CPB Netherlands Bureau for Economic Policy Analysis

CPB Discussion Paper | 319

Do Extrinsically Motivated Mental Health Care Providers Have Better Treatment Outcomes?

Rudy Douven
Minke Remmerswaal
Robin Zoutenbier

Do Extrinsically Motivated Mental Health Care Providers Have Better Treatment Outcomes?*

Rudy Douven†, Minke Remmerswaal‡, and Robin Zoutenbier§

December 17, 2015

Abstract

In 2008, a step-wise increasing compensation scheme was introduced for self-employed mental health care providers in the Netherlands. Using a large administrative dataset, we exploit the discontinuities in the compensation scheme to separate more extrinsically from non-extrinsically motivated providers. We find that the majority of the providers are, to some degree, extrinsically motivated and strategically set treatment duration to exploit the discontinuities in the compensation scheme. In addition, non-extrinsically motivated providers treat mental health patients shorter, receive less compensation and report better treatment outcomes, as measured by the improvement in Global Assessment of Functioning. This suggests that the compensation scheme rewards inefficient or low quality providers.

Keywords: physician incentives, compensation scheme, extrinsic motivation, treatment outcomes

JEL: H51, I11, J22, M52

*We gratefully acknowledge comments and suggestions by Peter Alders, Leon Bettendorf, Roy van Egmond, Erik Schut, and seminar participants at Brandeis University, CPB Bureau for Economic Policy Analysis, Dutch Healthcare Authority, Dutch Ministry of Finance, and Harvard Medical School.

†CPB Bureau for Economic Policy Analysis and Erasmus University Rotterdam. Email: R.Douven@cpb.nl

‡CPB Bureau for Economic Policy Analysis. Email: M.Remmerswaal@cpb.nl

§CPB Bureau for Economic Policy Analysis. Email: R.Zoutenbier@cpb.nl

1 Introduction

The effectiveness of a treatment depends on many different factors that are often difficult to observe in practice, particularly so in mental health care (Frank and McGuire 2000). It can be hard for mental health care providers to determine the best treatment for a patient's condition or determine the optimal treatment duration. Treatment variation in health care may, therefore, reflect differences in decision making by different types of providers (Chandra et al. 2012). We are interested in understanding the differences in motivations between mental health care providers and the effect of these differences on treatment outcomes.

Imperfect agent models describe two important motivations for providers (Ellis and McGuire 1986, 1990); providers care for the well-being of their patients and for the compensation they receive for their services. Survey evidence by Campbell et al. (2007) shows that almost all physicians rate the well-being of their patients as the most important of the two.[1] However, when faced with different scenario's where providers have to trade-off patient well-being to economic interests, some of these providers reveal that earnings and costs of providing services are also important factors. In line with this finding, many studies in health economics invariably find that providers do respond to financial incentives and are to some degree extrinsically motivated (see McGuire 2000, Christianson and Conrad 2011, McClellan 2011, and Chandra et al. 2012 for an overview of the literature).

This paper empirically studies differences in provider‘s sensitivity to financial incentives and how these differences affect their decisions and treatment outcomes. To do so, we make use of a unique feature of the compensation scheme for providers of mental health care in the Netherlands: their compensation increases in treatment duration with large discontinuous steps. This creates an opportunity for providers to increase their total compensation by strategically choosing the duration of treatment. A simple imperfect agency model predicts that, with such a compensation scheme, providers that are more extrinsically motivated are more likely to prolong treatment to the start of a higher tariff as compared to non-extrinsically motivated providers. Moreover, we predict that longer treatment duration may also result in a higher patient benefits for patients of more extrinsically motivated

[1] In addition, Kesternich et al. (2015) experimentally show that professional norms that induce altruism, e.g. the Hippocratic oath for doctors, increase the weight providers place on patient benefits instead of their financial compensation.

providers.

We empirically test our predictions using a proprietary administrative dataset that covers the Dutch curative secondary mental health care. Our data covers the majority of treatments performed during the years 2008 to 2010. We find large variation between providers in their sensitivity to the discontinuities in the compensation scheme. The majority of the providers strategically set, at least to some degree, treatment duration to receive a higher compensation for a treatment. The extent to which providers make use of this opportunity differs between providers. Some providers only prolong some treatments while others prolong all treatments to a higher tariff. A smaller group of non-extrinsically motivated providers is not sensitive to the discontinuities. Moreover, we find that non-extrinsically motivated providers treat patients significantly shorter, report better treatment outcomes, and receive less compensation as compared to more extrinsically motivated providers. Our findings suggest an additional argument for outcome based payment. Outcome or value based payments will incentivize providers to improve treatment value and, in addition, has the advantage that it rewards the higher quality providers.

Our main contribution to the literature is that the compensation scheme allows us to test in a clean way whether providers respond to the financial incentive by providing more care. In practice, such opportunities are very rare because in most situations many contemporary factors, such as differences between provider characteristics, may also play a role (see Chandra et al. 2012).

Our work is also related to the literature that studies treatment variation. Recent empirical work shows that motivations of doctors are important and can lead to different outcomes. For example, outcomes may depend on differences in physician practice styles (Abaluck et al. 2014, Currie and Macleod 2013, and Currie et al. 2015), the knowledge of doctors (Leonard et al. 2010), the specific context (Brock 2014), the size of the payment (Clemens and Gottlieb 2014) and the type of payment (Douven et al. 2015a). The verdict about the precise interaction between various extrinsic and intrinsic motivations is still open. Some argue that strong extrinsic incentives are necessary for efficiency while others argue that when intrinsic motivation is high, such as in health professionals, strong extrinsic incentives are not necessary and may even undermine agents' intrinsic motivations (Rebitzer and Taylor, 2011).

We continue as follows. The next section provides a short overview of the

Dutch mental health care system. Section 3 shows our theoretical framework and our most important predictions. The data is explained in section 4 and our empirical strategy in section 5. Next, the results of the empirical analysis are reported and discussed in section 6 and 7. Finally, section 8 concludes.

2 The Dutch mental health care system

The Dutch mental health care sector is split into primary care, curative secondary care and long-term secondary care. Primary mental health care covers the most basic care, usually mild issues such as mild depressions or temporary mental difficulties. Secondary mental health care compromises more severe or lasting mental health problems (e.g. depression, anxiety disorders or schizophrenia). Patients treated for less than a year are treated in curative secondary care while patients treated for a year or more are treated in long-term secondary care. This paper covers curative secondary care only. Secondary curative care is the largest segment of mental health care, covering more than 1 million patients a year and accounting for almost 4 billion euro in annual costs (NZa, 2013).

Patients need a referral from their general practitioner to have access to curative secondary care. After receiving the referral, the patient is free to choose any mental health care provider. However, in practice most patients follow the advice of their general practitioner. The out-of-pocket costs for the use of mental health care were limited during the years of our study.[2] As of 2008 the mental health care is part of the health insurance scheme with regulated competition in the Netherlands. This implies that health insurers had to contract mental health services for their insured. However, during the period we study, insurers did not have an incentive to control costs because they did not incur the financial risks of the provision of mental health care.[3]

Most mental health care providers work in large regional institutions such

[2] The annual deductible was 150 euro in 2008, 155 euro in 2009, and 165 euro in 2010 for all curative services (including mental health services and excluding general practitioner care and obstetrics).

[3] During the period studied in this paper, insurers did not receive information about the exact treatment duration but only on the tariff that is declared by the provider. Hence, insurers where not able to analyse the response of providers to the compensation scheme. As of 2014 insurers do get information on the exact treatment duration. Government policy at the time under study was that a proper risk adjustment system should be implemented first, before health insurers could bear the financial risks of mental health expenditure.

as a regional facility for ambulatory care or a specialized psychiatric hospital. These institutions were compensated with fixed budgets and employees received a fixed salary. Next to the large institutions there were also a lot of self-employed specialists who work in smaller private practices. A private practice consists of a single specialist or a few cooperating self-employed specialists. These self-employed specialists cover about ten percent of the curative market. In contrast to the large institutions, these self-employed providers were compensated according to a step-wise increasing scheme.

The compensation scheme for self-employed providers is of prime interest for our study. Self-employed providers are paid for each DBC (Diagnosis Treatment Combination) they complete. DBC's are combinations of a specific diagnosis and a specific treatment for the diagnosis.[4] For example, a DBC may consist of the diagnosis 'depression' with an accompanying treatment of 1000 minutes of therapy. A DBC is closed after treatment has been completed or when one year has passed since the start of the DBC (then a new DBC may be started for the second year). The fee for a DBC is determined on the basis of the diagnosis and the total time spent on all medical activities within the DBC and the fee covers all costs, such as labor and capital costs.

The fee structure follows a discontinuous step-wise function. The total fee increases at 800, 1800 or 3000 minutes and is flat in between.[5] Fee increments are large. For example, a treatment for mood disorder is rewarded with 877 euro (2008 prices) when treatment duration is less than 800 minutes and with 1786 euro when treatment lasts between 800 to 1800 minutes.[6] Douven et al. (2015b) show that providers respond strongly to this incentive. However, they did not study possible provider variation among self-employed mental health care providers and how this affects health outcomes, as we do in this paper.

[4] The DBC is similar to a Diagnostic-Related Group (DRG) that is used in many other countries. See Westerdijk et al. (2012) for an extensive discussion on DBC's.

[5] There are more tariff thresholds at higher levels of treatment duration. Treatment durations of more than 4000 minutes are very sparse in secondary mental health care. We therefore focus at the thresholds below 4000 minutes.

[6] Note that financial rewards are even higher for other thresholds, with a fee of 3297 euro for a treatment duration between 1800 and 3000 minutes and 5611 euro for more than 3000 minutes of treatment. The thresholds are the same for all mental health diagnosis but threshold fees may differ slightly across diagnosis.

3 Theoretical framework

We formulate a simple imperfect agency model to study how financial incentives affect a provider's choice of treatment duration. Our model builds on previous models by Ellis and McGuire (1986, 1990), Rebitzer and Taylor (2011) and Chandra et al. (2012). In our model, a provider j decides on the duration of treatment, denoted by x_i, for patient i. The provider's utility depends on three components: the benefit to the patient B, the financial compensation π, and the cost of providing treatment c. We thus implicitly assume that all providers are intrinsically motivated and, therefore, take the patient's benefits into account when deciding on a treatment plan. The utility for provider j when treating patient i is given by:

$$U_{ij} = B_j(x_i, \theta_i) + \alpha_j \pi(x_i) - c_j x_i. \tag{1}$$

The patient's benefit from treatment as assessed by provider j is given by $B_j(x_i, \theta_i)$, where θ_i represents the severity of a patient's condition. Subset j indicates that B_j may differ across providers, for example because providers differ in treatment quality or in their subjective assessment about patient's benefit. In addition, we assume that $B_j^{'}(x) \geq 0$ and $B_j^{''}(x) < 0$. That is, every patient benefits from some treatment and longer treatment leads to higher patient benefit but at a decreasing rate.[7]

Providers additionally receive a compensation $\pi(x_i)$ that depends on the duration of a treatment. This compensation scheme is the same for all providers. Providers are compensated according to a step-wise compensation scheme (as in the actual compensation scheme explained in the previous section). The compensation scheme is modelled as follows:

$$\pi(x_i) = \pi_l \text{ for } k_l \leq x_i < k_{l+1}. \tag{2}$$

Figure 1 illustrates a step-wise increasing compensation scheme. Treating a patient for x minutes, where $k_l \leq x < k_{l+1}$, results in a compensation of π_l. Increasing the duration to $x > k_{l+1}$ (but lower than k_{l+2}) increases

[7] The assumption of $B_j^{'}(x) \geq 0$ implies that longer treatment never results in worse outcomes. This, for example, may not hold in hospital care where infections or wrong procedures may result in a decline in patient benefit. However, treatment in mental health care consists mostly of face-to-face contact, with very little risks of complications. Moreover, Owen et al. (2014) provide some evidence for the assumption $B_j^{''}(x) < 0$ in mental health care.

the compensation to $\pi_{l+1} > \pi_l$. Additionally, we assume that providers differ their valuation of financial rewards. We denote provider j's in their sensitivity to financial incentives by $\alpha_j \geq 0$. In this theoretical exercise we label providers who put no weight to the compensation scheme as non-extrinsically motivated providers, $\alpha_j = 0$. The higher $\alpha_j > 0$, the more extrinsically motivated the provider.

Finally, the cost of providing the treatment is represented by a simple linear cost function $c_j x_i$. Each provider has costs for improving the patient's well-being and may additionally have other production costs or indirect costs related to treatment duration. Subscript j captures that providers may differ in their cost of providing treatment.

Prediction 1
If the benefit to the patient and the cost of treatment are equal across providers (i.e. $B_j = B$ and $c_j = c$) then for a given set of patients $\{\theta_i | i \in I\}$, more extrinsically motivated providers (i.e. providers with higher α_j's) are more likely to prolong treatment to reach the next tariff threshold.

A utility maximizing provider chooses optimal treatment duration by comparing marginal benefits to marginal costs. Non-extrinsically motivated providers only take patient benefits and the cost of treatment into account. These providers do not take the compensation scheme into account when deciding on the optimal treatment duration. Optimal treatment duration for a non-extrinsically motivated provider is given by:

$$\frac{\partial B(x_i, \theta_i)}{\partial x_i} = c. \tag{3}$$

This results in treatment duration x_i^N, with corresponding patient benefit $B_i^N = B(x_i^N, \theta_i)$ and compensation $\pi_i^N = \pi(x_i^N)$.

Extrinsically motivated providers do take the compensation scheme into account when deciding on optimal treatment duration:

$$\max_x \left\{ \frac{\partial B(x_i, \theta_i)}{\partial x_i} = c;\ x_i = k_{l \in L} \right\}. \tag{4}$$

That is, an extrinsically motivated provider either chooses the same level treatment duration as the non-extrinsically motivated provider or chooses to

prolong treatment to the next treatment duration threshold of the compensation scheme.[8] Suppose that a non-extrinsically motivated provider decides to treat a patient for $x^N < k_{l+1}$ minutes. An extrinsically motivated provider will decide to treat the same patient for $x^E = k_{l+1}$ minutes as long as:

$$B(k_{l+1}, \theta_i) - B(x_i^N, \theta_i) + \alpha_j(\pi(k_{l+1}) - \pi(k_l)) \geq c(k_{l+1} - x_i^N) \qquad (5)$$

Since $B'(x) > 0$ and $(\pi(k_{l+1}) - \pi(k_l))$ is positive and constant, it is straightforward to show that more extrinsically motivated providers (those with higher α_j's) will prolong treatment for a larger subset of patients in $\{\theta_i | i \in I\}$.[9]

Prediction 2
If the benefit to the patient and the cost of treatment are equal across providers (i.e. $B_j = B$ and $c_j = c$) then for a given set of patients $\{\theta_i | i \in I\}$, more extrinsically motived providers (i.e. higher α_j's) have equal or higher patient benefits.

Figure 2 illustrates the difference between an extrinsically motivated provider and a non-extrinsically motivated provider. Following prediction 1, a more extrinsically motivated provider will in more cases prolong treatment to $x_i^E = k_{l+1}$ and thus on average will treat longer and have higher patient benefit outcomes (because B_j is monotonically increasing). However, the prolongation of treatment may come at a welfare loss if the benefit to the patient by treating longer is smaller than additional treatment costs: $B_i^E - B_i^N < \pi(k_{l+1}) - \pi(k_l) + c(x_i^E - x_i^N)$.[10]

4 Data and descriptive statistics

To test our predictions we use a large administrative data set from the Dutch Health Care Authority (NZa). This data set covers all secondary mental health care treatments that took place in the Netherlands between 2008 and 2010. We restrict our sample in a number of ways. First, we consider

[8] Note that because $\pi(x)$ changes in discontinuous steps, it has no derivative at the points $x = k_{l \in L}$ and a derivative of zero between these points.

[9] Note that providers only have an incentive to prolong treatment to k_{l+1} and no incentive to shorten treatment to k_l. This follows from equation 5 and the fact that x_i^N is optimal for the non-extrinsic provider.

[10] For example, if x_i^N is already on the "flat of the curve" then all $x_i^E > x_i^N$ will result in a welfare loss.

only regular DBC's. Treatments with emergency care or short treatments are excluded, as these were compensated differently. Moreover, we exclude treatments that last less than 200 minutes or more than 4000 minutes from our sample as these are very rare. Since we are interested in the decision of the provider, we also exclude treatments that were ended because the patient decided to stop the treatment. Lastly, to obtain enough power to discriminate across providers, we exclude all self-employed providers for whom we have less than ten observed treatments per year. This results in a sample of 1,440 self-employed providers covering 187,976 treatments.

Table 1 reports some descriptive statistics of the sample. Providers are obligated to record all time spent on a patient within a single treatment. The average treatment duration in our sample is 1085 minutes. If a patient suffers from multiple mental illnesses then the provider will have to record each illness as a separate DBC. The recorded time may consist of direct face-to-face time with the patient (both individual and in a group), indirect time (e.g., administration or consulting colleagues), daytime activities or an entire day of in-house guidance. Total treatment duration is determined by adding up weighted treatment time in the various categories: the number of minutes spend on therapy (both direct and indirect time), 15 minutes of time for each hour spend on daytime activities and 60 minutes for each whole day that the patient receives in-house guidance (without staying overnight). In addition, Figure 3 shows the distribution of treatment duration in the sample. There are clear gaps and bunches around the compensation tariff thresholds of 800, 1800 and 3000 minutes. The density is lower just before the start of a new tariff and higher just after. Douven et al. (2015b) showed that these gaps and bunches were not present for the providers in the Netherlands that were not paid according to the step-wise compensation scheme. This suggests that the duration of some treatments is set strategically by providers so as to increase total compensation.[11]

Another key variable of interest is the outcome of a treatment. We approximate the health outcome using the widely applied Global Assessment of Functioning (GAF). The GAF is a subjective assessment by the provider regarding the mental well-being of the patient and the patient's functioning

[11] Note that there are strong similarities with the bunching literature as discussed by Kleeven (2015). According to that literature we study notch points. Our utility function is more complex than in general because it includes benefit to the patient, but compared to the tax literature for example we have more observations for single persons since providers treat many patients during a year.

in daily life. The GAF is measured on a ten-point scale ranging from "[1] Persistent danger of severely hurting self or others (e.g., recurrent violence), persistent inability to maintain minimal personal hygiene or serious suicidal act with clear expectation of death." to "[10] Superior functioning in a wide range of activities, life's problems never seem to get out of hand, is sought out by others because of his or her many positive qualities. No symptoms." (American Psychiatric Association, 2000).[12] A provider scores each patient's GAF once at the start of a treatment (simultaneous to determining the diagnosis for the treatment) and once after treatment has ended. Reporting the GAF for every patient is mandatory in the Netherlands (DBC Maintenance, 2009).

The average GAF in our sample at the start of a treatment is 6.19 on a scale to ten. The average GAF increases with 0.98 point to 7.17 after a treatment is closed. Figure 4 plots the distribution of the difference in GAF before and after the treatment. There is some variation in the difference in GAF before and after treatment, but almost 99 percent of all treatments have a difference in GAF between minus one and plus four.

Table 1 also shows some other treatment characteristics. Almost 75 percent of all treatments cover one of the five main diagnosis: mood, anxiety, personality, adjustment, or early childhood disorders. Mental health care providers also record the reason for ending a treatment. Almost 52 percent of treatments is closed for administrative reasons. This implies that the treatment has lasted for more than a year and a new DBC is started in the next year. Next, 44 percent of treatments is closed after treatment is completed and 4 percent of treatments are ended because of other case related reasons (such as a referral to another specialist). The number of treatments provided by self-employed providers increases over the years, 30 percent of all treatments in our sample were provided in 2008 while 35 percent were provided in 2010. Table 1 also shows some provider characteristics. There are two types of providers in our sample: psychologists and psychiatrists. An important difference between these specialists is that a psychiatrist has a medical degree and is legally allowed to prescribe medicine. Roughly 67 percent of all treatments are provided by psychologists and 33 percent by psychiatrists. Finally, providers perform, on average, 54 treatments per year. There are some large providers in the data, but for 90 percent of all providers there are

[12] The GAF ranges from 1 to 100, but is measured in ten categories of ten GAF points (1-10, 11-20, ... , 91-100).

100 observations or less each year.

5 Empirical strategy

To test our two predictions, we first construct a measure of each provider's degree of extrinsic motivation, similar to α_j in (1). We do not estimate each providers' utility function separately, but we examine how each provider responds to the tariff thresholds of the step-wise increasing compensation scheme. First, we look at the entire distribution of treatment duration for each individual provider each year. Following our theoretical framework, we expect that providers who are more likely to choose treatment duration just before a tariff threshold are less extrinsically motivated while providers that choose more often treatment durations just after a tariff threshold are more extrinsically motivated. We denote all treatments of duration t by provider j as n_{jt} and the total number of observed treatments of provider j as N_j. Next, we count all treatments in the interval $[T-250, T-1]$ minutes before the thresholds T, denoted by $\sum_T \sum_{T-250}^{T-1} n_{jt}$, and all treatments in the interval $[T, T+250]$ minutes after the thresholds, denoted by $\sum_T \sum_{T}^{T+250} n_{jt}$. Finally, we estimate provider j's sensitivity to financial incentives by:

$$\hat{\alpha}_j = (1 - \frac{\sum_T \sum_{T-250}^{T-1} n_{jt}}{N_j})(\frac{\sum_T \sum_{T}^{T+250} n_{jt}}{N_j}), \text{ with } T \in 800, 1800, 3000. \quad (6)$$

That is, for each provider we multiply one minus the fraction of treatments in the interval $[T-250, T-1]$ before the threshold with the fraction of treatments in the interval $[T, T+250]$ after the threshold. Providers with relatively many treatments in $[T-250, T-1]$ score low on the extrinsic measure, $\hat{\alpha}_j$ and providers with relatively many treatments in $[T, T+250]$ score high on the extrinsic measure. Note that within the interval $[T-250, T+250]$ extrinsically motivated providers have on average longer treatment durations than non-extrinsically motivated providers, but this doesn't necessarily mean that they treat longer overall. For example, we may find the opposite result if more extrinsically motivated providers decide to end most treatments in $[T, T+250]$ where less extrinsically motivated providers decide to prolong treatment duration.[13]

[13] In general using smaller intervals to measure $\hat{\alpha}_j$ in (1), of for example $[T-50, T+50]$ minutes instead of $[T-250, T+250]$ minutes, would be preferable. In section 7 we show that

Next, we test prediction 1 by estimating differences in treatment duration between extrinsically and non-extrinsically motivated providers using OLS. The independent variable is the observed treatment duration x_{ij} for patient i performed by provider j. The main variable of interest is the degree of extrinsic motivation. We will use a dummy variable E_j that is equal to one when providers are extrinsically motivated ($\hat{\alpha}_j > \bar{\alpha}$) and equal to zero when non-extrinsically motivated ($\hat{\alpha}_j \leq \bar{\alpha}$), where $\bar{\alpha}$ is the value of $\hat{\alpha}_j$ based on a smoothed distribution without gaps and bunches (see Appendix A).[14] The difference in treatment duration between extrinsically and non-extrinsically motivated providers is denoted by ψ. We control for a number of treatment characteristics and provider characteristics denoted by y_i and z_j respectively.

$$x_{ij} = \psi E_j + y_i'\beta + z_j'\delta + \varepsilon_{ij}. \tag{7}$$

Lastly, we test prediction 2 by examining whether there is a difference between extrinsically and non-extrinsically motivated providers in treatment outcomes, as measured by the improvement or deterioration in a patient's GAF. We use an ordered probit regression to estimate the difference between these two groups, as the change in GAF is measured on an ordered categorical scale. The effect of the extrinsic dummy E on a latent variable ΔGAF_{ij}^* is estimated by:

$$\Delta GAF_{ij}^* = \gamma E_j + y_i'\beta + z_j'\delta + \varepsilon_{ij}, \text{where } \Delta GAF_{ij} = \begin{cases} -9 \text{ if } \Delta GAF_{ij}^* < \tau_1 \\ -8 \text{ if } \tau_1 \leq \Delta GAF_{ij}^* < \tau_2 \\ \vdots \\ 8 \text{ if } \tau_{18} \leq \Delta GAF_{ij}^* < \tau_{17} \\ 9 \text{ if } \tau_{19} \leq \Delta GAF_{ij}^* < \tau_{18} \end{cases} \tag{8}$$

The effect of being an extrinsically motivated provider on the latent GAF difference is denoted by γ. Furthermore, a higher score on the latent difference in GAF translates to a higher probability of being in a higher ΔGAF

using the shorter intervals of 50 minutes, and to a lesser extent also 150 minutes, has the disadvantage of wrongly classifying extrinsically motivated providers as non-extrinsically motivated. We also performed a visual inspection of each providers' distribution function to investigate whether our measure $\hat{\alpha}_j$ was plausibly estimated. We show some examples of provider distributions in Appendix B. In total there are 1440 different graphs; one for every provider. These graphs are available from the authors upon request.

[14] Classifying providers in two groups allows us to compare patient characteristics of both groups. In section 7, we show that our results also hold when using the continuous variable $\hat{\alpha}$ in our estimations.

category. The difference in actual GAF can take any integer value between -9 and 9 (because the start GAF and end GAF range between 1 and 10). The latent GAF difference is translated to the actual GAF difference using the estimated ordered category thresholds, denoted by τ_k.

6 Results

The distribution of providers' estimated responsiveness to financial incentives (represented by $\hat{\alpha}_j$ in equation 6) is reported in Figure 5. The providers responsiveness to financial incentives ranges from zero to one. The dashed vertical line shows the estimated responsiveness for a provider who does not take the compensation scheme into account when deciding how long to treat a patient (see Appendix A for the estimation). A provider who does not take the financial incentives into account scores $\bar{\alpha} = 0.15$ on our measure of responsiveness.[15] We assume that providers positioned to the right of the dashed line ($\hat{\alpha}_j > \bar{\alpha} = 0.15$) are extrinsically motivated, while providers positioned to the left of the dashed line ($\hat{\alpha}_j \leq \bar{\alpha} = 0.15$) are non-extrinsically motivated.

We find a lot of variation in sensitivity to financial incentives.[16] The distribution of sensitivity is slightly centred to the right side of the dashed line, implying that the majority of providers do, to some extent, take the discontinuities into account when deciding on the duration of treatment. However, there are very little observations in the extreme right side of the distribution, implying that very little providers always choose a treatment duration at the start of a new tariff. The distribution of responsiveness to financial incentives seems very stable over the years. This suggests that there is no learning over time, for instance because some providers were not aware in the first year of the possibilities to increase the total payment. Roughly 300 providers are classified as non-extrinsically motivated and 900 providers as extrinsically motivated.

Figure 6 illustrates the distribution of treatment duration split by extrinsically and non-extrinsically motivated providers. The Figure shows that

[15] In section 7 we show our results also for other choices, such as a higher cut-off point of $\bar{\alpha} = 0.21$.

[16] Appendix B shows some distributions of treatment duration for individual providers with different levels of $\hat{\alpha}_j$. These distributions show that the peaks at the start of the thresholds are clearer for providers with higher levels of $\hat{\alpha}_j$.

our procedure for splitting the providers into two groups has been successful. The distribution of treatment duration for extrinsically motivated providers shows clear gaps before the start of a new tariff and bunches right after the start of a new tariff. In contrast, the distribution of treatment duration for non-extrinsically motivated providers is fairly smooth, without gaps and bunches.

Table 2 compares both groups of providers on several treatment and provider characteristics. First, we find that patients of extrinsically motivated providers are very similar to those of non-extrinsically motivated providers. There is no significant difference in reported GAF before the start of treatment, indicating that the average severity of patients' mental health care problems is similar in both groups. There are, however, significant differences between both groups of providers in the reported GAF after completing the treatment. Extrinsically motivated providers report 0.11 points higher GAF as compared to non-extrinsically motivated providers. We find only small differences in the type of diagnosis for both groups. Extrinsically motivated providers are slightly more likely to treat patients with personality or early childhood disorders and less likely to treat patients with adjustment disorders.

Both groups of providers are also very similar on other treatment and patient characteristics. There are slightly more treatments being performed by extrinsically than by non-extrinsically motivated providers over the years. We also find that extrinsically motivated providers treat more patients in a year than non-extrinsically motivated providers. The difference is 10 patients per year and is highly significant. Finally, patients of extrinsically motivated providers are slightly younger as compared to those of non-extrinsically motivated providers.

Table 3 reports the estimated difference in treatment duration between both groups of providers. The reported standard errors are clustered at the provider-level to account for correlation of errors between treatments of the same provider. The first column includes only a dummy indicating whether the provider belongs to the extrinsic group or not. We find a positive and highly significant effect; extrinsically motivated providers treat patients 222 minutes longer than non-extrinsically motivated providers. The size of this effect is substantial and corresponds to roughly 20 percent of the average treatment duration in the sample.[17] Column 2 includes year dummies and

[17] This also results in a higher average compensation of 2106 euro (112 euro per hour) for

some provider characteristics as control variables. The difference in treatment duration between extrinsic and non-extrinsic providers reduces slightly but remains highly significant. The year dummies also contribute significantly; average treatment duration in the sample increases over the years. Moreover, we find that psychologists treat longer than by psychiatrists, which might capture the fact that psychologists handle more difficult patients or psychologists are prone to treat longer and psychiatrists rely more on medicine. Finally, the last column of Table 3 includes some treatment characteristics as control variables. Again the difference in treatment duration between both groups of providers is robust in both sign and significance. The difference reduces slightly to 172 minutes, but remains highly significant. The control variables also contribute significantly. We find that females and younger patients are treated longer. Additionally, completed treatments and treatments that are closed for case related reasons are shorter than treatments that are ended for administrative reasons. The start GAF dummies and diagnosis dummies are jointly significant. Patients that start treatment with a higher GAF (and thus a less severe diagnosis) are treated shorter than patients with a low GAF (and thus a more severe diagnosis).[18]

Next, Table 4 reports the difference in treatment outcomes, as measured by the change in GAF, between extrinsic and non-extrinsic providers. The first estimation includes only the variable extrinsic as an explanatory variable. Results show that patients that are treated by an extrinsic provider gain less in GAF as compared to patients treated by a non-extrinsic provider. This difference is small but significant at $p = .03$. The size of the effect is interpreted by comparing the estimated effect on the latent variable with the difference in ordinal category thresholds, denoted by τ_k. The difference between the estimated thresholds ranges somewhere around 1.000. This implies that the effect of extrinsic (-0.111) is comparable to around one tenth of a category in GAF difference.

Column 2 of Table 4 includes a number of provider characteristics and year dummies as control variables. The difference in GAF changes between extrinsic and non-extrinsic providers is robust both in sign and significance. Remarkably, the change in GAF seems to decline over time. Moreover, pa-

extrinsically motivated providers and 1614 euro (107 euro per hour) for non-extrinsically motivated providers.

[18] We also performed our main regressions with interactions between the start GAF dummies and main diagnosis dummies as control variables. This did not influence our main results.

tients treated by a psychologist are rated with lower changes in GAF as compared to patients treated by psychiatrists. The change in GAF is also lower for larger providers (as measured by the number of patients).

The third column includes a set of treatment specific control variables. The difference between extrinsic and non-extrinsic providers is robust, the effect decreases marginally from -0.101 to -0.093 and remains significant ($p = .03$). The patient characteristics, such as gender and age, seem to matter very little for the change in GAF. The reason for ending the treatment is important. Patients for whom the treatment is completed are scored with an increase in GAF of almost one point more than patients whose treatment is ended administratively. Those treatments that are ended for case related reasons score between the completed treatments and administratively ended treatments. Last, the start GAF dummies and diagnosis dummies have a strong and significant impact on the change in GAF. Treatments that start with a low start GAF score show a strong increase in GAF, while those with a high start GAF show a low increase. This might reflect that it is easier to increase the GAF when the start GAF is low. In addition, there is a lot of variation between diagnosis groups in the change in GAF. A possible explanation may be that some diagnoses are easier to treat than others.

7 Robustness and extensions

We perform a number of different robustness checks. First, we investigate how sensitive our results are for choosing different intervals before and after the discontinuity threshold when measuring provider's sensitivity to financial incentives $\hat{\alpha}_j$. Table 5 reports the results when using different intervals. The first column classifies providers as extrinsically motivated when they have more treatments up to 50 minutes after the thresholds and less treatments 50 minutes before the thresholds. We find much weaker results as compared to our main results (see column 3 of Table 5). The average treatment duration in the non-extrinsically motivated group offers an explanation. In column 1, average treatment duration is the same for extrinsically and non-extrinsically motivated providers. However, for intervals of 150, 250, or 350 minutes, the average treatment duration in the non-extrinsically motivated group drops but remains stable in the extrinsically motivated group. This indicates that the specification in column 1 does not capture all extrinsically motivated

providers and, that longer intervals are needed to capture more providers.[19]

Next, we investigate how robust our findings are for the choice of the cut-off value ($\bar{\alpha} = 0.15$) that we use to split providers into extrinsically and non-extrinsically motivated providers. There may be some uncertainty in our cut-off value. For instance, providers with an $\hat{\alpha}_j$ just above the cut-off value may not be a lot different from those just below the cut-off value. To investigate such issues we perform our analyses using a cut-off value of $\bar{\alpha}+1/2\text{SD}= 0.22$. The first column of Table 6 compares the extrinsically and non-extrinsically motivated providers based on this higher cut-off value. The difference in treatment duration and GAF-score are smaller when using the higher cut-off value. This suggests that providers with $\bar{\alpha} < \hat{\alpha} < \bar{\alpha}+1/2\text{SD}$ are more similar to the extrinsically motivated providers than the non-extrinsically motivated providers.[20] We also performed our analysis while excluding providers with an $\hat{\alpha}_j$ just above the threshold value. Column 2 of Table 6 compares non-extrinsically motivated providers (with $\hat{\alpha}_j < \bar{\alpha} = 0.15$) with very extrinsically motivated providers (with $\hat{\alpha}_j > \bar{\alpha}+1/2\text{SD}_{\hat{\alpha}} = 0.22$), omitting all providers in between. We find that omitting the providers with $\hat{\alpha}_j$ just above the cut-off does not significantly change our main results. Very extrinsically motivated providers treat, on average, 192 minutes longer. This difference is slightly higher as compared to the main regression. In addition, non-extrinsically motivated providers report slightly higher increases in GAF scores. Finally, in the last column of Table 6 we do not separate both groups but include the continuous variable $\hat{\alpha}_j$ in our regression. Again the results are robust in both sign and significance, indicating that our results do not depend on the binary specification E_j. We have additionally tested a quadratic term of $\hat{\alpha}_j$ in our estimation, but we did not find a significant effect of the quadratic term.

Table 7 investigates whether our results are driven by providers with very

[19] This is also visible in the distribution of treatment duration for extrinsically and non-extrinsically motivated providers. When we use an interval of 50 minutes, the distribution of treatment duration for both the extrinsically and non-extrinsically motivated providers shows clear bunches after the tariff thresholds. We also tested for other screening devices to separate extrinsically from non-extrinsically motivated providers. For example, by comparing only the number of treatments after the tariff threshold. This does not result in large differences in the classification of providers.

[20] This is also clear from the distribution of treatment duration for these providers. The distribution for providers with $\bar{\alpha} < \hat{\alpha} < \bar{\alpha}+1/2\text{SD}$ shows clear bunches after the start of a higher tariff. This indicates that these providers are, at least to some extent, extrinsically motivated.

few or with many observed treatments in the data. Column 1 reports the regression results for providers with at least 20 observations (treatments) per year. The difference in treatment duration between extrinsically and non-extrinsically motivated providers is robust in sign and significance. The size of the effect increases slightly to 181 minutes (as compared to 172 minutes in the main regression). Likewise, the difference in the change in GAF is also robust; extrinsically motivated providers report a slightly lower increase in GAF as compared to non-extrinsically motivated providers. The second column includes all providers with 40 observations or more in a year. Again, both the difference in treatment duration and in change in GAF are robust in sign and significance. The differences are slightly stronger than in the first column, indicating that the effect of the extrinsic variable is stronger for larger providers than for smaller providers. This is also supported by the results in column 3. Column 3 of Table 7 excludes the largest providers in our sample of self-employed, those with more than 200 observations in a year. The results excluding the large providers are slightly smaller than in the main regression but still significant.[21]

So far, we did not discuss the possibility that providers fill in artificial treatment durations. This is difficult to prove with our dataset, but one way to obtain more information about this possibility is to study treatments that were not ended by providers but by patients. In our main results we have omitted treatments ended by the patient because we are mainly interested in the provider's behaviour. There may be various reasons for a patient to end treatment, for example, when the patient moves to another location, when a patient does not show up for a year or when a patient passes away. Figure 7 shows the distribution of treatment duration of all treatments ended by the patient. Remarkably, we find that extrinsically motivated providers still prolong duration up until the start of a threshold (as can be seen from the clear peaks) while non-extrinsically motivated providers do not. These peaks might also indicate supplier induced demand, in its very specific meaning whereby providers increase utilization by changing patients' desired treatment (see Chandra et al. 2012). Using the same separation de-

[21] We additionally performed a robustness check excluding all providers with 70 percent or more of all observations below 700 minutes of treatment duration. These providers may be difficult to classify as extrinsically motivated or non-extrinsically motivated. We find that only few providers have more 70 percent or more of there observations below 700 minutes. Excluding these providers does not significantly change the sign and significance of the main results.

vice as before, but now for treatment durations that were ended by patients, we identify 13% of extrinsically motivated providers with presumably many artificial treatment durations.[22] Running our regressions without this group of extrinsically motivated providers did not change our main results.

8 Concluding remarks

This paper has used discontinuities in the compensation scheme for secondary mental health care to separate extrinsically motivated from non-extrinsically motivated providers. These discontinuities provide a strong financial incentive to strategically prolong treatment duration beyond the start of a higher tariff. Our results show that many providers are sensitive to the discontinuities in the compensation scheme and prolong treatment duration so as to increase the total compensation. There is, however, also a smaller group of providers that disregards the discontinuities in the compensation scheme despite the large financial incentive. These providers may, for instance, be unaware of the incentive scheme, found it too costly to figure out how the financial incentive scheme works, or were intrinsically motivated to provide the best treatment to the patient irrespective of financial incentives or their own benefit.

In contrast to the predictions of our simple imperfect agency model, we do not find that extrinsically motivated providers have better treatment outcomes (as measured by the change in GAF during treatment). There may be several explanations for our findings. First, non-extrinsically motivated providers may produce better treatment outcomes, i.e. $B'_N > B'_E$ in our agency model. This implies that there could be a strong correlation between being an non-extrinsically motivated provider and being a high quality provider. Important to note here is that we have no information about provider differences in assessing patients benefits, for example some providers may be more pessimistic than others when assessing GAF improvements.

An additional explanation may be that the cost of treatment for extrinsically motivated providers is higher (that is: $c_N < c_E$). This might be a plausible assumption when extrinsically motivated providers also value their financial compensation higher. As a result, extrinsically motivated providers may treat some patients shorter (when they are not near the thresholds) and

[22] Many extrinsically motivated providers had very few or no treatment durations ended by patients making it impossible to judge the possibility of artificial treatment durations.

other patients longer (to reach the next threshold) than non-extrinsically motivated providers. This could, on average, result in longer treatment durations and lower health outcomes for extrinsically motivated providers.

An important limitation of our study is that we only have one outcome indicator (changes in GAF) as a proxy for treatment outcomes. The GAF is a widely used measure and its reliability and validity have been studied extensively. There has been some critique on the GAF as measure of mental well-being, however there are also a number of studies that show that the GAF is a reliable indicator, especially when comparing groups of patients (Jones et al. 1995, Soderberg et al. 2014). In recent years, more quality indicators for Dutch mental health care have become available, such as routine outcome monitoring. Unfortunately these alternative outcome measures are not yet available for research. However, in future research it would be interesting to see whether our results also hold for different measures.

In 2013 the Dutch government decided that all providers, including the large mental health institutions with salaried employees, should be paid according to the step-wise increasing compensation scheme. This may create the risk of crowding out intrinsic motivation when employees have to alter their treatment duration to obtain more income for their employer. A recent article in a Dutch newspaper provides some anecdotal evidence. A mental health care provider indicated in an interview that she considered to quit her practice because the compensation system was too much focussed at financial earnings instead of on patient benefits.[23]

The uneven compensation between low and high quality providers could be solved in a number of ways. First, starting in 2014, health insurers are able to observe the exact duration of individual treatments (before 2014 they were only able to observe the declared compensation). As a result, insurers are better able to discriminate between individual providers and address possible bunching by providers. Another way is to improve the step-wise compensation scheme by changing the location (and the number) of tariff thresholds (see also Douven et al. 2015b). Lastly, population based payment mechanisms could be introduced, preferably with suitable quality indicators (see McClellan et al. 2014). However, finding suitable performance measures to base payment on is difficult in mental health care. For example, the

[23] M. Effting (February 13th, 2015), Tjak, tjak, volgende patient: de ontsporing van de GGZ (in Dutch). *de Volkskrant*. Retrieved from http://www.volkskrant.nl/binnenland/tjak-tjak-volgende-patient-de-ontsporing-van-de-ggz a3850376/

provider reported GAF score is not suitable because it is sensitive to gaming by the provider.

References

[1] Abaluck, J., L. Agha, C. Kabrhel, A. Raja, and A. Venkatesh (2014), Negative Tests and the Efficiency of Medical Care: What Determines Heterogeneity in Imaging Behavior? NBER Working Paper No.19956.

[2] American Psychiatric Association (2000), *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition. Washington, DC: American Psychiatric Association.

[3] Brock, J.M., A. Lange, and K.L. Leonard (2015), Generosity Norms and Intrinsic Motivation in Health Care Provision: Evidence from the Laboratory and Field. EBRD working paper No. 147, Forthcoming at the *Journal of Human Resources*.

[4] Campbell, E.G. et al. (2007), Professionalism in Medicine: Results of a National Survey of Physicians, *Annals of Internal Medicine*, Vol. 147(11): 795-802.

[5] Chandra, A., D. Cutler, and Z. Song (2012), Who ordered that? The Economics of Treatment Choices in Medical Care. In: *Handbook of Health Economics Vol. II*, Pauly, M.V., T.G. McGuire, and P.P. Barros (Eds.), Elsevier, pp.397-432.

[6] Christianson, J.B. and D. Conrad (2011), Provider Payment and Incentives. In: *The Oxford Handbook of Health Economics*, S.A. Glied and P.C. Smith (Eds.), Oxford University Press, pp. 624-628.

[7] Clemens, J. and J. D. Gottlieb (2014), Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health, *American Economic Review*, Vol.104(4) pp. 1320-1349.

[8] Currie, J. and W. B. MacLeod (2013). Diagnosis and unnecessary procedure use: Evidence from C-section. Technical report, National Bureau of Economic Research.

[9] Currie, J., W. B. MacLeod, and J. Van Parys (2015), Physician Practice Style and Patient Health Outcomes: The Case of Heart Attacks, NBER Working Paper No.21218.

[10] DBC Maintenance (2009), Spelregels DBC-registratie GGZ V20091001 (in Dutch).

[11] Douven, R., R. Mocking, and I. Mosca (2015a), The Effect of Physician Remuneration on Regional Variation in Hospital Treatments, *International Journal of Health Economics and Management*, Vol.15(2): 215-240.

[12] Douven, R., M. Remmerswaal and I. Mosca (2015b), Unintended Effects of Reimbursement Schedules in Mental Health Care (2015), *Journal of Health Economics*, Vol.42: 139-150.

[13] Ellis R.P. and T.G. McGuire (1986), Provider Behavior Under Prospective Reimbursement: Cost Sharing and Supply, *Journal of Health Economics*, Vol.5: 129-151.

[14] Ellis R.P. and T.G. McGuire (1990), Optimal Payment Systems for Health Services, *Journal of Health Economics*, Vol.9(4): 375-396.

[15] Frank, R.G. and T.G. McGuire (2000), Economics and Mental Health. In: *Handbook of Health Economics Vol. I*, Culyer, A.J. and J.P. Newhouse (Eds.), Elsevier, pp.893-954.

[16] Jones, S.H., G. Thornicroft, M. Coffey, and G. Dunn (1995), A Brief Mental Health Outcome Scale-reliability and Validity of the Global Assessment of Functioning (GAF), *British Journal of Psychiatry: The Journal of Mental Science*, Vol.166(5): 654-659.

[17] Kesternich, I., H. Schumacher, and J. Winter (2015), Professional Norms and Physician Behavior: Homo Oeconomicus or Homo Hippocratus?, *Journal of Public Economics*, Vol.131: 1-11.

[18] Kleeven, H. (2015), Bunching, Draft paper prepared for the *Annual Review of Economics*, Vol. 8, 2016.

[19] Leonard K.L. and M. C. Masatu (2010), Using the Hawthorne Effect to Examine the Gap Between a Doctors Best Possible Practice and Actual Performance, *Journal of Development Economics*, Vol.93(2): 226-243.

[20] McClellan, M. (2011), Reforming Payments to Healthcare Providers: The Key to Slowing Healthcare Cost Growth While Improving Quality?, *Journal of Economic Perspectives*, Vol. 25(2): 69-92.

[21] McClellan, M., J. Kent, S.J. Beales, S.I.A. Cohen, M. Macdonnell, A. Thoumi, M. Abdulmalik and A. Darzi (2014), Accountable Care Around The World: A Framework To Guide Reform Strategies, *Health Affairs*, Vol.33(9): 1507-1515.

[22] McGuire, T.G. (2000), Physician Agency. In: *Handbook of Health Economics Vol. 1A*, A.J. Culyer and J.P. Newhouse (Eds.), Elsevier, pp. 461-536.

[23] NZa (2013), Marktscan geestelijke gezondheidszorg: Weergave van de markt 2008-2013, Report from the Dutch Healthcare Authority (in Dutch).

[24] Owen, J.J., J. Adelson, S. Budge, S.M. Kopta, and R.J. Reese (2014), Good-enough Level and Dose-effect Models: Variation Among Outcomes and Therapists, *Psychotherapy Research*.

[25] Rebitzer, J.B. and L. J. Taylor (2011), Extrinsic Rewards and Intrinsic Motives: Standard and Behavioral Approaches to Agency and Labor Markets. In: *Handbook of Labor Economics*, O. Ashenfelter and D. Card (Eds.), Elsevier, pp. 701-772.

[26] Soderberg, P., S. Tungstrom, and B.A. Armelium (2014), Special Section on the GAF: Reliability of Global Assessment of Functioning Ratings made by Clinical Psychiatric Staff, *Psychiatric Services*.

[27] Westerdijk, M., J. Zuurbier, M. Ludwig, and S. Prins, (2012), Defining Care Products to Finance Health Care in the Netherlands, *The European Journal of Health Economics*, Vol.13(2): 203-221.

Tables

Table 1: Descriptive statistics

	Mean	Min	Max
Treatment duration (in minutes)	1085	200	4000
	(666)		
Global Assessment of Functioning (GAF)			
-Before treatment	6.19	1	10
	(0.90)		
-After treatment	7.17	1	10
	(1.09)		
-ΔGAF	0.98	-7	8
	(0.96)		
Main diagnosis (%)			
- Mood	0.26	0	1
- Anxiety	0.14	0	1
- Personality	0.12	0	1
- Adjustment	0.12	0	1
- Early childhood	0.10	0	1
- Other childhood	0.05	0	1
- Somatoform	0.02	0	1
- Other	0.18	0	1
Reason for ending treatment (%)			
- Administrative reasons	0.52	0	1
- Treatment completed	0.44	0	1
- Case related reason	0.04	0	1
Treatment year (%)			
-2008	0.30	0	1
-2009	0.34	0	1
-2010	0.35	0	1
Treatments by provider type (%)			
- Psychologist	0.67	0	1
- Psychiatrist	0.33	0	1
Treatments per provider-year	53.74	11	917
	(59.59)		

Notes: Standard deviations between parentheses. There are 1,440 providers in the sample covering 187,976 different treatments.

Table 2: Descriptives split by providers' responsiveness to incentives

	Extrinsic	Non-extrinsic	Difference
Global Assessment of Functioning (GAF)			
-Before treatment	6.19	6.20	-0.01
	(0.90)	(0.91)	
-After treatment	7.15	7.26	-0.11*
	(1.10)	(1.05)	
Main diagnosis (%)			
- Mood	0.26	0.28	-0.02
- Anxiety	0.14	0.15	0.00
- Personality	0.12	0.10	0.02**
- Adjustment	0.12	0.13	-0.01*
- Early childhood	0.10	0.06	0.04**
- Other childhood	0.05	0.04	0.00
- Somatoform	0.02	0.03	-0.01*
- Other	0.18	0.20	-0.02**
Reason for ending treatment (%)			
- Administrative reasons	0.52	0.53	-0.01
- Treatment completed	0.44	0.43	0.01
- Case related reason	0.04	0.04	0.01
Treatment year (%)			
-2008	0.30	0.33	-0.04
-2009	0.34	0.35	-0.00
-2010	0.36	0.32	0.04*
Treatments by provider type (%)			
- Psychologist	0.68	0.63	0.05
- Psychiatrist	0.32	0.37	-0.05
Treatments per provider-year	55.81	45.84	9.97***
	(61.98)	(48.66)	
Patient's gender (% female)	0.61	0.61	0.01
	(0.49)	(0.49)	
Patient's age (in years)	38.65	41.62	-2.97***
	(15.69)	(14.64)	

Notes: Standard deviations between parentheses. There are 154,694 treatments (82%) that are conducted by an extrinsically motivated provider and 33,278 treatments (18%) that are conducted by an non-extrinsically motivated provider. *, **, and *** indicate significance based on a two sided test with standard errors clustered at the provider level at the .10, .05, and .01 levels, respectively.

Table 3: Results of the regression on treatment duration

	(1)	(2)	(3)
Dependent variable: treatment duration			
Extrinsic motivation	221.869***	202.628***	171.775***
	(25.058)	(21.688)	(24.790)
Treatment year:[a]			
- 2009		71.366***	53.970***
		(11.880)	(8.726)
- 2010		116.887***	103.940***
		(15.657)	(11.993)
Psychologist:[b]		-185.925***	-270.492***
		(28.192)	(27.671)
Provider size (no. of patients)		0.063	0.009
		(0.170)	(0.146)
Female patient			36.571***
			(6.277)
Patient's age			-5.308***
			(0.351)
Reason for ending treatment:[c]			
- Treatment completed			-386.410***
			(13.780)
- Case related reason			-350.033***
			(18.895)
Start GAF dummies	No	No	Yes
Diagnosis dummies	No	No	Yes
Location dummies	No	Yes	Yes
Constant	Yes	Yes	Yes
Observations	187,976	187,976	181,755
Providers	1440	1440	1420
R-squared	0.016	0.041	0.175

Notes: Standard errors between parentheses are clustered at the provider level. [a]Reference category: "2008", [b]Reference category: "psychiatrist", and [c]Reference category: "administrative reasons". *, **, and *** indicate significance based on a two sided test at the .10, .05, and .01 levels, respectively.

Table 4: Results of the ordered probit regression on GAF outcomes

	(1)	(2)	(3)
Dependent variable: ΔGAF			
Extrinsic motivation	-0.111**	-0.101**	-0.093**
	(0.051)	(0.046)	(0.042)
Treatment year:[a]			
- 2009		-0.069***	-0.058***
		(0.016)	(0.017)
- 2010		-0.088***	-0.071***
		(0.022)	(0.024)
Psychologist:[b]		-0.176***	-0.141**
		(0.050)	(0.062)
Provider size (no. of patients)		-0.000	-0.001**
		(0.000)	(0.000)
Female patient			0.006
			(0.013)
Patient's age			-0.001
			(0.001)
Reason for ending treatment:[c]			
- Treatment completed			0.723***
			(0.031)
- Case related reason			-0.301***
			(0.041)
Start GAF dummies	No	No	Yes
Diagnosis dummies	No	No	Yes
Location dummies	No	Yes	Yes
$\tau_{[\Delta GAF=-1\rightarrow 0]}$	-2.240	-2.407	-4.567
$\tau_{[\Delta GAF=0\rightarrow 1]}$	-0.548	-0.700	-2.643
$\tau_{[\Delta GAF=1\rightarrow 2]}$	0.609	0.466	-1.312
$\tau_{[\Delta GAF=2\rightarrow 3]}$	1.438	1.298	-0.343
$\tau_{[\Delta GAF=3\rightarrow 4]}$	2.151	2.016	0.514
Observations	179,730	179,730	178,662
Providers	1424	1424	1415
Log-likelihood	-237409	-236204	-213649

Notes: Standard errors between parentheses are clustered at the provider level. [a]Reference category: "2008", [b]Reference category: "psychiatrist", and [c]Reference category: "administrative reasons". The reported cut-off values τ, represent 99 percent of the distribution of ΔGAF. This ranges from a ΔGAF of minus one up to plus four. *, **, and *** indicate significance based on a two sided test at the .10, .05, and .01 levels, respectively.

Table 5: Robustness to different intervals around the tariff thresholds

Interval (in minutes):	50	150	250	350
Panel A: dependent variable treatment duration				
Extrinsic motivation	6.540	102.756***	171.775***	246.322***
	(19.504)	(23.540)	(24.790)	(22.782)
Control variables	Yes	Yes	Yes	Yes
Observations	181,755	181,755	181,755	181,755
R-squared	0.165	0.169	0.175	0.185
Providers	1420	1420	1420	1420
% Extrinsic	0.690	0.777	0.794	0.803
Mean duration (in minutes):				
- Extrinsic	1073	1099	1113	1131
- Non-extrinsic	1086	977	903	827
Interval (in minutes):	50	150	250	350
Panel B: dependent variable ΔGAF				
Extrinsic motivation	-0.004	-0.068*	-0.093**	-0.076*
	(0.035)	(0.039)	(0.042)	(0.045)
Control variables	Yes	Yes	Yes	Yes
Observations	178,662	178,662	178,662	178,662
Log-likelihood	-213742	-213690	-213649	-213678
Providers	1415	1415	1415	1415
% Extrinsic	0.690	0.778	0.794	0.803
Mean ΔGAF:				
- Extrinsic	0.970	0.964	0.959	0.958
- Non-extrinsic	1.003	1.040	1.066	1.065

Notes: Standard errors between parentheses are clustered at the provider level. 'Interval (in minutes)' refers to the number of minutes before and after the threshold when calculating equation 6. *, **, and *** indicate significance based on a two sided test at the .10, .05, and .01 levels, respectively.

Table 6: Robustness to different specifications of the extrinsic variable

Specification:	$\bar{\alpha} + 1/2\text{SD}_{\hat{\alpha}}$	$1/2\text{SD}_{\hat{\alpha}}$ gap	Continuous
Panel A: dependent variable treatment duration			
Extrinsic motivation	110.924***	191.969***	414.909***
	(19.086)	(25.540)	(69.403)
Control variables	Yes	Yes	Yes
Observations	181,755	136,030	181,755
R-squared	0.172	0.180	0.174
Providers	1420	1305	1420
% Extrinsic	0.553	0.728	n.a.
Specification:	$\bar{\alpha} + 1/2\text{SD}_{\hat{\alpha}}$	$1/2\text{SD}_{\hat{\alpha}}$ gap	Continuous
Panel B: dependent variable ΔGAF			
Extrinsic motivation	-0.067	-0.108**	-0.317**
	(0.043)	(0.048)	(0.144)
Control variables	Yes	Yes	Yes
Observations	178,662	133,813	178,662
Log-likelihood	-213661	-160917	-213578
Providers	1415	1300	1415
% Extrinsic	0.553	0.729	n.a.

Notes: Standard errors between parentheses are clustered at the provider level. The column $\bar{\alpha}+1/2\text{SD}_{\hat{\alpha}}$ refers to the specification where all providers with $\hat{\alpha}_j \leq \bar{\alpha} + 1/2SD_{\hat{\alpha}}$ are classified as non-extrinsically motivated and those with $\hat{\alpha}_j > \bar{\alpha} + 1/2SD_{\hat{\alpha}}$ are classified as extrinsically motivated, the column $1/2SD_{\hat{\alpha}}$ gap refers to the specification where all providers with $\bar{\alpha} < \hat{\alpha}_j < \bar{\alpha}+1/2SD_{\hat{\alpha}}$ are omitted, and the column 'continuous' reports the results where the continuous version of $\hat{\alpha}_j$ is used in estimation. *, **, and *** indicate significance based on a two sided test at the .10, .05, and .01 levels, respectively.

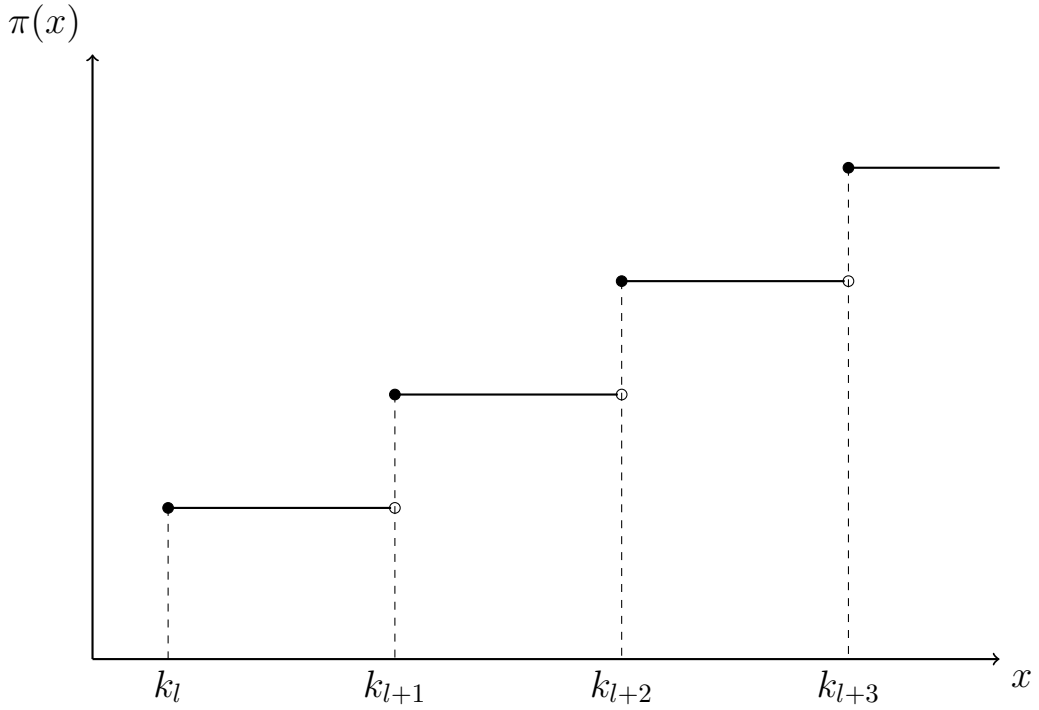
Table 7: Robustness based on samples with providers selected on size (number of treatments per year)

Observations per provider-year:	> 20	> 40	< 200
Panel A: dependent variable treatment duration			
Extrinsic motivation	180.658***	208.720***	167.900***
	(26.328)	(31.985)	(16.934)
Control variables	Yes	Yes	Yes
Observations	171,101	135,156	156,246
R-squared	0.174	0.162	0.192
Providers	1189	727	1406
% Extrinsic	0.810	0.830	0.792
Observations per provider-year:	> 20	> 40	< 200
Panel B: dependent variable ΔGAF			
Extrinsic motivation	-0.107**	-0.114**	-0.076*
	(0.046)	(0.057)	(0.045)
Control variables	Yes	Yes	Yes
Observations	168,112	132,496	154,296
Log-likelihood	-201678	-161381	-184470
Providers	1184	723	1401
% Extrinsic	0.811	0.829	0.792

Notes: Standard errors between parentheses are clustered at the provider level. Column '> 20' reports the results for a sample with all providers with more than 20 observations a year, column '> 40' reports the results for a sample with all providers with more than 40 observations a year, and column '< 200' reports the results for a sample with all providers with less than 200 observations a year. *, **, and *** indicate significance based on a two sided test at the .10, .05, and .01 levels, respectively.

Figures

Figure 1: Step-wise compensation scheme



Note: The treatment duration thresholds k_l are the same for all diagnoses where $k_1 = 800$, $k_2 = 1800$ and $k_3 = 3000$ minutes. Compensation schemes may slightly differ across diagnoses. For example, for Schizofrenia provider compensations are $\pi(k_l) = 960$ euros for $k_l < 800$ minutes, $\pi(k_l) = 1930$ euros for $800 \leq k_l < 1800$ minutes, $\pi(k_l) = 3580$ euros for $1800 \leq k_l < 3000$ minutes, $\pi(k_l) = 5880$ euros for $3000 \leq k_l < 6000$ minutes.

Figure 2: Patient's benefit from treatment

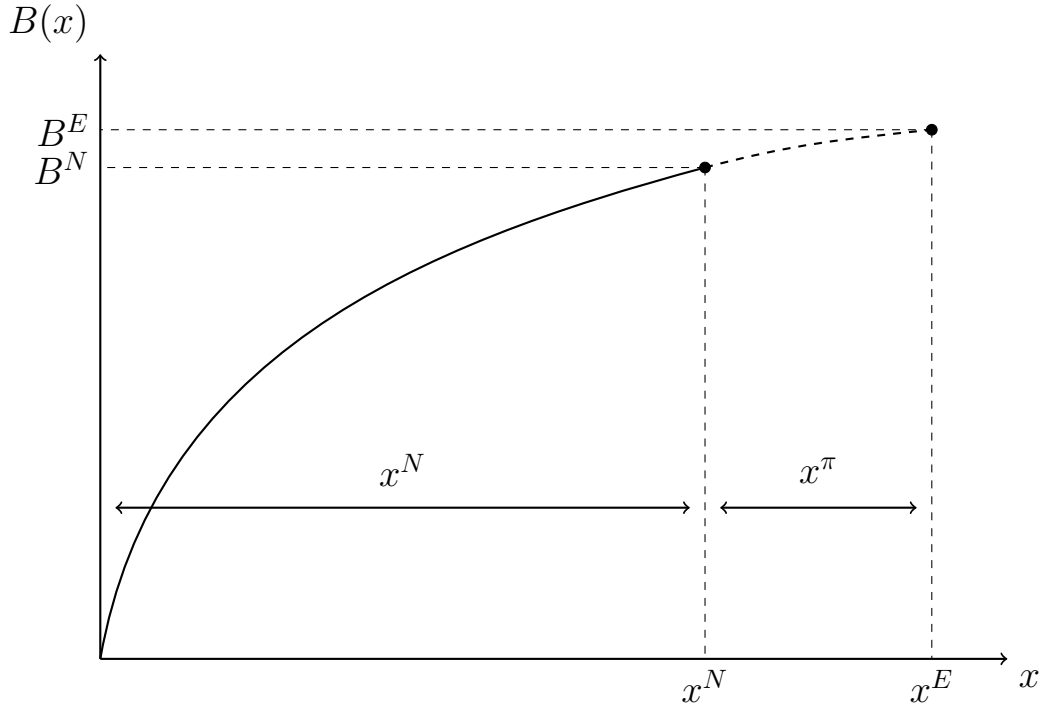
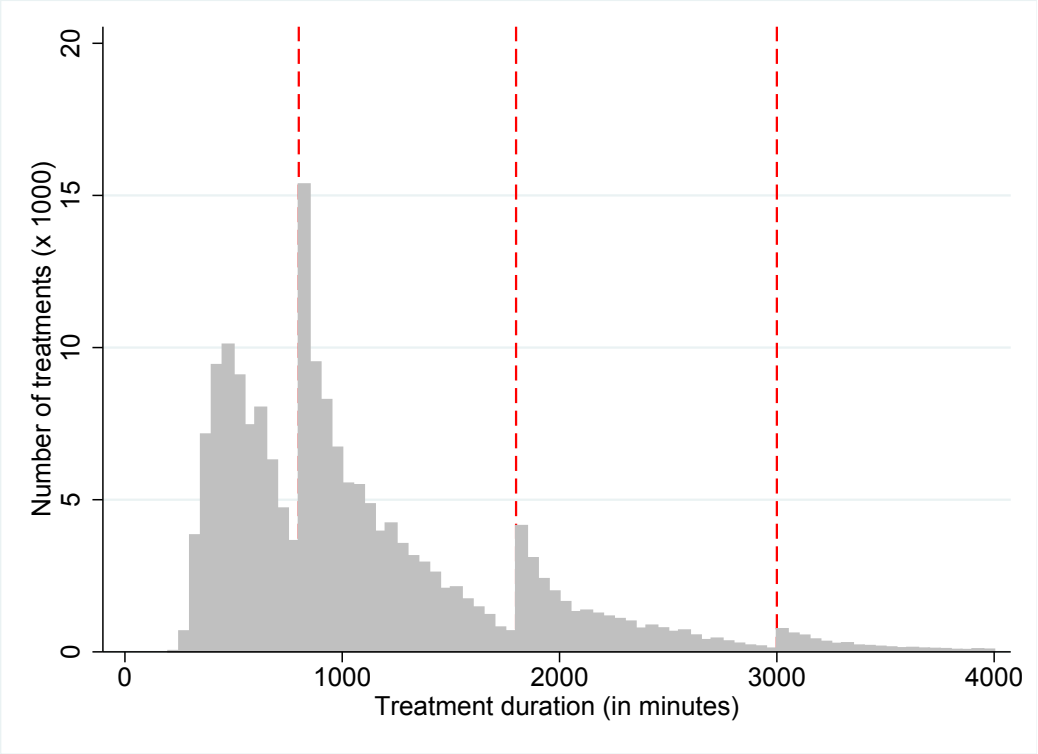


Figure 3: Distribution of treatment duration



Note: Distribution reported in bins of 50 minutes.

Figure 4: Distribution of ΔGAF

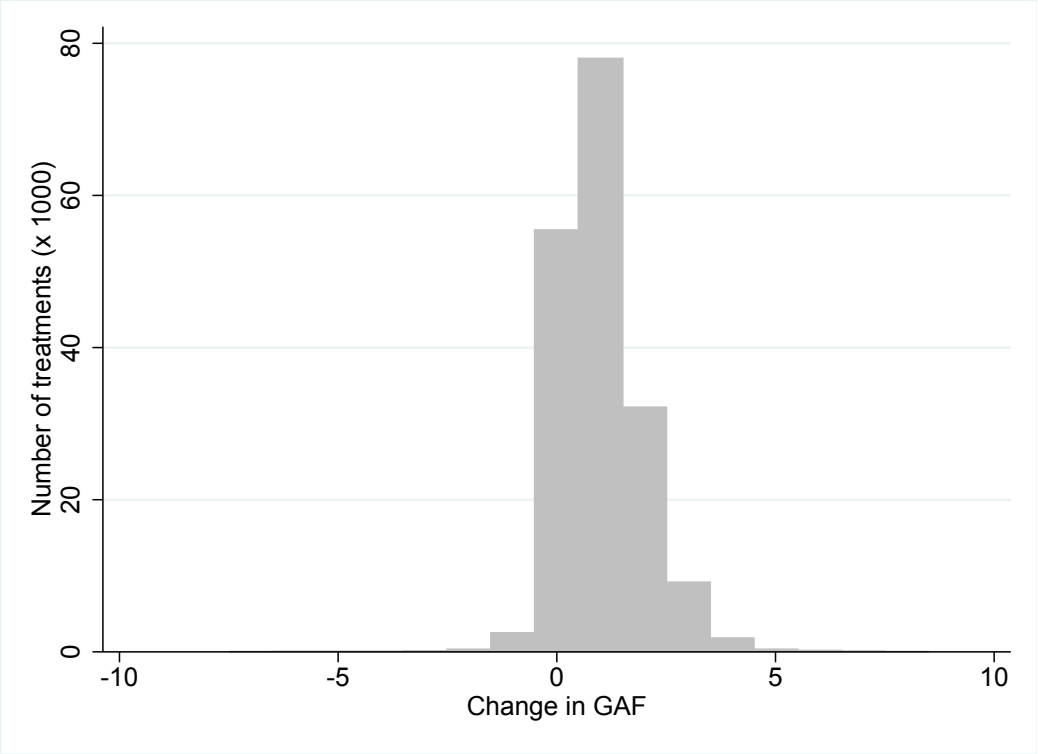
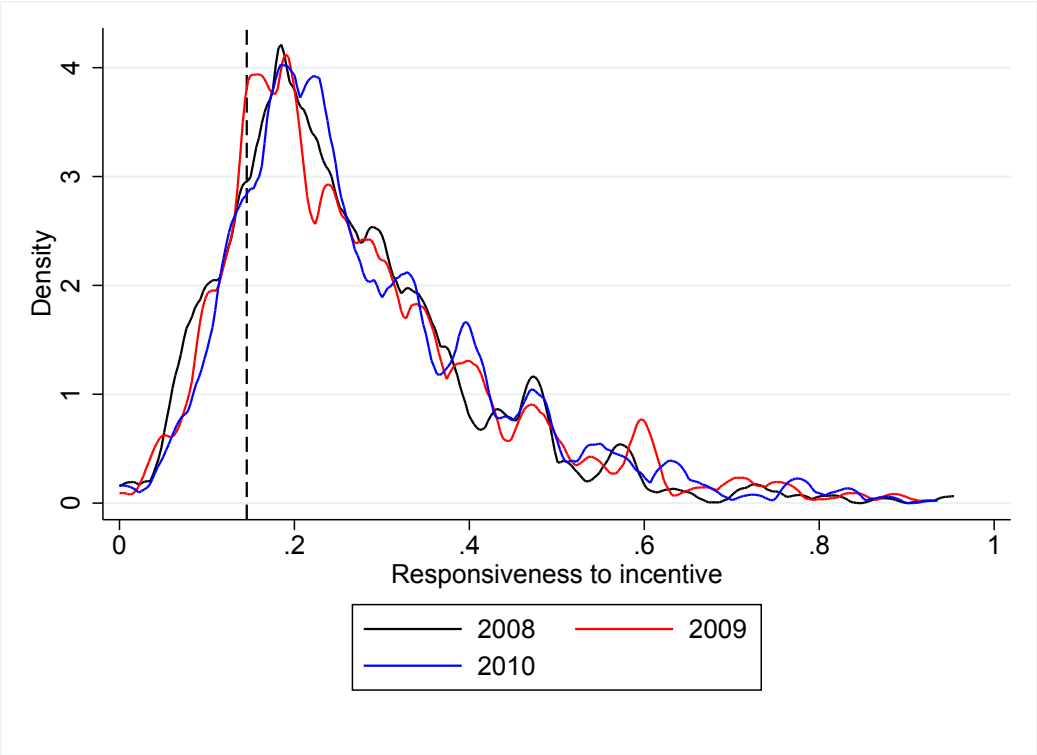


Figure 5: Distribution of responsiveness to the discontinuity in the compensation scheme



Notes: A value of $\hat{\alpha}_j = 0$ implies that a provider treats patients only before the tariff threshold, while a value of $\hat{\alpha}_j = 1$ implies that a provider treats only after the threshold. Dashed line shows the expected value for a provider that does not take the compensation scheme into account ($\bar{\alpha} = 0.15$.).

Figure 6: Distributions of treatment duration for extrinsic providers and non-extrinsic providers

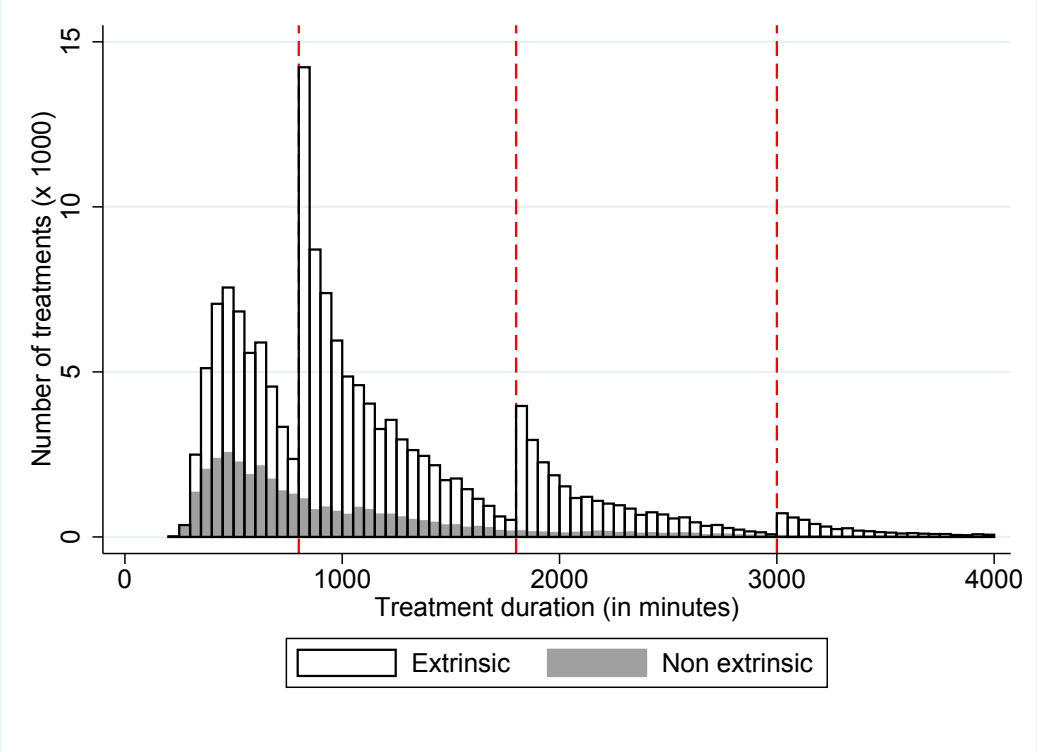
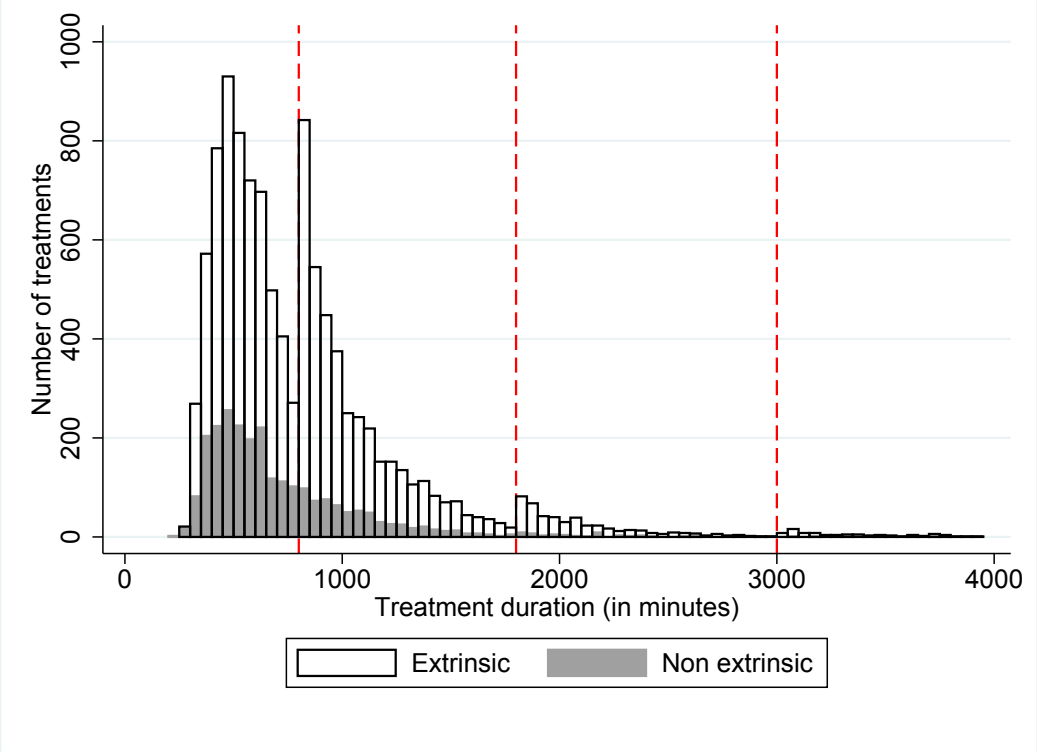


Figure 7: Distributions of treatment duration for treatments ended by the patient



Appendix A

The level of $\bar{\alpha}$ denotes the expected value of α when a provider does not take the compensation scheme into account. Following Douven et al. (2015a), we first estimate a smooth distribution of of treatment duration. In this estimation procedure we assume that the size of the gaps in treatment duration before a treatment duration threshold cancel out against the size of the bunches after the threshold (see Figure 8). This distribution is estimated by minimizing the following weighted non-linear equation:

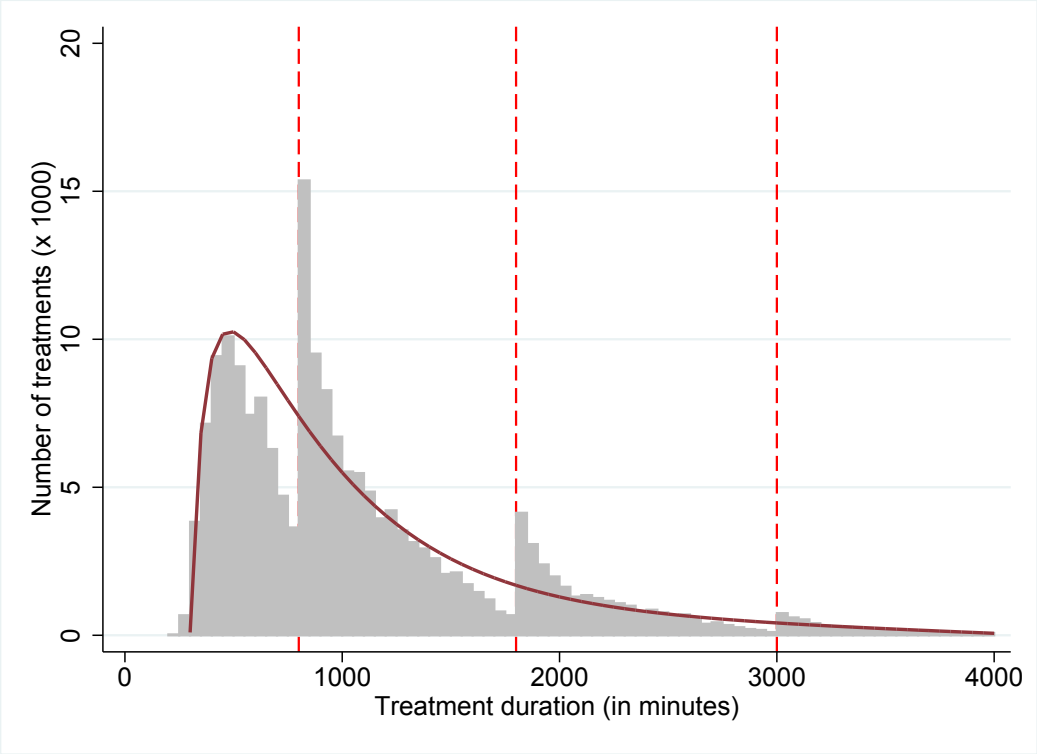
$$\min_{\beta} \sum w_t [Y_t - \beta_1 + \beta_2 t + \beta_3 / t + \beta_4 e^{-\beta_5 t}]^2, \text{ s.t.} \tag{9}$$

$$\sum [Y_t - \beta_1 + \beta_2 t + \beta_3 / t + \beta_4 e^{-\beta_5 t}] = 0 \text{ for } t \in [4, 16], [17, 36], [37, 60], [1, 74]$$

Where t indicates duration intervals of 50 minutes (e.g. t=1 covers all treatments with duration up to 50 minutes). Furthermore, $t \in [4, 16]$, $t \in [17, 36]$, and $t \in [37, 60]$ correspond to the area around the tariff thresholds of 800, 1800, and 3000 minutes, respectively. We additionally force the sum of all residuals to be equal to zero. Lastly, individual weights are denoted by w_t.

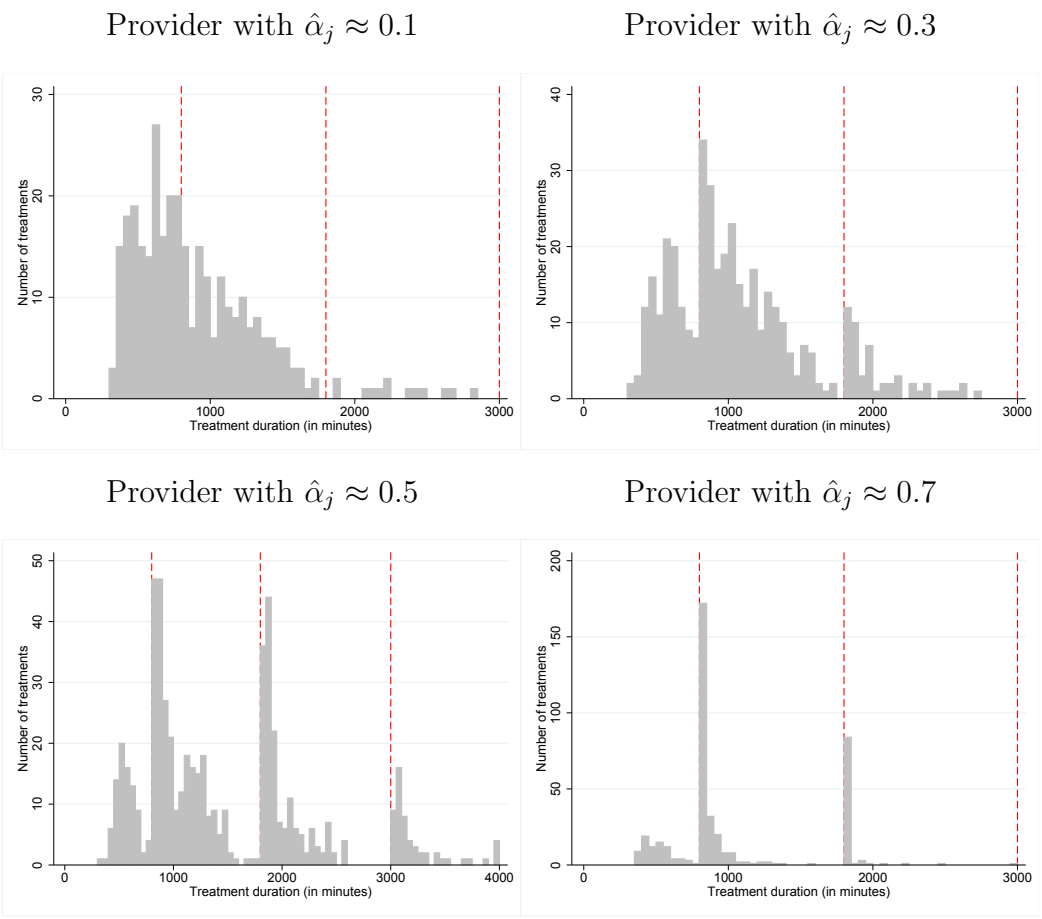
Second, the estimated distribution is entered in equation (1) to obtain a value for α. This value of α indicates how many observations are expected before the threshold and after the threshold when a provider does not take the higher tariff into account. This implies that all providers with $\alpha_j < \bar{\alpha}$ do not seem to take the tariff threshold into account and all providers with $\alpha_j > \bar{\alpha}$ do take the tariff threshold into account when deciding how long to treat.

Figure 8: Estimation of the smoothed distribution



Appendix B

Figure 9: Examples of distributions of treatment duration for different providers



Publisher:

CPB Netherlands Bureau for Economic Policy Analysis
P.O. Box 80510 | 2508 GM The Hague
T (070) 3383 380

December 2015 | ISBN 978-90-5833-712-2