[24] As in Inspectie SZW (2021), eligibility for students and people above pension-age is noisier and so these groups are excluded.

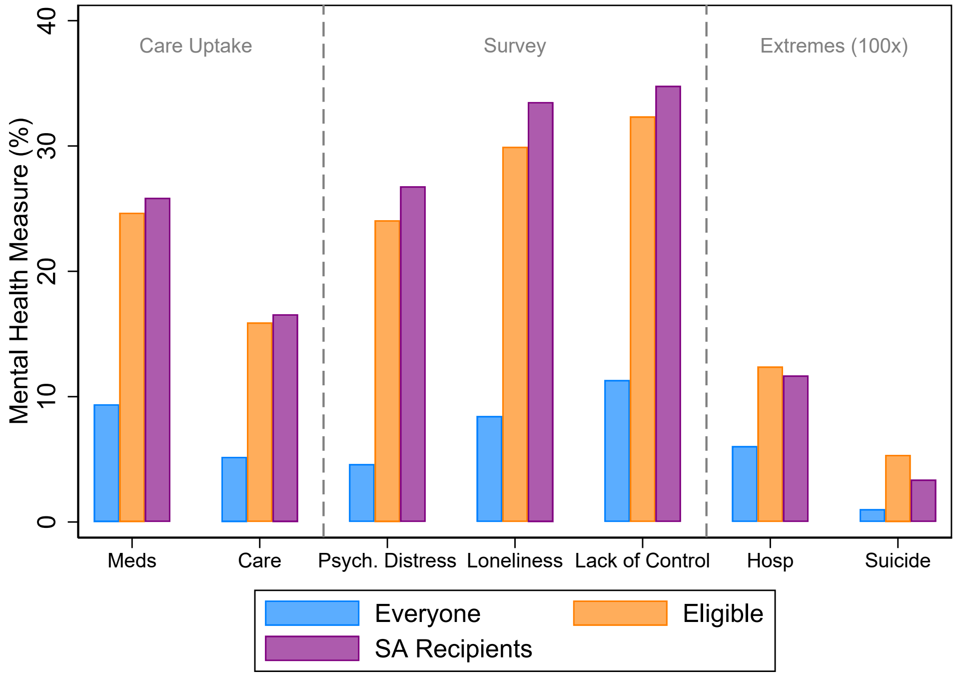



FIGURE 1. Prevalence of poor mental health in the Netherlands. This graph shows raw means of the seven mental health measures, across three different populations. All measures are in terms of percentages and are probabilities of the following: dispensed psychotropic medications, $> 0$ mental healthcare 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 assistaance, and the social assistance recipients, from 2011-2020 in each case.

This suggests limited self-targeting; if self-targeting were effective, the more vulnerable group would be over-represented among recipients. This implies that people with mental disorders either do not value benefits more than those with good mental health or that they do but find barriers costlier to overcome. Section 2 says the first step to distinguishing between these explanations is the difference in average take-up levels.

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 $SA_{it} = \mathbb{1}\{i$ receives SA in period $t\}$. For almost all results, this will refer to a stock. How should we measure poor mental health? I define Poor $\text{MH}_{it} = \mathbb{1}\{i$ dispensed psychotropic medications in year $t\}$.

In a related paper, we show that this is an accurate proxy for poor mental health status (Kolstad et al., 2024). 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 (Kolstad et al., 2024). Prescriptions are done by GPs, who are the first access point to healthcare. In general, access to healthcare in the Netherlands is excellent, and people often still receive care even if they default on their premia (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 verify that all findings about mental health measured by dispensations of psychotropic drugs are consistent when mental health is measured in the survey.

## 4. Average Take-up Levels

Average levels of the take-up of social assistance by mental health are useful descriptives to examine targeting and important inputs to identification of need and cost.

First, in terms of raw levels, figure Figure 2 shows the baseline probability of being eligible for social assistance by mental health, measured by psychotropic drug dispensation, as well as the take-up levels conditional on eligibility. People with poor mental health are three times more likely to be eligible, but conditional on eligibility seem to take-up around the same rate as those with good mental health.

4.1. **Design.** Do 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)? This is **Step 1** in the three-step identification from Section 2.3.1.

For each individual $i$ and year $t$, define $SA_{it} = \mathbb{1}\{i$ receives SA in year $t\}$. Poor $\text{MH}_{it} = \mathbb{1}\{i$ dispensed psychopharma. in year $t\}$. Equation (4.1) represents a *correlation test* measuring the overall extent of selection.[25]

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

[25]Throughout, I use a linear-probability model, but the results are not substantially different using logit or probit.

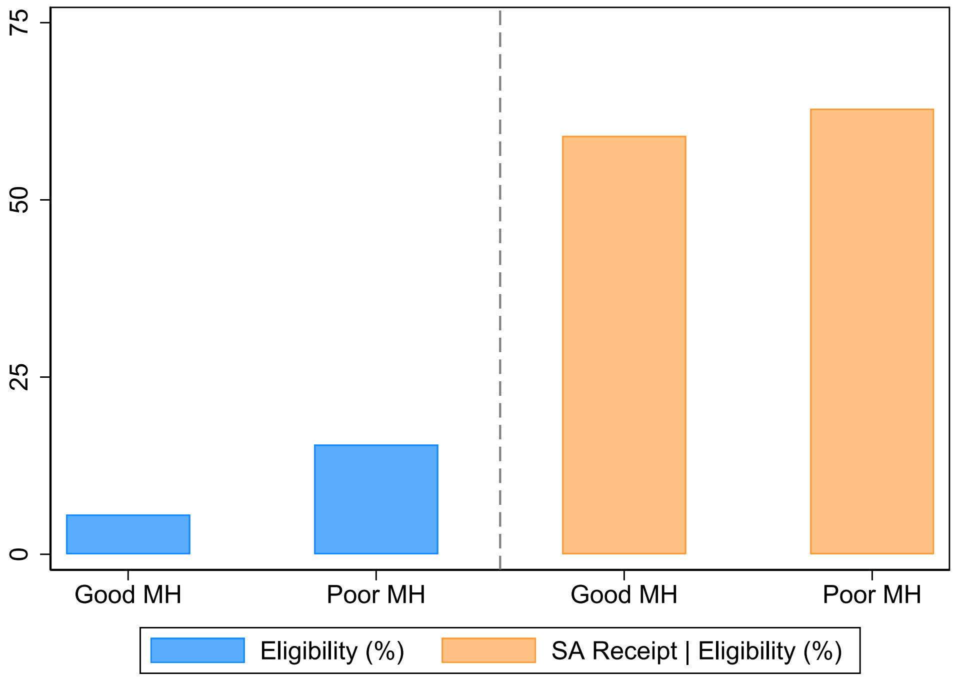



FIGURE 2. $\mathbb{P}[\text{Eligible}]$ and $\mathbb{P}[SA|\text{Eligible}]$, compared for people with poor mental health (dispensed psychopharma in year previously) vs good mental health (not). Underlying population: 2011-2020 in each case.

$X_{it}$ are flexible controls of income,[26] wealth, education, hh composition, work status, work sector and year, age, gender and municipality fixed effects. $\varepsilon_{it}$ is an idiosyncratic error term. $\beta$ measures the selection of social assistance recipients with respect to mental health and is the coefficient of interest.

Importantly, I estimate the correlation test 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.[27]

4.2. **Results.** Table C.2 shows the main results using Poor $\text{MH}_{it} = \mathbb{1}\{i$ dispensed psychopharma. in year $t\}$. Aligned with Figures 1 and 2, I find that people with poor mental health take-up social assistance only slightly more than those with good mental health. The table shows estimates of $\beta$, for seven specifications of increasing saturation.

[26]I include household standardised income percentile (moving average $t-4 \rightarrow t-2$) fixed effects.
[27]Indeed, Muilwijk-Vriend et al. (2019) show that $\hat{\beta}$ is positive for the overall population, but of course, this could be due to people with poor mental health being more likely to be eligible.

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 between groups ranges from around -1 to +1p.p, depending on inclusion of additional controls. Estimates are statistically significant but economically small. In the full specification (Column 5), people with poor mental health have social assistance receipt rates less than 1% higher than those with good mental health, holding all else constant.

Two controls explain a large portion of the variation. Adding lagged income controls increases $R^2$ from 0.05 to 0.16, as people with poor mental health have different income levels, which determine benefits-level entitlement. Adding lagged work status (including past social assistance receipt) further increases $R^2$ from 0.16 to 0.64, with $\hat{\theta}_{SA_{t-2}=1} = 42.35$, showing strong autocorrelation in social assistance take-up. The positive $\hat{\beta}$ with individual fixed effects supports this.

The increase in the point estimate when controlling for lagged work status suggests social assistance may improve mental health, as it indicates that part of the initial negative association between assistance receipt and mental health stems from people who have consistently not used social assistance having poorer mental health. Once past work status is controlled, the analysis shows that among those with similar social assistance histories, individuals with poorer mental health are more likely to use assistance now, suggesting social assistance could help mitigate some mental health challenges.

How does $\hat{\beta}$ compare to the coefficients on other covariates, $\hat{\theta}$? The income percentile fixed effects range from 20 to -20, age fixed effects from 0 to 10 and municipality fixed effects from -15 to 5. This reinforces the idea that $\hat{\beta}$ is economically small.

Figure 3 presents the results of the correlation test 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 with each other and show a small positive difference in rate of receipt by people with poor mental health vs people with good mental health. Table C.3 shows the full results.

Of course, there is a broad spectrum of different mental disorders. Using psychotropic drug dispensations to measure poor mental health provides a practical approach to distinguish these differences. Figure 4 shows coefficients from a regression of SA receipt on dummies for *type* of

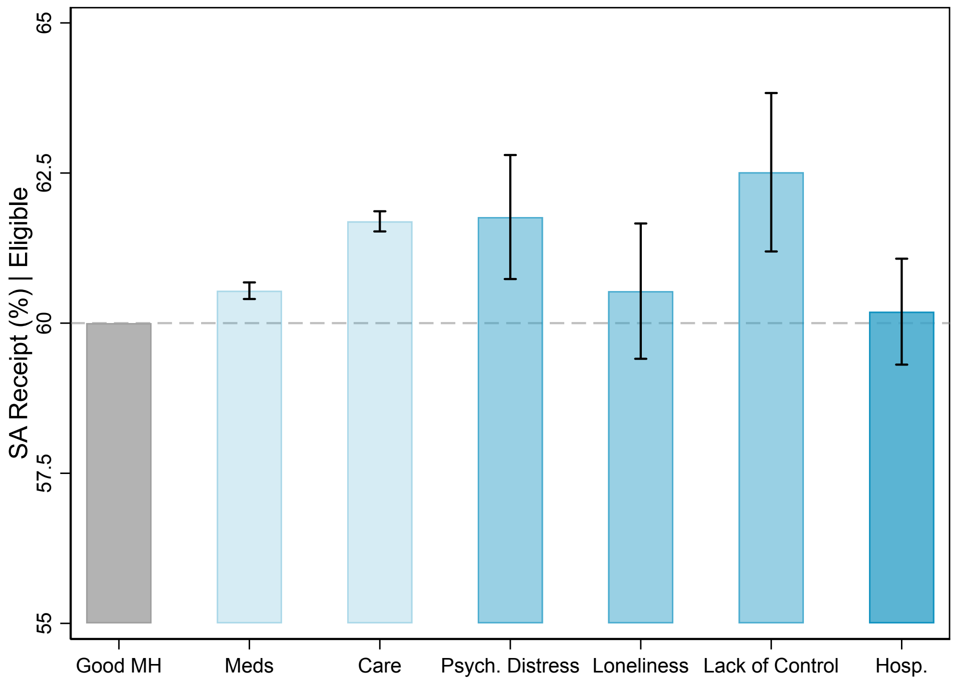



FIGURE 3. Coefficients of social assistance take-up regressed on mental health status–indicators of: psychotropic drugs, mental healthcare, severe surveyed psychological distress/loneliness/lack of control over own life, or mental health hospitalisation. Point estimates added to the control mean, with 95% confidence intervals. Lagged controls include 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.

psychotropic drug dispensed, and all controls. People using ADHD medication, hypnotics / sedatives and anxiolytics are no more likely to receive social assistance than those not using any psychotropic medications. Anti-depressant dispensation is associated with a higher rate-of-receipt, whereas anti-psychotic dispensation is associated with a *lower* rate-of-receipt. Table C.1 shows the full results.

## 5. BARRIER SCREENING EFFECTS

Moving past average take-up levels, recall from Section 2 that we need to identify take-up responses to changes in barriers to assistance to reveal how costly people with poor mental health find overcoming these barriers and thus start to examine targeting effectiveness.

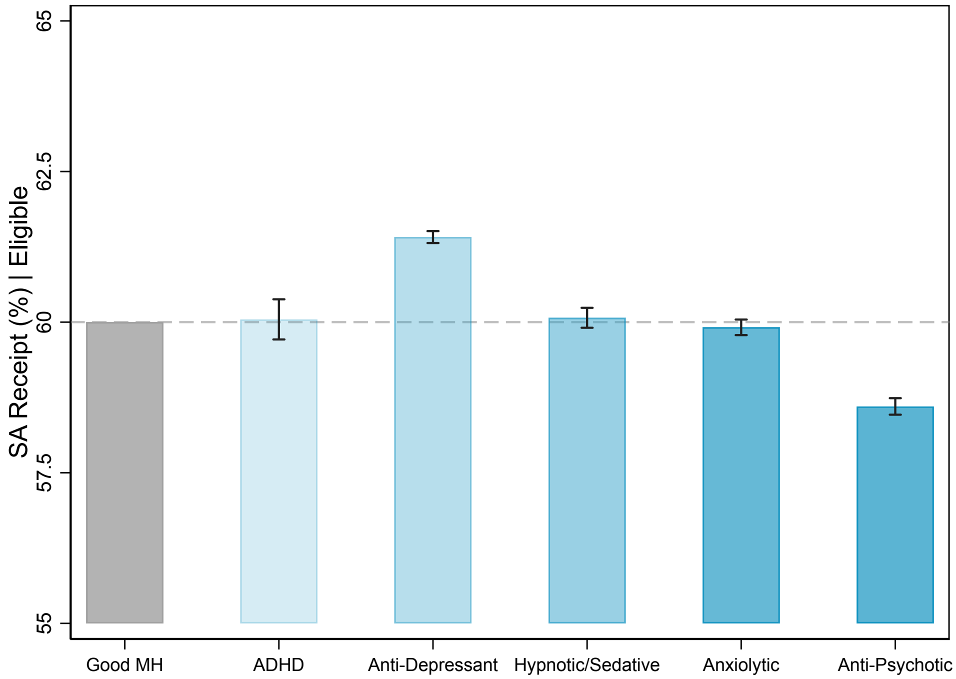



FIGURE 4. Coefficients of social assistance take-up regressed on psychopharmacology dispensation fixed effects (by type: ADHD medications, anti-depressants, hypnotics/sedatives, anti-anxiety medications and anti-psychotics). Point estimates added to the control mean, with 95% confidence intervals. Lagged controls include 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.

I examine the effects of the Participation Act, a large reform to social assistance design in the Netherlands, which increase 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). The Participation Act intensified obligations for recipients and incentivized municipalities to restrict inflow through (threat of) sanctions (SCP, 2019; van der Veen, 2019).[28]

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

[28]For more details, see Appendix D.1.

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.[29]

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

$\eta_i$ and $\gamma_t$ are individual and year fixed-effects respectively. $X$ is a vector of time-varying controls including income, education and muncipality, hh composition and sector fixed effects. $\delta_t$ for $t \geq 2013$ are the coefficients of interest and represent the heterogeneous treatment effect of the policy by baseline mental health. Poor 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.

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 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.[30] 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.[31] Sensitivity analyses confirm that the findings remain robust across various specifications and assumptions, as detailed in Section 5.5.

5.2. **Main Results.** Figure 5 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. [32]

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

[29]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).
[30]Figure D.6 shows the results when including adults aged 35-45. Although this group is more contaminated by Wajong entrants, the results are similar, suggesting the main estimates are not driven by Wajong entrants.
[31]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.
[32]Note: as the policy happens in the aftermath of the GFC, I expect $M \ll 1$ in the framework of Rambachan and Roth (2023). In this case, statistically insignificant pre-trends are meaningful.

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. This is comparable in magnitude (but opposite sign) to Finkelstein and Notowidigdo (2019) who estimate that barriers to SNAP in the US screen-out low-earnings potential types 2 percentage points *less* than high-types.

What do we learn from this through the lens of the theory? Section 4 suggests that take-up levels are roughly equalised on average. Therefore people with poor mental health have similar average value of benefits net of average cost of ordeals to people with good mental health. The barrier screening effects estimated show ordeals are more costly for those with poor mental health.[33]

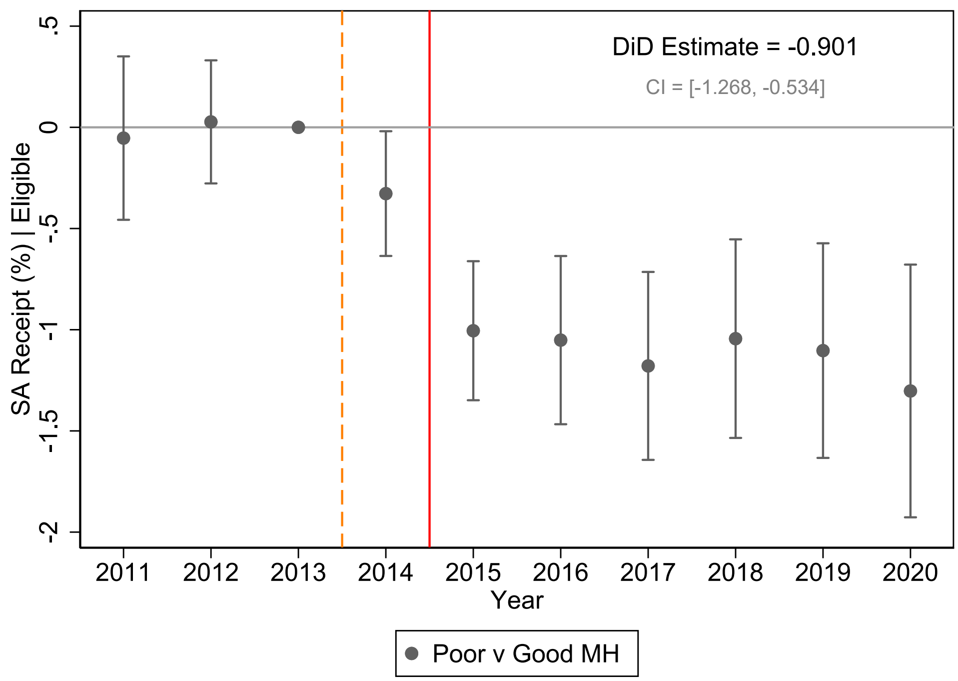



FIGURE 5. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in pre-period. Controls include individual fixed effects, 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.

[33]The idea is that **Step 1** shows the likelihood of being marginal is similar across types. Therefore, large heterogeneous barrier screening effects are informative of differences in cost.

5.3. **Mechanism.** Figure 6 shows the effects on inflow and outflow. Outflow is not mechanically zero, the estimates are just tiny and have tight confidence intervals. This shows that the main results comes exclusively from a deterrence of inflow, aligning with Cook and East (2024) who suggest 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).

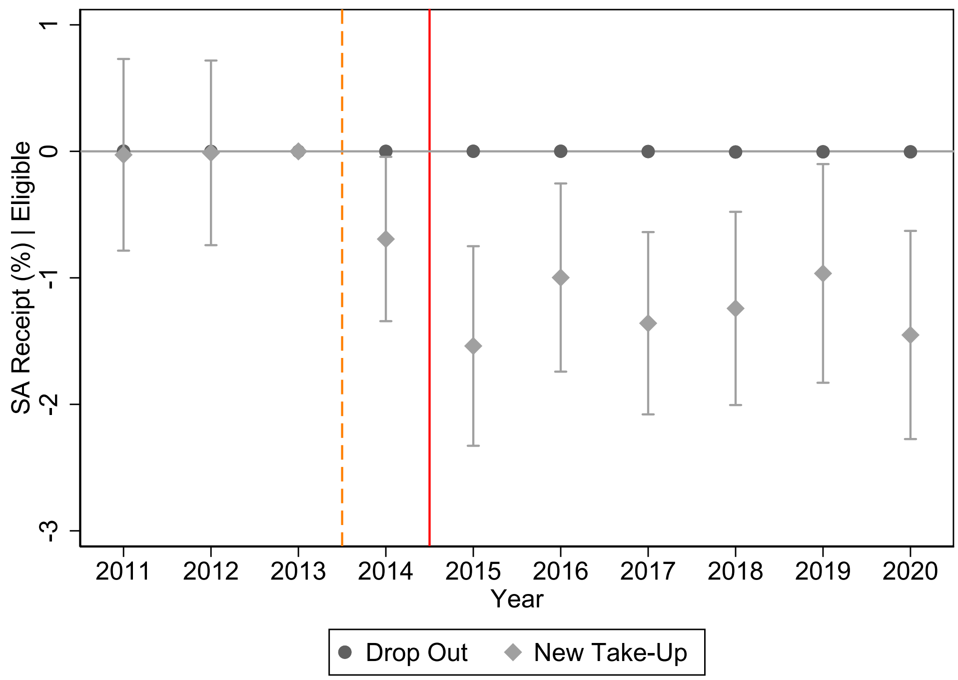



FIGURE 6. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. Here, I split by inflow (receipt conditional on being ineligible last period), and drop-out (non-receipt conditional on receipt last period). 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, 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.

The reduction of inflow suggests people with poor mental health are deterred by increased obligations and also from more unpleasant interactions with the municipality, given incentives to reduce 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".[34] Differential screening at entry, without differential outflow, suggests these anticipatory psychological costs dominate actual 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 7 shows that the Participation Act screens-out those using anti-psychotics twice as much as those using anti-depressants. Figure 8 shows that the barrier screening effects are more pronounced when poor mental health is measured by surveyed severe psychological distress. This likely reflects the fact that the main estimates are a lower-bound since some mental disorders are not diagnosed. Taking these results together suggests that severity of mental disorders—rather than simply their presence—exacerbates the disutility from ordeals.

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 the main result might be driven by differential take-up rates among new entrants and exiters to eligibility. However, Figure D.3 shows that the results are consistent for the always-eligible population—couples who remain eligible throughout the sampling period, suggesting that the main results are not driven by differential selection of the newly eligible.[35] The point estimates being smaller for the always-eligible is 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 issues was changed to be under the remit of municipalities starting in 2015. Figure D.7 shows the WMO reform does not drive the results.

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

[35] While this may seem like a stark restriction, 25% of the eligible are always-eligible.

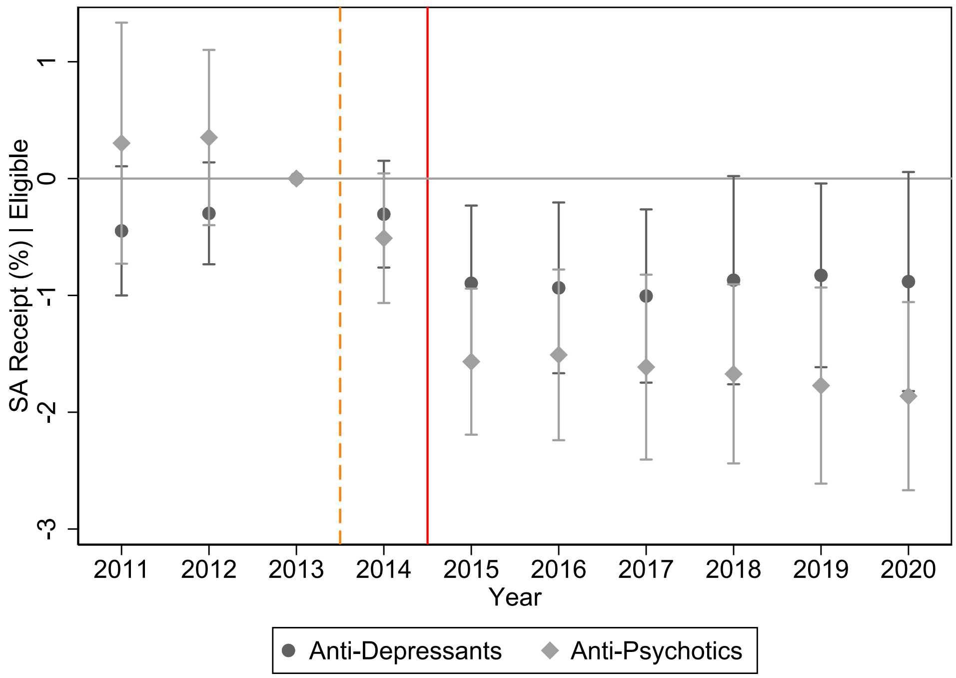



FIGURE 7. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. Here, Poor $\text{MH}_i$ can now take 3 values: 0 (control), 1 (anti-depressants) or 2 (anti-psychotics). The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in pre-period. Controls include individual fixed effects, income, education and muncipality, hh composition and sector fixed effects. I plot the estimate $\hat{\delta}_t^{\text{Dep.}}$ and $\hat{\delta}_t^{\text{Psycho.}}$. Standard-errors are clustered at the level of municipality of residence in 2013. Coefficient estimates are contained in Table D.2.

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.6 shows that the estimates are unchanged when including adults aged 35-45, confirming that the age restriction effectively controls for the potential contamination.

Thirdly, the Participation Act could have affected people with poor and good mental health differently due to its differential implementation, particularly the change in how eligibility was

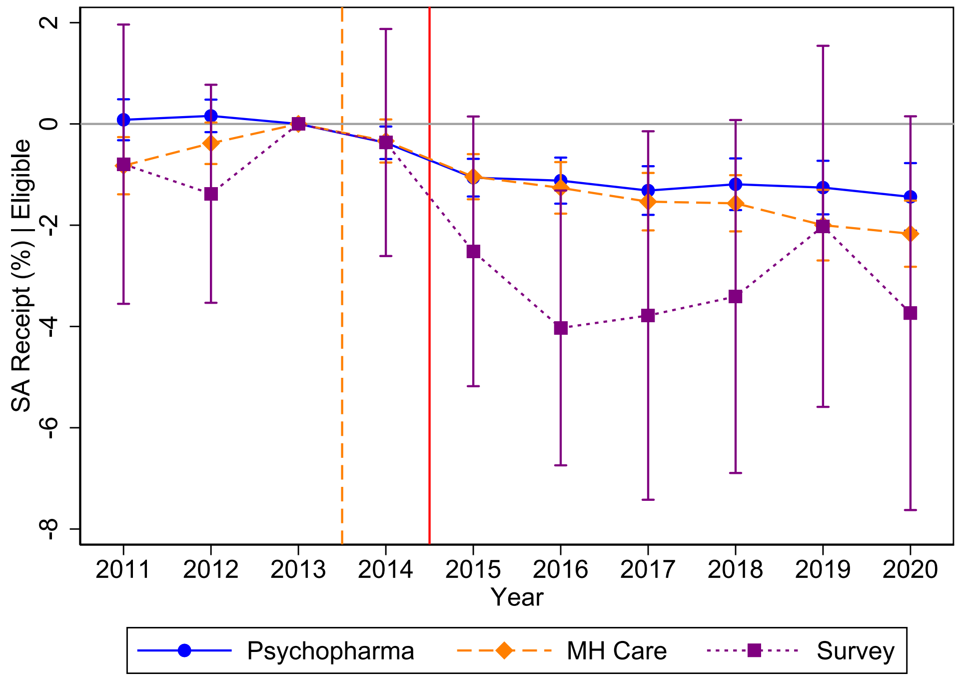



FIGURE 8. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. Here, Poor $\text{MH}_i$ is defined in 3 ways: dispensations of psychotropic drugs in pre-period, $> 0$ mental healthcare costs in pre-period, surveyed severe psychological distress in 2012. The analysis population is eligible middle-age couples. Controls include individual fixed effects, 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.

calculated based on household composition. To control for this, I include flexible controls for household size, and Figure D.4 shows that this does not drive the results.

Fourth, a concern is that the observed heterogeneous treatment effects could be due to pre-existing differences in take-up rates, rather than baseline mental health. However, when splitting the poor mental health group into subgroups—moderate (anti-depressants and anti-anxiety) and severe (anti-psychotics)—we find that both groups' take-up rates diverge from the good mental health group after 2015. Prior to 2015, the severely poor mental health group had lower receipt levels to the good mental health group, supporting the hypothesis that mental health differentially affects responses to barriers, and not simply pre-policy take-up levels.

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 8 shows that results when defining poor mental health based on self-reported symptoms, not prescriptions, are even stronger. Additionally, Figure D.8 shows that people dispensed drugs exhibit significantly worse mental health both before and after dispensation, with scores consistently above the threshold for moderate mental illness.

A related concern might be that the results could reflect long-term positive effects on social assistance take-up following a mental health shock. However, Figure D.9 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.10 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.[36]

## 6. Benefit Take-up Effects

In the final empirical part of the paper, I estimate the take-up response to exogenous variation in benefits. 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 9.

Before diving into the details of identification, Figure 10 shows graphical evidence of the behavioural response to a change in benefit level 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* income.[37] The take-up functions diverge starkly at the threshold for poor vs good mental health. This is confirmed by fitting polynomials on either side of the threshold and testing for differences in their slopes. The results show that the slope change is almost twice as large for individuals with poor mental health compared to those with good mental health.

[36] Overall, these findings suggest that being prescribed psychopharma at any point is a consistent indicator of mental health status, which remains significantly worse than the good mental health group throughout the study period.

[37] Granular analysis is critical - hence the switch to monthly data. We can reconcile the findings in Figure 10 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.

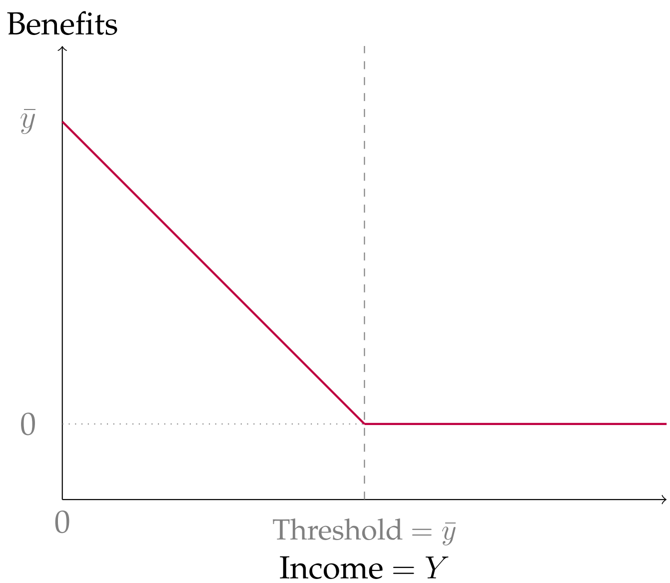



FIGURE 9. Benefits schedule as a function of income

People with income above the threshold take-up primarily due to measurement error: some sources of income do not count towards the eligibility threshold. Therefore, I need to calculate the income concept used to determine eligibility. This $Y_{\text{calc}}$ differs from $Y_{\text{true}}$ 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. $Y_{\text{true}}$ is observed for the selected sample: recipients.

To minimise attenuation from measurement error, I focus on single employees.[38] The data contain monthly income information for employees–minimising error due to (a) and singles are immune from issues in (b). Figure E.1 shows a histogram of $Y_{\text{true}} - Y_{\text{calc}}$ for the analysis population. $Y_{\text{true}}$ is negatively selected for recipients, so we expect the distribution to be left-skewed. Measurement error has significant mass around 0 and both mean and median are small (-€51, -€13 respectively).

6.1. **Identification.**

[38] 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.11 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.

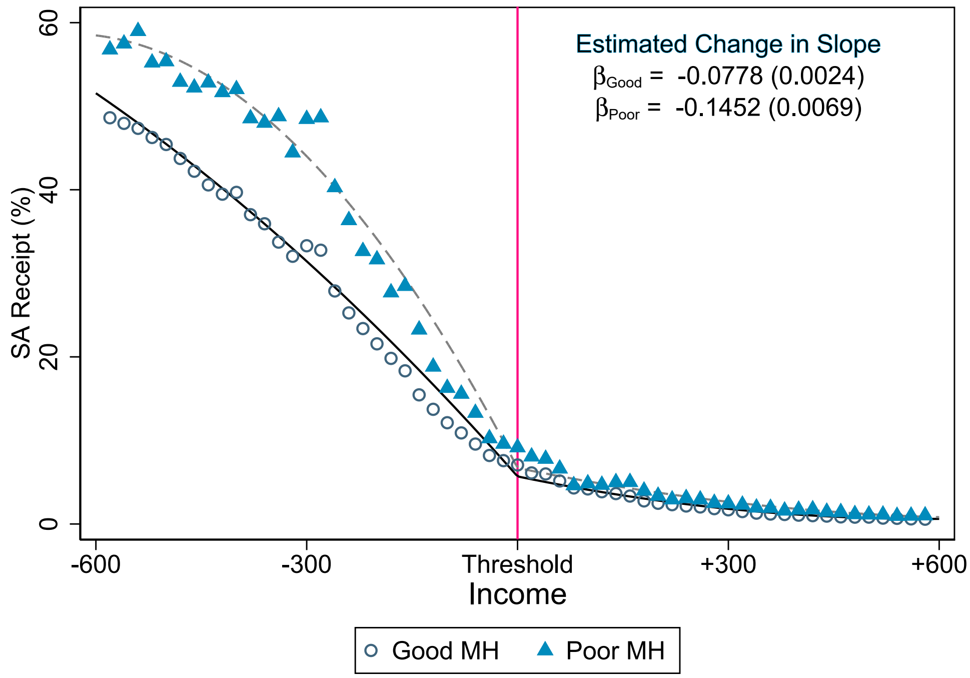



FIGURE 10. Average rate of receipt within income slice in a large window of income either side of the eligibility threshold. Income in this plot is monthly. Poor mental health is defined as receiving psychopharma in the year previously. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions.

6.1.1. *Theory.* Figure 10 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 $B$. Appendix E.1 sets out my theoretical approach to impute the ex-ante benefits schedule (i.e. the expected benefits a potential applicant is eligible for conditional on their income) accounting for exemptions. Figure E.3 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) version of $B$: $B^* \triangleq B + U_B$. As discussed above, $Y$ is also measured with error: $Y^* \triangleq Y + U_Y$. 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.[39]

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

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

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).

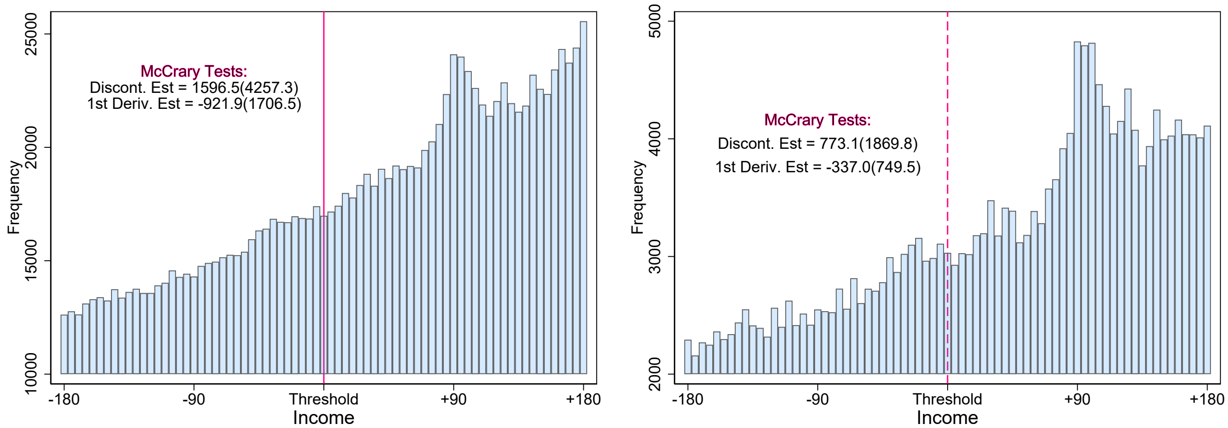
**Support for Identification Assumptions:** The key identification assumption is that there is no manipulation of income around the threshold. Figure 11 and Figure E.13 shows no evidence for strategic income targeting around the eligibility threshold: McCrary (2008) tests with seventh-order polynomials show no statistically significant bunching. Although the threshold equals the full-time monthly minimum wage, the sample works much less than full-time (around 100 hours per month) and income used for eligibility does not only come from labour. Adjustment frictions are likely a key reason for lack-of-bunching (Kleven, 2016).

6.2. **Main Results.** First, I pool people with good and poor mental health together. Figure E.7 shows that employees 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 12. I estimate $(\hat{\beta}_{1H}, \hat{\beta}_{1L}) = (-0.0218, -0.0508)$.[40] 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.8 to re-scale the above reduced-form and account for measurement error, we obtain IV estimates (and associated confidence intervals):

[39] See Appendix E.2 for more details.

[40] 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. Here, the zoomed-out version in Figure 10 gives us confidence that the take-up response is indeed twice as large for those with poor mental health. Moreover, Figures E.17 and E.18 confirm that for less extreme restrictions to the bandwidth, the estimates are very similar, but more precise.




(A) Good Mental Health

(B) Poor Mental Health

FIGURE 11. Density of income around the eligibility threshold. McCrary (2008) tests for discontinuity in levels and slopes around the threshold are shown. Income is monthly. Poor mental health is defined as receiving psychopharma in the year previously. The sample contains singles employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions.

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

Translating these to take-up elasticities with respect to changes in benefits yields 0.16, 0.38 for good and poor mental health, respectively. The elasticity for good mental health is on the lower-end of the range of previously estimated take-up elasticities of social insurance (Krueger and Meyer, 2002; McGarry, 1996), whereas for poor mental health lies within-range. The full set of reduced-form and IV estimates (with and without controls) are contained in Table E.1.

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.14 and E.15 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.16 displays a permutation test (Ganong and Jäger, 2018), and shows no evidence for worrying non-linearities. Figure E.17 and Figure E.18 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.

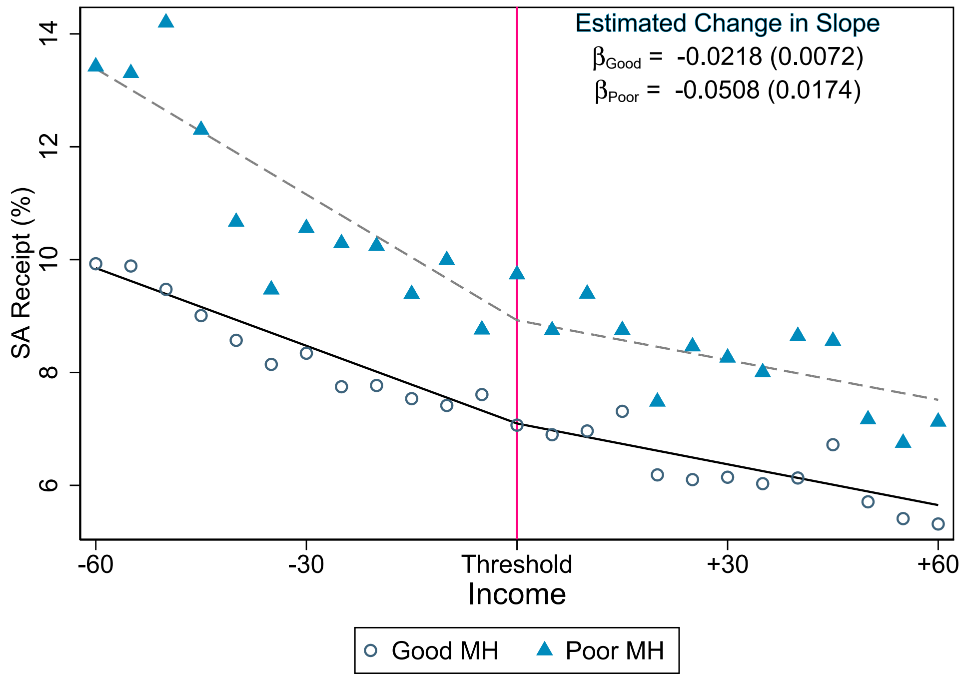



FIGURE 12. Average rate of receipt within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly. Poor mental health is defined as receiving psychopharma in the year previously. The sample contains singles employees, years 2011-2014. See Section 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.

6.2.2. *Different Mental Health Measures.* How do the results compare when mental health is measured differently or across different disorders? Figure E.9 and Figure E.10 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 0.0704 for anti-depressants and 0.120 for anti-psychotics. This suggests that the magnitude of the take-up response to a change in benefits is increasing in severity of mental disorder.

Moreover, Figure E.11 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.

6.3. **Mechanisms.** Why do people with poor mental health react more to a change in benefits, conditional on having the same income? Similarly to Section 5, this larger sensitivity despite similar average take-up levels reflects higher need, i.e. a larger marginal value of income from social assistance. There are two main reasons why this might be the case.

First, cash transfers improve mental health (Haushofer et al., 2020). If people with poor mental health anticipate the protective effect of social assistance income on their mental health, it could cause them to value €1 more than people with good mental health and thus have a higher behavioural response.

However, I find no strong support for this mechanism in my setting. The reduced-form RKD induces exogenous variation in social assistance receipt, which I then regress future psychotropic drug dispensations on to estimate $\frac{\partial MH}{\partial SA}$. Figure E.12 shows a (noisy) 0. I cannot rule out social assistance improving mental health,[41] but it does not seem to be the main driver. This is perhaps not surprising - Miller and Bairoliya (2023); Silver and Zhang (2022) also do not find strong evidence that cash improves mental health. Indeed, (Solmi et al., 2022) argue many mental illnesses start early in life.

Instead, I interpret these results through the psychology literature studying mental disorders. This literature often refers to the impairment of everyday functioning as a key mechanism in the difficulties this population face. Of course, the cognitive burden of mental illness, including effects on information processing, attention, memory and executive function can clearly hinder psycho-social functioning (Kessler et al., 2003; Evans et al., 2014). Mental disorders can also affect everyday functioning through impaired emotion regulation - this can affect work, relationships and self-image (Gross and Muñoz, 1995).

The cognitive burden and emotion resilience tax imposed by mental disorders seemingly increases the difficulty handling common everyday stressors amongst the low-income population, thus increasing the value of support.[42] This idea aligns closely with Amartya Sen's "capabilities approach" (Sen, 1999, 2008); people with poor mental health need more income to get by. Section 6.2.2 shows that the results are consistent when poor mental health is measured through surveyed psychological distress. Moreover, the estimate associated with anti-psychotics—typically

[41]The descriptive results of Section 4 where the point estimate on the correlation test increases when controlling for lagged social assistance receipt.

[42]Indeed, people with poor mental health work less than those without, and limits on earnings capacity are indicative of higher marginal utility of benefits (Deshpande and Lockwood, 2022). An economic model of scarcity would resonate closely with this interpretation (Mullainathan and Shafir, 2013). Given limited mental resources, people with poor mental health will have a higher value of releasing resources through additional money compared to people with good mental health with the same initial income.

prescribed for more severe mental health conditions—is larger than that for anti-depressants. These findings are consistent with the idea that the increased need for assistance stems from the common component of psychological distress inherent in poor mental health, characterized by impaired cognition and emotion regulation.

Perhaps most interestingly, this higher need is estimated through revealed-preference. Not only do people with poor mental health need benefits more, they *think* that they need benefits more. This suggests that impaired functioning seems to dominate anhedonia and other psychological mechanisms lowering the perceived value of help. If anything, this is likely an under-estimate of true need given pessimism characterises depression, one of the most common mental disorders. I return to this point in the welfare calibration.

## 7. Calibration of Welfare Effects

In this section, I use the empirical results of Section 4, Section 5 and Section 6 to quantify need for benefits and cost of overcoming barriers, heterogeneously by mental health. These key primitives are important determinants to the effectiveness social assistance targeting using barriers. For example, I calculate the welfare effects derived in Proposition 2.1. To be clear, this is not the only way of measuring effectiveness. However *any* measure will need to trade-off the differential need for benefits by people with poor mental health against differential cost of overcoming barriers.

The sufficient statistics for these welfare effects are need ($v'_\theta$), cost ($\kappa'_\theta$), benefit take-up effects $\left(\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}\right)$ and barrier screening effects $\left(\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}\right)$.

7.1. **Quantifying Sufficient Statistics.** For the calibration, I assume $\theta \in \{L, H\}$: mental health is either poor or good. Throughout the empirical sections, I examine take-up conditional on eligibility. However, welfare-estimates should reflect the general population - for example because the government budget constraint should reflect the fact that the ineligible fund benefits for the recipients, and not the eligible non-takers. Appendix F.1 shows how to rescale take-up levels and responses estimated on the eligible to reflect the overall population. The tax rate $\tau \approx 37\%$

7.1.1. *Identifying Need and Cost.* I employ the three-step identification method set out in Section 2.3.1. Appendix F.2 shows the full set of results of this calibration. First, Section 4 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 between need and 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 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 $v'_L = 2.22$.[43]

**Cost:** Finally, I use the difference-in-differences results of Section 5 to calibrate $\kappa'_\theta(\Lambda)$. I use the raw descriptive drop in inflow for people with good mental health (see Figure D.2) to calibrate $\frac{\partial \mathbb{P}[SA]_H}{\partial \Lambda} = -0.014$. The main results of Section 5 thus imply $\frac{\partial \mathbb{P}[SA]_L}{\partial \Lambda} = -0.023$. I then apply **Step 3** of the identification method. These results are combined with $f_\varepsilon(v_L - \kappa_L) = f_\varepsilon(v_H - \kappa_H) = 0.63 \times 0.000227$, and imply that that $\kappa'_H = 98$ and $\kappa'_L = 161$.

7.1.2. *General Policy Implications.*

**Under revealed preference:** These estimates suggest that people with poor mental health have more than twice as high a marginal value of additional income (need) versus those with good mental health. The differences are not explained by differences in income, as the regression kink design estimates $\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}$ conditional on income being equal to the eligibility threshold for both groups. This implies a strong redistributive motive towards people with poor mental health. Moreover, need is revealed from behaviour, suggesting that policy instruments to transfer income to those with poor mental health can be effective in practice.

[43]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 10.

The fact that people with poor mental health need money more, but take-up at the same rate as those with good mental health suggests they are inefficiently excluded from social assistance by barriers.

The quantities also imply ordeals impose a 64% higher cost on those suffering from mental disorders relative to people with good mental health. This suggests that although governments have an incentive to redistribute money towards people with poor mental health, barriers to access are a costly way target. These statements are formalised with a specific example policy experiment in Section 7.2.

**The Role of Bias:** What if revealed preference fails? In this case, take-up responses reflect *perceived* need and cost, and the government may have an incentive to paternalistically On the cost side, the baseline model suggests substantial out-screening by barriers reveals large costs for the infra-marginals, but in reality $\kappa'_\theta$ calibrated above can be thought of "as-if" costs of barriers (Goldin and Reck, 2022) which could differ from the truth. The costs are revealed by a policy which intensified obligations and increased unpleasant and stigmatising interactions with the municipality. The policy disproportionately deters people with poor mental health from flowing-in to the benefit. Outflow does not change differentially.

These findings imply that perceived costs may overestimate actual costs amongst those with poor mental health. While the psychological impact of fear and shame, as discussed by Ministerie van SZW (2022), is likely non-negligible, the absence of differential outflow suggests that compliance costs from obligations may be over-stated. I explore how much this matters for welfare effects in Section 7.2.2 for the specific policy experiment studied, following Naik and Reck (2025).

On the need-side, pessimism is a common symptom of depression (Alloy and Ahrens, 1987), a common mental disorder. This suggests that the disproportionate perceived need for people with poor mental health is likely an *under-estimate*. Bias likely only increases the welfare effects of redistributing income to people with poor mental health.

7.1.3. *General Psychological Implications.* Taking the results together, my findings suggest that people with poor mental health both reveal a higher marginal value of additional income and a higher marginal cost of administrative and psychological barriers. The observed patterns in Section 5 and Section 6 indicate that these differences reflect general psychological distress rather than being driven by specific characteristics of particular conditions or the effects of mental health treatment.

Heterogeneous take-up responses, as proxied by dispensations of psychotropic drugs, are amplified when poor mental health is measured through surveyed psychological distress and increase in magnitude with disorder severity (e.g., anti-psychotics are associated with higher take-up responses than anti-depressants). This pattern suggests a common underlying component of poor mental health across disorders.

The psychology literature identifies impairments in cognitive function and emotion regulation as central mechanisms underlying this common component of poor mental health (Bierman et al., 2008; Hammar and Årdal, 2009; Hyman et al., 2006; Gross and Muñoz, 1995). My results align with a model in which deficits in these areas not only exacerbate the challenges of navigating access barriers to social assistance but also impede everyday functioning in other domains, such as work and social life (Kessler et al., 2003; Evans et al., 2014). These broader functional impairments plausibly explain why the revealed marginal value of additional income support is particularly high for this group.

7.2. **Quantifying Welfare Effects.** In this section, I calibrate the welfare effects of marginal changes in benefits and barriers as a function of the sufficient statistics. In the data, the prevalence of poor mental health conditional on eligibility is $\mu(L) = 0.25$.[44] I set $\mathbb{P}[SA]_L = \mathbb{P}[SA]_H = 0.6$. I start from the baseline case of no social welfare preference for poor mental health. This means that the heterogeneous monthly net fiscal externalities $FE_\theta = \tau(y^{SA=0} - y^{SA=1}) + (1-\tau)B$ are, on average:

$$FE_L = 0.37 \times (€512.22 - €331.27) + (1 - 0.63) \times €972.22 = €679.45 \tag{7.1}$$

$$FE_H = 0.37 \times (€574.29 - €390.95) + (1 - 0.63) \times €916.29 = €645.09 \tag{7.2}$$

The fiscal externality of inducing someone with poor mental health to apply is larger than for good mental health. People with poor mental health receive more benefits than those with good mental health because they earn less when on social assistance. Here, the fact that $y_L^{SA=0} \approx y_H^{SA=0}$ comes from restricting to the eligible population. Intuitively, the change in policy induces the eligible to change their take-up rather than the ineligible. Focussing on the general population would likely imply $FE_H \gg FE_L$ as $y_L^{SA=0} \ll y_H^{SA=0}$.

[44] This is a sizeable fraction. The prevalence among eligibles is more than double the general population, as discussed in Section 4. This highlights how the economic vulnerability of people with poor mental health contributes to the welfare effects, as their overrepresentation among the eligible population amplifies the importance of addressing their needs.

7.2.1. *MVPFs of Ordeals and Benefits under Revealed Preference.* For the calibration, I recast welfare effects in terms of the Marginal Value of Public Funds (MVPF) (Hendren and Sprung-Keyser, 2020). This is defined as the willingness-to-pay for a policy change divided by the cost to the government's budget constraint. I estimate $MVPF$s for barrier and benefit changes.

In the baseline case, I follow Hendren and Sprung-Keyser (2020) and write the direct effects of policy changes in terms of each type's *own* willingness-to-pay. This money-metric social welfare function has the advantage of having interpretable units (€'s) for inter-personal utility comparisons. However, it does not capture any heterogeneity in marginal value of income across types - a factor which is crucial in this context. Proposition A.1 derives the formula for the MVPF of a change in barriers ($d\Lambda$) as follows.

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

The numerator reflects the fact that $d\Lambda$ imposes dis-utility on infra-marginals and the denominator reflects the government saving money due to lower take-up. The direct costs of ordeals are rescaled by need so that they are measured in each type's WTP (in €'s). I estimate these quantities on the eligible population but extrapolate to the general population, as shown in Proposition F.1. For the numerator, this requires integrating against the conditional density of mental health for the eligible. For the denominator, we rescale the eligible take-up response by a constant for each $\theta$, representing "effective eligibility." The intuition is that we must adjust for baseline incomplete take-up and the fact that some ineligible individuals may be on the margin of take-up, as they could be just indifferent between earning slightly above the threshold or below it to qualify for social assistance.

$$MVPF_{d\Lambda} = \frac{-0.6 \times \frac{161}{2.2} \times 0.25 - 0.6 \times \frac{98}{1} \times 0.75}{679.45 \times 0.25 \times (-0.023) \times \underbrace{\frac{1}{1 - 0.6 \times 0.907}}_{\text{Effective Eligibility}_L} + 645.09 \times 0.75 \times (-0.014) \times \underbrace{\frac{1}{1 - 0.6 \times 0.954}}_{\text{Effective Eligibility}_H}}$$

$$= 2.34$$

An $MVPF_{d\Lambda}$ of 2.34 means that ordeals impose a direct cost of €2.34 on infra-marginals for every €1 saved by the government through lower take-up. $MVPF_{d\Lambda} \gg 1$ suggests that $d\Lambda$ is a costly way to raise government revenue. Notice the *money-metric* barrier costs of people with poor mental health are €72.6, whereas €98 for good mental health. However, €1 is more than twice as valuable to the person struggling with a mental disorder - which means that the monetary cost does not reflect the much greater dis-utility imposed by ordeals on individuals with mental illness.

Proposition A.1 derives the formula for the MVPF of a change in benefits as follows, again extrapolating from the eligible population. The numerator reflects the value of the transfer $dB$ to infra-marginals and the denominator captures the mechanical (government must pay for the transfer) and behavioural (government must also pay for increased take-up) revenue effects.

An $MVPF_{dB} < 1$ is to be expected since social assistance is a re-distributive program. It means means that beneficiaries gain 91 cents for every €1 spent raising the benefit level. The estimated value lies in the range surveyed by Hendren and Sprung-Keyser (2020).

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

$$\begin{aligned} MVPF_{dB} &= \frac{0.6 \times 0.25 + 0.6 \times 1 \times 0.75}{0.63 \times (0.6 \times 0.25 + 0.6 \times 0.73) + 679.45 \times 0.25 \times \frac{0.63 \times 0.000503}{1-0.6 \times 0.907} + 645.09 \times 0.75 \times \frac{0.63 \times 0.000227}{1-0.6 \times 0.954}} \\ &\approx 0.91 \end{aligned}$$

Comparing $MVPF_{d\Lambda}$ to $MVPF_{dB}$ suggests that reducing ordeals is a $2.4\times$ more effective policy than increasing benefits.[45] The *comparison* characterises the welfare effects derived in Proposition 2.1 and implies that the government is actually willing to reduce benefit levels to finance a reduction in take-up barriers.[46]

[45]Note that the social marginal utility of the beneficiaries of the two policies should be taken into account when comparing the $MVPF$'s (Hendren, 2016). In Appendix F, I show that the social marginal utility of beneficiaries of $dB$ is 1.30 and of $d\Lambda = 1.24$

[46]These $MVPF$s do not directly capture redistribution as social welfare is money-metric. If I calculate $MVPF$'s where the numerator is written using utilitarian social welfare functions, I set social welfare-weights to correspond to the marginal value of income $\lambda_\theta = v'_\theta(B)$. This effectively writes the numerators in constant units of people with good mental health's WTP for €1. Then, $MVPF_{d\Lambda}^{\text{Utilitarian}} = 2.91$ and $MVPF_{dB}^{\text{Utilitarian}} = 1.19$. The latter being above 1 highlights the redistributive motive. This exercise assumes comparability of utility across individuals, and cannot allow for type-specific scalars multiplying utility.

I find that people with poor mental health have twice the need than those with good mental health, but only a 64% higher cost. Why, then, does reducing barriers dominate increasing benefits? The reason is that poor targeting through barriers actually *reduces* the effectiveness of increasing *benefits*. Social assistance is poorly targeted on average, since take-up is similar across mental health states. This makes $dB$ *costly* as it redistributes to all infra-marginal individuals. If people with good mental health had a much lower take-up rate, $MVPF_{dB}$ would be higher.

7.2.2. *The Role of Bias.* The finding that reducing barriers is more effective than increasing benefits is robust to whether ordeal costs reflect true welfare burdens or behavioral frictions. The key intuition is that poor targeting fundamentally limits the effectiveness of benefit increases, *regardless* of the normative relevance of ordeal costs (Naik and Reck, 2025).

Following my empirical results, consider the case where individuals overstate barrier costs: a share $\psi$ of revealed costs represent true welfare burdens, while $1 - \psi$ reflect pure hassle costs that affect behavior but not welfare. Appendix A.3 shows that the government must believe that less than 35% of perceived ordeal costs represent true welfare costs in order to reverse the ranking $MVPF_{d\Lambda} > MVPF_{dB}$. Put differently, the policy conclusion is robust to substantial overstatement of barrier costs: as long as more than one-third of revealed costs reflect true welfare, reducing barriers remains more effective than increasing benefits.

7.2.3. *Relaxing other Assumptions.* I conclude with a discussion of two key identification assumptions, how to relax them through the use of additional structure and the effect this has on welfare consequences. The two key modelling assumptions are (i) Take-up depends on an additive independent choice shock, (ii) $\mathbb{P}$[Marginal to Barrier Change]$_\theta$ = $\mathbb{P}$[Marginal to Benefits Change]$_\theta$. For full details, see Appendix F.3.

(i) Relaxing independence involves adopting models from Rafkin et al. (2023) or Finkelstein and Notowidigdo (2019) where $v'_\theta(B)$ is independent of $\theta$ conditional on income. These models do not fit my context appropriately because they would imply that people with poor mental health have an easier time overcoming barriers, and are substantially *less* pessimistic about the benefit level. Both of these results contradict psychological evidence (Martin et al., 2023b; Evans et al., 2014; Alloy and Ahrens, 1987).

(ii) For internal validity, I focus on subsamples in Section 5 and Section 6 which are different and call into question the extent to which marginal take-up responses can be compared. However, **Step 2** of the 3-step identification Section 2.3.1 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 $64\%$

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 that relaxing (ii) through additional structure on the take-up equation suggests that $\mathbb{P}[\text{Marginal to Barrier Change}]_\theta < \mathbb{P}[\text{Marginal to Benefits Change}]_\theta$. This suggests that ordeal costs are a *lower-bound* and pushes further in favour of reducing barriers over increasing benefits.

# 8. Conclusion

This paper shows that people with poor mental health are high-need, yet inefficiently excluded from low-income welfare assistance due to high cost of overcoming barriers. I use a theoretical framework to show how to disentangle need for benefits and cost of barriers using take-up levels and how take-up responds to changes in benefits and barriers. Empirically, I use Dutch administrative data containing detailed information on social assistance take-up and mental health.

Descriptively, while people with poor mental health are three times more likely to be eligible for low-income benefits, conditional on eligibility, they take-up at around the same rate as those with good mental health. A policy which increases barriers disproportionately screens out those with poor mental health, while they also take-up more in response to a change in benefits. This is identified with a regression kink design on the kinked benefits schedule. Combining theory and empirics shows that reducing barriers is twice as effective as increasing benefits.

**Future work:** Throughout, I have assumed a static model where mental health is not directly affected by ordeals. This simplification could mean that my estimates of the welfare effect of a change in ordeals is underestimated because barriers likely worsen mental health directly (Brewer et al., 2022). In this context, mental health is unique, for example in comparison to income or education, because of its potential to respond to aspects of the take-up environment.

Due to these issues, work in progress calibrates a dynamic structural model of evolving mental health type affecting and responding to receipt of social assistance. Through this exercise, I aim to quantify the discrepancies between welfare effects under a static model with those under a dynamic setup. For example, people with poor mental health are more likely to be screened out. If this directly worsens their mental health, there would be evidence of a psychological poverty trap (Haushofer, 2019; Ridley et al., 2020) which could decrease welfare effects.

Finally, the theoretical framework described above is designed for analysing the targeting of social assistance, however can easily be applied to study the welfare consequences of people with poor mental health being screened out of other programs. One program of particular relevance is mental healthcare itself. There is evidence of forgoing mental health treatment by people with serious mental disorders. For example, Cronin et al. (2024) develop a discrete choice model which suggests that people with poor mental health could have increased psychological cost of talk therapy, despite needing it more, which could cause them to forgo. My framework can be applied to evaluate the welfare consequences of this, and determine whether those suffering from mental disorders take-up mental healthcare at the optimum rate. Work is underway along these lines.

# References


Abramson, B., J. Boerma, and A. Tsyvinski (2024): "Macroeconomics of Mental Health," Tech. rep., National Bureau of Economic Research. [*Cited on page 5.*]

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 4.*]

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 pages 38, 42, and 106.*]

Anders, J. and C. Rafkin (2022): "The Welfare Effects of Eligibility Expansions: Theory and Evidence from SNAP." *SSRN*. [*Cited on page 106.*]

Angelucci, M. and D. Bennett (2024a): "Depression, Pharmacotherapy, and the Demand for a Preventive Health Product," *Available at SSRN 4808853*. [*Cited on page 5.*]

——— (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 5.*]

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 5.*]

Baily, M. N. (1978): "Some aspects of optimal unemployment insurance," *Journal of public Economics*, 10, 379–402. [*Cited on page 10.*]

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 5.*]

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 5.*]

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 pages 1 and 5.*]

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 14.*]

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 5.*]

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 5 and 39.*]

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 43.*]

CALONICO, S., M. D. CATTANEO, AND R. TITIUNIK (2014): "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 82, 2295–2326. [*Cited on pages 32, 37, 83, and 103.*]

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 29, 31, 32, 82, 83, 85, and 87.*]

CHETTY, R. (2008): "Moral hazard versus liquidity and optimal unemployment insurance," *Journal of political Economy*, 116, 173–234. [*Cited on page 10.*]

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 5.*]

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 25.*]

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, 5, and 44.*]

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 page 18.*]

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 23.*]

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 4.*]

DESHPANDE, M. AND L. M. LOCKWOOD (2022): "Beyond health: Nonhealth risk and the value of disability insurance," *Econometrica*, 90, 1781–1810. [*Cited on page 35.*]

EINAV, L. AND A. FINKELSTEIN (2011): "Selection in insurance markets: Theory and empirics in pictures," *Journal of Economic perspectives*, 25, 115–38. [*Cited on page 10.*]

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 5, 35, 39, 42, and 106.*]

FEHR, D., G. FINK, AND B. JACK (2022): "Poor and Rational: Decision-Making under Scarcity," *Journal of Political Economy*, forthcoming. [*Cited on page 5.*]

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 2, 4, 6, 12, 13, 24, 42, 53, 105, and 106.*]

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 33 and 100.*]

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 82.*]

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 4.*]

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 1.*]

GOLDIN, J. AND D. RECK (2022): "Optimal defaults with normative ambiguity," *Review of Economics and Statistics*, 104, 17–33. [*Cited on page 38.*]

GROSS, J. J. AND R. F. MUÑOZ (1995): "Emotion regulation and mental health." *Clinical psychology: Science and practice*, 2, 151. [*Cited on pages 5, 35, and 39.*]

HALLER, A. AND S. STAUBLI (2024): "Measuring the Value of Disability Insurance from Take-Up Decisions," . [*Cited on pages 11 and 104.*]

HAMMAR, Å. AND G. ÅRDAL (2009): "Cognitive functioning in major depression-a summary," *Frontiers in human neuroscience*, 3, 26. [*Cited on pages 5 and 39.*]

HAUSHOFER, J. (2019): "Is there a Psychological Poverty Trap?" Tech. rep., Working Paper. [*Cited on page 43.*]

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 35.*]

HEEKELAAR, M. (2021): "Beschikbare en benodigde financiële middelen voor de Participatiewet: Analyse," . [*Cited on page 22.*]

HENDREN, N. (2016): "The policy elasticity," *Tax Policy and the Economy*, 30, 51–89. [*Cited on page 41.*]

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, 40, 41, and 54.*]

HOMONOFF, T. AND J. SOMERVILLE (2021): "Program Recertification Costs: Evidence from SNAP," *American Economic Journal: Economic Policy*, 13, 271–98. [*Cited on page 4.*]

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 39.*]

INSPECTIE SZW (2021): "Niet-gebruik van de algemene bijstand: Een onderzoek naar de omvang, kenmerken, langdurigheid en aanpak," *Den Haag: Inspectie SZW*. [*Cited on page 16.*]

KAUR, S., S. MULLAINATHAN, S. OH, AND F. SCHILBACH (2021): "Do Financial Concerns Make Workers Less Productive?" Tech. rep., National Bureau of Economic Research. [*Cited on page 5.*]

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 5, 35, and 39.*]

KLEVEN, H. J. (2016): "Bunching," *Annual Review of Economics*, 8, 435–464. [*Cited on pages 32 and 56.*]

KLINE, P. AND C. R. WALTERS (2019): "On Heckits, LATE, and numerical equivalence," *Econometrica*, 87, 677–696. [*Cited on page 12.*]

KO, W. AND R. A. MOFFITT (2024): "Take-up of social benefits," *Handbook of Labor, Human Resources and Population Economics*, 1–42. [*Cited on pages 1, 2, 3, and 8.*]

KOLSTAD, J., C. NAIK, W. PARKER, AND J. SPINNEWIJN (2024): "Social Determinants of Mental Health," . [*Cited on pages 3, 15, and 18.*]

KROMHOUT, M., N. KORNALIJNSLIJPER, AND M. DE KLERK (2018): "Summary changing care and support for people with disabilities," . [*Cited on page 26.*]

KRUEGER, A. B. AND B. D. MEYER (2002): "Labor supply effects of social insurance," *Handbook of public economics*, 4, 2327–2392. [*Cited on page 33.*]

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 106.*]

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 15.*]

MANI, A., S. MULLAINATHAN, E. SHAFIR, AND J. ZHAO (2013): "Poverty impedes cognitive function," *Science*, 341, 976–980. [*Cited on page 5.*]

MARTIN, L., L. DELANEY, AND O. DOYLE (2023a): "Everyday administrative burdens and inequality," *Public Administration Review*. [*Cited on page 5.*]

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 pages 5, 42, and 106.*]

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 32, 33, and 97.*]

MCFADDEN, D. (1981): "Econometric Models of Probabilistic Choice"," . [*Cited on page 12.*]

MCGARRY, K. (1996): "Factors determining participation of the elderly in supplemental security income," *Journal of Human Resources*, 331–358. [*Cited on page 33.*]

MILLER, R. AND N. BAIROLIYA (2023): "Health, longevity, and welfare inequality of older americans," *Review of Economics and Statistics*, 105, 1145–1160. [*Cited on page 35.*]

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 5.*]

MINISTERIE VAN SZW (2015): "Participatiewet," https://wetten.overheid.nl/BWBR0015703/2015-01-01/. [*Cited on pages 3, 15, 16, and 81.*]

——— (2022): "Participatiewet in Balans: uitkomsten beleidsanalyse," https://www.rijksoverheid.nl/documenten/kamerstukken/2022/06/21/bijlage-rapport-participatiewet-in-balans. [*Cited on pages 3, 25, 26, 38, and 67.*]

MOGSTAD, M., A. TORGOVITSKY, AND C. R. WALTERS (2024): "Policy evaluation with multiple instrumental variables," *Journal of Econometrics*, 105718. [*Cited on page 12.*]

MUILWIJK-VRIEND, S., C. TEMPELMAN, L. KROON, M. LAMMERS, R. PONDS, C. VAN WOERKENS, AND P. KONING (2019): *Gezondheidsproblemen in WW en bijstand*, SEO Economisch Onderzoek. [*Cited on page 19.*]

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 pages 5 and 57.*]

MULLAINATHAN, S. AND E. SHAFIR (2013): *Scarcity: Why having too little means so much*, Macmillan. [*Cited on page 35.*]

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 3, 13, 38, and 42.*]

NICHOLS, A. L. AND R. J. ZECKHAUSER (1982): "Targeting transfers through restrictions on recipients," *The American Economic Review*, 72, 372–377. [*Cited on pages 1 and 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 87.*]

RAFKIN, C., A. SOLOMON, AND E. SOLTAS (2023): "Self-Targeting in US Transfer Programs," *Available at SSRN 4495537*. [*Cited on pages 2, 4, 9, 12, 42, 105, and 106.*]

RAMBACHAN, A. AND J. ROTH (2023): "A more credible approach to parallel trends," *Review of Economic Studies*, 90, 2555–2591. [*Cited on page 23.*]

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 and 43.*]

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 18.*]

ROTH, C., P. SCHWARDMANN, AND E. TRIPODI (2024): "Misperceived effectiveness and the demand for psychotherapy," . [*Cited on page 5.*]

RUSSO, A. (2023): "Waiting or Paying for Healthcare: Evidence from the Veterans Health Administration," Tech. rep., Working Paper. [*Cited on page 11.*]

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 5.*]

SCP (2019): "Eindevaluatie van de Participatiewet," . [*Cited on pages 3, 22, 26, and 66.*]

SEN, A. (1999): *Development as Freedom*, New York: Oxford University Press. [*Cited on pages 6 and 35.*]

——— (2008): "The idea of justice," *Journal of human development*, 9, 331–342. [*Cited on pages 6 and 35.*]

SERENA, B. L. (2024): "The Causal Effect of Scaling up Access to Psychotherapy," . [*Cited on page 5.*]

SHAH, A. K., S. MULLAINATHAN, AND E. SHAFIR (2012): "Some consequences of having too little," *Science*, 338, 682–685. [*Cited on page 5.*]

SHAHN, Z. (2023): "Subgroup Difference in Differences to Identify Effect Modification Without a Control Group," *arXiv preprint arXiv:2306.11030*. [*Cited on page 23.*]

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 pages 4 and 6.*]

SHREEKUMAR, A. AND P.-L. VAUTREY (2021): "Managing Emotions: The Effects of Online Mindfulness Meditation on Mental Health and Economic Behavior," . [*Cited on page 5.*]

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 pages 5 and 35.*]

SOLMI, M., J. RADUA, M. OLIVOLA, E. CROCE, L. SOARDO, G. SALAZAR DE PABLO, J. IL SHIN, J. B. KIRKBRIDE, P. JONES, J. H. KIM, ET AL. (2022): "Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies," *Molecular psychiatry*, 27, 281–295. [*Cited on page 35.*]

VAN DER VEEN, R. (2019): "Basic income experiments in the Netherlands?" *Basic Income Studies*, 14. [*Cited on pages 22 and 66.*]

VERLAAT, T. AND A. ZULKARNAIN (2022): "Evaluatie experimenten Participatiewet: effecten op brede baten," . [*Cited on page 14.*]

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

WOODFORD, M. (2020): "Modeling imprecision in perception, valuation, and choice," *Annual Review of Economics*, 12, 579–601. [*Cited on page 12.*]

WOOLDRIDGE, J. M. (2019): "Correlated random effects models with unbalanced panels," *Journal of Econometrics*, 211, 137–150. [*Cited on page 71.*]

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 4.*]

ZECKHAUSER, R. (2021): "Strategic sorting: the role of ordeals in health care," *Economics & Philosophy*, 37, 64–81. [*Cited on page 2.*]

# Appendix A. Theory Appendix

Let $\theta$ 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:[47] 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 $\tau$. Thus, $v_\theta(B)$ is given by:

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

Thus, value is the net-utility gain from social assistance and comes from two main sources. First, if $y_\theta^{SA=0} \leq y_\theta^{SA=1} + B$, $\theta$ derives utility from the top-up in consumption $(1-\tau)y^{SA=0} \rightarrow (1-\tau)\cdot\left[y_\theta^{SA=1}+B\right]$. Second, if $y^{SA=1} < y^{SA=0}$, $\theta$ 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 $y_\theta$, the utility functions $u_\theta$ 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 (2.5). 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.

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

[47]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 $\approx 0$ 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 $\theta$ subscripts:

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

Rearranging gives Equation (2.6). □

A.1. **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 written in the units of each types' willingness-to-pay. Hendren and Sprung-Keyser (2020) show that the composite policy increasing $\Lambda$ ($B$ adjusts) is social-welfare improving, if the gains from increasing spending on $dB$ exceed the losses from reducing spending through an increase $d\Lambda$.

Let $\eta_\theta$ denote each individual's social marginal utility of income. Therefore, $\eta_\theta = \lambda_\theta \cdot v'_\theta$: social marginal utility is equal to social marginal welfare weight $\times$ individual marginal utility of income. Let $WTP^P_\theta = \frac{d\mathcal{U}_\theta}{dP} \cdot \frac{1}{v'_\theta}$ be $\theta$'s willingness-to-pay for a policy $P$: the direct welfare effect divided by the marginal utility of income.

**Proposition A.1.** *(Hendren and Sprung-Keyser, 2020) Let $\bar{\eta}_P$ be the average social marginal utility of the beneficiaries a policy $P$:*

$$\bar{\eta}_P = \int \eta_\theta \frac{WTP^P_\theta}{\int WTP^P_\theta d\mu} d\mu \tag{A.2}$$

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

$$\bar{\eta}_{dB} \cdot MVPF_{dB} > \bar{\eta}_{d\Lambda} \cdot MVPF_{d\Lambda} \tag{A.3}$$

*where:*

$$\bar{\eta}_{dB} = \int \eta_\theta d\mu \tag{A.4}$$

$$\bar{\eta}_{d\Lambda} = \int \eta_\theta \frac{\kappa'_\theta / v'_\theta}{\int \kappa'_\theta / v'_\theta d\mu} d\mu \tag{A.5}$$

*and the MVPF of an increase in ordeals is given by Equation (A.6).*

$$MVPF_{d\Lambda} = \frac{\overbrace{-\int \lambda_\theta \cdot \mathbb{P}[SA]_\theta \cdot \frac{\kappa'_\theta(\Lambda)}{v'_\theta(B)}\, d\mu}^{\text{Direct Effect } <0}}{\underbrace{\int FE_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda} d\mu}_{\text{behavioural Revenue Effect } <0}} \tag{A.6}$$

*and the MVPF of an increase in benefit level is given by Equation (A.7).*

$$MVPF_{dB} = \frac{\overbrace{\int \lambda_\theta \cdot \mathbb{P}[SA]_\theta \, d\mu}^{\text{Direct Effect } >0}}{\underbrace{(1-\tau) \cdot \int \mathbb{P}[SA]_\theta \, d\mu}_{\text{Mechanical Revenue Effect } >0} + \underbrace{\int FE_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial B} d\mu}_{\text{behavioural Revenue Effect } >0}} \tag{A.7}$$

The direct effect of an increase in ordeals $d\Lambda$ is that it imposes dis-utility on infra-marginal individuals $\kappa'_\theta$. Written in terms of € cost, this is $\frac{\kappa'_\theta}{v'_\theta}$. 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,

$$\frac{\partial W}{\partial \Lambda} = -\int \lambda_\theta \mathbb{P}[SA]_\theta \kappa'_\theta d\mu \tag{A.8}$$

$$\frac{\partial W}{\partial B} = \int \lambda_\theta \mathbb{P}[SA]_\theta v'_\theta d\mu \tag{A.9}$$

$$\frac{\partial G}{\partial \Lambda} = \int FE_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda} d\mu \tag{A.10}$$

$$\frac{\partial G}{\partial B} = (1-\tau) \int \mathbb{P}[SA]_\theta \, d\mu + \int FE \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial B} d\mu \tag{A.11}$$

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 $n_\theta$. 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,

$$SA = 1 \iff u\left((1-\tau)\cdot\left(B + y^{SA=1}\right), y^{SA=1}\right) - \kappa \cdot \Lambda + \kappa_0 - \varepsilon \geq u\left((1-\tau)y^{SA=0}, y^{SA=0}\right) \tag{A.12}$$

Then:

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

$$\mathbb{P}[SA]_L - \mathbb{P}[SA]_H \approx \left(\Psi \frac{\partial \mathbb{P}[SA]_L}{\partial B} + \Lambda \frac{\partial \mathbb{P}[SA]_L}{\partial \Lambda}\right) \cdot \left(\frac{f_L}{f_H} - 1\right) \tag{A.13}$$

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

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

*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 $\Lambda$.

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:

$$v(B) = v \cdot (1-\tau) \cdot \underbrace{\left\{B + y^{SA=1} - \frac{y^{SA=0}}{1+e} - \frac{\left(y^{SA=1}\right)^{1+1/e}}{\left(y^{SA=0}\right)^{1/e} 1+e}\right\}}_{\triangleq \Psi} \tag{A.14}$$

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 (2.8) and (2.9) 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.13).

□

A.3. **Robustness to Bias.** Suppose a share $\psi$ 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:

$$\mathcal{U}_\theta = \int_{-\infty}^{\varepsilon^*_\theta} \left[v_\theta(B) - \kappa_\theta(\Lambda) + MI_\theta - \varepsilon\right] dF(\varepsilon) \tag{A.15}$$

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 $\Lambda$ 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.*

$$\frac{d\mathcal{U}_\theta}{d\Lambda} = -\psi \cdot \kappa'_\theta(\Lambda) \cdot \mathbb{P}[SA]_\theta + MI_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda} \tag{A.16}$$

$$\frac{d\mathcal{U}_\theta}{dB} = v'_\theta(B) \cdot \mathbb{P}[SA]_\theta + MI_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial B} \tag{A.17}$$

*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.16). 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 $MI_\theta$ 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.16) and (A.17).

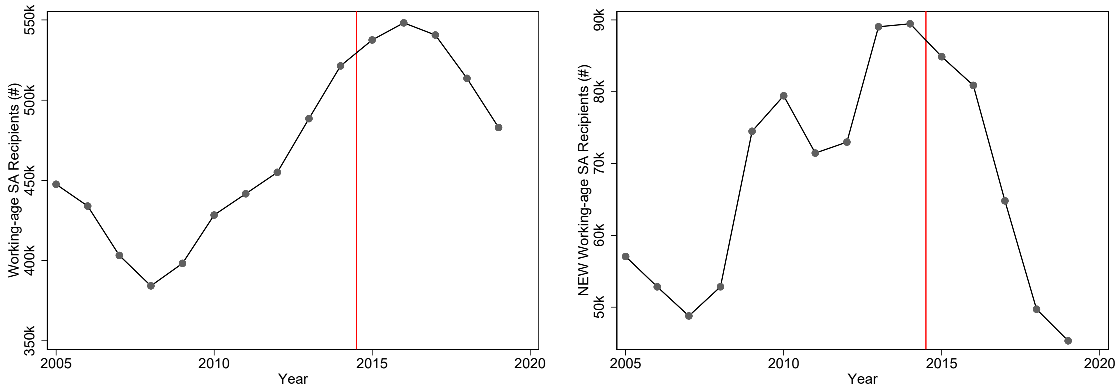
**Corollary A.1.** *With bias:*

$$MVPF_{d\Lambda} = \frac{-\psi \cdot \int \lambda \cdot \frac{\kappa'(\Lambda)}{v'(B)}\mathbb{P}[SA]d\mu - (1-\psi) \cdot \int \lambda \frac{\kappa(\Lambda)}{v'(B)}\frac{\partial \mathbb{P}[SA]}{\partial \Lambda}d\mu}{\int FE \cdot \frac{\partial \mathbb{P}[SA]}{\partial \Lambda}d\mu} \tag{A.18}$$

$$MVPF_{dB} = \frac{\int \lambda \mathbb{P}[SA]d\mu + (1-\psi) \cdot \int \lambda \frac{\kappa(\Lambda)}{v'(B)}\frac{\partial \mathbb{P}[SA]}{\partial B}d\mu}{\int FE \cdot \frac{\partial \mathbb{P}[SA]}{\partial B}d\mu} \tag{A.19}$$

# 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 Appendix B.1




(A) Caseload (B) Inflow

FIGURE B.1. Take-up of SA (%) is plotted over time for 2005 - 2021. Both absolute caseload and inflow are shown. Two time periods are defined by an important policies: before/after the Participation Act of 2015 as discussed above.

B.1. **Eligibility Pseudocode.**

| 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

**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 ↦ Add yearly income from business, assets, sickness/disability benefits /12
      Gross Income ↦ Add unemployment benefits over periods with no employment income
    Deductions = payroll taxes + national insurance contributions + employee insurance contributions
      Deductions ↦ 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 ≤ Threshold, wealth ≤ wealth limit, and house value ≤ 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}
```

## Appendix C. Average Take-up Levels: Additional Material

| $\hat{\beta}$: 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 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.1. 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 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.

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

TABLE C.2. 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 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.

| | (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 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.3. 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 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.

## APPENDIX D. BARRIER SCREENING EFFECTS: ADDITIONAL MATERIAL

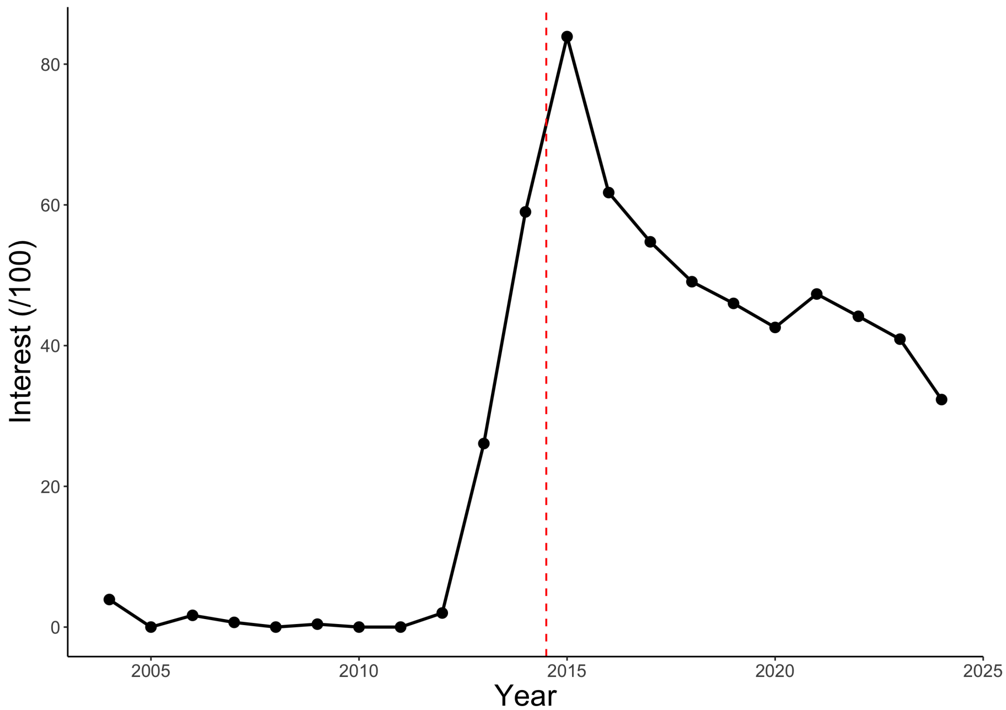



FIGURE D.1. Google Trends for *Participatiewet*, the Dutch translation of the "Participation Act" over time in the Netherlands.

D.1. **Detail on the Participation Act.** I argue that the Participation Act increased barriers to accessing social assistance. The policy intensified the obligations associated with receiving social assistance and incentivised municipalities to reduce caseload which they did via (threat of) sanctions.

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. Indeed, van der Veen (2019) state that "the PA 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'".

| **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.

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 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."[48]

D.2. **Results.**

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

| $\hat{\delta}_t$ (p.p.) for different specifications | Figure 5 | Figure 6 | | Figure 7 | | Figure 8 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 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 5 through 8. For more details see those graphs.

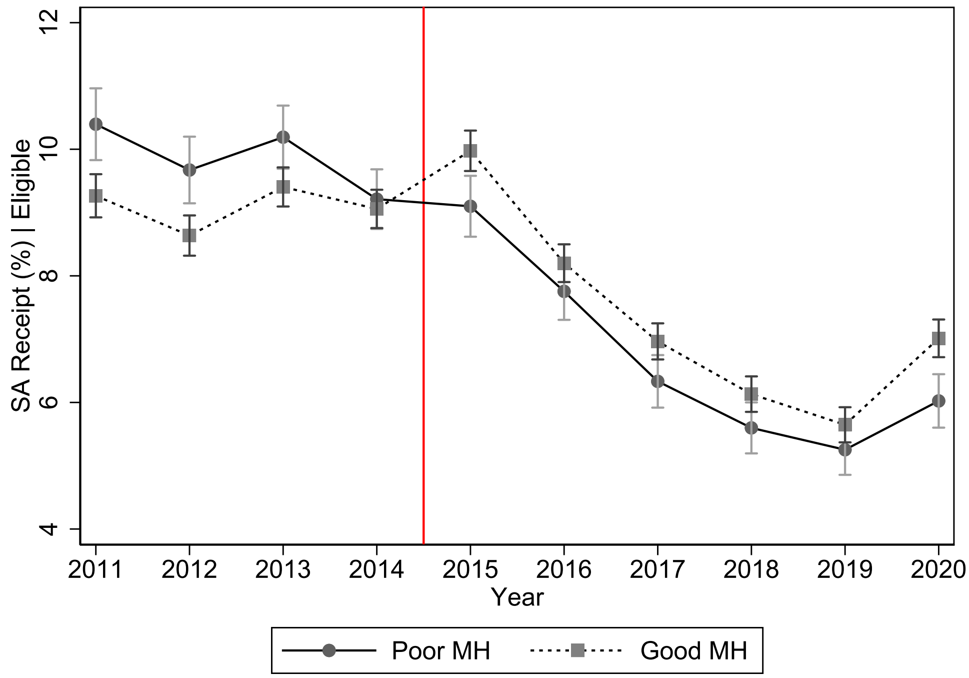



FIGURE D.2. Evolution of inflow of social assistance over time, split by people with poor mental health in the pre-period vs those with good mental health in this period. Raw means and respective 95% confidence intervals are shown. The introduction of the Participation Act in 2015 is shown by the red 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.6 | Figure D.7 | Figure D.10 | Figure D.11 |
|---|---|---|---|---|---|---|---|---|
| | 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.11. For more details see those graphs.

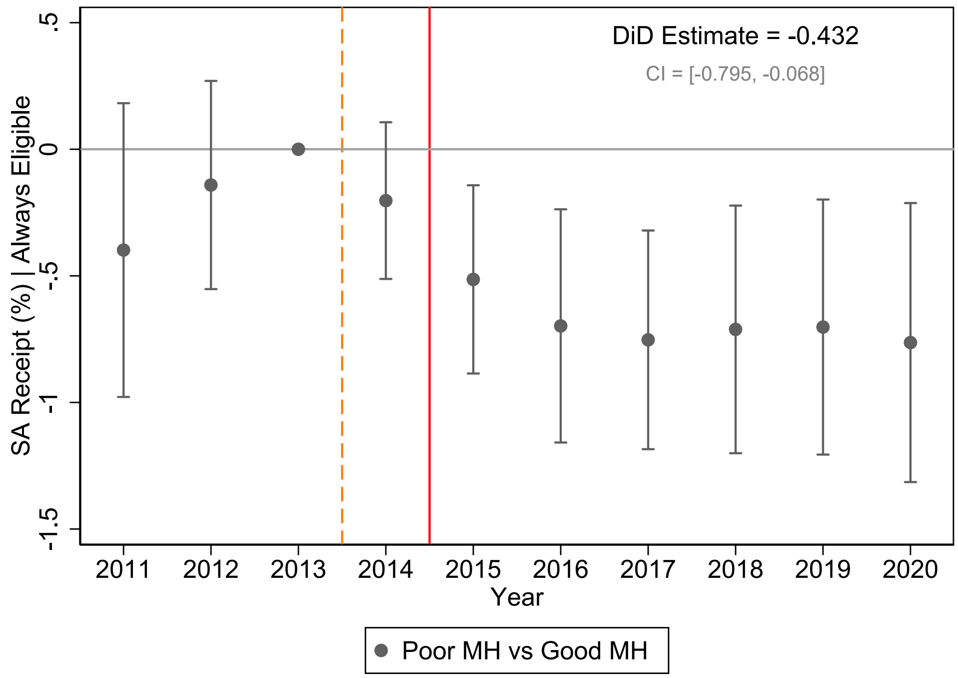



FIGURE D.3. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is always-eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the pre-period. Controls include individual fixed effects, 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).

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

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

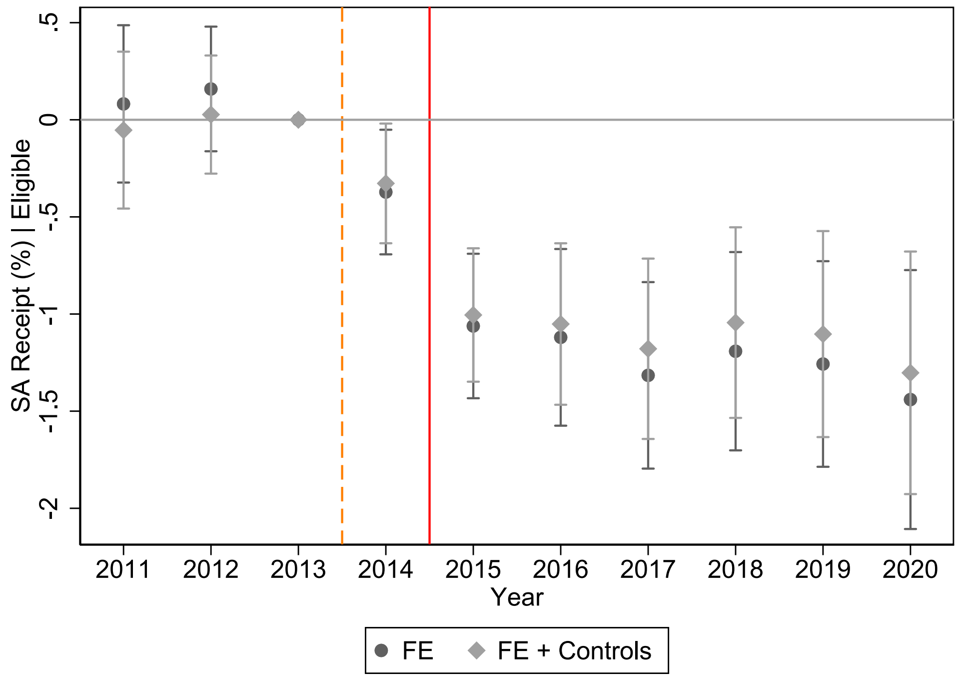



FIGURE D.4. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the 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, income, education and muncipality, hh composition and sector fixed effects. Standard-errors are clustered at the level of municipality of residence in 2013.

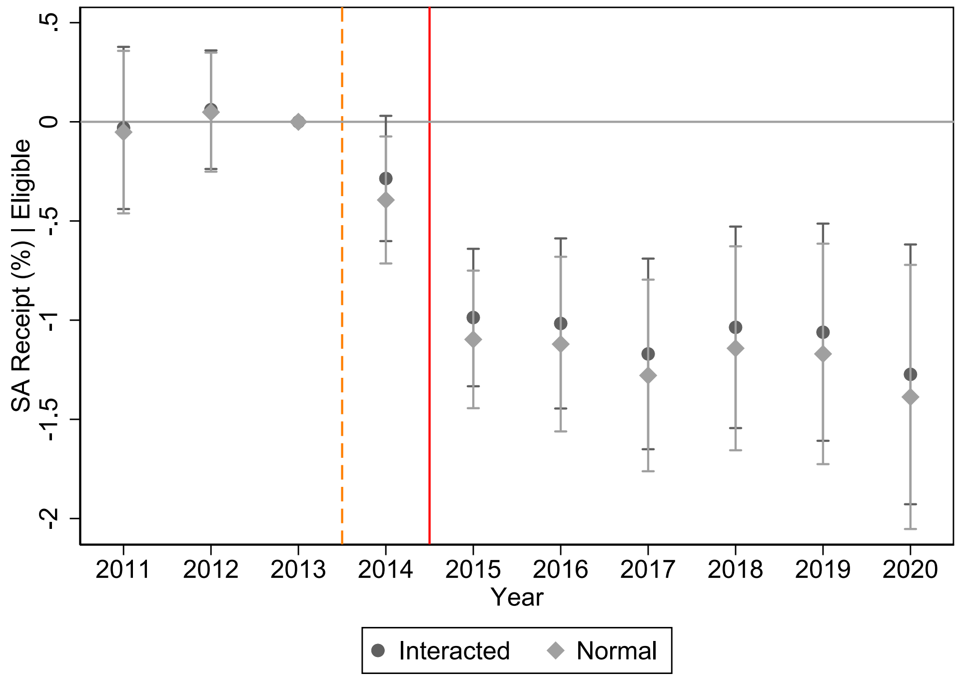



FIGURE D.5. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the pre-period. $\hat{\delta}_t$ are shown for two specifications - one with all time-varying controls - individual fixed effects, 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.

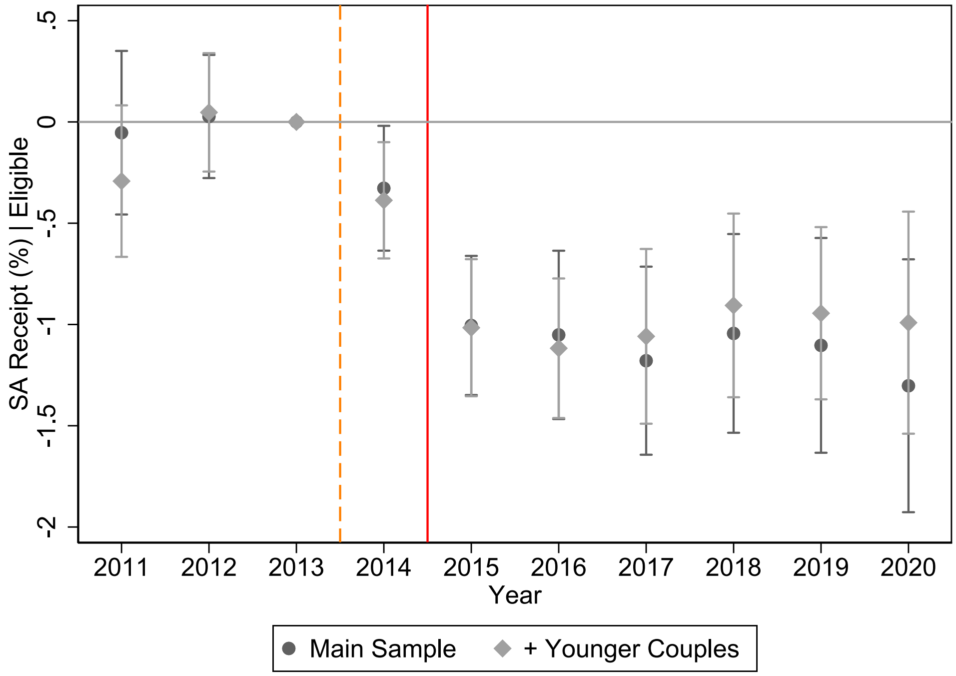



FIGURE D.6. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the 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, income, education and muncipality, hh composition and sector fixed effects. Standard-errors are clustered at the level of municipality of residence in 2013.

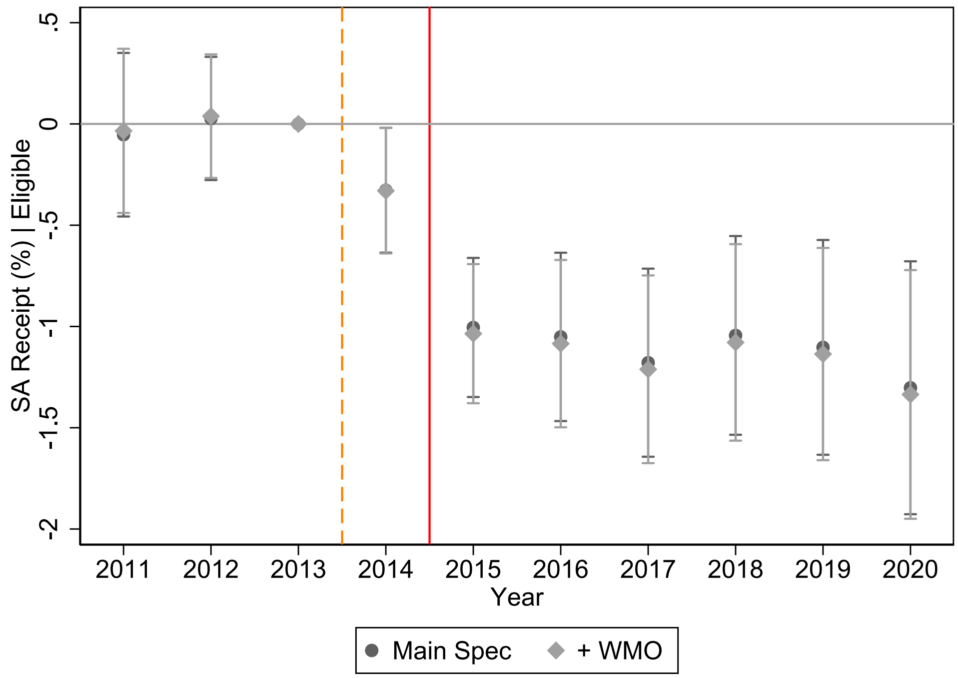



FIGURE D.7. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the pre-period. $\hat{\delta}_t$ are shown for two specifications - one with the main time-varying controls–individual fixed effects, 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.

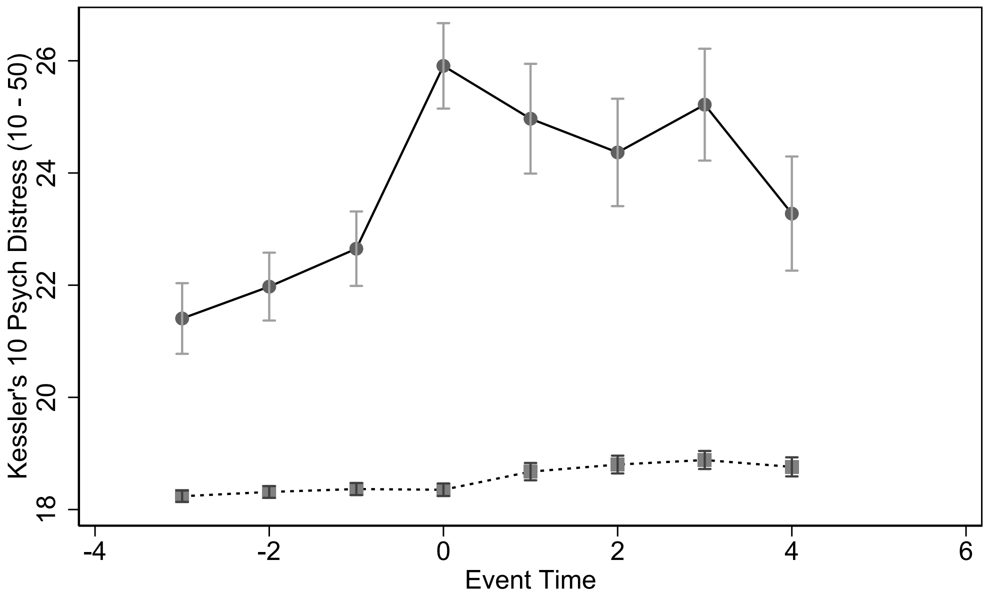



FIGURE D.8. This plot shows the mean subjective mental health (measured by Kessler's 10 Psychological Distress) for two groups: one group is prescribed psychopharma for the first time in Event Time 0, the other group has no prescriptions for all event times $t \leq 0$. Standard-errors are clustered at the level of municipality of residence in 2013.

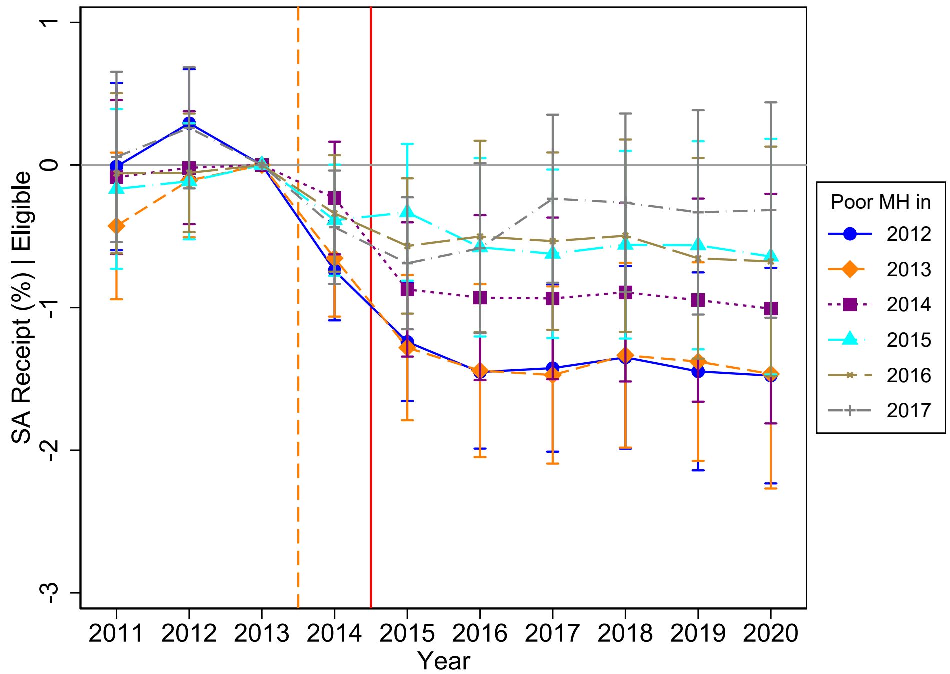



FIGURE D.9. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the pre-period. $\hat{\delta}_t$ are shown changing the definition of Poor MH$_i$. The different definitions are Poor MH$_i = \mathbb{1}\{\text{Prescribed Psychopharma in year } y\}$, showing estimates for $y \in \{2012, ..., 2017\}$. Controls include individual fixed effects, 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.4

| $\hat{\delta}_t$ (p.p.) for different timing of MH diagnosis | Figure D.9: 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.4. Coefficient Estimates of Figure D.9

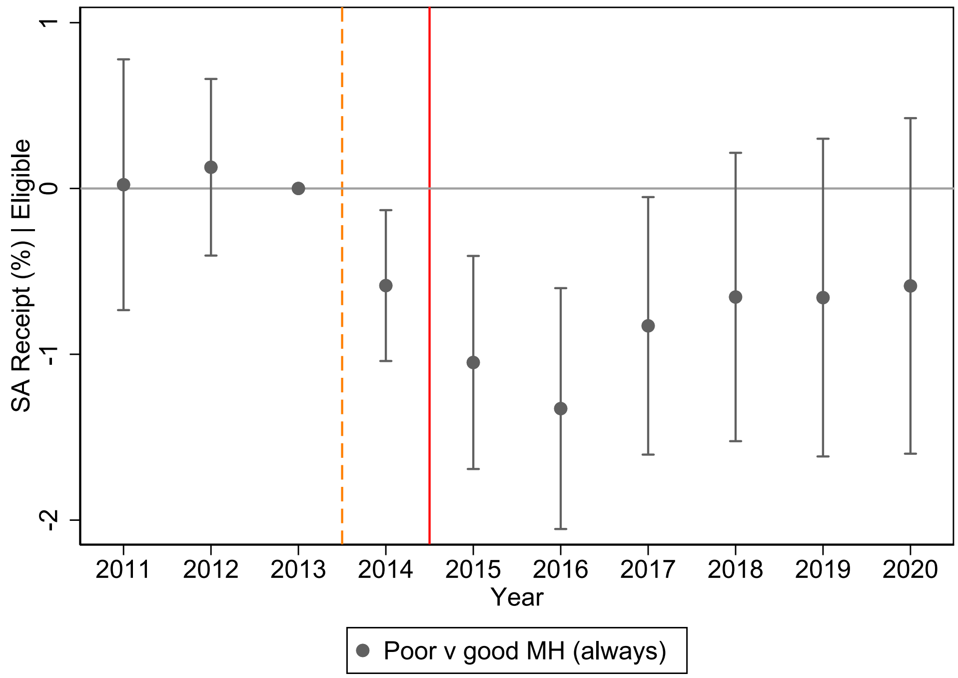



FIGURE D.10. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the pre-period. $\hat{\delta}$ for the main specification, where Poor MH$_i$: dispensed psychopharma in every year 2011-2020, vs good mental health throughout. Controls include individual fixed effects, income, education and muncipality, hh composition and sector fixed effects. Standard-errors are clustered at the level of municipality of residence in 2013.

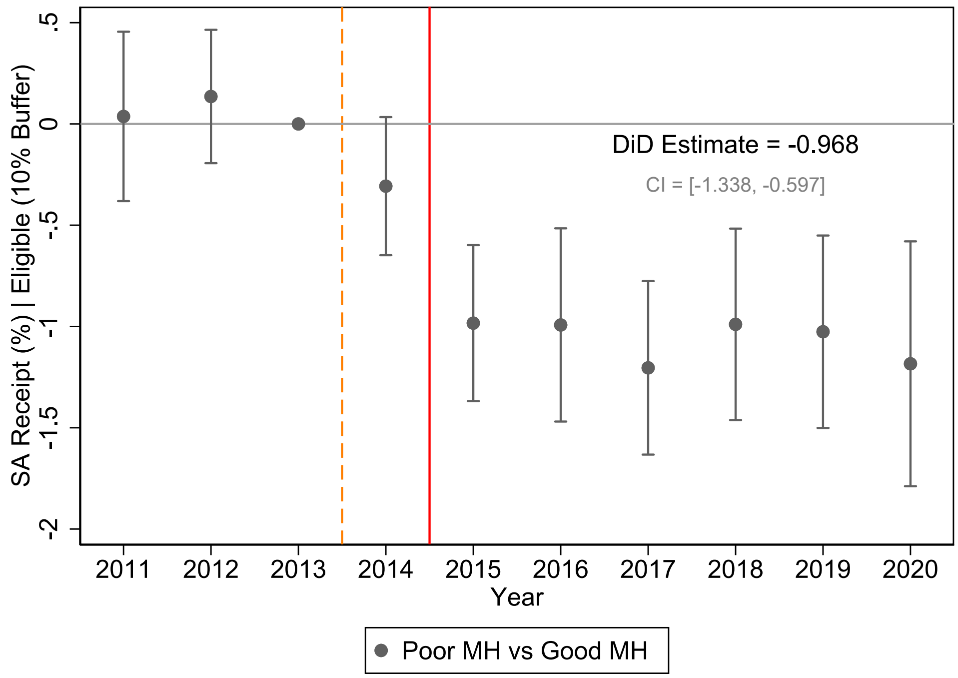



FIGURE D.11. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in the pre-period. $\hat{\delta}$ for the main specification for individuals with income below 90% of the eligibility threshold. Poor $\text{MH}_i = \mathbb{1}\{$Prescribed Psychopharma in pre-period. Controls include individual fixed effects, 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).

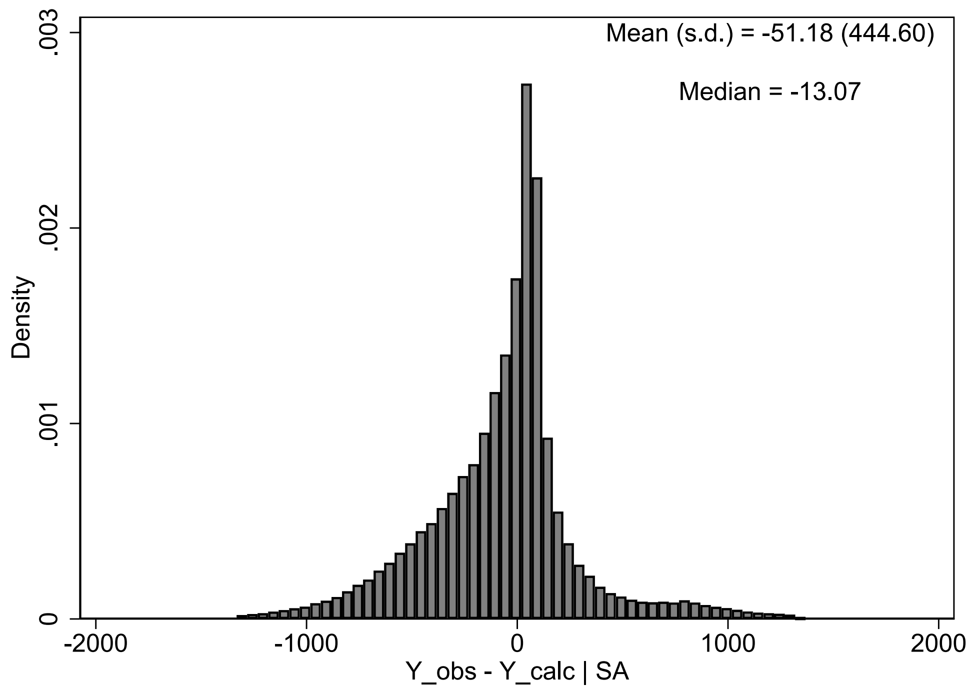



FIGURE E.1. Histogram of $Y_{\text{true}} - Y_{\text{calc}}$ for the analysis population of the RKD.

E.1. **Theory.** Income is exempted "insofar as, in the judgment of the [municipality], it contributes to [their] employment opportunities" (Ministerie van SZW, 2015). Appendix E.3 contains some descriptive facts about income-exemptions. 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 $\nu$ 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.2 shows that observed benefits conditional on receipt departs from the benefits schedule, particularly at and above the threshold. In this region, applicants really only 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 $\mathbb{E}[B|Y = y]$.

I impute the benefits schedule using a theoretical approach.[49] 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.3. 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.

Figure E.3 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.

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 $\varepsilon$: $\mathbb{P}[SA] = p(B, Y, \varepsilon)$. Let $I_X$ be the support of random variable $X$ which is potentially multi-dimensional, in which case represents a product space.

**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)

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

*where weighting function*

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

[49] Gelber et al. (2020) also use imputation for the first-stage of their RKD. The key idea, as in their paper, is that this imputation generates measurement error in the first-stage as well. The Card et al. (2015) framework can account for this measurement error.

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.4. 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 $\implies$ $\mathbb{P}[\text{Exemption}|Y = y]$ continuous at $\bar{y}$.

Figure E.13 and Figure 11 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.2 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. **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 individuals working more than zero hours and whose primary income is from work, to avoid notches in the benefit schedule (e.g., disability benefits) tied to the social assistance (SA) eligibility threshold. This threshold corresponds to the social minimum, which links to other government programs. Therefore, individuals who derive all their income from other benefits are ineligible for SA and are excluded. The typical person at the threshold earns most of their income from work/self-employment, with potential supplementary benefits, making them likely to move above or below the threshold at any point.

I further restrict the sample to 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 ordeal requirements.

**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 €80. 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.[50] The non-regularized CCT bandwidth delivers a

[50]Indeed, the CCT robust bandwidth without regularization performs poorly in simulations (see Appendix E.5).

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.

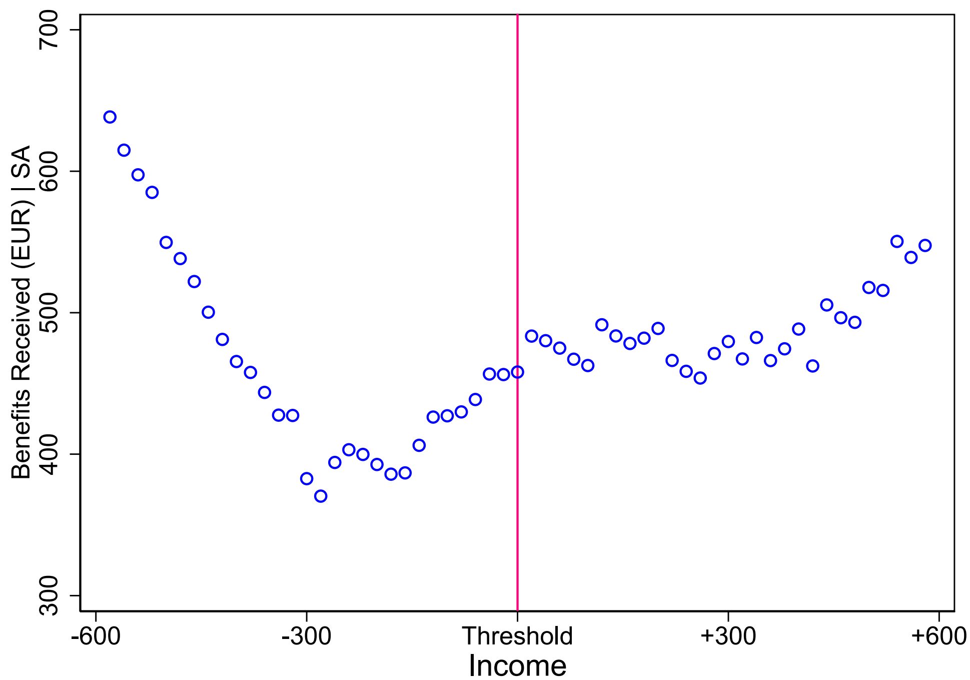



FIGURE E.2. Plot of benefits received conditional on receipt, averaged within income slice (€10 bins). A window of €1000 either side of the threshold is shown.

E.3. **Income Exemptions.** I model the unobserved benefits schedule as Equation (E.3).

$$B = b(y, \nu) = \begin{cases} \bar{y} & \text{if exemption, } \nu = 1 \\ \max\{\bar{y} - y, 0\} & \text{if exemption, } \nu = 0 \end{cases} \tag{E.3}$$

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.3). Then,* $\mathbb{E}[B|Y = y] = p(y) \cdot \bar{y} + \big(1 - p(y)\big) \cdot \max\{\bar{y} - y, 0\}$ *where:*

$$p(y) = \begin{cases} \frac{\left(\mathbb{E}[B|SA, Y=y] - (\bar{y}-y)\right)\cdot \mathbb{P}[SA|Y=y]}{y\cdot \mathbb{P}[SA|Y=y,\nu=1]} & \textit{if } y \leq \bar{y} \\ \frac{\mathbb{E}[B|SA,Y=y]\cdot \mathbb{P}[SA|Y=y]}{\bar{y}\cdot \mathbb{P}[SA|Y=y,\nu=1]} & \textit{if } y \geq \bar{y} \end{cases} \tag{E.4}$$

The proof is a simple application of Bayes-rule. I proxy $\mathbb{P}[SA|Y=y,\nu=1] \approx \mathbb{P}[SA|Y=0]$: the take-up rate conditional on exemption is equal to the take-up rate for people who have no income ($\approx 100\%$).

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

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

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

$$\begin{aligned} &= \bar{y}\cdot p(y)\cdot \mathbb{P}_y[SA|\nu=1] + \max\{\bar{y}-y,0\}\cdot \left[\mathbb{P}_y[SA] - p(y)\cdot \mathbb{P}_y[SA|\nu=1]\right] \\ &= \begin{cases} [\bar{y}-y]\cdot \mathbb{P}_y[SA] + y\cdot p(y)\cdot \mathbb{P}_y[SA|\nu=1] & \text{if } y\leq \bar{y} \\ \bar{y}\cdot p(y)\cdot \mathbb{P}_y[SA|\nu=1] & \text{if } y\geq\bar{y} \end{cases} \end{aligned}$$

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.4).

□

E.4. **Card et al. (2015) assumptions for validity of fuzzy RKD.**

(1) **Regularity:** $(\varepsilon,\nu)$ 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

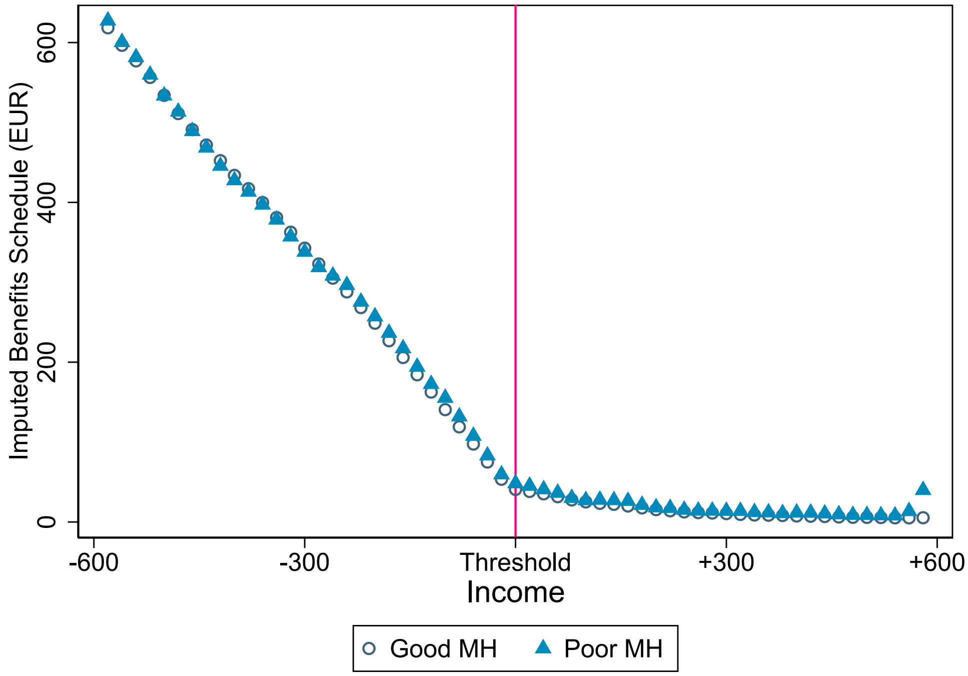



FIGURE E.3. Results of Imputation from Proposition E.2, calculated separately for people dispensed anti-depressants in the year previously (poor mental health) versus those who were not (good mental health).

(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:

$$\forall \mathbb{P}[SA | B = b, Y \leq \bar{y}] > 0 \tag{E.5}$$

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

$$\exists \Delta > 0 \text{ s.t. } \mathbb{P}[SA | Y = y] > 0 \; \forall y \in [\bar{y}, \bar{y} + \Delta] \tag{E.6}$$

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

$$\mathbb{P}[\text{Exemption} | Y = y] \text{ continuous at } \bar{y} \tag{E.7}$$

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 $Y_{\text{true}} > \bar{y}$. Assumption 2 is a corollary of $b(y, v)$ being continuous.

E.5. **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.5.1. *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. $\varepsilon$ 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:

$$B(y) \geq \kappa \tag{E.8}$$

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.

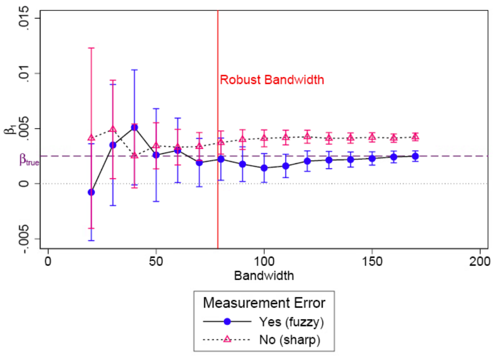
E.5.2. *Results.*

**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.

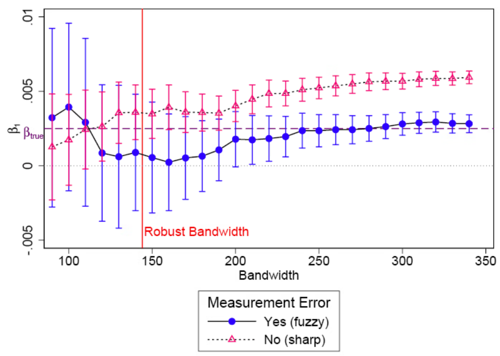
**Bandwidth:** for linear estimation, CCT bandwidths seem to perform well, but estimates become noisy for lower values with measurement error. For the identification of heterogeneous effects under measurement error, CCT performs poorly: I now assume that half of my simulated individuals have value $\alpha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

$$\alpha \cdot B(y) \geq \kappa \tag{E.9}$$

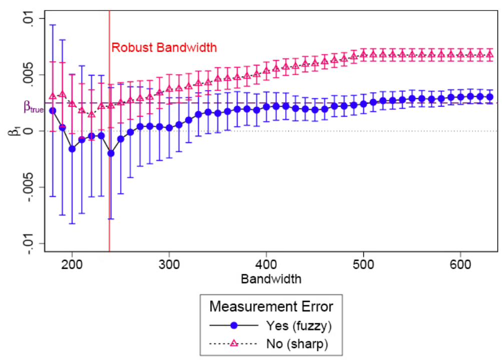
and rate of receipt $\mathbb{P}[SA | Y = y] = F_\varepsilon\big(\alpha \cdot B(y) - \kappa\big)$. I estimate the RKDs separately for $\alpha = 1, 2$ and test for a difference in the RKD estimates at different bandwidths. The estimates are shown in




(A) Local polynomial order $p = 1$




(B) Local polynomial order $p = 2$




(C) Local polynomial order $p = 3$

FIGURE E.4. Results of simulations showing estimates from RKDs using different bandwidths and different local polynomial orders. In each, the CCT robust bandwidth is shown.

Figure E.5. 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. **Other:** use standard triangular kernel.

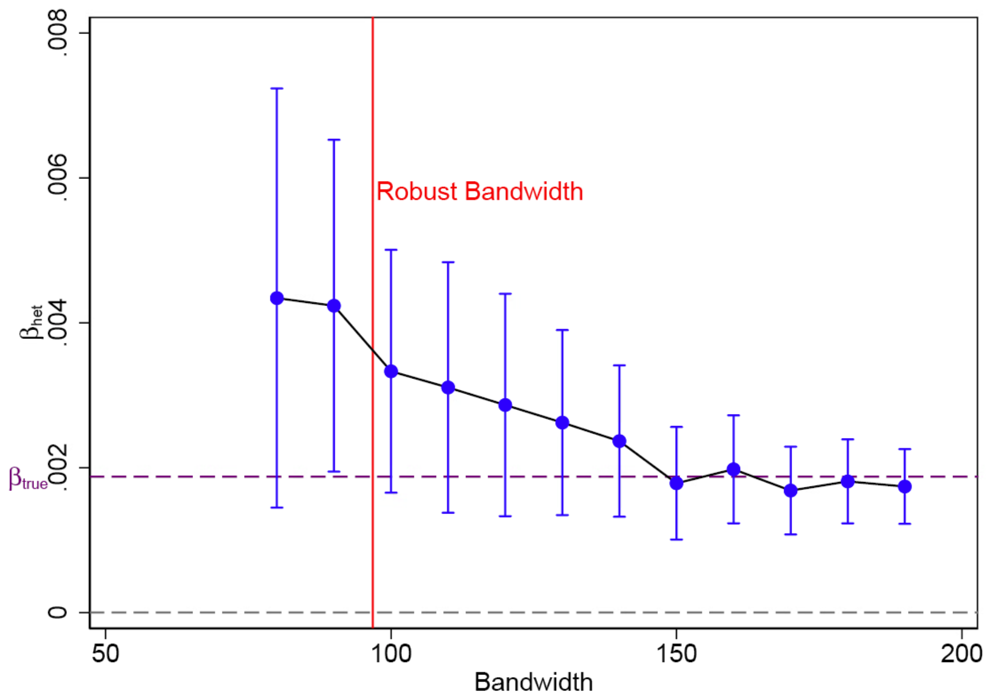



FIGURE E.5. Results of simulations showing estimates from heterogeneous RKD ($\alpha = 1$ vs $2$) using different bandwidths. CCT robust bandwidth is shown.

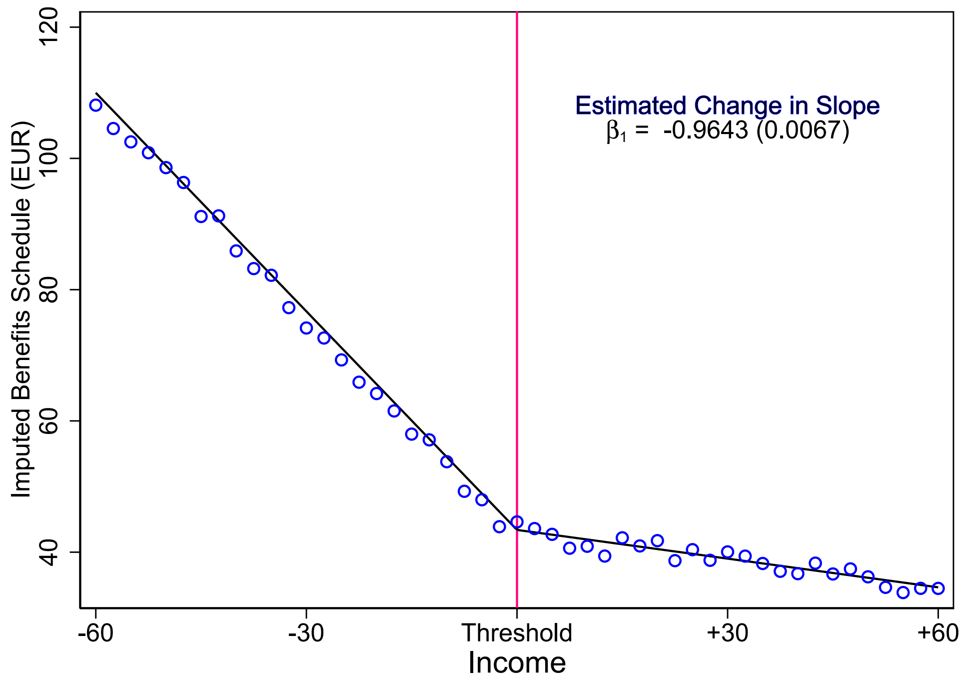



FIGURE E.6. Average imputed benefits schedule within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly. The sample contains singles employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Regression lines are shown 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.

E.6. **Results.**

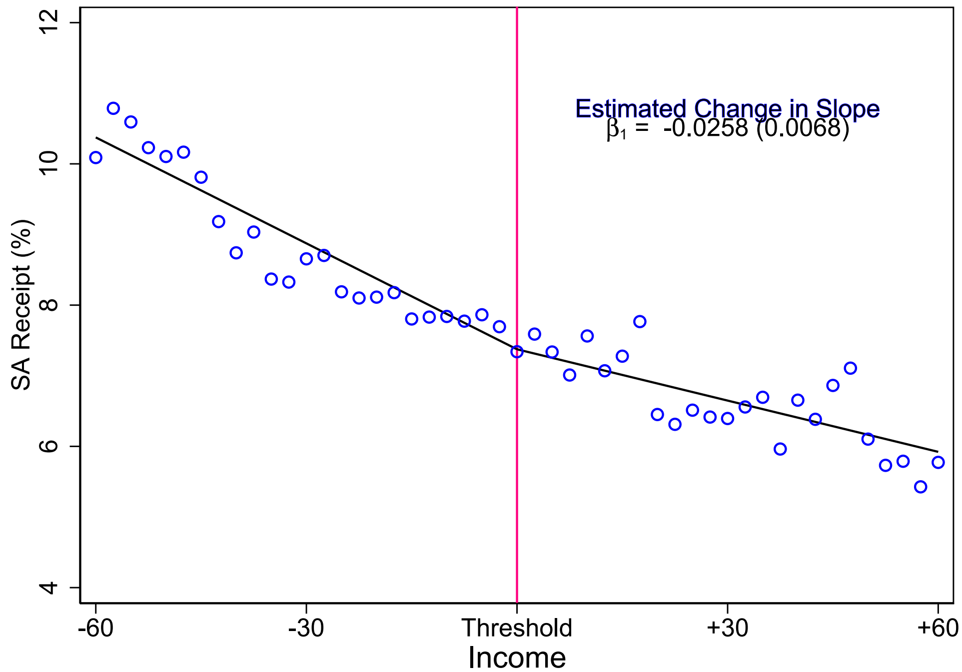



FIGURE E.7. Average rate of receipt within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly. The sample contains singles employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Regression lines are shown following 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.

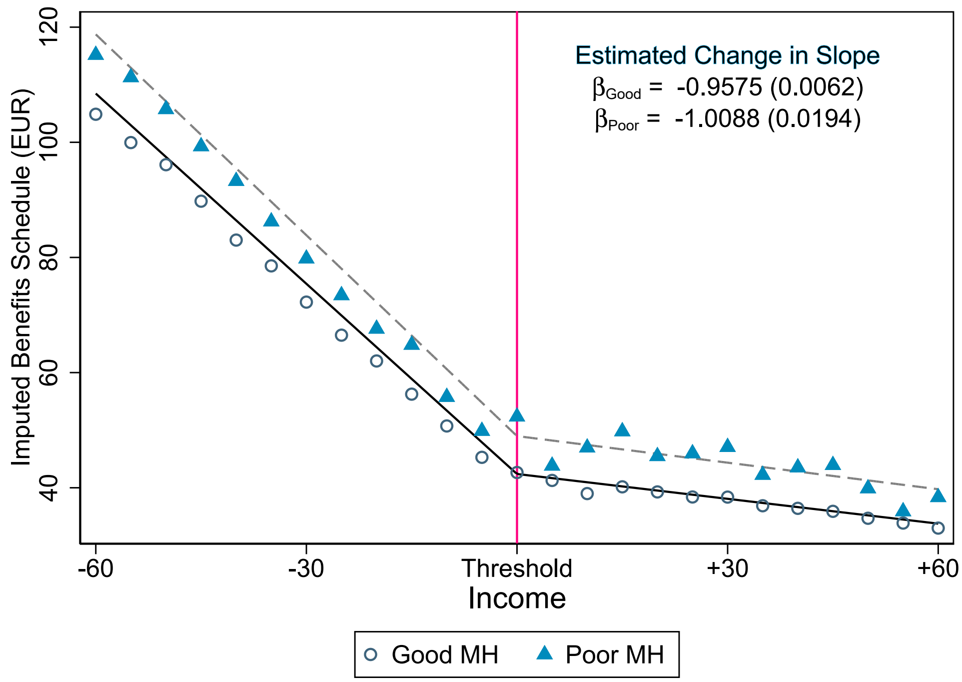



FIGURE E.8. Average imputed benefits schedule within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly. Poor mental health is defined as receiving anti-depressants in the year previously. The sample contains singles employees, years 2011-2014. 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 | | | |
|---|---|---|---|---|---|---|---|---|
| | Overall | | Heterogeneous by MH | | Overall | | Heterogeneous by MH | |
| | Raw | + Controls | Raw | + Controls | Raw | + Controls | Raw | + Controls |
| Income - Threshold<br>*Overall: everyone, Het: 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) |
| Income - Threshold (het)<br>*Het: poor vs good MH* | | | 0.00058 | -0.00703 | | | 0.00506 | -0.00513 |
| | | | (0.00938) | (0.00826) | | | (0.0119) | (0.0104) |
| $\min\{\text{Income - Threshold}, 0\}$<br>*Overall: everyone, Het: good MH* | -0.0258*** | -0.0207*** | -0.0218** | -0.0187** | | | | |
| | (0.00684) | (0.00585) | (0.00720) | (0.00653) | | | | |
| $\min\{\text{Income - Threshold}, 0\}$ (het)<br>*Het: poor vs good MH* | | | -0.0290 | -0.0136 | | | | |
| | | | (0.0182) | (0.0164) | | | | |
| Benefits<br>*Overall: everyone, Het: good MH* | | | | | 0.0213*** | 0.0267*** | 0.0227** | 0.0194** |
| | | | | | (0.00600) | (0.00707) | (0.00751) | (0.00677) |
| Benefits (het)<br>*Het: poor vs good MH* | | | | | | | 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 in parentheses
$^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE E.1. Estimates from the regression kink design using a bandwidth of €60. Columns 2 - 5 show the reduced-form and 6 - 9 IV. Columns 2 and 3 show estimates of the slope of receipt of SA w.r.t income - both below and above the threshold. Column 2 reflects estimates without any covariates. In Column 3, I add controls for month, year, age, gender, wealth, education, municipality, hh composition and sector fixed effects. Columns 4 and 5 show heterogeneous estimates by mental health (raw / + controls respectively). "Income - Threshold" reflects the results for people with good mental health, and "Income - Threshold (het)" that of poor mental health (relative to good). Analogously columns 6 - 9 are the IV estimates using the imputed benefits schedule as a first-stage. Throughout, standard-errors are clustered at the municipality level. Underlying population = singles 2011-2014 who get most income from work. See text for details.

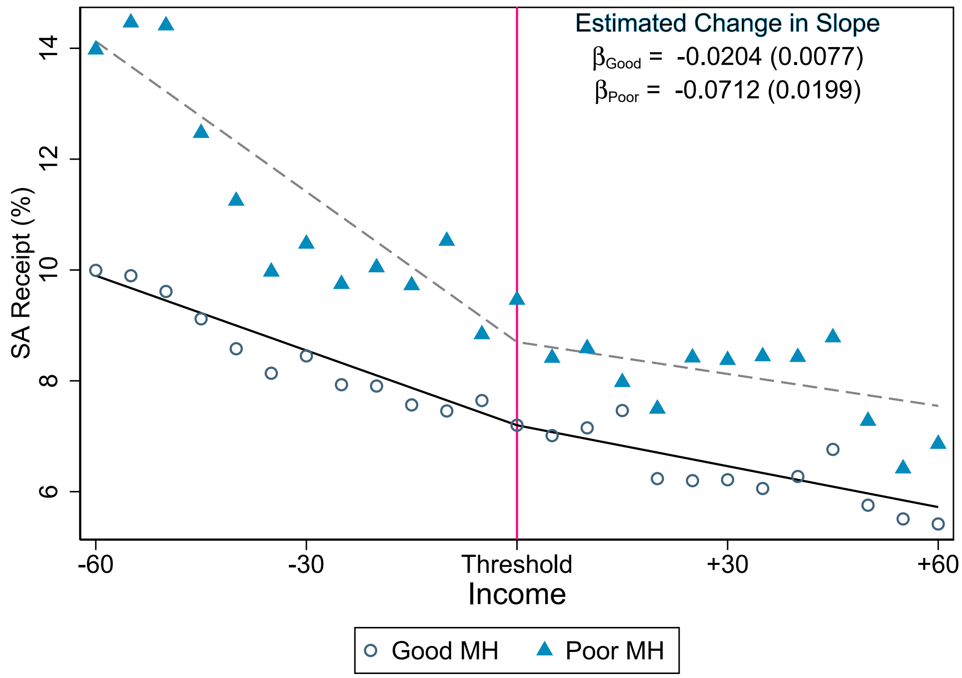



FIGURE E.9. Average rate of receipt within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly. Poor mental health is defined as receiving anti-depressants in the year previously. The sample contains singles employees, years 2011-2014. See 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.

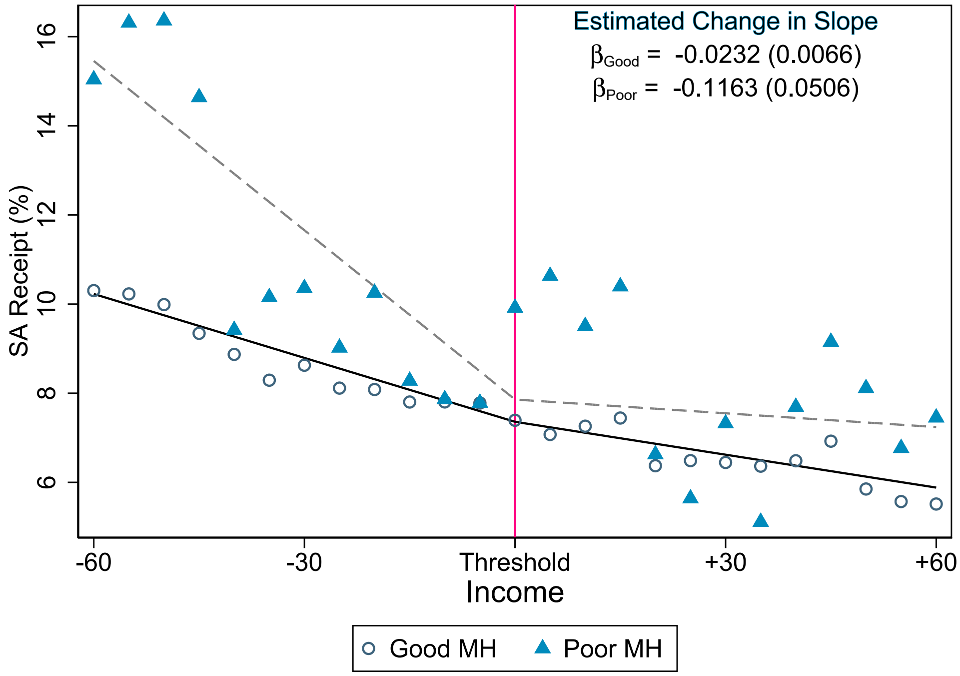



FIGURE E.10. Average rate of receipt within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly. Poor mental health is defined as receiving anti-psychotics in the year previously. The sample contains singles employees, years 2011-2014. See 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.

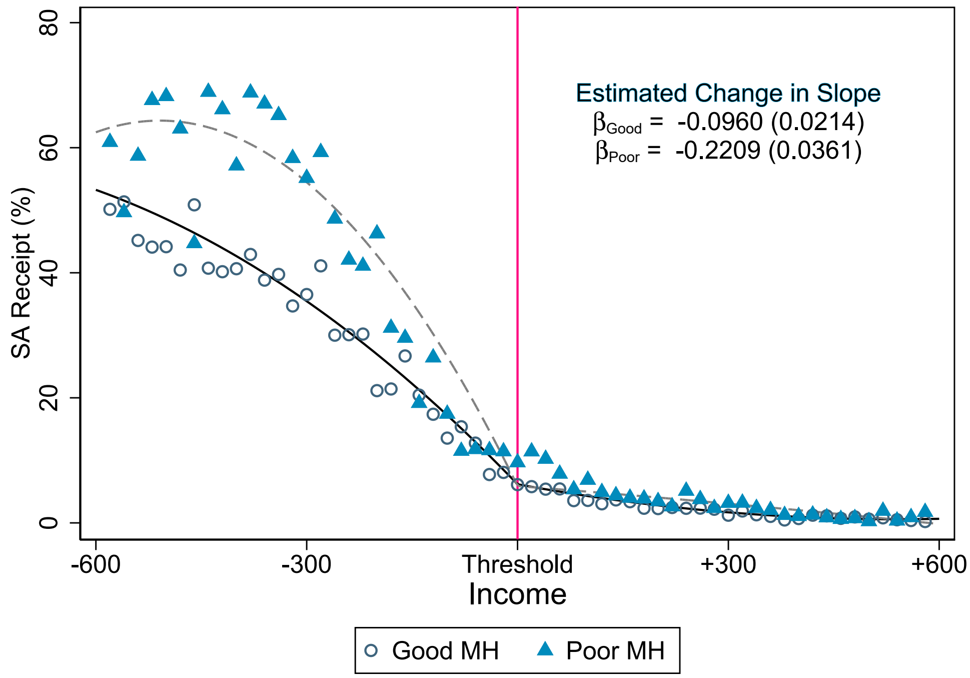
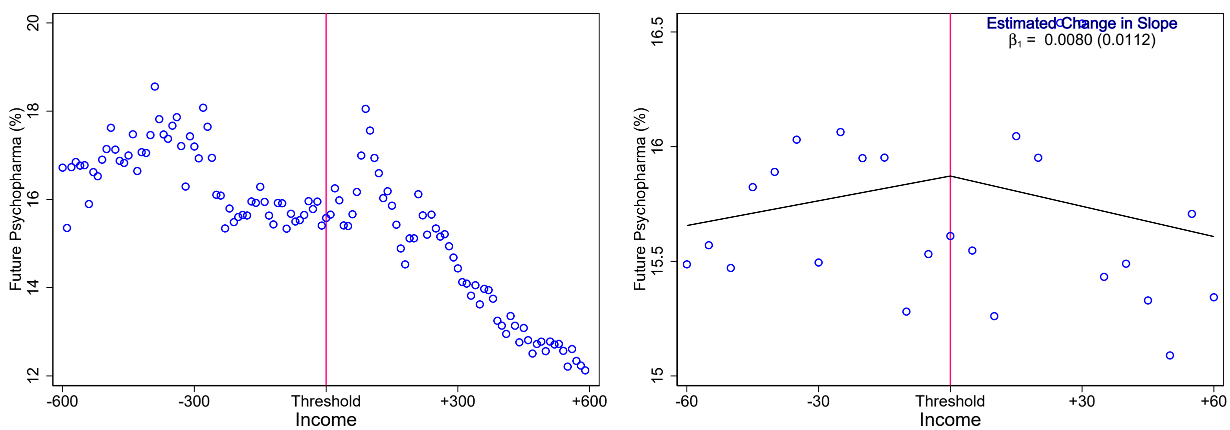



FIGURE E.11. Average rate of receipt within income slice in a large window of income either side of the eligibility threshold. Income in this plot is monthly. Poor mental health is defined as reporting severe psychological distress in the survey in 2012. The sample contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions.




(A) Overview

(B) Zoomed

FIGURE E.12. Average dispensations of psychotropic drugs in the future year within current income slice in a window around the eligibility threshold. Income in this plot is monthly. The sameple contains single employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions.

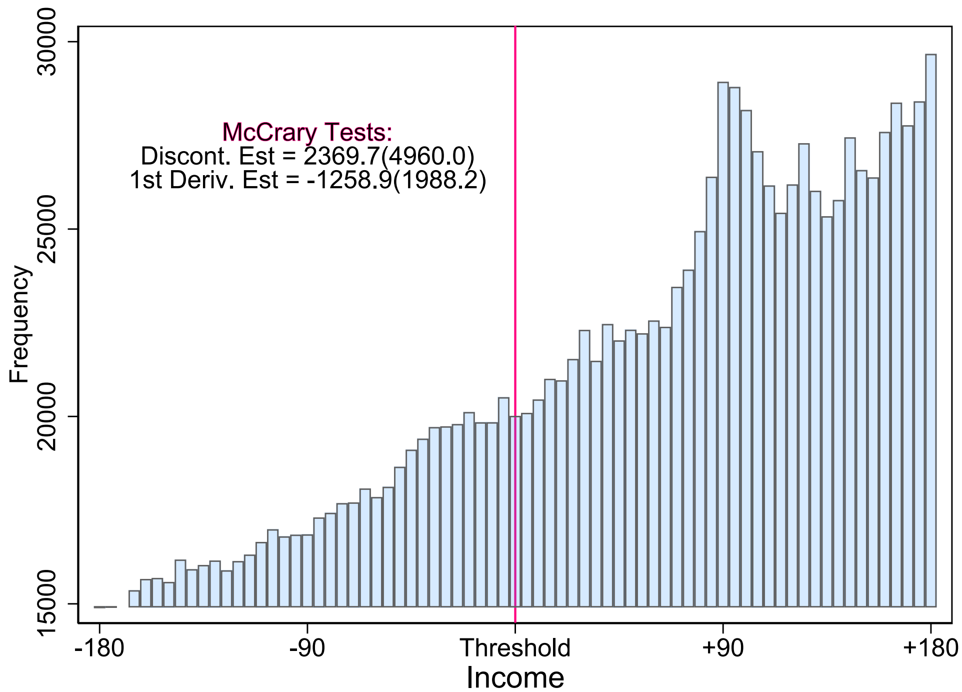



FIGURE E.13. Density of income around the eligibility threshold. McCrary (2008) tests for discontinuity in levels and slopes around the threshold are shown. Income in this plot is monthly. The sample contains singles employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions.

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

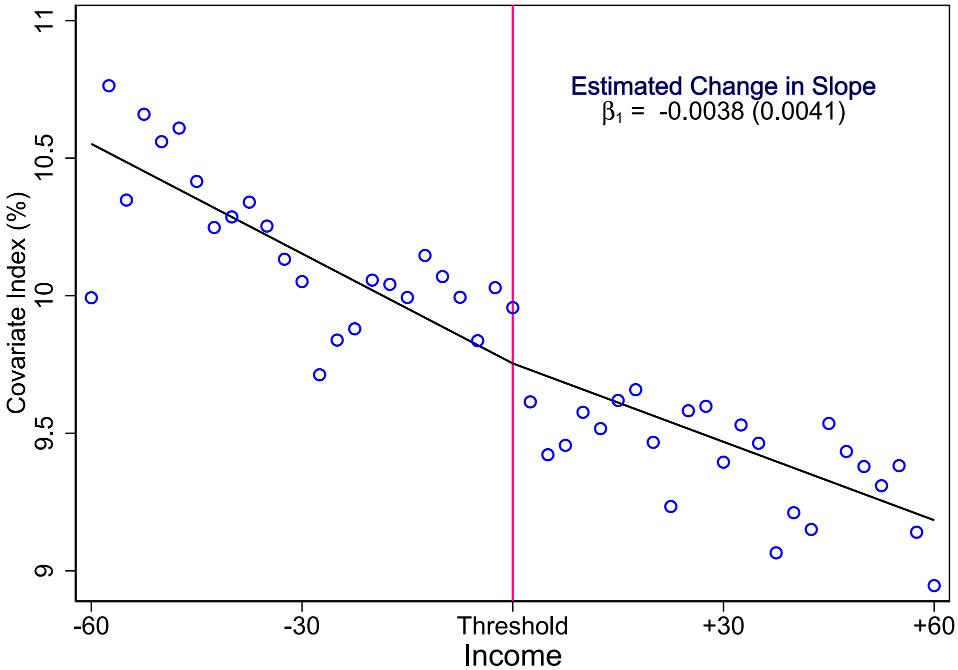



FIGURE E.14. Covariate Test: plot shows fitted values of a regression of social assistance take-up on all pre-determined controls used throughout this paper including income, education, hh composition, municipality FEs. These fitted values form a "Covariate Index" which is binned. An RKD estimate with income as the running variable is also shown. Income in this plot is monthly. The sample contains singles employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Standard-errors are clustered at the municipality level.

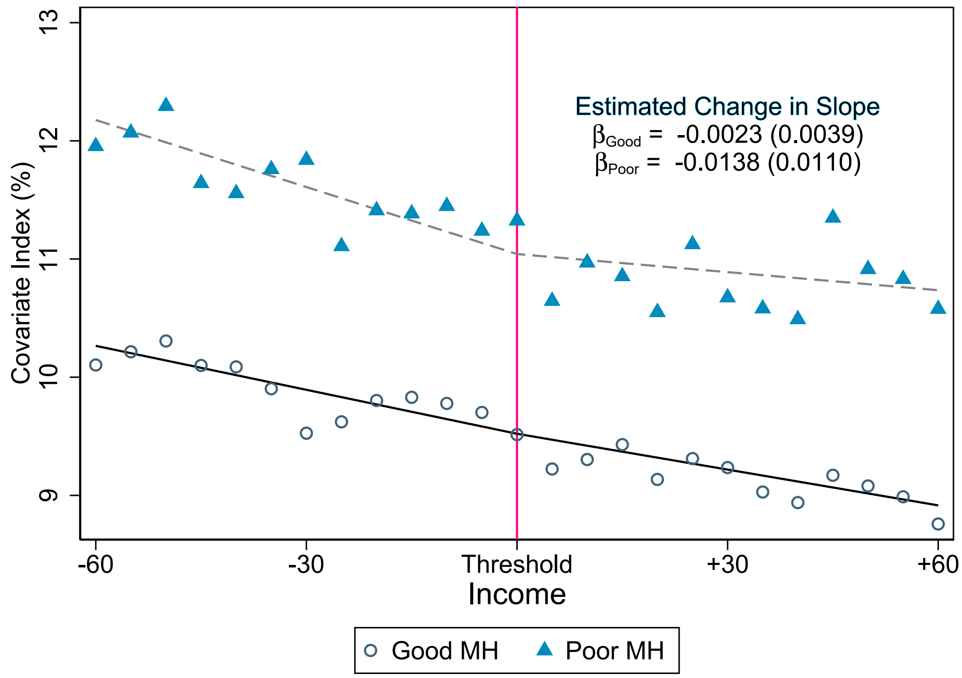



FIGURE E.15. Covariate Test: plot shows fitted values of a regression of social assistance take-up on all pre-determined controls used throughout this paper including income, education, hh composition, municipality FEs. These fitted values form a "Covariate Index" which is binned. An RKD estimate with income as the running variable is also shown. Separated by mental health. Income in this plot is monthly. Poor mental health is defined as receiving psychopharma in the year previously. The sample contains singles 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.

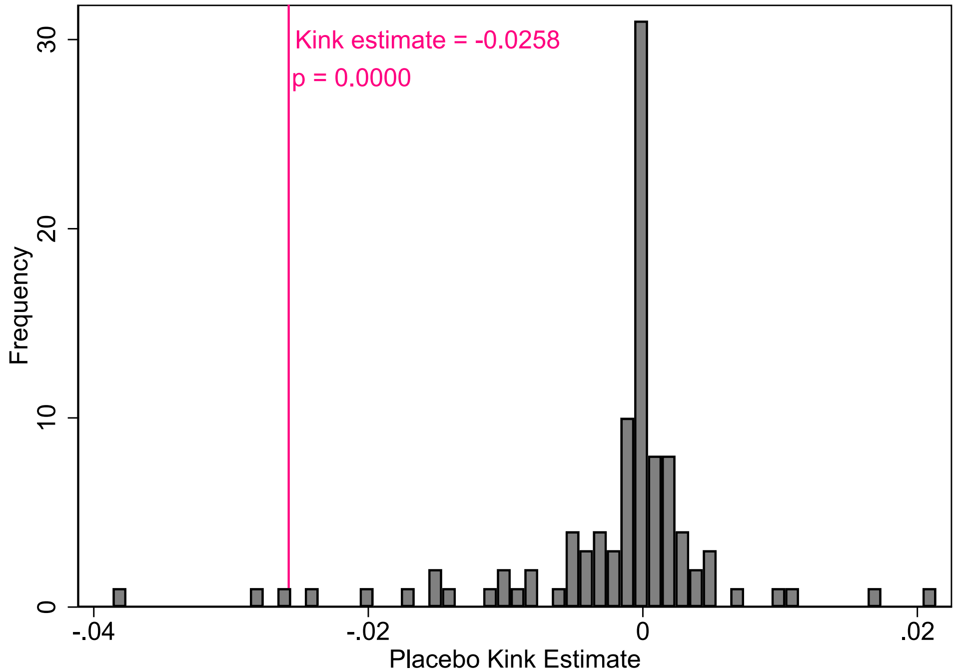



FIGURE E.16. Results of permutation test à la Ganong and Jäger (2018). I estimate RKDs on 100 placebo kinks in the range $[\bar{y} - 600, \bar{y} + 600]$ and plot a histogram of the estimates. A binomial test is used to check whether the true estimate is an outlier. The sample contains singles employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Standard-errors are clustered at the municipality level.

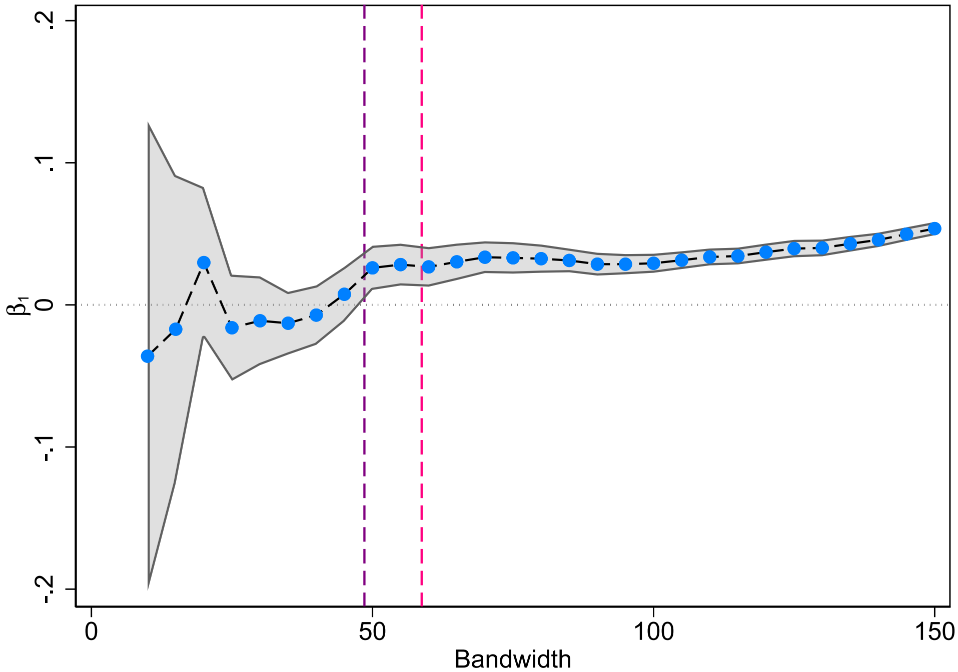



FIGURE E.17. Results of test of sensitivity to changes in bandwidth. I estimate RKDs changing the bandwidth, with the CCT robust bandwidth displayed. The lower purple dashed line indicates the CCT robust bandwidth with regularization, and the upper pink dashed line indicates the CCT robust bandwidth without regularization. This plot shows the estimates and confidence intervals. The sample contains singles employees, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Standard-errors are clustered at the municipality level.

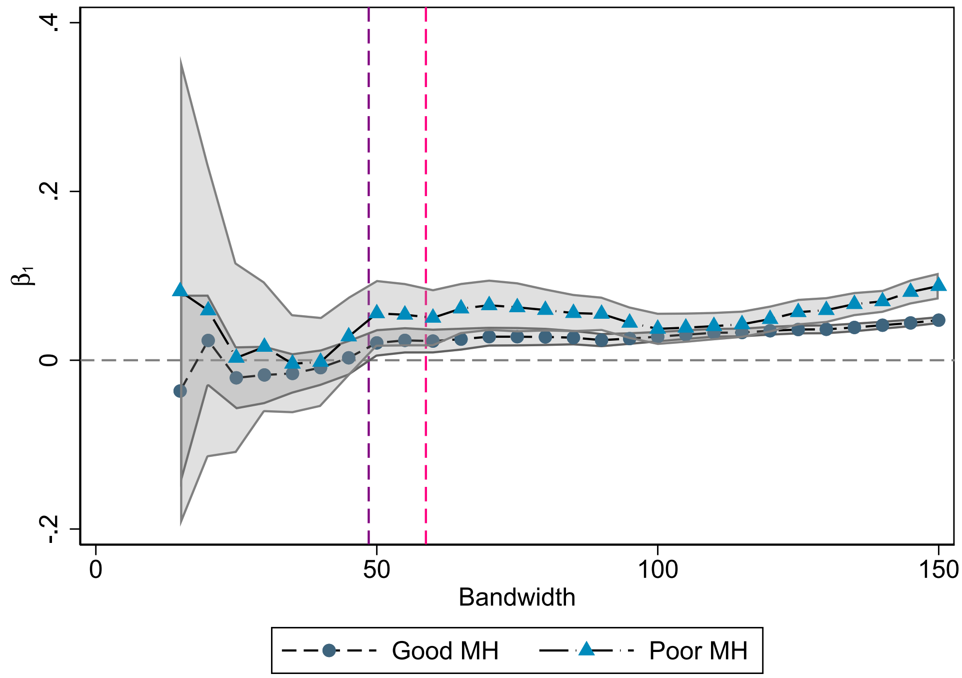



FIGURE E.18. Results of test of sensitivity to changes in bandwidth. I estimate heterogeneous RKDs changing the bandwidth, with the CCT robust bandwidth displayed. The lower purple dashed line indicates the CCT robust bandwidth with regularization, and the upper pink dashed line indicates the CCT robust 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 singles 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 (2.4) makes much more sense if it holds for $\theta$ 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:*

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

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 labor supply responses to $dB$ or $d\Lambda$. 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 $\mu_{\text{Eligible}}$ is the conditional density of types $\theta$. Similarly, Bayes Rule $\rightarrow \int \frac{\partial \mathbb{P}[SA]_\theta}{\partial X} \cdot H_\theta d\mu = \mathbb{P}[\text{Eligible}] \cdot \int \frac{\partial \mathbb{P}[SA|\text{Eligible}]_\theta}{\partial X} \cdot \frac{1}{1 - \mathbb{P}[SA \mid \text{Eligible}]_\theta \cdot \mathbb{P}[\text{Ineligible} \mid \text{No } SA]_\theta} \cdot H_\theta d\mu_{\text{Eligible}}$.

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

| Sufficient Statistics | Method | Estimated Value |
|---|---|---|
| $\left(\frac{\partial\hat{\mathbb{P}}[SA]_H}{\partial B}, \frac{\partial\hat{\mathbb{P}}[SA]_L}{\partial B}\right)$ | RKD | $(0.000227, 0.000503)$ |
| $v'_H$<br>$f(v_H - \kappa_H)$ | Normalization<br>RKD$_H$ + $v'_H$ | 1<br>0.000227 |
| $f(v_L - \kappa_L)$ | $\hat{\mathbb{P}}[SA]_H = \hat{\mathbb{P}}[SA]_L$<br>+ shortcut: $f(\cdot)_L = f(\cdot)_H$ | 0.000227 |
| $v'_L$ | RKD$_L$ + $f(v_L - \kappa_L)$ | 2.2 |
| $\left(\frac{\partial\hat{\mathbb{P}}[SA]_H}{\partial \Lambda}, \frac{\partial\hat{\mathbb{P}}[SA]_L}{\partial \Lambda}\right)$ | (Diff, Diff-in-Diff) | $(-0.014, -0.023)$ |
| $(\kappa'_H, \kappa'_L)$ | (Diff-in-)Diff + $f(\cdot)_L = f(\cdot)_H$ | $(98, 161)$ |

TABLE F.1. Table summarising the calibration of the key sufficient statistics

are smaller than those estimated on the larger bandwidth of €600 either side of the threshold as in Figure 10, $(0.000778, 0.00145)$ for good and poor mental health, respectively. If I alternatively use these larger estimates for the calibration, I find that $v'_L = 1.86$, $\kappa'_H = 28.6$ and $\kappa'_L = 46.9$. These estimates imply:

$$MVPF_{dB} = 0.47$$

$$MVPF_{d\Lambda} = 0.71$$

and increasing barriers is concluded as 48% 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 Additivity.* 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 $\varepsilon$, behaviour will follow a threshold-rule: $SA = 1 \iff \varepsilon \leq \varepsilon^*$ where $\varepsilon^*_\theta$ 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 (2.11) becomes

$$\frac{-\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}}{\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}} = \frac{\mathbb{E}\left[\kappa'_\theta(\Lambda, \varepsilon) | \varepsilon = \varepsilon^*_\theta\right]}{\mathbb{E}\left[v'_\theta(B, \varepsilon) | \varepsilon = \varepsilon^*_\theta\right]} \tag{F.3}$$

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*. Therefore, approximating $\mathbb{E}\left[v'_\theta(B, \varepsilon) | \text{Inframarginal}\right]$ as $\mathbb{E}\left[v'_\theta(B, \varepsilon) | \text{Marginal to } dB\right]$ seems reasonable.

On the other hand, those people screened out by the change of ordeals are likely less needy on average than the inframarginals. 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 would suggest that my estimate of $MVPF_{dB}$ is about right, but $MVPF_{d\Lambda}$ is an underestimate.

F.3.2. *Relaxing Independence.* Relaxing independence of $\varepsilon$ leads to the following. Suppose we used the model of Rafkin et al. (2023) where $v'_\theta(B)$ is independent of $\theta$ conditional on income, but $\varepsilon_\theta \sim F_\theta$. Then, same average take-up levels combined with the difference-in-differences results would suggest $f_{\varepsilon_L}(v_L - \kappa_L) = 1.65 \times f_{\varepsilon_H}(v_H - \kappa_H)$, inconsistent with the regression kink design results. Using Finkelstein and Notowidigdo (2019)'s model, $\kappa'_\theta$ are assumed to be opportunity costs of time, the only reason why need would vary across individuals with the same income is due to misperceptions of the benefit level and $\varepsilon_\theta \sim F_\theta$. My results would then suggest $\kappa'_L = w_L =$

€11.7 and $\kappa'_H = w_H =$ €13.7, where $w_\theta$ is $\theta$'s wage. Then, the regression kink design estimates would suggest $v'_L = 1.8 \times v'_H$. This would imply that people with poor mental health have an easier time overcoming barriers, and are substantially *less* pessimistic about the benefit level. Both of these results contradict psychological evidence (Martin et al., 2023b; Evans et al., 2014; Alloy and Ahrens, 1987).

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, my framework is flexible enough to relax this assumption under structural assumptions. 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 $64\%$ 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). I find that in the structural model, the probability of being marginal to a barrier instrument less than $1/10^{\text{th}}$ to that of a benefits instrument - this only pushes the welfare comparison that I explore in the next section *more* to the side of reducing barriers.

**Details:** Assume linearity: then, and $\kappa_\theta(\Lambda) = \kappa_\theta \cdot \Lambda$ and as in Equation (A.14),

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

Note that one or other of these assumptions is assumed throughout in Anders and Rafkin (2022), Finkelstein and Notowidigdo (2019) and Rafkin et al. (2023). In this case: $\hat{\mathbb{P}}[SA]_L \approx \hat{\mathbb{P}}[SA]_H \implies$

| 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<br>Mean (SD) | Single Employees<br>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)

$\upsilon_L \cdot \Psi_L - \kappa_L \cdot \Lambda = \upsilon_H \cdot \Psi_H - \kappa_H \cdot \Lambda$. This means that $\kappa_L - \kappa_H = \frac{\upsilon_L \cdot \Psi_L - \upsilon_H \cdot \Psi_H}{\Lambda}$. Recall that $\frac{\partial \hat{\mathbb{P}}[SA]_H}{\partial B} = 0.000227$ and $\frac{\partial \hat{\mathbb{P}}[SA]_L}{\partial B} = 2.2 \times 0.000227$. These estimates imply $\kappa_L - \kappa_H = \frac{2.2 \times 520 - 516}{\Lambda}$.

Let $f_\varepsilon^{d\Lambda} = \alpha \cdot f_\varepsilon^{dB}$ - i.e. assume that the ratio of the probability of being marginal to a benefits-level instrument over probability of being marginal to an ordeal change is constant across mental health types. In this case, $\frac{\partial \hat{\mathbb{P}}[SA]_L - \hat{\mathbb{P}}[SA]_H}{\partial \Lambda} = -(\kappa_L - \kappa_H) \cdot \alpha \cdot 0.000503$, as $\frac{\partial \hat{\mathbb{P}}[SA]_H}{\partial B} = f_\varepsilon^{dB}$. Rearranging for $\alpha$,

$$\hat{\alpha} = \frac{-\Lambda \frac{\partial \hat{\mathbb{P}}[SA]_L - \hat{\mathbb{P}}[SA]_H}{\partial \Lambda}}{0.000503 \times 636.9} \tag{F.4}$$

Therefore as long as we can estimate the heterogeneous semi-elasticity $-\Lambda \frac{\partial \hat{\mathbb{P}}[SA]_L - \hat{\mathbb{P}}[SA]_H}{\partial \Lambda}$, we are done.

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 $100\%$. Therefore,

$$-\Lambda \frac{\partial \hat{\mathbb{P}}[SA]_L - \hat{\mathbb{P}}[SA]_H}{\partial \Lambda} = \frac{0.09}{0.221} = 0.040$$
$$\implies \hat{\alpha} = \frac{0.040}{0.35} = 0.114$$

In particular, $f^{d\Lambda} < f^{dB}$ - which only pushes in the direction of $MVPF_{d\Lambda} > MVPF_{dB}$.

F.4. **Welfare Effects.** Calculating social marginal utilities of beneficiaries of benefit and barrier change instruments:

$$\begin{aligned}
\bar{\eta}_{dB} &= 0.25 \times 2.07 + 0.73 \times 1 \\
&\approx 1.29 \\
\bar{\eta}_{d\Lambda} &= 0.25 \times 2.07 \times \frac{36.2/2.07}{0.25 \times 36.2/2.07 + 0.73 \times 20.2} + 0.73 \times 1 \times \frac{20.2}{0.25 \times 36.2/2.07 + 0.73 \times 20.2} \\
&\approx 1.26
\end{aligned}$$

F.4.1. *Calibration with bias.* How does bias affect the quantification of welfare effects? This requires us to evaluate the size of $MI_\theta$, 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 linearization $\kappa_\theta(\Lambda) = \kappa_\theta \cdot \Lambda$. Therefore, evaluating the new $MVPF$ formulas requires taking a stance on what $\Lambda$ is. As discussed in Appendix F.3, qualitative evidence from municipalities suggests the percent change in $\Lambda$ due to to the Participation Act is an increase of 22.1%. Further, I assume that the Participation Act represented an absolute change in $\Lambda$ of 1 unit. Therefore, $\Lambda = 1/0.221 = 4.52$. For example, $\Lambda$ could represent number of hours spent on obligations, and $\kappa_\theta$ 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:

$$MI_L = (1 - \psi) \cdot 4.52 \cdot 161 = (1 - \psi) \cdot 728$$
$$MI_H = (1 - \psi) \cdot 4.52 \cdot 98 = (1 - \psi) \cdot 442$$

These estimates mean that we can quantify how large the $MVPF$ formulas are for different values of $\psi$. 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:

$$MVPF_{d\Lambda}^{\psi=0} = 0.30$$
$$MVPF_{dB}^{\psi=0} = 0.96$$

$MVPF_{d\Lambda}^{\psi=0} < MVPF_{d\Lambda}^{\psi=1}$ 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 $\psi^*$ 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 56% 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 $dB$.