Mental Health and the Targeting of Social Assistance

Canishk Naik

LSE

November 2024

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 65% 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.

London School of Economics, Department of Economics. c.naik@lse.ac.uk.

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 Sonia Bhalotra, 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, 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 seminar participants at the CPB, IFS, IIPF and LSE. 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 the exclusion of 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. Poor mental health affects *cost* of barriers as well as, in principle, *need* for support. The challenge is that both need and cost affect take-up, however separating need from cost is essential for assessing effectiveness. This is because barriers target well if the more needy can afford the cost and the less needy cannot. 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.[1] Empirically, measuring mental health at scale is challenging with survey data (due to small samples and under-reporting because of stigma (Bharadwaj et al., 2017)) but also with administrative data (extreme outcomes are unlikely and people with poor mental health often do not use care (Cronin et al., 2024)).

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 effects 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 barriers, either they have higher cost (sensitivity to barriers) 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) Similarly, large responses to changes in benefits reflect either high need (high *marginal* value) or average value net of cost ≈ 0, but the latter can be separated-out using difference in take-up levels.

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

[1]The exception is disability insurance. Here, Godard et al. (2022); Haller and Staubli (2024) emphasize that the key policy challenge with mental disorders is that they are hard for case-workers to observe.

[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).

transferring €1 from someone with good mental health to someone with poor mental health. The latter reflects the welfare costs that overcoming barriers imposes 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]

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 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

[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).

[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).

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 estimate the effect of the Participation Act (Ministerie van SZW, 2015) on the take-up of social assistance, a policy which increased access barriers by intensifying the obligations that recipients have to satisfy and increasing incentives for municipalities to restrict inflow (SCP, 2019). The act increased the compliance costs from obligations and the psychological costs from unpleasant interactions with the municipality (Ministerie van SZW, 2022). Using a difference-in-differences design, I show that the policy reduces receipt of social assistance by people with poor mental health relative to good. These results are driven exclusively by deterrence of inflow, which decreases by 10% vs baseline inflow for good mental health.

Third, people with poor mental health respond twice as much to a change in benefits than those with good mental health. Social assistance in the Netherlands tops-up income to an eligibility threshold, which implies a kinked benefits-schedule as a function of income (100% marginal tax rate below the threshold, 0% above). I exploit this kinked schedule using a regression kink design to identify the effect of benefit level on take-up, heterogeneously by mental health. I estimate elasticities of social assistance receipt with respect to benefits of 0.76 and 0.29 for poor and good mental health, respectively.

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 65% 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 effect of the policy divided by government cost, of a reduction in barriers as 1.88 and of an increase in benefits as 0.82. This implies reducing barriers is an effective use of government funds, $2.1\times$ more so than increasing benefits.

Contribution to the Literature: This paper contributes to the public economics literature on the targeting of government programs. Nichols and Zeckhauser (1982) argue that

ordeals, or costly barriers to take-up, can screen for need if the correlation between need and these costs is weakly negative. Some empirical studies support this, showing barriers effectively target aid (Alatas et al., 2016; Giannella et al., 2023; Wu and Meyer, 2023). Indeed, Rafkin et al. (2023) find that self-targeting is valuable from a social welfare perspective. However, targeting through ordeals can be undermined by take-up frictions among the needy (Deshpande and Li, 2019; Finkelstein and Notowidigdo, 2019; Homonoff and Somerville, 2021) or adverse selection (Shepard and Wagner, 2022). Generally in this literature, need is seen as arising from low-income.

I take a new approach and focus specifically on mental health. To my knowledge, this is the first paper to assess the welfare consequences of excluding those with poor mental health from assistance. Although the behavioural public policy literature has examined the descriptive relationship between mental health and the take-up of social programs (Arulsamy and Delaney, 2022; Bell et al., 2022; Martin et al., 2023a,b), welfare effects require an assessment of need. I show that people suffering from mental disorders need benefits more than those with good mental health, even conditional on having the same income. This highlights that vulnerability is often multi-dimensional and goes beyond poverty. My results show that despite their higher need, people with poor mental health are disproportionately screened-out by barriers, implying inefficient targeting.

The idea that mental disorders both increase need and make it harder to navigate barriers to assistance 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 develops a theoretical framework to discipline these opposing forces and implements the identification empirically using rich administrative 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). In economics, studies demonstrate that mental healthcare interventions, such as psychotherapy and

mindfulness, 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) and how income impacts mental health (Christian et al., 2019; Miller et al., 2024; Schmidt et al., 2021; Silver and Zhang, 2022).

I quantify the policy relevance of the cognitive and emotional burdens that mental disorders impose on individuals by empirically estimating the welfare costs of ordeals for these people. Moreover, I use a revealed-preference approach as in Deshpande and Lockwood (2022); Haller and Staubli (2024) 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 meet 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.

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). I allow for heterogeneous marginal value of €1 (need), even across individuals with the same income or consumption. This generalisation is motivated by the vulnerability of people with poor mental health going beyond their risk of poverty. I propose a method for separately isolating need from the cost of overcoming barriers using take-up responses to changes in benefits

and barriers.[7] The framework yields empirically-implementable formulas for the welfare effects of targeting. Proofs and extensions are in Appendix A.

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

2.1.1. *Setup.* Individuals are indexed by θ.[8] Social assistance is defined by two policy parameters. B is the (monetary) benefit, Λ is the barrier that individuals have to overcome to receive B. Each θ 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 Λ. 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 θ 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 psychological factors and are reflected by behaviour. Given the limited evidence on welfare effects for individuals with poor mental health, simplicity is crucial. As such, I minimise structural assumptions and focus on identifying the key statistics that are sufficient for policymakers to assess targeting effectiveness.

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

[8] In my empirical setting, θ 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.

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 τ.

2.2. **Welfare.**

2.2.1. *Individual Welfare.* Denote U_θ as θ's utility (which depends on take-up), and $\mathcal{U}_\theta$ expected utility. Following the setup in Section 2.1.1, I normalise utility to 0 if $SA = 0$.

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

where $\varepsilon_\theta^* = v_\theta(B) - \kappa_\theta(\Lambda)$. Importantly, this formulation assumes rationality.[9]

2.2.2. *Social Welfare.* Let $\mu(\theta)$ be the distribution of types, and λ_θ 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] τ is also fixed. The government does not observe individuals' private types (θ, ε), making targeted policy design challenging. Instead, it must rely on blunt instruments—benefit levels (B) and barriers to access (Λ)—which do not vary by θ to indirectly target those most in need. The policy-maker's goal is to allocate benefits to

[9] See Section 2.3.2 for a discussion of all key assumptions.
[10] I discuss how to explicitly model eligibility in detail in Appendix A.

individuals with a high, unobservable marginal value of benefits ($v'_\theta(B)$, i.e., need). Barriers (Λ) 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.

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 Λ (B adjusts).

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

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

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 θ's with the highest need (v'_θ large) and *lowest* ordeal-costs (κ'_θ 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_θ is the fiscal externality of θ applying:

[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."

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 θ.

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*, $\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}$.

$\frac{dW}{d\Lambda}$ is my overall metric for the social welfare consequences of targeting using ordeal mechanisms. However, the units are hard to interpret. 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 θ's having different marginal values of income unless social welfare weights are included. This is why I derive both formulations.

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'_θ), cost (κ'_θ), benefit take-up effects $\left(\frac{\partial \mathbb{P}[SA]_\theta}{\partial B}\right)$ and barrier screening effects $\left(\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}\right)$.[12]

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'_θ). Similarly, barrier screening effects $\left(\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}\right)$ depend on the cost of barriers (κ'_θ).

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

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

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

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

Intuitively, Λ is a price of taking up. Therefore, responsiveness to take-up is large when consumers are price-responsive (κ' 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 extrapolation term allowing for the inference of infra-marginal costs/benefits through marginal take-up responses.

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'_θ), cost (κ'_θ) 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_θ) 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 κ' 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: $\mathbb{P}[SA]_\theta = \mathbb{P}[SA]_{\tilde{\theta}}$.[13] 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

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

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 θ_0 we calculate v'_θ for all other θ using Equation (2.10). This normalization is without loss, and effectively scales all welfare effects in terms of θ_0's WTP for €1.[14]

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 κ'_θ for all θ as follows:[15]

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

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 ε 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.[16] Without independence, the model is not identified and either v_θ constant across θ or κ_θ constant

[14] v'_θ is then understood as θ's need relative to θ_0.

[15] This within-type identification method is as in Haller and Staubli (2024), but the across-type identification 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.

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

across θ 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.[17] Additivity allows me to separate need/cost from the extrapolation term $f_\varepsilon(v_\theta - \kappa_\theta)$ 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 θ and ε 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 E shows welfare effects in this case.

In the theory, θ 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.

Finally, I use a revealed-preference framework. Assuming rationality allows me to reveal need and cost from take-up responses, and to use the Envelope Theorem when deriving welfare effects. In Section 7, I follow the framework developed by Naik and Reck (2024) to characterise how confident the government needs to be about bias to reverse the estimated sign of the welfare effects.

The rest of the paper is devoted to evaluating the key sufficient statistics and discussing 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.

[17] 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.

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.[18] 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

[18]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.

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.[19]

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.[20] 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.

[19]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.

[20]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. **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.

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.[21] 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.[22] 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.

Mental health information: Finally, the data contain three classes of mental health measures: take-up of mental healthcare (mental healthcare expenditures and dispensations of

[21] 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.

[22] As in Inspectie SZW (2021), eligibility for students and people above pension-age is noisier.

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).

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.

This suggests that there does not seem to be substantial self-targeting. Otherwise, the more vulnerable group would be over-represented among the recipients. It would appear then that those with mental disorders face additional challenges accessing benefits.

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

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.

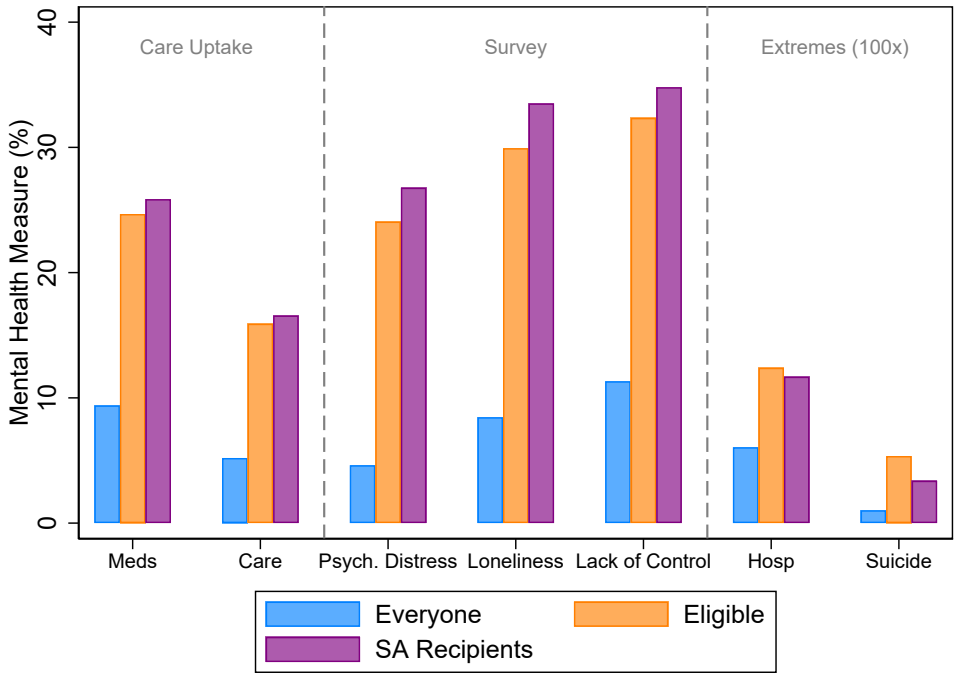


FIGURE 1. Prevalence of poor mental health in the Netherlands. This graph shows raw means of the seven mental health measures, across three different populations. All measures are in terms of percentages and are probabilities of the following: dispensed psychotropic medications, > 0 mental healthcare 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.

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.

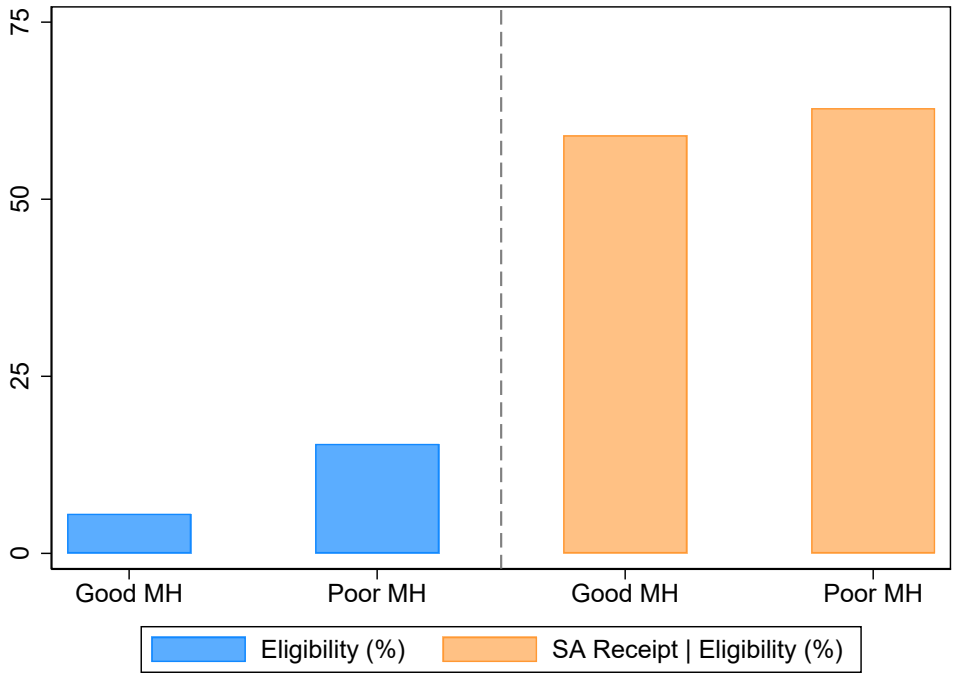


FIGURE 2. $\mathbb{P}[\text{Eligible}]$ and $\mathbb{P}[SA|\text{Eligible}]$, compared for people with poor mental health (dispensed psychopharma in year previously) vs good mental health (not). Underlying population: 2011-2020 in each case.

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 $SA_{it} = \mathbb{1}\{i \text{ receives SA in year } t\}$. Poor $\text{MH}_{it} = \mathbb{1}\{i \text{ dispensed psychopharma. in year } t\}$. Equation (4.1) presents a specification to answer this question.[23]

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

X_{it} are flexible controls of income,[24] wealth, education, hh composition, work status, work sector and year, age, gender and municipality fixed effects. ε_{it} is an idiosyncratic error term. β measures the selection of social assistance recipients with respect to mental health and is the coefficient of interest.

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

[24]I include household standardised income percentile (moving average $t-4 \to t-2$) fixed effects.

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.[25]

4.2. **Results.** Table 1 shows the main results of the correlation test, using Poor $\text{MH}_{it} = \mathbb{1}\{i \text{ dispensed psychopharma. in year } t\}$. The table shows estimates of β, the association between lagged mental health and receipt of social assistance for various regression specifications. First, I show raw differences in take-up between good and poor mental health. Column 2 shows the main specification - which additionally includes lagged controls for income, wealth, education, work status, hh composition and municipality sector FEs. The estimates indicate that people with poor mental health receive social assistance slightly more than people with good mental health (≈ 1 p.p. is about 1.5% of the baseline mean $\approx 60\%$). These estimates are statistically significant, but economically small. Adding a rich set of control variables does not change the association substantially.

	Raw	**Main Spec.**
$\hat{\beta}$: Receipt of SA, poor vs good MH (p.p.)	3.072***	0.978***
	(0.810)	(0.074)
Observations (people-years)	5,671,855	5,187,572
R^2	0.001	0.654
Baseline Mean	59.97	62.45
Regressors	49	356

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

TABLE 1. Results of a regression of receipt of social assistance on mental health status (measured by dispensation of psychotropic meds). The first column shows the results with no controls. The second column shows the main specification, which additionally includes lagged controls for income, wealth, education, work status, hh composition and municipality sector FEs. The sample contains the calculated eligible for SA in 2011-2020. Standard-errors are clustered at the municipality-level.

[25]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.

Figure 3 presents the results of the correlation test varying the measure of mental health used. $\hat{\beta}$ are plotted for each mental health status variable: $\mathbb{1}\{$dispensed of psychotropic meds$\}$, log 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$\}$. The estimates vary, but qualitatively speaking show a small positive difference in rate of receipt by people with poor mental health vs people with good mental health.

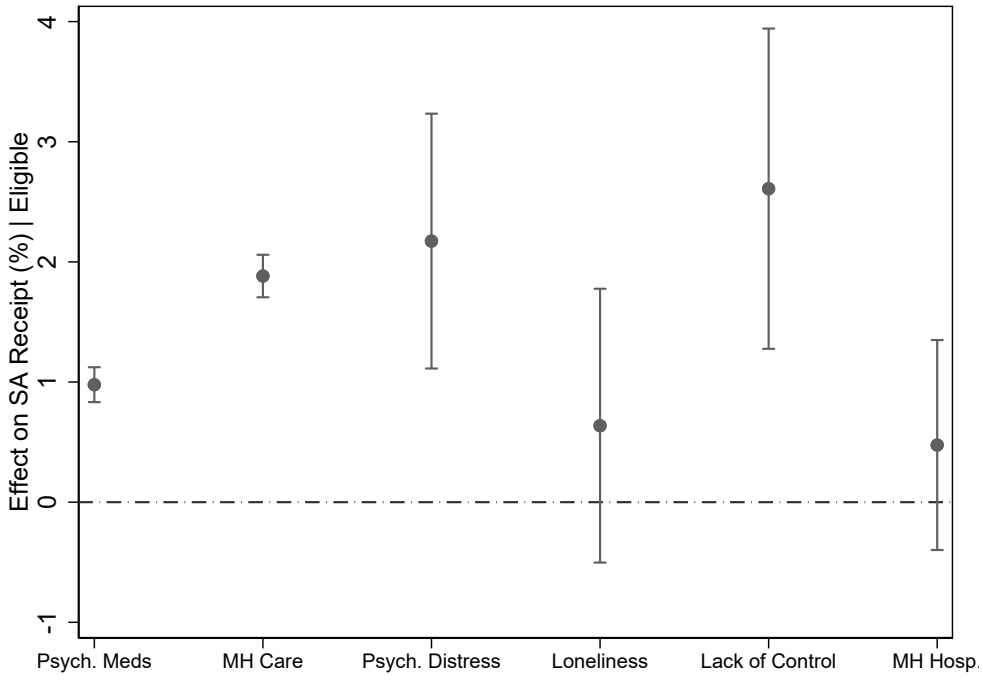


FIGURE 3. Results of a regression of take-up of social assistance on mental health status (measured by psychopharma prescription, log mental healthcare costs, being hospitalized for a mental health condition and three surveyed measures of subjective mental health). Lagged Controls: income, wealth, education, work status, hh composition and municipality, year, age and sector FEs. The sample contains the calculated eligible for SA in 2011-2020. Standard-errors are clustered at the municipality-level.

Some interesting heterogeneity appears when we split the psychopharma results by type. Figure 4 shows the results of a regression of SA receipt on psychopharmacology-type dummies and controls. People being prescribed ADHD, hypnotics and sedatives and anxiolytics are no more likely to receive social assistance than the mentally healthy.

Anti-depressant prescription is associated with a higher rate-of-receipt, whereas anti-psychotic prescription is associated with a *lower* rate-of-receipt.

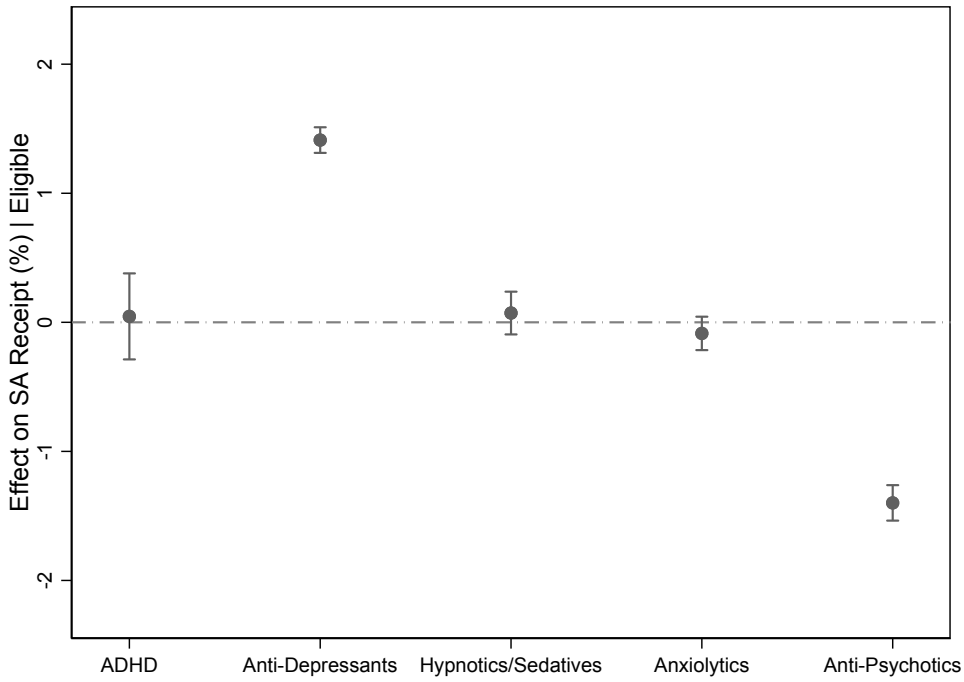


FIGURE 4. Results of a regression of take-up of social assistance on psychopharmacology prescription fixed effects (by type: ADHD medications, anti-depressants, hypnotics/sedatives, anti-anxiety medications and anti-psychotics). Controls: income, wealth, education, lagged work status, hh composition and municipality, year, age and sector FEs. In the plot, drugs are ranked by how much they predict suicide in the following year from ADHD (not predictive) to anti-psychotics (highly predictive). The sample contains the calculated eligible for SA in 2011-2020. Standard-errors are clustered at the municipality-level.

No strong evidence for selection either way potentially obfuscates two counteracting forces: people with poor mental health could value social assistance more, but at the same time find it more difficult to receive. The former would increase their rate-of-receipt, while the latter would decrease it. Recall from Section 2 that the correlation test is not a sufficient statistic for welfare in the case where another variable affects marginal utility apart from level-of-consumption (here, mental health). Separating need from ordeal-costs is required to evaluate the welfare consequences of targeting. Take-up responses to changes in benefits and barriers can achieve this.

5. Barrier Screening Effects

I examine the effects of a large reform to the design of social assistance in the Netherlands to assess the impact of ordeals on the receipt of social assistance by people with good and poor mental health. The reform was called the Participation Act (Ministerie van SZW, 2015). It was announced in 2014, and implemented starting 1st January 2015. Figure C.1 shows that the policy was being discussed in the public starting even in 2013, discussion starts rapidly increasing in 2014.

The Participation Act was an austerity policy passed in the context of steadily rising caseload since the Great Financial Crisis (see Figure B.1). In response, the municipal social assistance budget was reduced from 1,400 million euros in 2010 to about 500 million euros from 2018 (Heekelaar, 2021). The Participation Act increased barriers to access. The policy intensified the obligations that recipients had to satisfy (SCP, 2019) and incentivised municipalities to restrict inflow through the (threat of) sanctions (van der Veen, 2019).

The policy represents an exogenous increase in barriers for middle-age couples. I look at couples because the eligibility threshold for single parents was cut in 2015, and incentivised single parents to re-classify as single households.[26] I restrict to middle-age because people who would have been eligible for the young-person's disability benefit (Wajong) became eligible for social assistance in 2015, which could affect take-up differentially by mental health. $< 3\%$ of the inflow into Wajong in 2014, who would be eligible for social assistance from 2015 on, is above the age of 35 (en Werkgelegenheid, 2017). Sensitivity analyses show that these assumptions and restrictions do not drive the results. All robustness checks are in Appendix C.

5.1. **Identification.** I use the ordeal variation generated by the Participation Act to identify the statistic $\frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda}$ which is necessary for evaluating welfare. This take-up response is the heterogeneous treatment effect by baseline mental health to a change in barriers.

The Participation Act affected everyone - there was no control group. However, because we are only interested in the heterogeneous treatment effect by mental health, a

[26] There was no incentive to become a couple after 2015, and the results are robust to restricting to a sample of households who were classified as couples throughout 2011-2020.

control group is not necessary under the strong parallel trends assumption 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).

Therefore, the analysis can be seen as a Difference-in Difference with people with poor mental health as the treatment group. The interpretation of the treatment effects is the heterogeneous effect $\frac{\partial \mathbb{P}[SA]_L}{\partial \Lambda} - \frac{\partial \mathbb{P}[SA]_H}{\partial \Lambda}$. The identification assumption is that people with poor mental health's receipt would have evolved in parallel to those with good mental health.

The main specification I use is given by Equation (5.1). η_i and γ_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. δ_t for $t \geq 2014$ are the coefficients of interest and represent the heterogeneous treatment effect of the policy by baseline mental health Poor MH_i. Poor MH_i is defined as having psychopharma dispensed at some point in the pre-period (2011 - 2014). Throughout, I cluster standard-errors at the level of municipality of residence in 2013.

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

I estimate Equation (5.1) for eligible middle-aged (45-65) couples, for whom the policy is an exogenous increase in barriers. I focus on the eligible because the take-up responses for this group can be attributed to the change in barriers, and not underlying changes in eligibility.[27]

5.2. **Main Results.** Figure 5 shows the results of the estimation according to Equation (5.1). The groups are on parallel trends before the policy announcement.[28]

Then, the rate of receipt by people with poor mental health decreases relative to that of people with good mental health in response to the Participation Act. The effect starts

[27]The effect holds also for those people which are "always-eligible" (eligible throughout 2011-2020) bringing confidence that the main results are not driven by eligibility churn. Figure C.6 shows the results when including also adults aged 35-45. This group is contaminated by the Wajong entrants to a greater extent, but have similar results, suggesting the estimates of the main specification are not driven by Wajong entrants.

[28]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.

when the Act is announced, and then is especially pronounced in 2015. The overall difference-in-difference estimate of $\frac{\partial \mathbb{P}[SA]_L}{\partial \Lambda} - \frac{\partial \mathbb{P}[SA]_H}{\partial \Lambda} \approx -1$p.p. This is comparable in magnitude (but opposite sign) to Finkelstein and Notowidigdo (2019) who estimate $\frac{\partial \mathbb{P}[SA]_L}{\partial \Lambda} - \frac{\partial \mathbb{P}[SA]_H}{\partial \Lambda} = 2$p.p. for an information + assistance treatment in the context of SNAP in the US.

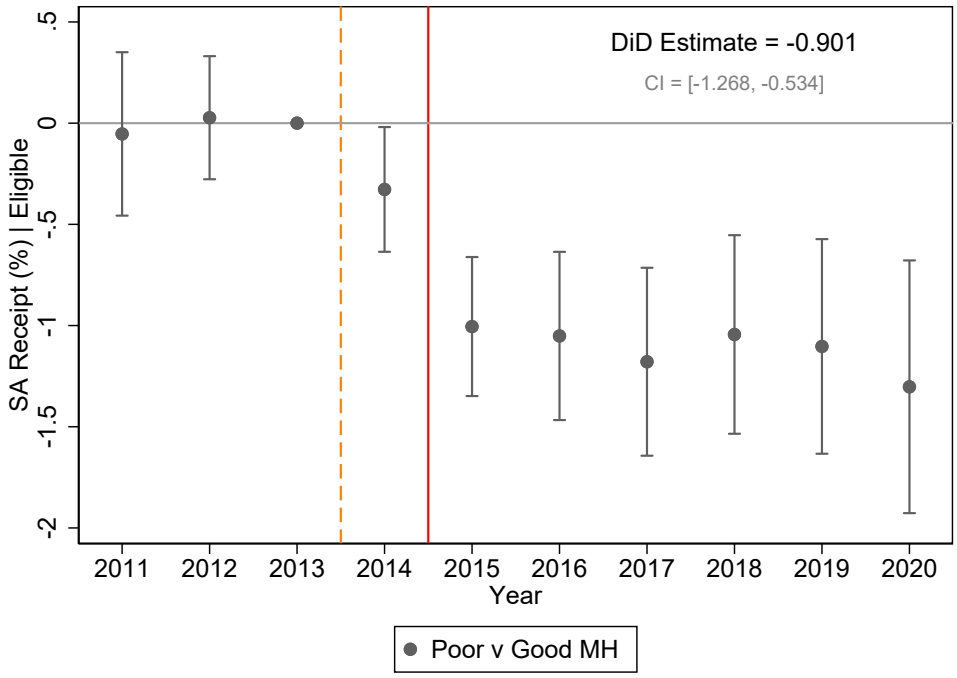


FIGURE 5. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in pre-period. Controls include individual fixed effects, income, education and muncipality, hh composition and sector fixed effects. The TWFE estimate $\hat{\delta}$ in the regression $SA_{it} = \alpha + \eta_i + \gamma_t + \delta \cdot \mathbb{1}\{t \geq 2014\} \times \text{Poor MH}_i + X_{it}'\theta + \varepsilon_{it}$ is also shown. Standard-errors are clustered at the level of municipality of residence in 2013.

Figure 6 shows the effects on inflow and outflow. The estimates on drop-out are not mechanical - they are just indistinguishable from 0 and have tight confidence intervals. The effects show that the disproportionate screening-out of people with good mental health comes exclusively from a deterrence of inflow. These results align with Cook and East (2024) who suggest work requirements can screen-out individuals at the extensive margin

in the US. The disproportionate reduction in inflow for people with poor mental health (1p.p.) is around 10% of the baseline control mean (see Figure C.2).

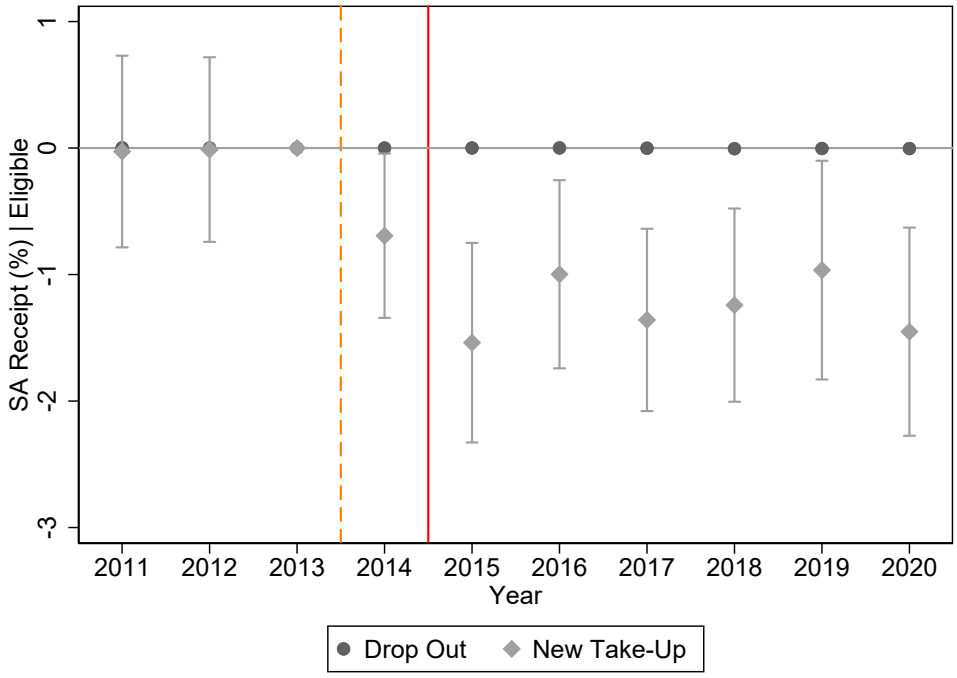


FIGURE 6. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. Here, I split by inflow (receipt conditional on being ineligible last period), and drop-out (non-receipt conditional on receipt 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. The TWFE estimate $\hat{\delta}$ in the regression $SA_{it} = \alpha + \eta_i + \gamma_t + \delta \cdot \mathbb{1}\{t \geq 2014\} \times \text{Poor MH}_i + X'_{it}\theta + \varepsilon_{it}$ is also shown. Standard-errors are clustered at the level of municipality of residence in 2013.

This corroborates SCP (2019) who find no effect of the Participation Act on outflow into paid work - I find no effect on outflow into non-take-up. All of the effect comes from reduced inflow. This aligns with the qualitative evidence of Ministerie van SZW (2022), who state that the main effect was a "scaring" effect by the municipality which reduces inflow. The authors describe the issue as follows.

> "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."[29]

5.3. **Different Mental Health Measures.** Figure 7 shows the heterogeneous barrier screening effects where I split by moderate and serious mental illness. The rate-of-receipt both by people on anti-depressants and anti-psychotics falls relative to the good mental health - the DiD estimate for anti-psychotics is more than double that for anti-depressants.

Figure 8 shows the heterogeneous barrier screening effects for different definitions of poor mental health: dispensations of psychotropic drugs in pre-period, > 0 mental healthcare costs in pre-period, surveyed severe psychological distress in 2012. For each measure, receipt for people with poor mental health falls relative to that of good mental health. The effects are similar for those using mental healthcare and those dispensed psychotropic drugs. People with surveyed severe psychological distress are screened out even more. This likely reflects the fact that the main estimates are a lower-bound since some mental disorders go un-diagnosed.

Taking Figures 7 and 8 together, it seems the disproportionate screening-out of an increase in barriers is more intense the more severe the underlying mental disorder.

5.4. **Robustness.** I undertake a series of robustness checks to support the results in Section 5.2. First, the sample population here is the couples who are eligible for social assistance in each year. This population is changing over time as people churn in and out of eligibility with income changes, as well as could be affected by the introduction of the

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

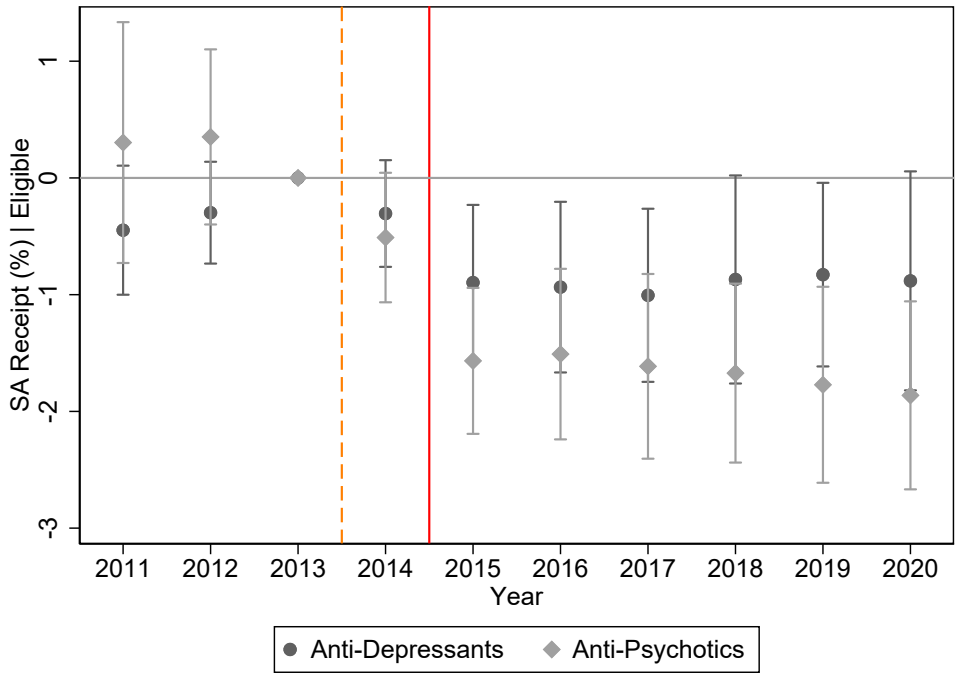


FIGURE 7. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. Here, Poor MH$_i$ can now take 3 values: 0 (control), 1 (anti-depressants) or 2 (anti-psychotics). The analysis population is eligible 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 $\hat{\delta}_t^{\text{Psycho.}}$. The TWFE estimate $\hat{\delta}^{\text{Dep.}}$ and $\hat{\delta}^{\text{Psycho.}}$ and in the regression $SA_{it} = \alpha + \eta_i + \gamma_t + \delta \cdot \mathbb{1}\{t \geq 2014\} \times \text{Poor MH}_i + X_{it}'\theta + \varepsilon_{it}$ is also shown. Standard-errors are clustered at the level of municipality of residence in 2013.

cost-sharing standard or inflow of people with youth disabilities. Therefore, we might be worried that the main result is being driven by differential take-up rates of the new entrants/exiters into eligibility by baseline mental health.

In order to control for this confounding factor, I first focus on the always-eligible population. This is couples who remain eligible throughout the sampling period. While this may seem like a stark restriction, 25% of the eligible are always-eligible. Figure C.3 shows the results of the estimation of Equation (5.1) on the always-eligible couples. The results are similar - suggesting that the overall targeting results are not driven by differential selection of the new entrants into / leavers from eligibility by baseline mental health. The

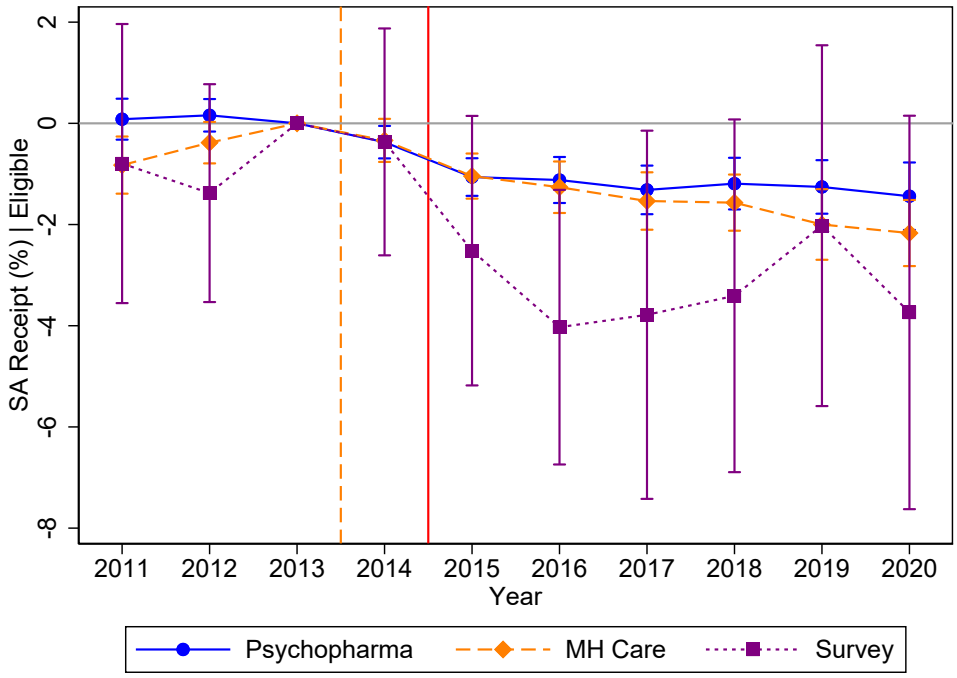


FIGURE 8. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. Here, Poor MH$_i$ is defined in 3 ways: dispensations of psychotropic drugs in pre-period, > 0 mental healthcare costs in pre-period, 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. I plot the estimate $\hat{\delta}_t^{\text{Dep.}}$ and $\hat{\delta}_t^{\text{Psycho.}}$. The TWFE estimate $\hat{\delta}^{\text{Dep.}}$ and $\hat{\delta}^{\text{Psycho.}}$ and in the regression $SA_{it} = \alpha + \eta_i + \gamma_t + \delta \cdot \mathbb{1}\{t \geq 2014\} \times \text{Poor MH}_i + X_{it}'\theta + \varepsilon_{it}$ is also shown. Standard-errors are clustered at the level of municipality of residence in 2013.

point estimates being smaller for the always-eligible is to be expected - as these people are less likely to be on the margin of take-up.

More formally, this 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 (5.2).

$$\mathbb{E}[\varepsilon_{it} \big| \mathbf{X}_i, \eta_i, \mathbf{e}_i] = 0 \tag{5.2}$$

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. However, Figure C.4 shows that the results are virtually unchanged when removing all time-varying covariates from the estimation.

Secondly, at the same time as the introduction of the Participation Act, the government of the Netherlands implemented a reform in the structuring of long-term care (WMO) (Kromhout et al., 2018). The change which is relevant to my setting is that the remit of home support for people with mental health issues was changed to be under the remit of municipalities starting in 2015. Figure C.5 shows the WMO reform does not drive the results. Controlling flexibly for receipt of WMO care does not affect the estimates.

Thirdly, the Participation Act could have affected people with poor mental health and good mental health differently because of its differential implementation. The policy introduced a new way of calculating eligibility depending on how many people lived in the applicants' address, which could have affected people with poor mental health differently. I control flexibly for household composition and size, thus controlling for these eligibility changes, but Figure C.4 shows that this does not drive the results.

The fourth threat to identification is that the heterogeneous treatment effects that we observe are not due to differences in baseline mental health - but instead due to different levels in the pre-period. It could be that the policy affected groups which take-up at a higher rate differently and the difference we see has nothing to do with mental health.

On the contrary, when we split the poor mental health group into two subgroups - moderate mental illnesses (those receiving anti-depressants and anti-anxiety medications) and severe mental illnesses (those receiving anti-psychotics), we see that both groups' take-up rate diverges from the good mental health group after 2015. Further, prior to 2015, the severely poor mental health group had similar absolute levels of receipt to the good mental health group.[30] This is in line with my hypothesis that mental health differentially affects

[30]Recall the descriptive results - people with poor mental health take-up *more* than people with good mental health, but people with poor mental health take-up *less* than people with moderately poor mental health.

the response to ordeals and contradicts the hypothesis that differences in responses are driven by differences in levels.

Finally, because the groups are defined based on pre-period prescription of psychopharmacology, we might be worried that the $\hat{\delta}$'s are capturing the treatment effect of psychopharma on social assistance receipt. Figure C.7 shows that the results are not driven by the treatment effect of mental health care on social assistance receipt - Poor MH_i defined by surveyed poor mental health produces similar (stronger, if anything) results. Indeed, Figure C.8 shows the evolution of subjective mental health around a prescription event. Subjective mental health is significantly worse amongst people who are prescribed psychopharma 3 years before *and* after prescription, and throughout is above the threshold for moderate mental ill-health ($K10 \geq 20$).

A related concern would be that the results could be interpreted as the long-term effects of a mental health shock. Figure C.9 shows the effects, varying the year in which prescription of psychopharma defines mental health status. In each year, even when mental health is defined after the policy, there is a drop in $\hat{\beta}$ at 2015. Taking these graphs together, I interpret being prescribed psychopharma *at some point* as an indicator of mental health status, which improves, but is still significantly worse than the good mental health group throughout the study period.

Proposition 5.1 (Partial Identification of Value and Cost)**.** *The effect of ordeals on targeting, alongside the raw take-up rates, is enough to partially identify value and cost, as long as* $\varepsilon \perp \theta$.[31] $\mathbb{P}[SA]_L \approx \mathbb{P}[SA]_H$ *and* $\frac{\partial \mathbb{P}[SA]_L - \mathbb{P}[SA]_H}{\partial \Lambda} < 0$ *imply that people with poor mental health have a higher dis-utility of ordeals in receipt.*

$$\kappa'_L(\Lambda) > \kappa'_H(\Lambda) \tag{5.3}$$

Under a linear functional form for κ_θ*, this implies people with poor mental health also have a higher average value for the benefit.*

$$v_L(B) > v_H(B) \tag{5.4}$$

[31] Note that, the fact that the time-varying controls included in the main specification do not affect the results supports this assumption.

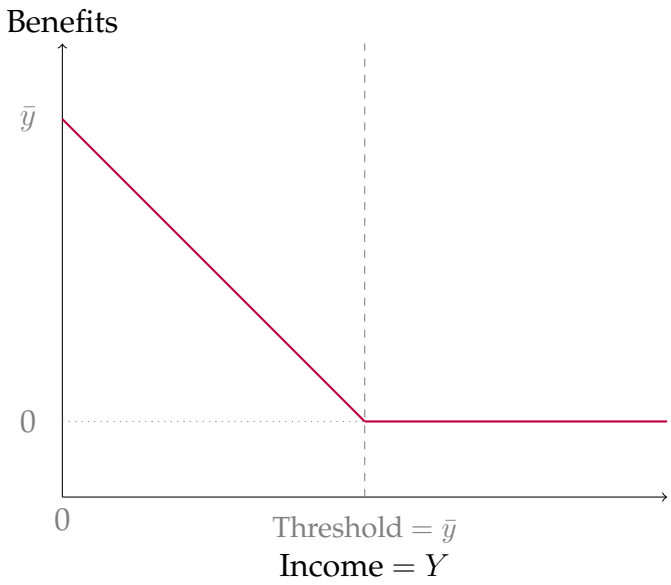


FIGURE 9. Benefits schedule as a function of income

Proof. See Appendix A. □

Proposition 5.1 shows that the results so far are sufficient to identify the sign of the covariance between need and cost. My estimate suggest need is positively correlated with cost. However, exact quantification of heterogeneous value and cost of ordeals requires another take-up response - to change in benefit level.

6. BENEFIT TAKE-UP EFFECTS

In order to separately identify need and ordeal-costs, I also estimate the effect of increases in benefit-level on take-up. I leverage quasi-experimental variation in the benefit-level coming from the kinked benefits schedule as a function of income, using a fuzzy regression kink design (RKD) as in Card et al. (2015). The statutory benefits schedule is displayed in Figure 9.

Before diving into identification, Figure 10 shows non-parametric evidence of the behavioural response to a change in benefit level by poor vs good mental health.

Why do people with income above the threshold take-up? This is primarily due to measurement error. Some sources of income do not count towards the calculation of eligibility $Y \leq \bar{y}$ (Ministerie van SZW, 2015). Therefore, I need to calculate the income concept used to determine eligibility. This Y_{calc} differs from Y_{true} because (a) some income

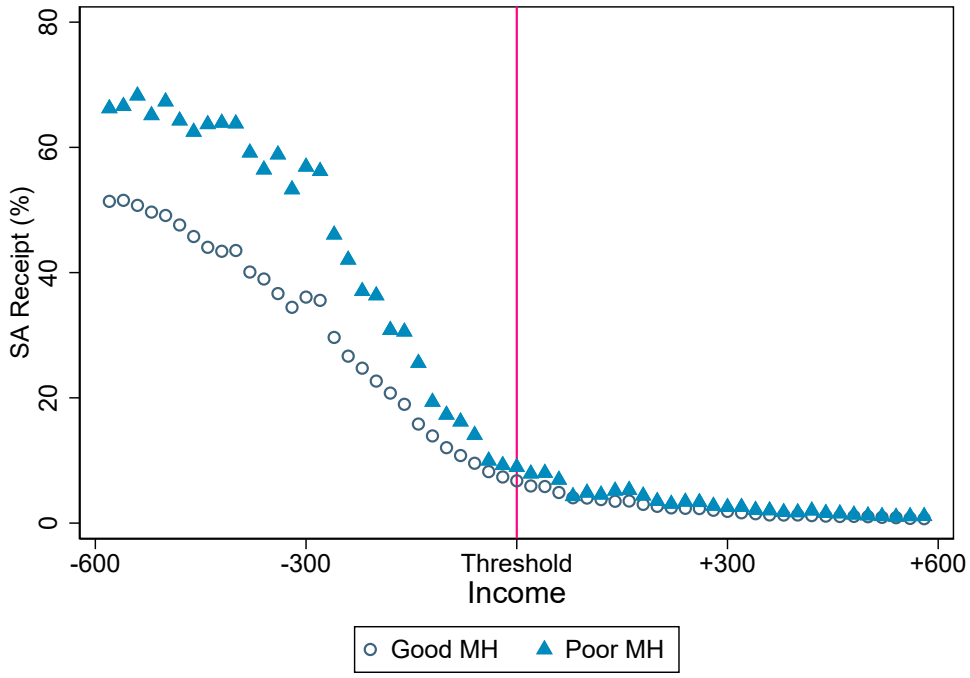


FIGURE 10. Average rate of receipt within income slice in a large window of income either side of the eligibility threshold. Income in this plot is monthly. Poor mental health is defined as receiving psychopharma in the year previously. The sample contains singles who get most income from labour, years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

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. Recall that familes are 1 or 2 adults (+ kids) who live together and share costs. Cost-sharing is unobservable / subject to case-worker discretion. (c) students are ineligible if they can claim student financing. This is unobserved, but I focus on working age individuals for this reason.

Y_{true} is observed for the selected sample: social assistance recipients. This means that I can observe measurement error for this selected sample. Figure D.1 shows a histogram of $Y_{\text{true}} - Y_{\text{calc}}$ for the analysis population of the RKD.[32] Y_{true} is of course negatively selected for recipients, so we expect the distribution to be left-skewed. However, both the mean and median are small (-29EUR, 7EUR respectively).

[32]For more details, see Appendix D.

Figure 10 shows that for incomes above the threshold, people suffering from mental disorders receive at roughly the same rate as those without, suggesting the extent of measurement error in eligibility income, as well as the probability of exemption is similar across mental health types. Below the threshold, people with poor mental health take-up more than those with good mental health within every income slice. The take-up functions diverge at the threshold. This suggests that $\frac{\partial \mathbb{P}[SA]_L}{\partial B} > \frac{\partial \mathbb{P}[SA]_H}{\partial B}$. They keep diverging up to around €300 below, at which point they start to converge.[33]

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

6.1.1. *Theory.* Figure 10 measures $\frac{d\mathbb{P}[SA]}{dy}$. 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. Income is exempted "insofar as, in the judgment of the [municipality], it contributes to [their] employment opportunities" (Ministerie van SZW, 2015). Appendix D.1 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 ν captures noncompliance with policy formula due to exemptions.

To properly re-scale the reduced-form estimates, we need to know how B depends on Y ex-ante. However, there is selection into social assistance with respect to exemptions. This makes sense because applicants receive more money with an exemption vs without, holding income fixed. Figure D.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

[33]This explains why take-up conditional on eligibility can be similar in levels for poor versus good mental health despite the difference in slopes. Note that away from the threshold, the slopes cannot be interpreted causally.

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.[34] 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 D.1. 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 D.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.

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 6.1. *(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)

$$\frac{\lim_{\xi \to \bar{y}^+} \frac{d\mathbb{P}[SA|Y^*=y^*]}{dy^*}\Big|_{y^*=\xi} - \lim_{\xi \to \bar{y}^-} \frac{d\mathbb{P}[SA|Y^*=y^*]}{dy^*}\Big|_{y^*=\xi}}{\lim_{\xi \to \bar{y}^+} \frac{d\mathbb{E}[B^*|Y^*=y^*]}{dy^*}\Big|_{y^*=\xi} - \lim_{\xi \to \bar{y}^-} \frac{d\mathbb{E}[B^*|Y^*=y^*]}{dy^*}\Big|_{y^*=\xi}} = \int \frac{\partial \mathbb{P}[SA \mid B = b(\bar{y}, v), Y = \bar{y}, \varepsilon = e]}{\partial B} \cdot \varphi(e, v)\, dF_{\varepsilon,\nu}(e, v) \tag{6.1}$$

[34] 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.

where weighting function

$$\varphi(e,v) = \frac{\mathbb{P}[U_Y = 0|Y = \bar{y}, \varepsilon = e, \nu = v]\big(b_1^+(v) - b_1^-(v)\big)\frac{f_{Y|\varepsilon=e,\nu=v}(\bar{y})}{f_Y(\bar{y})}}{\int \mathbb{P}[U_Y = 0|Y = \bar{y}, \varepsilon = e, \nu = \omega]\big(b_1^+(v) - b_1^-(v)\big)\frac{f_{Y|\varepsilon=e,\nu=\omega}(\bar{y})}{f_Y(\bar{y})}dF_\nu(\omega)} \tag{6.2}$$

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

The Card et al. (2015) identification assumptions are stated in full in Appendix D.2. Two are key to my setting. (a) the density of Y^* is continuously differentiable at the threshold $\bar{y}$, (b) the benefits-schedule is continuous $\implies$ $\mathbb{P}[\text{Exemption}|Y = y]$ continuous at $\bar{y}$.

Figure D.6 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 D.2 exhibits no such discontinuity. Moreover, there are no conditions in the law which state income below/above the threshold should be exempted differently.

6.1.2. *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.[35] The non-regularized CCT bandwidth delivers a 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.

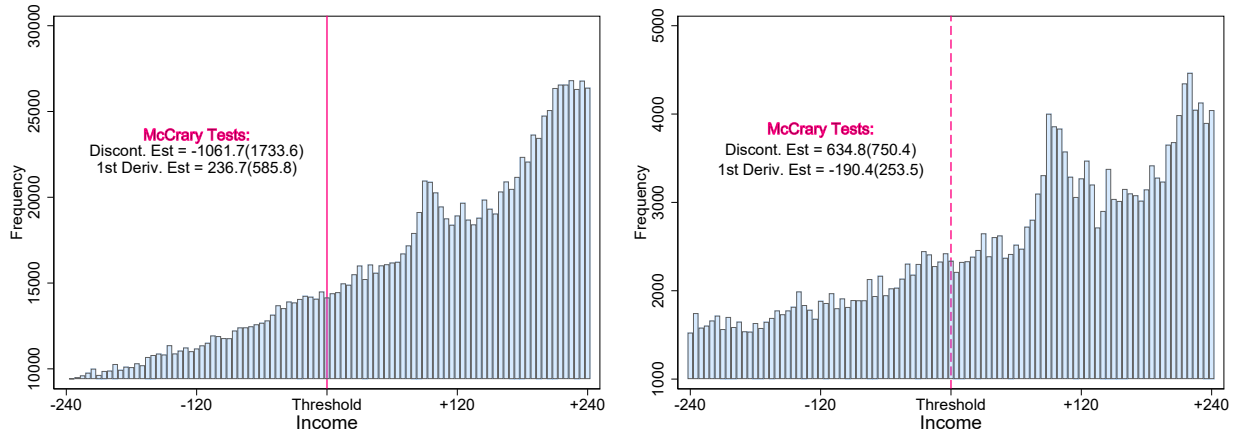
$$SA_{it} = \alpha + \beta_0 \cdot (y^*_{it} - \bar{y}_i) + \beta_1 \cdot \min\{y^*_{it} - \bar{y}_i, 0\} + \varepsilon_{it} \qquad \textbf{(Reduced Form)}$$

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

Support for Identification Assumptions: The key identification assumption is that there is no manipulation of income around the threshold. Figure 11 and Figure D.6 show that there is no evidence of lack-of-smoothness of the pdf of income around the eligibility threshold. Note that although the threshold equals the full-time monthly minimum wage, no bunching occurs, as the sample works much less than full-time (around 100 hours per month on average) and income used for eligibility comes not only from labour.

I assess the credibility of the design with standard robustness analyses whose results are described in Appendix D.4. Figure D.7 shows no strong evidence of selection along observable characteristics around the kink. While there is statistically distinguishable selection for poor mental health, Table D.1 shows that the addition of a rich set of covariates does not meaningfully affect the results. Figure D.10 displays a permutation test (Ganong and Jäger, 2018), and shows no evidence for worrying non-linearities. Figure D.10 and Figure D.11 explore sensitivity of the results to different bandwidths. Estimates are quite

[35] Indeed, the CCT robust bandwidth without regularization performs poorly in simulations (see Appendix D.3).



(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 in this plot is monthly. Poor mental health is defined as receiving psychopharma in the year previously. The sample contains singles who get most income from labour, years 2011-2014. See Section 6.1.2 for details on sample restrictions.

robust to lower bandwidths overall, and point estimates do not vary much in the heterogeneous case despite the confidence intervals overlapping with lower bandwidths.

6.2. **Results.** Figure D.12 shows overall (good and poor mental health together) reduced-form estimates. The estimated $\hat{\beta}_1 = -0.0338$ which translates to take-up going up by $\approx 3.4 p.p.$ for a €100 increase in the benefit level. There is no statistically distinguishable bunching around the threshold, which corresponds to adjustment frictions driving a large proportion of the choices of income around the threshold (Kleven, 2016).

The partial identification result in Section 5 suggests that people with poor mental health need the benefit more. I test this directly by performing the RKD heterogeneously by baseline mental health (measured again by prescription of psychopharma). This is shown in Figure 12. People with poor mental health react more in their rate-of-receipt to a change in benefit level - $(\hat{\beta}_{1H}, \hat{\beta}_{1L}) = (0.0263, 0.0819)$. 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.

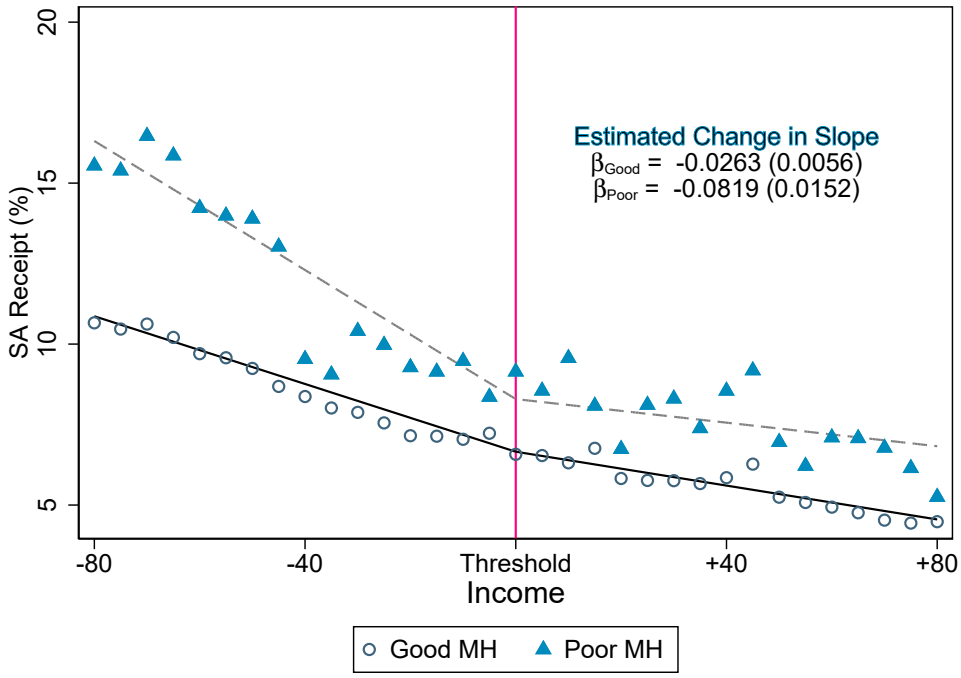


FIGURE 12. Average rate of receipt within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly. Poor mental health is defined as receiving psychopharma in the year previously. The sample contains singles who get most income from labour, years 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 D.3 shows the raw first stage, heterogeneously by mental health. Using these to re-scale the above reduced-form, we obtain IV estimates $\left[\frac{\hat{\beta}_1}{\hat{\delta}_1}\right]_{\text{Poor MH}} = 0.824$ and $\left[\frac{\hat{\beta}_1}{\hat{\delta}_1}\right]_{\text{Good MH}} = 0.280$.

As shown in Figure D.8, part of the reason why we might have such a stark difference between poor and good mental health is due to selection on other observables. Therefore, for the main results, I include controls for gender, age, year, month, foreign-born, household composition, education, municipality, work status, wealth and sector. These controls do not meaningfully change the results - even after controlling, I estimate people with poor mental health to react twice as much to a change in benefits than those with good mental health. My IV estimates are as follows.

$$\frac{\partial \hat{\mathbb{P}}[SA|Y=\bar{y}]}{\partial B} = \begin{cases} \underset{[0.27,0.029]}{0.028} \text{ p.p.} & \text{for Good MH} \\ \underset{[0.037,0.093]}{0.065} \text{ p.p.} & \text{for Poor MH} \end{cases} \tag{6.3}$$

with an implied elasticity of 0.29, 0.76 respectively. These elasticities lie at the lower (/upper) end of range of previously estimated take-up elasticities of social insurance for good (/poor) mental health respectively (Krueger and Meyer, 2002; McGarry, 1996). The elasticity for poor mental health is more than double that of good mental health. The full set of reduced-form and IV estimates (with and without controls) are contained in Table D.1.

6.3. **Mechanisms.** Why do people with poor mental health react more to a change in benefits, conditional on having the same income. I show in Section 7 that this comes from 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 psychopharma dispensations on to estimate $\frac{\partial MH}{\partial SA}$. Figure E.1 shows null results. This is perhaps not surprising - (Solmi et al., 2022) find that many mental illnesses start early in life - before people might enter my analysis sample.

Alternatively, psychology literature studying mental disorders often refer 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 by making it harder to regulate emotions - this can affect work, relationships and self-image (Gross and Muñoz, 1995).

These psychological theories can rationalize the results of people with poor mental health value money more, despite it not having any real protective effect. The cognitive burden and particularly the tax on emotional-resilience of mental disorders make it more difficult handle stressors which are common amongst this low-income population, and so increase the value of support. 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).[36]

7. Calibration of Welfare Effects

Finally, I quantify the welfare consequences of social assistance targeting with respect to mental health. I use the empirical results of Section 4, Section 5 and Section 6 to calculate how need and cost vary by mental health. These key primitives are important determinants to the effectiveness social assistance targeting using barriers. For example, in this section 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'_θ), cost (κ'_θ), 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 E.1 shows how to infer population take-up levels and responses using the eligible population.

7.1.1. *Identifying Need and Cost.*

I employ the three-step identification method set out in Section 2.3.1. Appendix E.2 shows the full set of results of this calibration. First, Section 4

[36] A simple 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.

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.

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

Thus, normalizing $v'_H = 1$ yields the extrapolation term - likelihood of being at the margin of take-up - for everyone. In Section 6, I estimate $\frac{\partial \mathbb{P}[SA]_H}{\partial B} = 0.00028$. Therefore, in the calibration: $f_\varepsilon(v_L - \kappa_L) = f_\varepsilon(v_H - \kappa_H) = 0.00028$. I estimate $\frac{\partial \mathbb{P}[SA]_L}{\partial B} = 0.00065$, which therefore implies $v'_L = 2.3$.

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 C.2) to calibrate $\frac{\partial \mathbb{P}[SA]_H}{\partial \Lambda} = -0.014$. The main results of Section 5 thus imply $\frac{\partial \mathbb{P}[SA]_L}{\partial \Lambda} = -0.023$. Combining these results with $f_\varepsilon(v_L - \kappa_L) = f_\varepsilon(v_H - \kappa_H) = 0.00028$, I find that $\kappa'_H = 79$ and $\kappa'_L = 130$.

These estimates show that people with poor mental health have a more than two times higher marginal value of benefits than those with good mental health. This difference is not explained by differences in income between the groups - because the regression kink design conditions on income. Moreover, people with poor mental health have a 65% higher cost of overcoming barriers than those with good mental health.

My findings highlight the significant disparities in the marginal value of benefits and the costs of access for individuals with poor mental health. The fact that this higher marginal value is not driven by differences in income underscores the multi-dimensionality of need. Furthermore, the 65% higher cost of overcoming barriers emphasizes the urgency for policymakers to address these access-issues.

7.1.2. *Relaxing Assumptions.* Recall that in order to estimate these sufficient statistics, there are three important identification assumptions. (i) Take-up depends on an additive independent choice shock, (ii) $\mathbb{P}$[Marginal to Barrier Change]$_\theta$ = $\mathbb{P}$[Marginal to Benefits Change]$_\theta$ and (iii) revealed preference. Below, I show how these assumptions can be relaxed through the use of additional structure, and discuss implications for how need and cost vary by mental health.

Relaxing independence of ε leads to the following. Suppose we used the model of Rafkin et al. (2023) where $v'_\theta(B)$ is independent of θ 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, κ'_θ 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_θ is θ'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).

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 people who earn most of their money from work - as there are notches in the schedules of other benefits at the social minimum = social assistance eligibility threshold. The samples for the two instruments are quite different, 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. Recall that the partial-identification result described in Proposition 5.1 does *not* rely on the benefit-level response estimated in Section 6. If people with poor mental health receive SA at the same rate as those with good mental health, but are more likely to be screened out by ordeals, they must value the benefit more. In Appendix E, 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 is about $1/4$ 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.

What if individuals are biased? In this case, take-up responses do not necessarily reveal true need or cost, but perceived need and cost. On the cost-side, the κ'_θ calibrated above can be thought of as-if costs of barriers (Goldin and Reck, 2022). Let ψ% of these costs be felt by the individual, where ψ represents the degree of normative ambiguity.

In the next section, I characterise the extent to which the government needs to believe people are biased in order to reverse the welfare conclusion, following Naik and Reck (2024). On the need-side I find that people with poor mental health have a larger perceived need than those with good mental health. If this were due to bias, it would imply those with poor mental health are *less* pessimistic than people with good mental health, contrary to psychological evidence (Alloy and Ahrens, 1987). Therefore, I interpret the RKD results as a reflection of truly larger need among those with poor mental health.

7.2. **Quantifying Welfare Effects.** In the data, the prevalence of poor mental health conditional on eligibility is $\mu(L) = 0.25$. 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. The tax rate $\tau \approx 37\%$. 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. This is driven by people with poor mental health also receive more benefits than those with good mental health - mostly driven by the fact that 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. If we were to focus on the general population, $FE_H \gg FE_L$ as $y_L^{SA=0} \ll y_H^{SA=0}$.

7.2.1. *MVPFs of Ordeals and Benefits.* These estimates plugged into Proposition A.1, assuming a utilitarian social welfare function imply the following:

$$
\begin{aligned}
MVPF_{d\Lambda} &= \frac{\overbrace{-\int \lambda_\theta \cdot \mathbb{P}[SA]_\theta \cdot \frac{\kappa'_\theta(\Lambda)}{v'_\theta(B)}\, d\mu}^{\text{Direct Effect } <0}}{\underbrace{\int FE_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda} d\mu}_{\text{Behavioral Revenue Effect } <0}} \\
&= \frac{-0.6 \times \frac{130}{2.3} \times 0.25 - 0.6 \times \frac{79}{1} \times 0.73}{679.45 \times 0.25 \times (-0.023) + 645.09 \times 0.73 \times (-0.014)} \\
&= 1.88
\end{aligned}
$$

An $MVPF_{d\Lambda}$ of 1.88 means that ordeals impose a direct cost of €1.88 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 €56.4, whereas €79 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.

$$
\begin{aligned}
MVPF_{dB} &= \frac{\overbrace{\int \lambda_\theta \cdot \mathbb{P}[SA]_\theta \, d\mu}^{\text{Direct Effect } >0}}{\underbrace{(1-\tau) \cdot \int \mathbb{P}[SA]_\theta \, d\mu}_{\text{Mechanical Revenue Effect } >0} + \underbrace{\int FE_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial B} d\mu}_{\text{Behavioral Revenue Effect } >0}} \\
MVPF_{dB} &= \frac{0.6 \times 0.25 + 0.6 \times 1 \times 0.73}{0.63 \times (0.58 \times 0.25 + 0.54 \times 0.73) + 679.45 \times 0.25 \times 0.00065 + 645.09 \times 0.73 \times 0.00028} \\
&\approx 0.82
\end{aligned}
$$

An $MVPF_{dB} < 1$ is to be expected since social assitance is a re-distributive program. It means means that beneficiaries gain 81 cents for every €1 spent raising the benefit level. The estimated value lies in the range surveyed by Hendren and Sprung-Keyser (2020).

Comparing $MVPF_{d\Lambda}$ to $MVPF_{dB}$ suggests that reducing ordeals is a $2.1\times$ more effective policy than increasing benefits.[37] Alternatively, the government is willing to *reduce* the benefit-level in order to finance a reduction in the take-up barriers. This clear welfare ordering depends crucially on the large positive correlation between need and cost across types.

Such a stark difference in $MVPF$s also generates confidence that the modelling assumptions are not what drives $MVPF_{d\Lambda} > MVPF_{dB}$. Note that $MVPF_{d\Lambda}$ is roughly proportional to κ' and $1/f_{\varepsilon}^{d\Lambda}(\cdot)$ holding all else fixed. This means that $\mathbb{P}[\text{Marginal to Ordeal Change}]_\theta > 2.1 \times \mathbb{P}[\text{Marginal to Benefits Change}]_\theta$ in order to reverse the welfare comparison.[38]

Finally, consider the case that individuals are biased, and perceive barrier costs to be larger than their true value. Namely, let a share ψ of the as-if cost revealed through take-up responses be a true cost, and $1-\psi$ pure hassle costs (which affect behaviour and not welfare). In this model, take-up is too low relative to the private optimum. Therefore, the $MVPF_{d\Lambda}$ scales down the direct cost of barriers by a factor ψ, but also includes a negative behavioral welfare effect $\times(1-\psi)$ since individuals are not privately-optimizing so the Envelope Theorem does not hold. $MVPF_{dB}$ now contains a new term in the numerator - namely an internality correction $\times(1-\psi)$ as increasing benefits helps individuals take-up closer to their private optimum. Appendix E shows the formulae.

When we calibrate these using the sufficient statistics estimated above, we find that the government must be confident that less than 44% of the perceived costs are true welfare costs in order to reverse the $MVPF_{d\Lambda} > MVPF_{dB}$. Here, I take the approach of Naik and Reck (2024): if revealed preference does not hold, the government does not know how

[37] 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 E, I show that the social marginal utility of beneficiaries of dB is 1.36 and of $d\Lambda = 1.27$

[38] In fact, as shown in Appendix E, $\mathbb{P}[\text{Marginal to Ordeal Change}]_\theta = \mathbb{P}[\text{Marginal to Benefits Change}]_\theta$ is actually a conservative assumption. Under linearity of v_θ and κ_θ, the correlation test leads me to estimate the probability of being marginal being lower ($\approx 15\%$) for the ordeal instrument than for the benefits-level change.

much behaviour reflects true welfare. However, policies still need to be set. In this case, it is optimal for the government to choose policies which are *robust* to normative ambiguity. The result states that reducing barriers being more effective than increasing benefits is robust as long as more than 44% of the as-if costs are normatively relevant.

8. Conclusion

This paper shows that people with poor mental health are high-need, yet inefficiently excluded from social 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 find that 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 estimated using a regression kink design on the kinked benefits schedule as a function of income. Combining theory and empirics shows that reducing barriers is twice as effective as increasing benefits.

Future work: As mentioned in Section 7, the policy recommendations depend on whether costs of overcoming barriers are true welfare costs, or just hassle costs which affect behaviour and not welfare. Therefore, in future work I plan to elicit behavioural biases for people with poor mental health and use this quantification to determine optimal policy.

Moreover, 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

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

Alloy, L. B. and A. H. Ahrens (1987): "Depression and pessimism for the future: biased use of statistically relevant information in predictions for self versus others." *Journal of personality and social psychology*, 52, 366. [*Cited on pages 43 and 44.*]

Anders, J. and C. Rafkin (2022): "The Welfare Effects of Eligibility Expansions: Theory and Evidence from SNAP." *SSRN*. [*Cited on page 88.*]

Angelucci, M. and D. Bennett (2024a): "Depression, Pharmacotherapy, and the Demand for a Preventive Health Product," *Available at SSRN 4808853*. [*Cited on page 6.*]

——— (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 6.*]

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 6.*]

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 6.*]

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 15.*]

BHARADWAJ, P., M. M. PAI, AND A. SUZIEDELYTE (2017): "Mental health stigma," *Economics Letters*, 159, 57–60. [*Cited on page 2.*]

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 6.*]

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 page 5.*]

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 47.*]

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

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 32, 34, 35, 36, 70, and 73.*]

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 6.*]

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 2 and 48.*]

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 17.*]

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 24.*]

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

DESHPANDE, M. AND L. M. LOCKWOOD (2022): "Beyond health: Nonhealth risk and the value of disability insurance," *Econometrica*, 90, 1781–1810. [*Cited on pages 6 and 41.*]

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.*]

EN WERKGELEGENHEID, M. V. S. Z. (2017): "UWV Monitor Arbeidsparticipatie 2016-Rapport-Kennisplatform Werk en Inkomen," . [*Cited on page 23.*]

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, 40, and 43.*]

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, 5, 6, 13, 25, 43, 58, and 88.*]

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 37 and 82.*]

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 35.*]

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

GODARD, M., P. KONING, AND M. LINDEBOOM (2022): "Application and award responses to stricter screening in disability insurance," *Journal of Human Resources*. [*Cited on page 2.*]

GOLDIN, J. AND D. RECK (2022): "Optimal defaults with normative ambiguity," *Review of Economics and Statistics*, 104, 17–33. [*Cited on page 44.*]

GROSS, J. J. AND R. F. MUÑOZ (1995): "Emotion regulation and mental health." *Clinical psychology: Science and practice*, 2, 151. [*Cited on pages 5 and 40.*]

HALLER, A. AND S. STAUBLI (2024): "Measuring the Value of Disability Insurance from Take-Up Decisions," . [*Cited on pages 2, 6, and 12.*]

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

HAUSHOFER, J. (2019): "Is there a Psychological Poverty Trap?" Tech. rep., Working Paper. [*Cited on page 48.*]

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 40.*]

HEEKELAAR, M. (2021): "Beschikbare en benodigde financiële middelen voor de Participatiewet: Analyse," . [*Cited on page 23.*]

HENDREN, N. (2016): "The policy elasticity," *Tax Policy and the Economy*, 30, 51–89. [*Cited on page 46.*]

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,*

46, and 59.]

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

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 and 40*.]

KLEVEN, H. J. (2016): "Bunching," *Annual Review of Economics*, 8, 435–464. [*Cited on pages 38 and 61*.]

KLINE, P. AND C. R. WALTERS (2019): "On Heckits, LATE, and numerical equivalence," *Econometrica*, 87, 677–696. [*Cited on page 13*.]

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, 3, and 9*.]

KOLSTAD, J., C. NAIK, W. PARKER, AND J. SPINNEWIJN (2024): "Social Determinants of Mental Health," . [*Cited on pages 3, 15, and 17*.]

KROMHOUT, M., N. KORNALIJNSLIJPER, AND M. DE KLERK (2018): "Summary changing care and support for people with disabilities," . [*Cited on page 30*.]

KRUEGER, A. B. AND B. D. MEYER (2002): "Labor supply effects of social insurance," *Handbook of public economics*, 4, 2327–2392. [*Cited on page 40*.]

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 43*.]

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 and 43.*]

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 38 and 79.*]

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 40.*]

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 6.*]

MINISTERIE VAN SZW (2015): "Participatiewet," https://wetten.overheid.nl/BWBR0015703/2015-01-01/. [*Cited on pages 4, 15, 16, 23, 32, and 34.*]

——— (2022): "Participatiewet in Balans: uitkomsten beleidsanalyse," https://www.rijksoverheid.nl/documenten/kamerstukken/2022/06/21/bijlage-rapport-participatiewet-in-balans. [*Cited on pages 4, 26, and 27.*]

MOGSTAD, M., A. TORGOVITSKY, AND C. R. WALTERS (2024): "Policy evaluation with multiple instrumental variables," *Journal of Econometrics*, 105718. [*Cited on page 13.*]

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 20.*]

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 90.*]

MULLAINATHAN, S. AND E. SHAFIR (2013): *Scarcity: Why having too little means so much*, Macmillan. [*Cited on page 41.*]

NAIK, C. AND D. RECK (2024): "Intrapersonal Utility Comparisons as Interpersonal Utility Comparisons: Welfare, Ambiguity, and Robustness in Behavioral Policy Problems," Tech. rep., National Bureau of Economic Research. [*Cited on pages 13, 44, and 46.*]

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 73.*]

RAFKIN, C., A. SOLOMON, AND E. SOLTAS (2023): "Self-Targeting in US Transfer Programs," *Available at SSRN 4495537*. [*Cited on pages 2, 5, 13, 43, and 89.*]

RAMBACHAN, A. AND J. ROTH (2023): "A more credible approach to parallel trends," *Review of Economic Studies*, 90, 2555–2591. [*Cited on page 24.*]

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 48.*]

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 17.*]

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 6.*]

SCP (2019): "Eindevaluatie van de Participatiewet," . [*Cited on pages 4, 23, 26, and 89.*]

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

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

SERENA, B. L. (2024): "The Causal Effect of Scaling up Access to Psychotherapy," . [*Cited on page 6.*]

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 24.*]

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 5 and 7.*]

SHREEKUMAR, A. AND P.-L. VAUTREY (2021): "Managing Emotions: The Effects of Online Mindfulness Meditation on Mental Health and Economic Behavior," . [*Cited on page 6.*]

SILVER, D. AND J. ZHANG (2022): "Impacts of Basic Income on Health and Economic Well-Being: Evidence from the VA's Disability Compensation Program," Tech. rep., National Bureau of Economic Research. [*Cited on page 6.*]

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 40.*]

VAN DER VEEN, R. (2019): "Basic income experiments in the Netherlands?" *Basic Income Studies*, 14. [*Cited on page 23.*]

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 29*.]

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

ZECKHAUSER, R. (2021): "Strategic sorting: the role of ordeals in health care," *Economics & Philosophy*, 37, 64–81. [*Cited on page 3*.]

APPENDIX A. THEORY APPENDIX

Let θ have a type-specific indirect utility functions: $u_\theta(c, y)$ is increasing in consumption c and decreasing in earned income y. Income depends on take-up but is fixed otherwise:[39] let $y_\theta^{SA=1}$ refer to income earned if on social assistance and $y_\theta^{SA=0}$ if not. All income (including benefits) is taxed at marginal tax rate τ. Thus, $v_\theta(B)$ is given by:

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

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

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

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

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

Using the chain rule: $\frac{dW}{d\Lambda} = \frac{\partial W}{\partial \Lambda} + \frac{\partial W}{\partial B}\cdot\frac{\partial B}{\partial \Lambda}$, and using the Leibniz rule to differentiate under the integral gives Equation (2.5). Here, the Envelope Theorem implies the behavioural welfare effect is 0. For example,

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

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

The above step is the Envelope Theorem at work.

$$= -\kappa'_\theta(\Lambda) \cdot F(\varepsilon^*_\theta)$$

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

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

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

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

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 Λ (B adjusts) is social-welfare improving, if the gains from increasing spending on dB exceed the losses from reducing spending through an increase $d\Lambda$.

Let η_θ 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 θ'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 Λ 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).

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

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

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

The direct effect of an increase in ordeals $d\Lambda$ is that it imposes dis-utility on infra-marginal individuals κ'_θ. 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 E.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.

□

Proof of Proposition 5.1. $\mathbb{P}[SA]_L = \mathbb{P}[SA]_H$ and $\varepsilon \perp \theta$ implies $F_\varepsilon\big(v_L(B) - \kappa_L(\Lambda)\big) = F_\varepsilon\big(v_H(B) - \kappa_H(\Lambda)\big) \implies v_L(B) - \kappa_L(\Lambda) = v_H(B) - \kappa_H(\Lambda) \implies f_\varepsilon\big(v_L(B) - \kappa_L(\Lambda)\big) = f_\varepsilon\big(v_H(B) - \kappa_H(\Lambda)\big)$.

Therefore, $\frac{\partial \mathbb{P}[SA]_L - \mathbb{P}[SA]_H}{\partial \Lambda} < 0 \implies \kappa'_L(\Lambda) > \kappa'(\Lambda)$. Under linearity $\kappa_\theta(\Lambda) = \kappa_\theta \cdot \Lambda$, this then implies $v_L(B) > v_H(B)$. □

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

This proposition requires some additional structure:

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

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

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:

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

We would like to approximate $v_L - v_H$ and $\kappa_L - \kappa_H$ using take-up responses to changes in B and Λ.

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

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

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. Therefore:

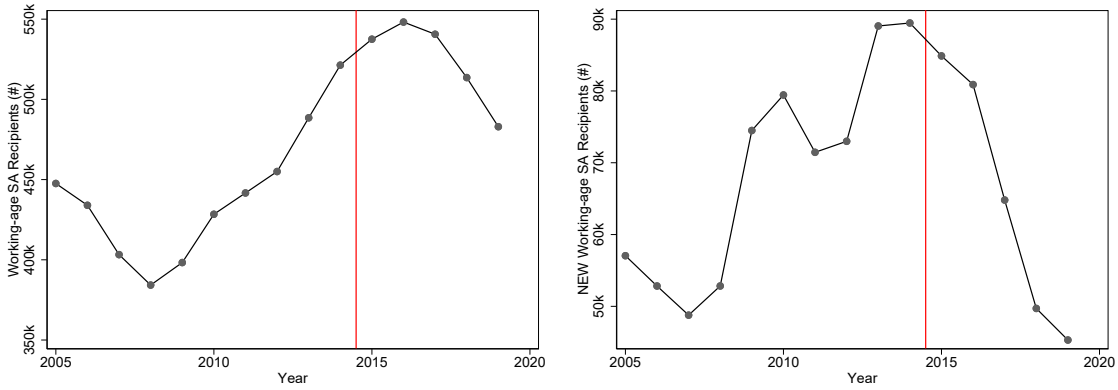
$$\begin{aligned} F(v_L - \kappa_L) - F(v_H - \kappa_H) &\approx [(v'_L(B) - v'_H(B)) \cdot \Psi - (\kappa'_L(\Lambda) - \kappa'_H(\Lambda)) \cdot \Lambda - \Delta\kappa_0] \cdot f_H \\ &= \left(\frac{\partial \mathbb{P}[SA]_L}{\partial B} \cdot \frac{f_H}{f_L} - \frac{\partial \mathbb{P}[SA]_H}{\partial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

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).

□

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.**

APPENDIX C. EFFECT OF ORDEALS ON TARGETING: ADDITIONAL MATERIAL

APPENDIX D. EFFECT OF BENEFIT-LEVEL ON TARGETING: 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).

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 (EUR)	66,949.4 (73,978.0)	13,125.2 (2,795.6)
Household Net Worth (EUR)	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

D.1. **Income Exemptions.** I model the unobserved benefits schedule as Equation (D.1).

(Mental) Health Information	General Population (Mean, SD)	Eligible (Mean, SD)
General		
All Care Spending (EUR)	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		
Mental Healthcare Spending (EUR)	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)

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

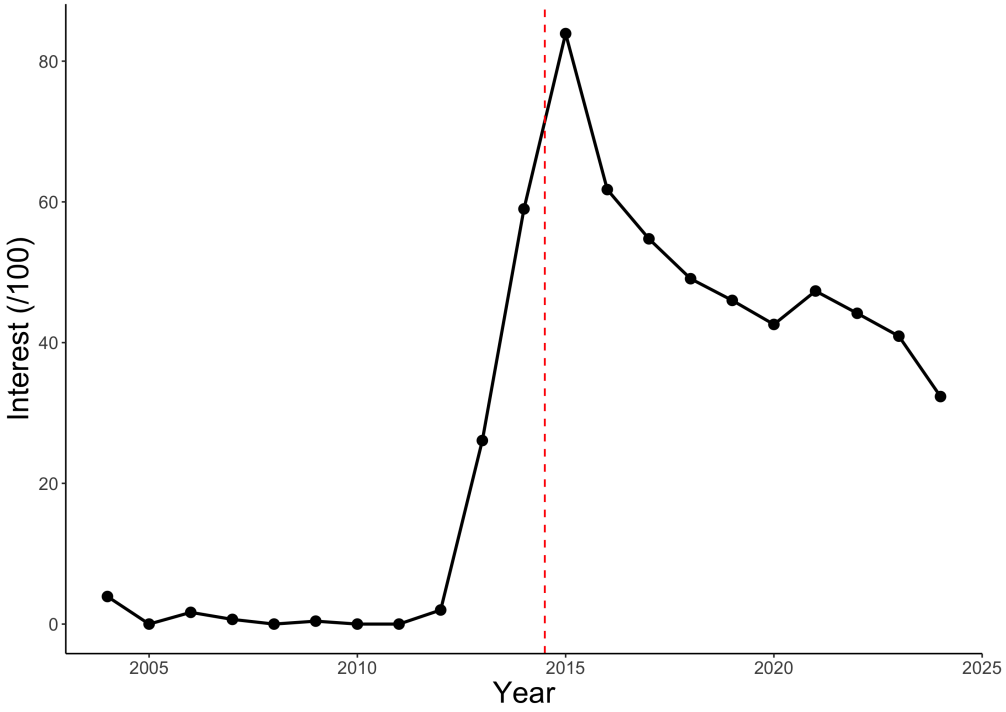


FIGURE C.1. Google Trends for *Participatiewet*, the Dutch translation of the "Participation Act" over time in the Netherlands.

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

Algorithm 1 Eligibility Calculation

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

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

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

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

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

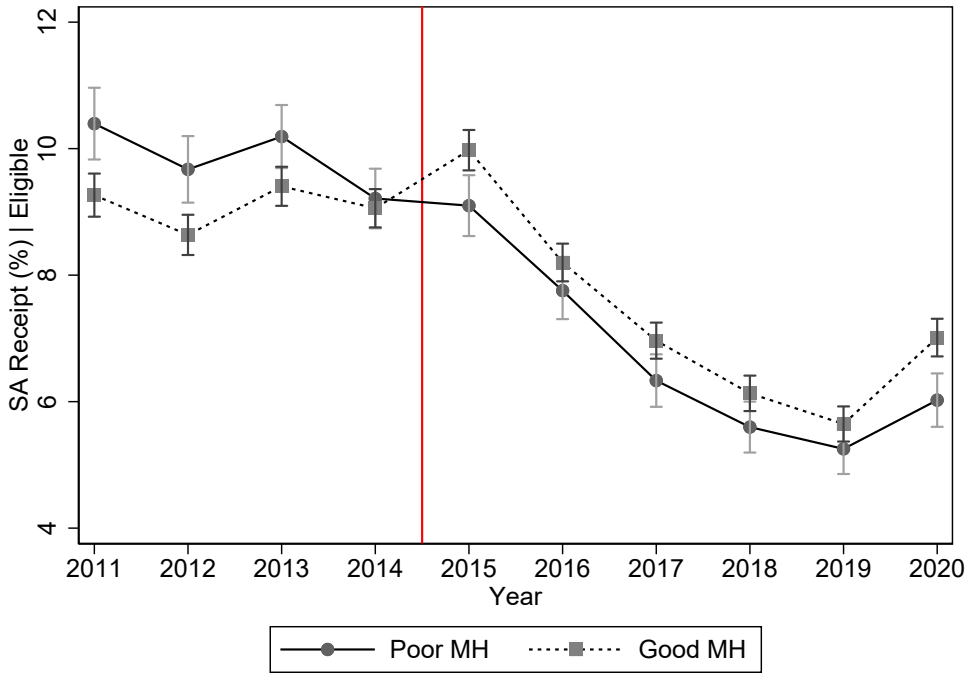


FIGURE C.2. Evolution of inflow of social assistance over time, split by people with poor mental health in the pre-period vs those with good mental health in this period. Raw means and respective 95% confidence intervals are shown. The introduction of the Participation Act in 2015 is shown by the red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

$$p(y) = \begin{cases} \frac{\left(\mathbb{E}[B|SA,Y=y]-(\bar{y}-y)\right)\cdot\mathbb{P}[SA|Y=y]}{y\cdot\mathbb{P}[SA|Y=y,\nu=1]} & \text{if } y \leq \bar{y} \\ \frac{\mathbb{E}[B|SA,Y=y]\cdot\mathbb{P}[SA|Y=y]}{\bar{y}\cdot\mathbb{P}[SA|Y=y,\nu=1]} & \text{if } y \geq \bar{y} \end{cases} \tag{D.2}$$

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 ($\approx 100\%$).

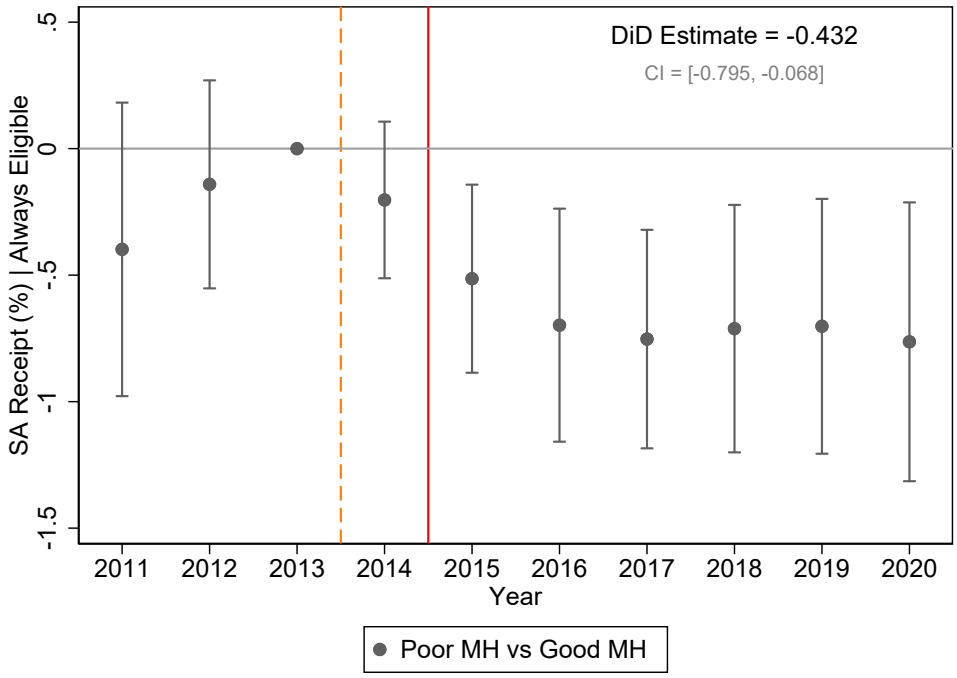


FIGURE C.3. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is always-eligible middle-age couples and poor mental health is defined by prescription of psychopharma in 2012, 2013 or 2014. Controls include individual fixed effects, income, education and muncipality, hh composition and sector fixed effects. The TWFE estimate $\hat{\delta}$ in the regression $SA_{it} = \alpha + \eta_i + \gamma_t + \delta \cdot \mathbb{1}\{t \geq 2014\} \times \text{Poor MH}_i + X_{it}'\theta + \varepsilon_{it}$ is also shown. Standard-errors are clustered at the level of municipality of residence in 2013.

Proof of Proposition D.1. Let $\mathbb{E}_y \triangleq \mathbb{E}[\cdot|Y = y]$ and $\mathbb{P}_y \triangleq \mathbb{P}(\cdot|Y = y)$

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

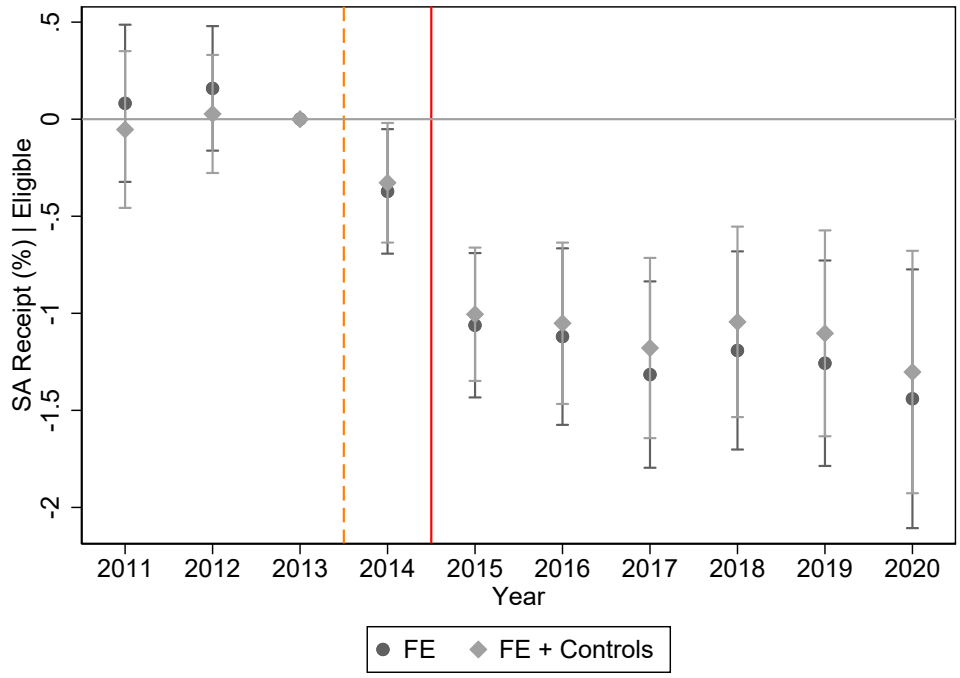


FIGURE C.4. Estimates $\hat{\delta}_t$ from Equation (5.1) showing the heterogeneous treatment effects of an increase in ordeals on rate-of-receipt by baseline mental health. The analysis population is eligible middle-age couples and poor mental health is defined by prescription of psychopharma in pre-period. $\hat{\delta}_t$ 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.

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

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

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

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

Rearranging for $p(y)$ gives the expression in Equation (D.2).

□

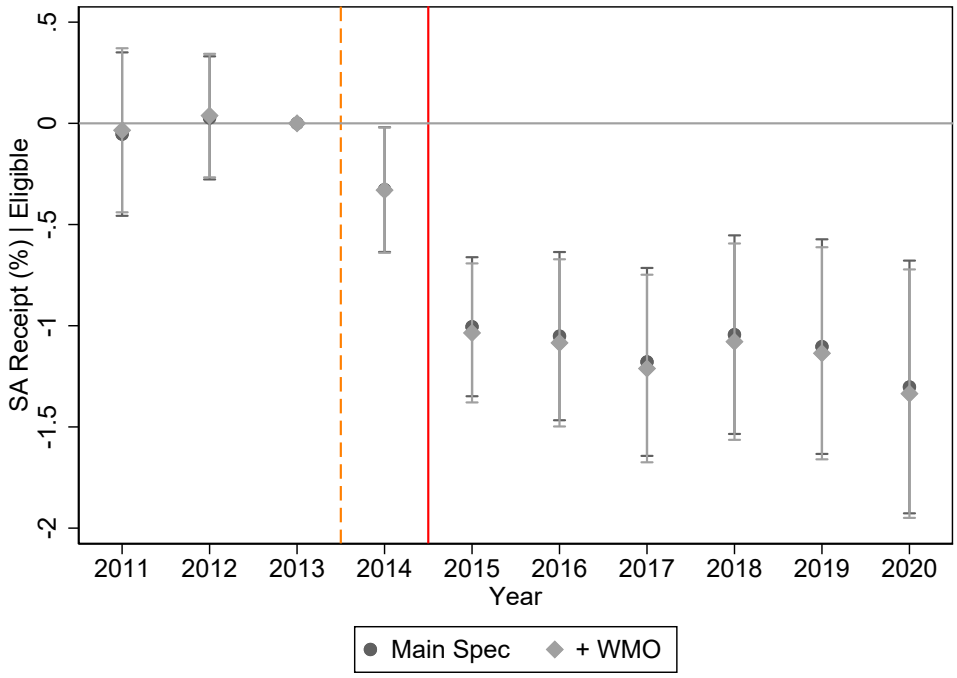


FIGURE C.5. TWFE Event Study estimates, adding controls for receiving long-term-care adjustments from the municipality, receipt of Wajong, and exemptions to obligations. Standard-errors are clustered at the level of municipality of residence in 2013.

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

(1) **Regularity:** (ε, ν) has bounded support. $p(\cdot,\cdot,\cdot)$ is **continuous** and partially differentiable w.r.t. first and second arguments. $p_1(b, y, e)$ continuous.

(2) **Smooth effect of Y:** $p_2(b, y, e)$ is continuous.

(3) **First Stage and Nonnegligible Population at Kink:** $b(y, v)$ continuous and $b_1(y, v)$ continuous apart from at $y = \bar{y}$. Positive mass at kink.

(4) **Smooth Density:** Density of Y is continuously differentiable

(5) **Smooth Probability of No Measurement Error:** $\mathbb{P}[U_Y = 0, U_B = 0|Y = y, \varepsilon, \nu]$ and partial derivative w.r.t. y is continuous.

(6) **Monotonicity:** Either $b_1^+(v) \geq b_1^-(v)$ for all v or $b_1^+(v) \leq b_1^-(v)$ for all v.

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

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

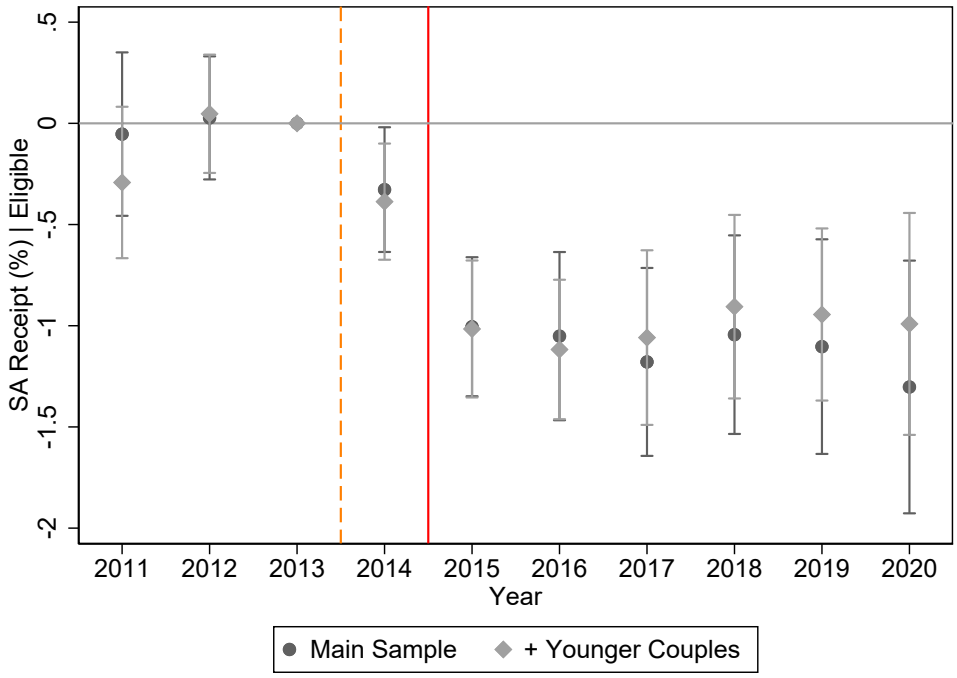


FIGURE C.6. TWFE Event Study estimates for different age ranges. Standard-errors are clustered at the level of municipality of residence in 2013.

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

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

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

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

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

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

Without both parts of Assumption 1, the numerator of the estimand in Equation (6.1) will be 0, while without one part only, regularity is violated. In my sample, around 8% of people receiving social assistance have $Y_{\text{true}} > \bar{y}$. **??** provides support that $\lim_{B \to 0} \mathbb{P}[SA|B] > 0$. Assumption 2 is a corollary of $b(y, v)$ being continuous.

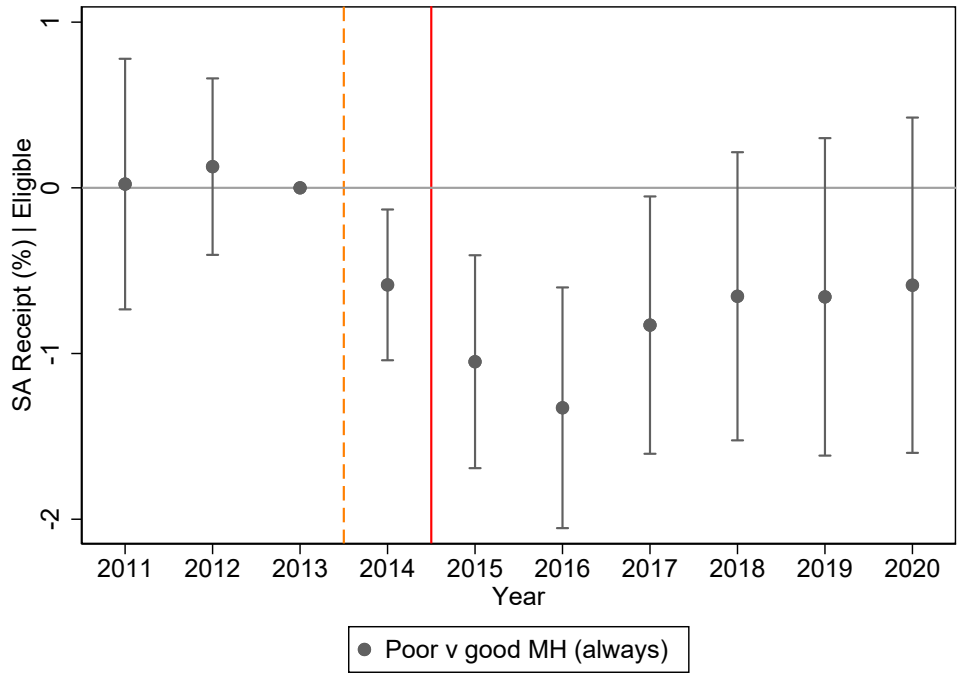


FIGURE C.7. $\hat{\delta}$ for the main specification, where Poor MH_i: dispensed psychopharma in every year 2011-2020, vs good mental health throughout. Standard-errors are clustered at the level of municipality of residence in 2013.

D.3. **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.

D.3.1. *Setup.* I simulate a million individuals which are characterised by ability $Y \sim U[500, 1500]$. This corresponds to their income. I set a fixed cost to be $\kappa = 150$ for everyone. Choice error $\varepsilon = \frac{U_1+U_2}{2}$ where $U_j \overset{\text{i.i.d.}}{\sim} U[-200, 200]$. I.e. ε follows a symmetric triangular distribution centered around 0. The threshold $\bar{y} = 1000$ for everyone. Benefits schedule $B(y)$ is programmed as $B(y) = \max\{\bar{y} - \nu \cdot y, 0\}$ where exemption $\nu \in \{0, 1\}$ and $\mathbb{P}[\nu = 1|Y = y] \equiv 0.1$. Individual y takes up iff:

$$B(y) \geq \kappa \tag{D.6}$$

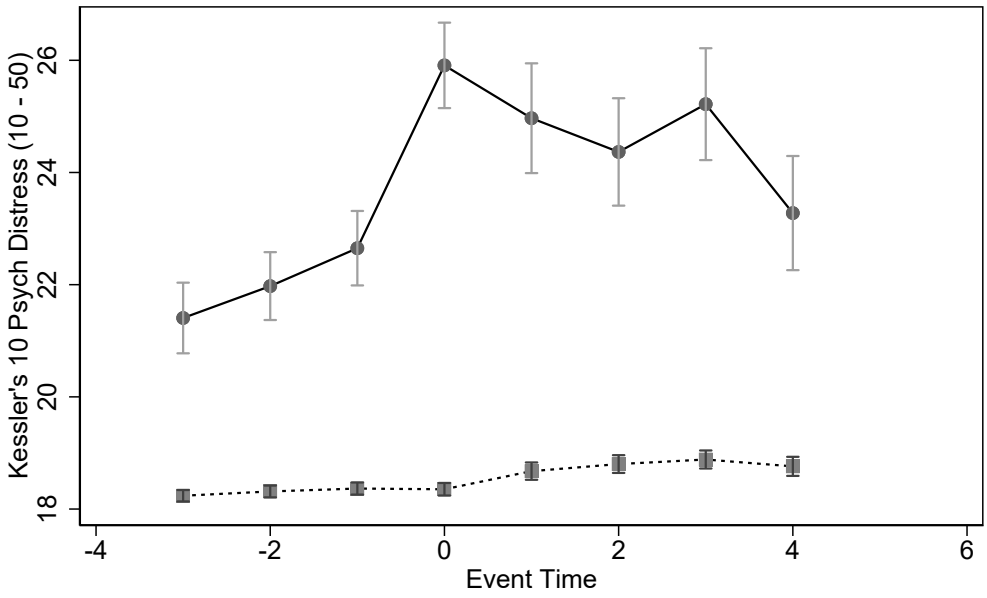


FIGURE C.8. This plot shows the mean subjective mental health (measured by Kessler's 10 Psychological Distress) for two groups: one group is prescribed psychopharma for the first time in Event Time 0, the other group has no prescriptions for all event times $t \leq 0$. Standard-errors are clustered at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

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 D.1. The results of this imputation are shown in **??**.

D.3.2. *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

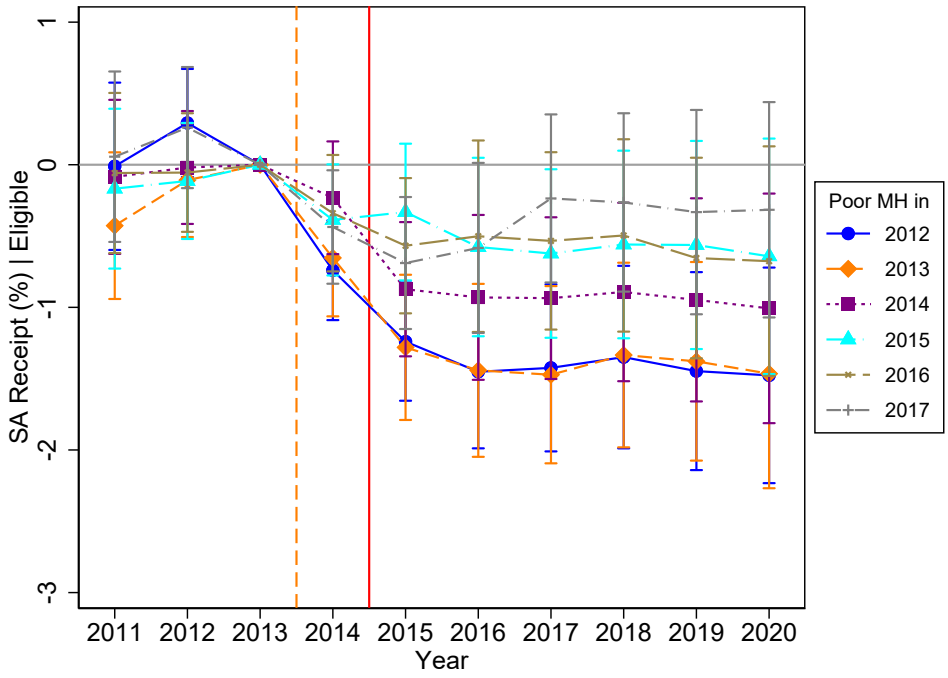


FIGURE C.9. $\hat{\delta}$ for the main specification, changing the definition of Poor MH$_i$. The different definitions are Poor MH$_i$ $= \mathbb{1}\{$Prescribed Psychopharma in year $y\}$, showing estimates for $y \in \{2012, ..., 2017\}$. Standard-errors are clustered at the level of municipality of residence in 2013.

of my simulated individuals have value $\alpha = 1$, and half $\alpha = 2$. Individuals take-up iff:

$$\alpha \cdot B(y) \geq \kappa \tag{D.7}$$

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

D.4. **Validity of RKD.**

D.5. **Results.**

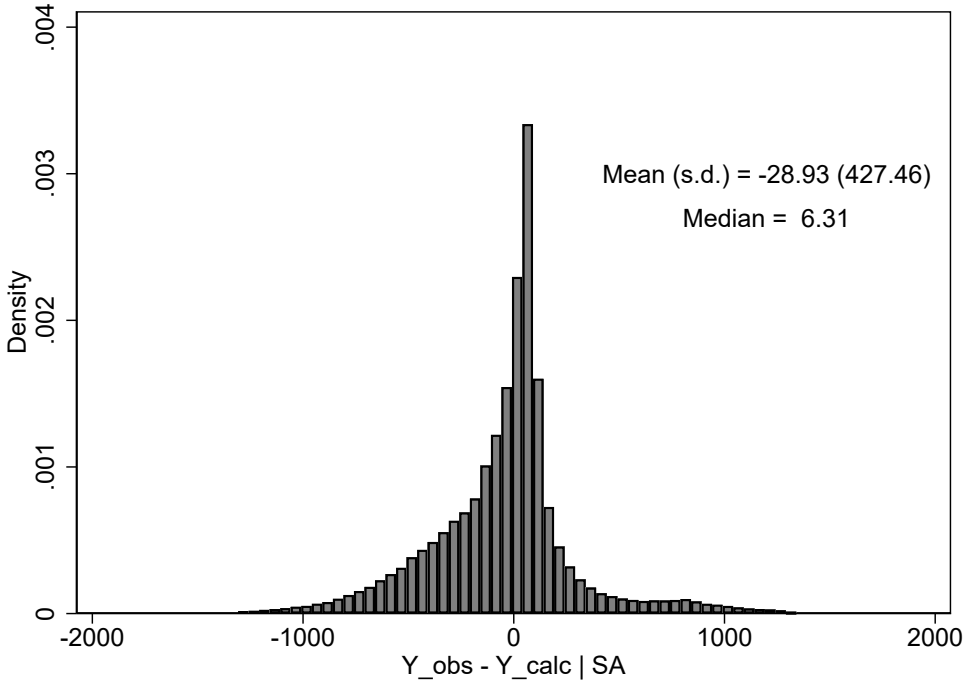


FIGURE D.1. Histogram of $Y_{\text{true}} - Y_{\text{calc}}$ for the analysis population of the RKD.

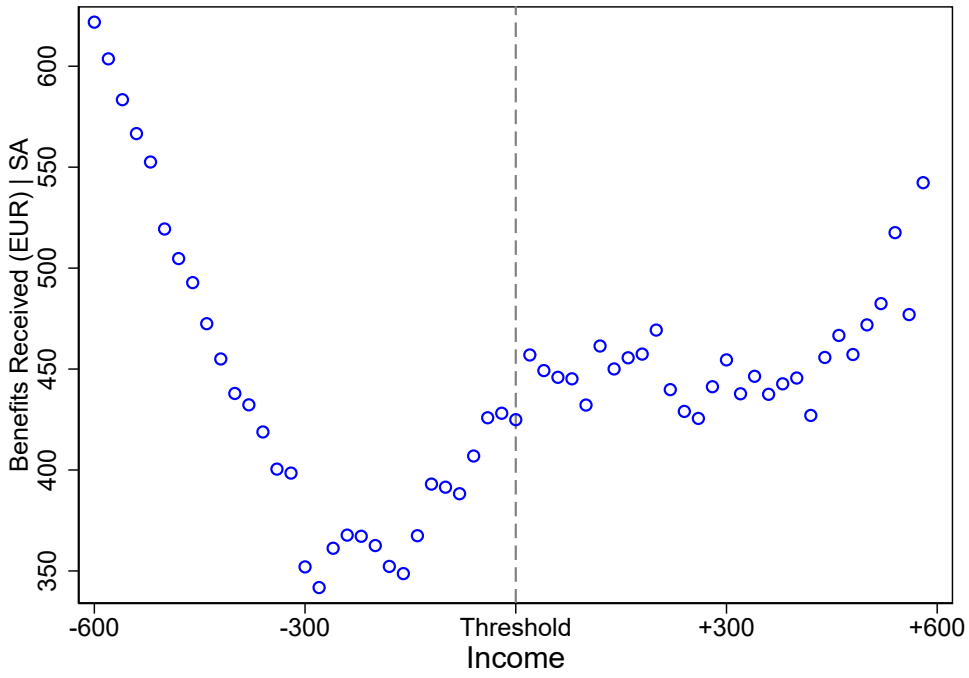


FIGURE D.2. Plot of benefits received conditional on receipt, averaged within income slice (10EUR bins). A window of 1000EUR either side of the threshold is shown.

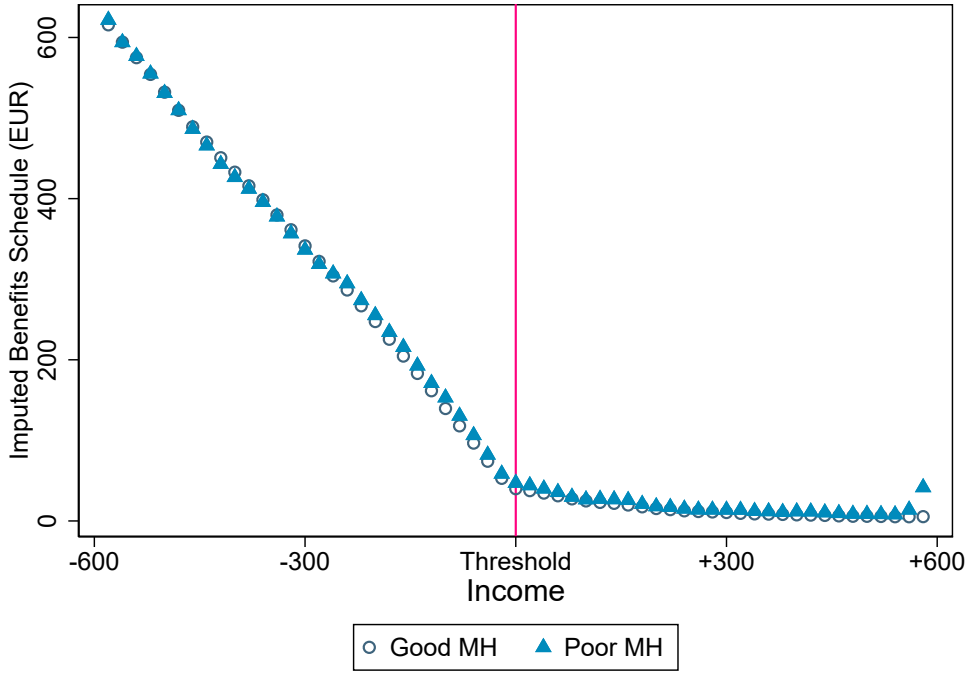
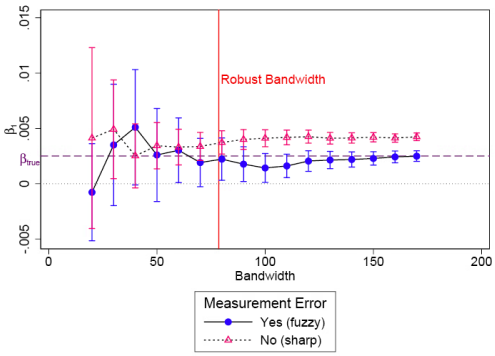
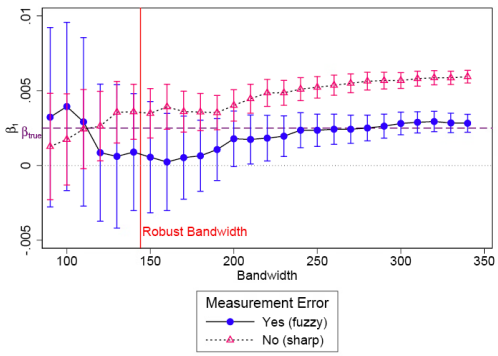


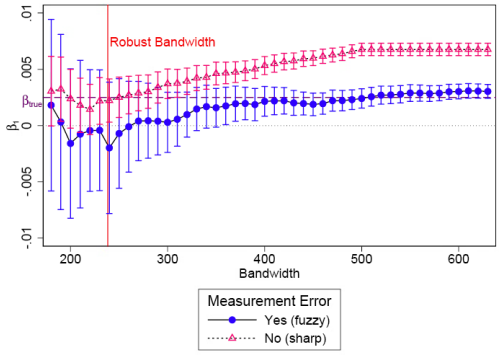
FIGURE D.3. Results of Imputation from Proposition D.1



(A) Local polynomial order $p = 1$



(B) Local polynomial order $p = 2$



(C) Local polynomial order $p = 3$

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

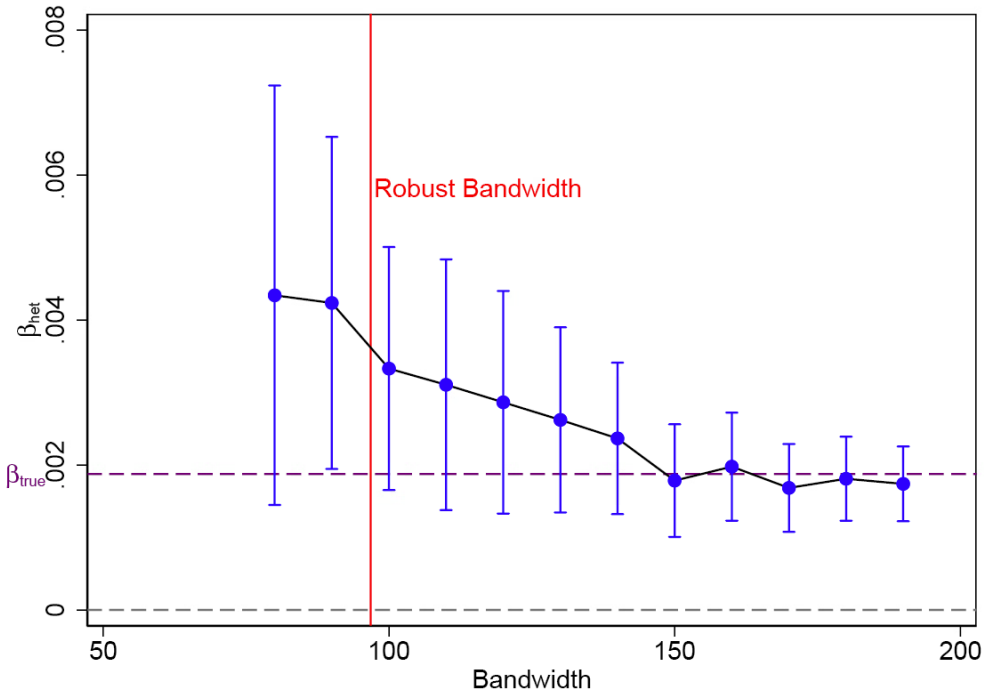


FIGURE D.5. Results of simulations showing estimates from heterogeneous RKD ($\alpha = 1$ vs 2) using different bandwidths. CCT robust bandwidth is shown.

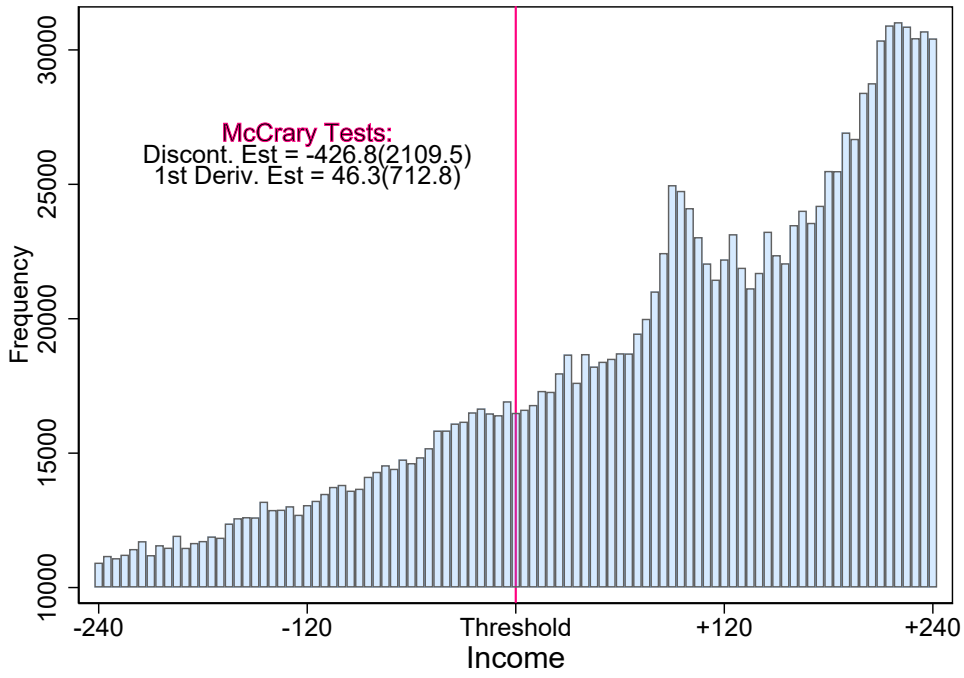


FIGURE D.6. Density of income around the eligibility threshold. McCrary (2008) tests for discontinuity in levels and slopes around the threshold are shown. Income in this plot is monthly. The sample contains singles who get most income from labour, years 2011-2014. See Section 6.1.2 for details on sample restrictions.

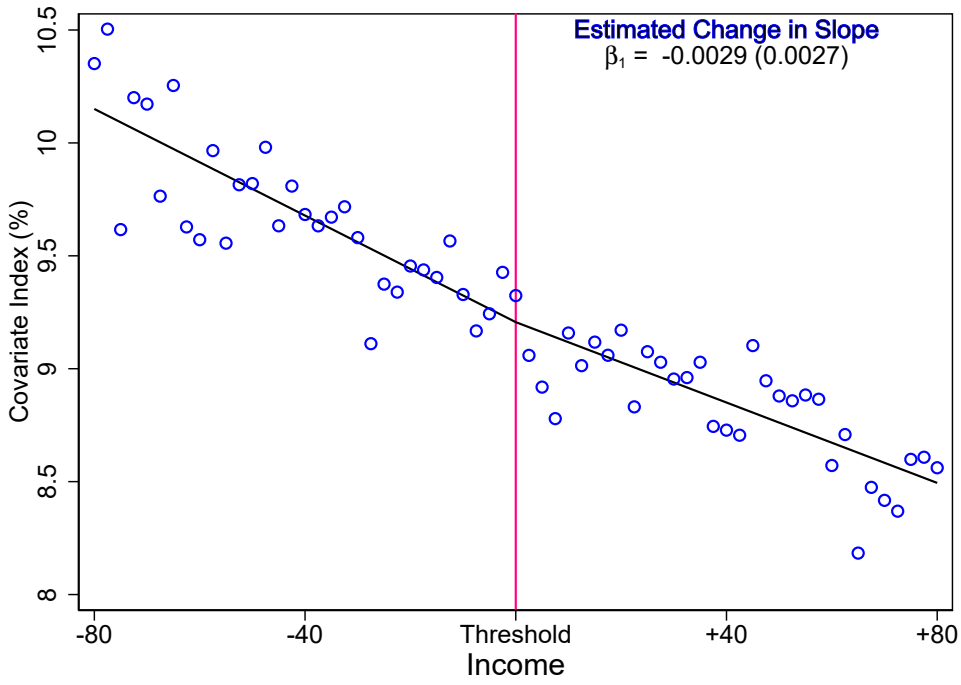


FIGURE D.7. Covariate Test: plot shows fitted values of a regression of social assistance take-up on all pre-determined controls used throughout this paper including income, education, hh composition, municipality FEs. These fitted values form a "Covariate Index" which is binned. An RKD estimate with income as the running variable is also shown. Income in this plot is monthly. The sample contains singles who get most income from labour, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Standard-errors are clustered at the municipality level.

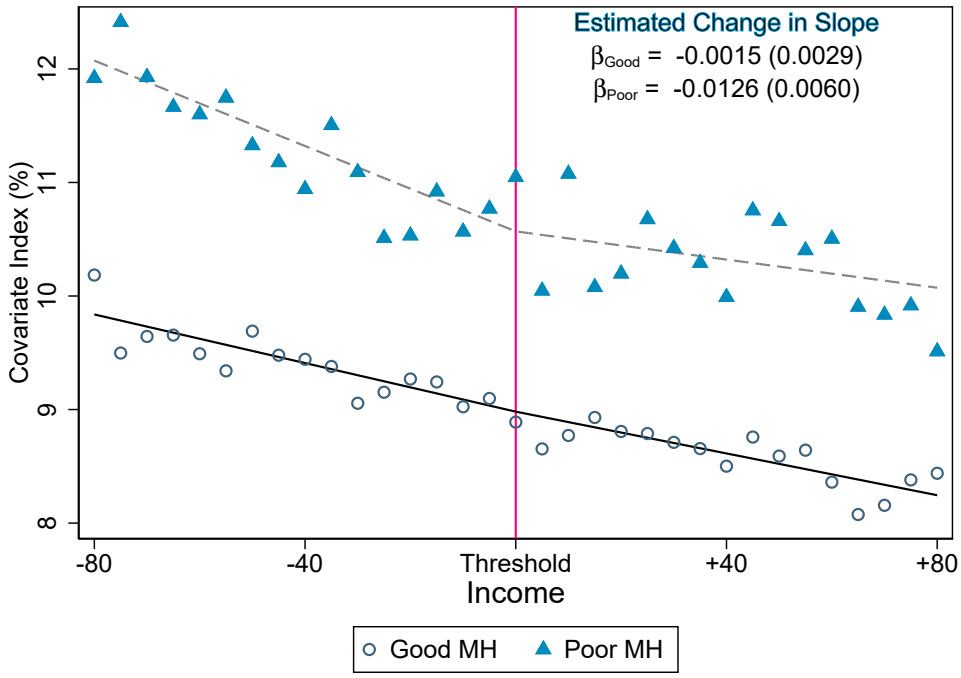


FIGURE D.8. Covariate Test: plot shows fitted values of a regression of social assistance take-up on all pre-determined controls used throughout this paper including income, education, hh composition, municipality FEs. These fitted values form a "Covariate Index" which is binned. An RKD estimate with 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 who get most income from labour, 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.

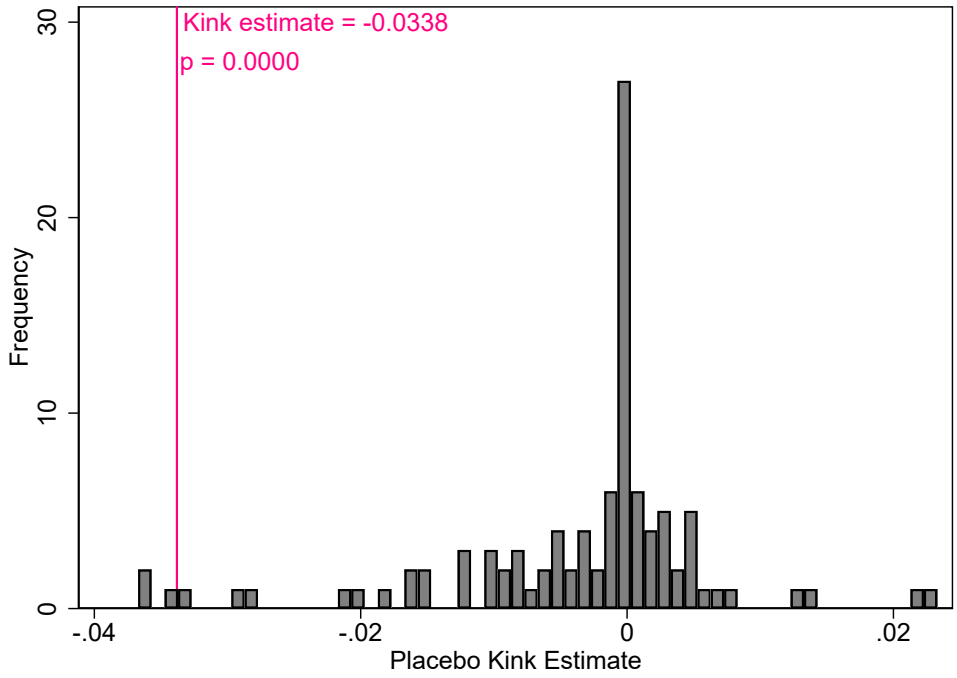


FIGURE D.9. Results of permutation test à la Ganong and Jäger (2018). I estimate RKDs on 100 placebo kinks in the range $[\bar{y} - 600, \bar{y} + 600]$ and plot a histogram of the estimates. A binomial test is used to check whether the true estimate is an outlier. The sample contains singles who get most income from labour, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Standard-errors are clustered at the municipality level.

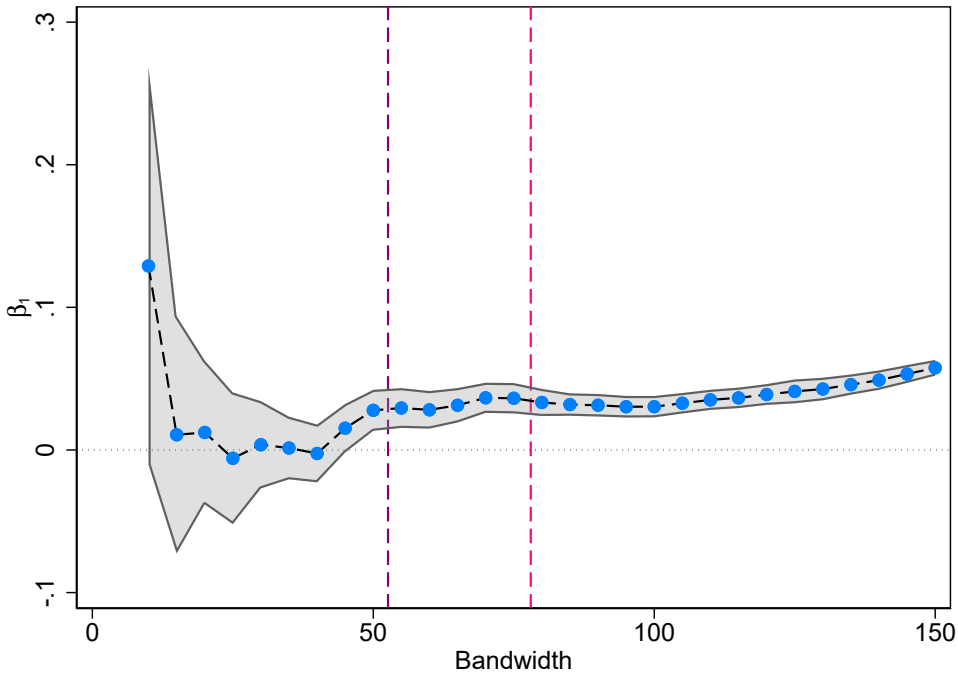


FIGURE D.10. Results of test of sensitivity to changes in bandwidth. I estimate RKDs changing the bandwidth, with the CCT robust bandwidth displayed. The lower purple dashed line indicates the CCT robust bandwidth with regularization, and the upper pink dashed line indicates the CCT robust bandwidth without regularization. This plot shows the estimates and confidence intervals. The sample contains singles who get most income from labour, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Standard-errors are clustered at the municipality level.

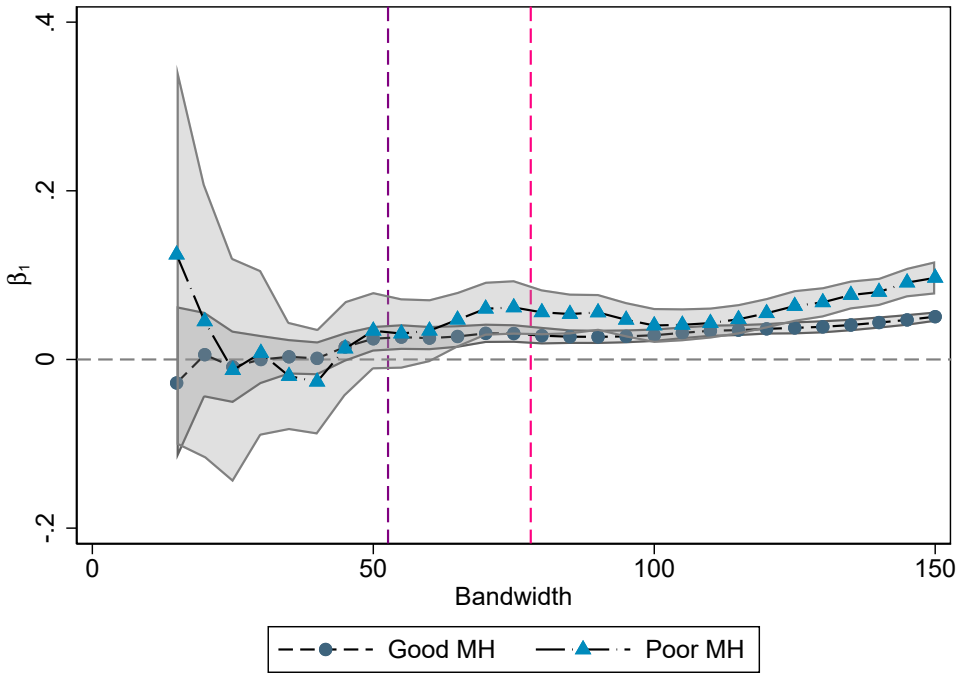


FIGURE D.11. Results of test of sensitivity to changes in bandwidth. I estimate heterogeneous RKDs changing the bandwidth, with the CCT robust bandwidth displayed. The lower purple dashed line indicates the CCT robust bandwidth with regularization, and the upper pink dashed line indicates the CCT robust 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 who get most income from labour, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Standard-errors are clustered at the municipality level.

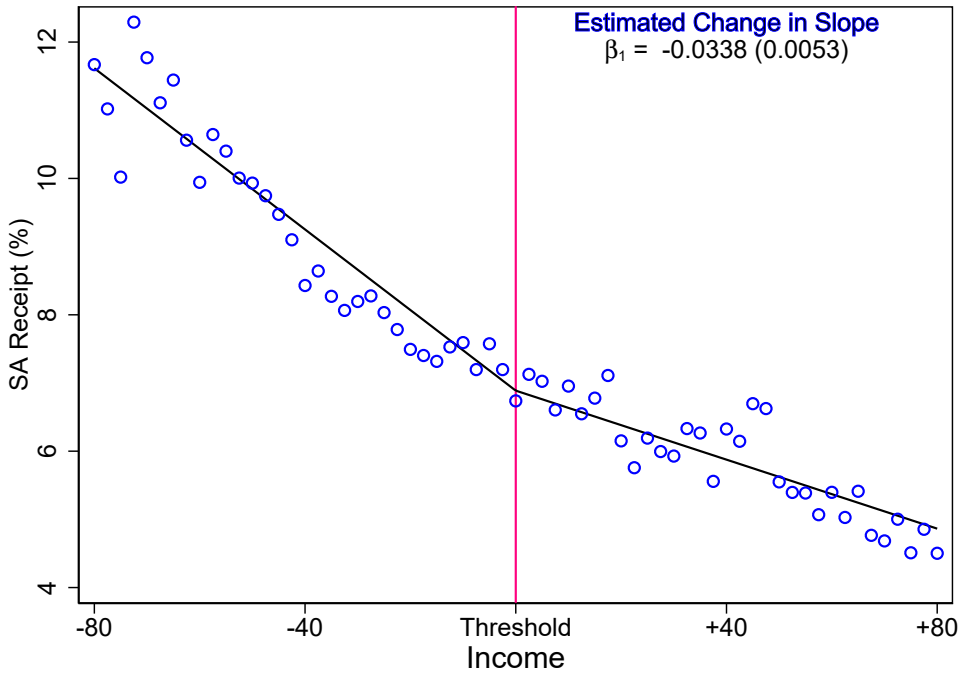


FIGURE D.12. Average rate of receipt within income slice in a small window of income either side of the eligibility threshold. Income in this plot is monthly. The sample contains singles who get most income from labour, years 2011-2014. See Section 6.1.2 for details on sample restrictions. Regression 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	-0.0253***	-0.0189***	-0.0263***	-0.0192***	-0.0199***	-0.0135***	-0.0221***	-0.0146**
Overall: everyone, Het: good MH	(0.00255)	(0.00221)	(0.00268)	(0.00240)	(0.00328)	(0.00288)	(0.00348)	(0.00313)
Income - Threshold (het)			0.00795	0.00276			0.0173	0.00928
Het: poor vs good MH			(0.00688)	(0.00678)			(0.0095)	(0.0091)
$\min\{$Income - Threshold$, 0\}$	-0.0338**	-0.0318***	-0.0263***	-0.0263***				
Overall: everyone, Het: good MH	(0.00530)	(0.00465)	(0.00559)	(0.00501)				
$\min\{$Income - Threshold$, 0\}$ (het)			-0.0556***	-0.0398*				
Het: poor vs good MH			(0.0161)	(0.0155)				
Benefits					0.0356***	0.0333***	0.0280***	0.0278***
Overall: everyone, Het: good MH					(0.0057)	(0.0049)	(0.0060)	(0.0053)
Benefits (het)							0.0814***	0.0653***
Het: poor vs good MH							(0.0156)	(0.0143)
Observations (people-months)	537,625	501,736	537,625	501,736	537,625	501,736	537,625	501,736
R^2	0.006	0.203	0.003	0.203				
Regressors	2	354	5	339	2	548	5	474

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

TABLE D.1. Estimates from the regression kink design using a bandwidth of €80. 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 (without / with 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.

APPENDIX E. CALIBRATION OF WELFARE EFFECTS: ADDITIONAL MATERIAL

E.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 θ in the general population, and not the eligible. In reality, the ineligible fund benefits for the recipients, and not the eligible non-takers.

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

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

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

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

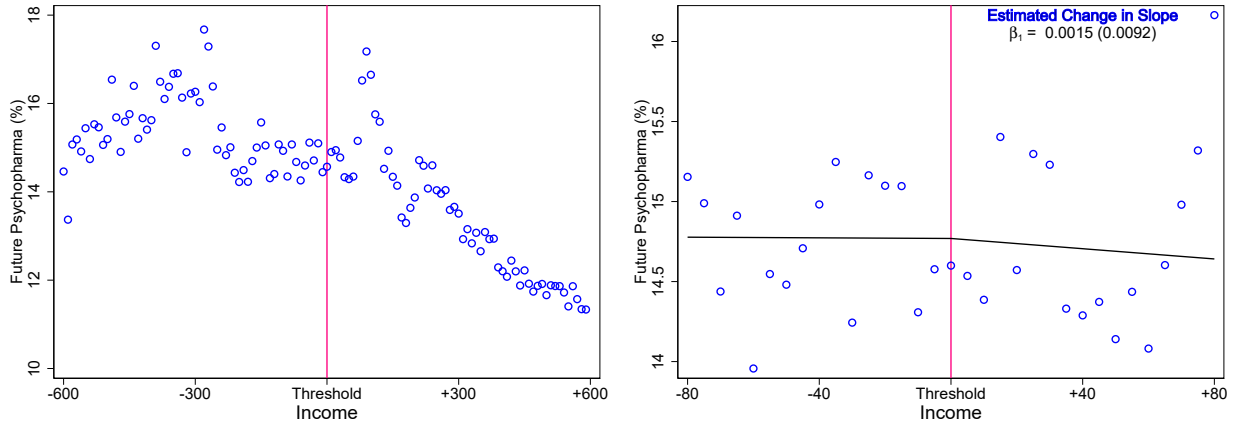
Proposition E.1 follows from Bayes Rule, the fact that eligibility is determined by $y \leq \bar{y}$ where $y = SA \cdot y^{SA=1} + (1 - SA) \cdot y^{SA=0}$ and from the fact we have assumed no labor supply responses to dB or $d\Lambda$. The intuition is as follows: we need to adjust for baseline incomplete 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 E.1 when calculating welfare effects? When integrating against average take-up levels, Bayes Rule $\rightarrow \int \mathbb{P}[SA]_\theta \cdot H_\theta d\mu = \mathbb{P}[\text{Eligible}] \cdot$

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

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	$(0.00028, 0.00065)$
v'_H	Normalization	1
$f(v_H - \kappa_H)$	RKD_H + v'_H	0.00028
$f(v_L - \kappa_L)$	$\hat{\mathbb{P}}[SA]_H = \hat{\mathbb{P}}[SA]_L$ + shortcut: $f(\cdot)_L = f(\cdot)_H$	0.00028
v'_L	RKD_L + $f(v_L - \kappa_L)$	2.3
$\left(\frac{\partial \hat{\mathbb{P}}[SA]_H}{\partial \Lambda}, \frac{\partial \hat{\mathbb{P}}[SA]_L}{\partial \Lambda}\right)$	(Diff, Diff-in-Diff)	$(-0.014, -0.023)$
(κ'_H, κ'_L)	(Diff-in-)Diff + $f(\cdot)_L = f(\cdot)_H$	$(79, 130)$

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



(A) Overview (B) Zoomed

FIGURE E.1. Comparison of RKD Overview and RKD Psych

E.2. **Identification.**

E.3. **Relaxing Modelling Assumptions.** Next, I set out how to relax the assumptions that the probability of being on the margin is the same for both instruments by imposing additional structure. Assume linearity: $v_\theta(B) = v_\theta \cdot B$ and $\kappa_\theta(\Lambda) = \kappa_\theta \cdot \Lambda$. 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: $\hat{\mathbb{P}}[SA]_L \approx \hat{\mathbb{P}}[SA]_H \implies v_L \cdot B - \kappa_L \cdot \Lambda = v_H \cdot B - \kappa_H \cdot \Lambda$. This means that $\kappa_L - \kappa_H = \frac{v_L \cdot B - v_H \cdot B}{\Lambda}$. Recall that $\frac{\partial \hat{\mathbb{P}}[SA]_H}{\partial B} = 0.00028$ and $\frac{\partial \hat{\mathbb{P}}[SA]_L}{\partial B} = 2.3 \times 0.00028$. These estimates imply $\kappa_L - \kappa_H = \frac{2.3 \times 874.54 - 865.27}{\Lambda}$.

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, $\frac{\partial \hat{\mathbb{P}}[SA]_L - \hat{\mathbb{P}}[SA]_H}{\partial \Lambda} = -(\kappa_L - \kappa_H) \cdot \alpha \cdot 0.00055$, as $\frac{\partial \hat{\mathbb{P}}[SA]_H}{\partial B} = f_\varepsilon^{dB}$. Rearranging for α,

$$\hat{\alpha} = \frac{-\Lambda \frac{\partial \hat{\mathbb{P}}[SA]_L - \hat{\mathbb{P}}[SA]_H}{\partial \Lambda}}{0.00028 \times 945.03} \tag{E.3}$$

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. The annexes to SCP (2019) contain a survey of 80 municipalities and questions on how often they impose obligations, and for each type of obligation how many impose these more after the introduction of the Participation Act. The results are shown in Table E.2.

Obligation	**Percent 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 E.2. Obligations and Percentages

I use this 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,

$$-\Lambda \frac{\partial \hat{\mathbb{P}}[SA]_L - \hat{\mathbb{P}}[SA]_H}{\partial \Lambda} = \frac{0.09}{0.221} = 0.040$$
$$\implies \hat{\alpha} = \frac{0.040}{0.52} = 0.077$$

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

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

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

E.5. **Robustness to Bias.** Suppose a share ψ of $\kappa_\theta(\Lambda)$ is a true cost, and $(1-\psi)$ is a hassle cost, which affects behaviour but not welfare. Then: $\mathbb{P}[SA]_\theta = F_\varepsilon\left[v_\theta(B) - \kappa_\theta(\Lambda)\right]$ still, but:

$$\mathcal{U}_\theta = \int_{-\infty}^{\varepsilon_\theta^*} \left[v_\theta(B) - \kappa_\theta(\Lambda) + MI_\theta - \varepsilon\right] dF(\varepsilon) \tag{E.4}$$

where $\varepsilon_\theta^* = v_\theta(B) - \kappa_\theta(\Lambda)$ and $MI_\theta = (1-\psi) \cdot \kappa_\theta(\Lambda)$ is the marginal internality (Mullainathan et al., 2012). Note that since the true cost $\psi \cdot \kappa \leq \kappa$, behaviour over-states the ordeal-cost, so take-up is too low relative to the private optimum. This means that a marginal increase in Λ has an extra negative behavioural welfare cost coming from people moving further away from the private optimum. A marginal increase in B has an extra positive behavioural welfare gain coming from the internality correction. This is shown in Proposition E.2.

Proposition E.2. *First order welfare effects when perceived cost differs from true cost.*

$$\frac{d\mathcal{U}_\theta}{d\Lambda} = -\psi \cdot \kappa'_\theta(\Lambda) \cdot \mathbb{P}[SA]_\theta + MI_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial \Lambda} \tag{E.5}$$

$$\frac{d\mathcal{U}_\theta}{dB} = v'_\theta(B) \cdot \mathbb{P}[SA]_\theta + MI_\theta \cdot \frac{\partial \mathbb{P}[SA]_\theta}{\partial B} \tag{E.6}$$

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 (E.5). Similarly,

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

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

These first order effects imply new MVPF formulas for the welfare effect of changing benefits and barriers. The fiscal externalities are unchanged - since they depend on behaviour only. However, the direct welfare effects reflect Equations (E.5) and (E.6).

Corollary E.1. *With bias:*

$$MVPF_{d\Lambda} = \frac{-\psi \cdot \int \lambda \cdot \frac{\kappa'(\Lambda)}{v'(B)} \mathbb{P}[SA] d\mu - (1-\psi) \cdot \int \lambda \frac{\kappa(\Lambda)}{v'(B)} \frac{\partial \mathbb{P}[SA]}{\partial \Lambda} d\mu}{\int FE \cdot \frac{\partial \mathbb{P}[SA]}{\partial \Lambda} d\mu} \tag{E.7}$$

$$MVPF_{dB} = \frac{\int \lambda \mathbb{P}[SA] d\mu + (1-\psi) \cdot \int \lambda \frac{\kappa(\Lambda)}{v'(B)} \frac{\partial \mathbb{P}[SA]}{\partial B} d\mu}{\int FE \cdot \frac{\partial \mathbb{P}[SA]}{\partial B} d\mu} \tag{E.8}$$

E.5.1. *Calibration.* How does bias affect the quantification of welfare effects? This requires us to evaluate the size of MI_θ, the marginal internality for each type. According to the theory,

$$MI_\theta = (1-\psi) \cdot \kappa_\theta(\Lambda) \tag{E.9}$$

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

Therefore, given the estimates from Section 7:

$$MI_L = (1-\psi) \cdot 4.52 \cdot 130 = (1-\psi) \cdot 590$$

$$MI_H = (1-\psi) \cdot 4.52 \cdot 79 = (1-\psi) \cdot 359$$

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

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

$$MVPF_{dB}^{\psi=0} = 0.96$$

$MVPF_{d\Lambda}^{\psi=0} < MVPF_{d\Lambda}^{\psi=1}$ 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 ψ^* required to reverse the welfare ordering $MVPF_{d\Lambda} > MVPF_{dB}$. This turns out to be $\psi^* = 44\%$. 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. Finally, note that $d\Lambda$ is unsurprisingly more sensitive to bias than dB.

LONDON SCHOOL OF ECONOMICS, DEPARTMENT OF ECONOMICS

E-mail address: c.naik@lse.ac.uk