CPB Netherlands Bureau for Economic Policy Analysis

CPB Discussion Paper | 292

Unintended effects of reimbursement schedules in mental health care

Rudy Douven
Minke Remmerswaal
Ilaria Mosca

Unintended effects of reimbursement schedules in mental health care

Rudy Douven

Minke Remmerswaal

Ilaria Mosca*

November 3, 2014

Abstract

We evaluate the introduction of a reimbursement schedule for self-employed mental health care providers in the Netherlands in 2008. The reimbursement schedule follows a discontinuous discrete step function —once the provider has passed a treatment duration threshold the fee is flat until a next threshold is reached. We use administrative mental health care data of the total Dutch population from 2008 to 2010. We find an efficiency effect: on the flat part of the fee schedule providers prolong treatment only if marginal benefits to patients outweigh marginal costs. We estimate a reduction in treatment duration by 2 to 6% and lower costs by 3 to 5% compared to a control group. However, we also find unintended effects: providers treat patients longer to reach a next threshold and obtain a higher fee. The data shows gaps and bunches in the distribution function of treatment durations, just before and after a threshold. In total, about 11 to 13% of treatments are shifted to over a next threshold, resulting in a cost increase of approximately 7 to 8%.

Keywords: mental health care, provider payment, regression discontinuity design
JEL classifications: I11, I12, I18

* Rudy Douven is affiliated with CPB Netherlands Bureau for Economic Policy Analysis and Erasmus University Rotterdam. When this paper was written, Douven was a Harkness Fellow 2013/2014 at Harvard Medical School, supported by the Commonwealth Fund, a private independent foundation based in New York City. The views presented here are those of the author and not necessarily those of The Commonwealth Fund, its directors, officers or staff. Email: R.Douven@cpb.nl , Minke Remmerswaal is affiliated with CPB Netherlands Bureau for Economic Policy Analysis. Email: M.Remmerswaal@cpb.nl. Ilaria Mosca is affiliated with Ecorys. Email: Ilaria.Mosca@ecorys.com. We would like to thank the Dutch Healthcare Authority (NZa) for providing the data. The data are not publicly available. We are grateful to the NZa and DBC-onderhoud for explaining the data. We would like to thank seminar participants at BU/Harvard/MIT Health Economics seminar in Boston at April 4, 2014, at NZa-seminar in Utrecht at May 8, 2014, Academy of Health in San Diego, June 8-10, 2014, IHEA in Dublin in July 13-16, 2014, CPB-seminar in The Hague, October 14, 2014, for comments. Furthermore, we are grateful to Pieter van Baal, Leon Bettendorf, Aaron Maras, Tom McGuire, Jan van Ours, Bastian Ravesteijn, Ingrid Seinen, Harry van Til and Gert Jan Verhoeven for providing comments on earlier versions of this paper.

Nederlandse samenvatting

In deze studie evalueren we de invoering van prestatiebekostiging bij zelfstandige zorgaanbieders in de curatieve geestelijke gezondheidszorg (cGGZ) in 2008. Het doel van dit beleid was om de efficiency in de cGGZ te vergroten. Bij prestatiebekostiging declareren zorgaanbieders hun behandeling op basis van Diagnose Behandelcombinaties (DBC's). Het tarief dat zorgaanbieders ontvangen, volgt een trapfunctie waarbij het tarief wordt verhoogd wanneer de behandelduur een bepaalde tijdsgrens overschrijdt. Voor de evaluatie gebruiken we data van alle behandelingen in de cGGZ die in Nederland hebben plaatsgevonden gedurende de periode 2008-2010. We vinden twee effecten. Ten eerste een efficiency-effect: prestatiebekostiging leidt tot ± 2-6% kortere behandelduur van patiënten in vergelijking met een controlegroep. Dit resulteerde in ± 3-5% lagere kosten. Ten tweede een onbedoeld effect: prestatiebekostiging leidt tot doorbehandelen van ± 11-13% patiënten totdat een hoger tarief is bereikt, hetgeen resulteerde in een kostenstijging van ± 7-8%. Gedurende de periode 2008-2010 waren, in vergelijking met een controlegroep, bij zelfstandige zorgaanbieders in de cGGZ de onbedoelde effecten dus groter dan het efficiency-effect.

Steekwoorden: Curatieve geestelijke gezondheidszorg, Doelmatige zorg, Regressie discontinuity design

1. Introduction

Before 2008, all mental health care in the Netherlands was organized and funded in a national insurance scheme (Exceptional Medical Expenses Act (AWBZ)). The AWBZ was paid for by income-differentiated premiums raised through taxes and provided long term and mental health care for all citizens. Mental health care providers were mainly funded with budgets. This changed in 2008, when the Dutch government placed a part of mental health care, the curative and acute mental health care, under the regime of regulated competition.[1] The main idea behind this policy change was to improve the efficiency in the sector by letting private insurers buy care on behalf of their enrollees. Providers no longer receive budgets, but replaced by a casemix-based reimbursement system that we will review in section two.[2] Mason and Goddard (2009) review the international literature on reimbursing mental health care providers and argue that casemix-based funding offers incentives for a range of objectives, including improvements in efficiency, quality of care and patient choice. They also criticize the Dutch reimbursement schedule and state: "…it […] therefore does not appear to encourage early discharge…" and "…could incentivize providers…to deliver medically unnecessary treatments…". Dutch policymakers also recognized that the reimbursement schedule in mental health care might create unintended incentives (VWS, 2010, NZa, 2010). This research aims to quantify these possible effects.

The design of a payment system is a complicated matter, especially in mental health care. Uncertainty and variations in treatments are likely to be greater in the mental health care market making the response of patients and providers to financial incentives larger than in other areas of health care (Frank and McGuire, 2000). A large body of the literature in health economics establishes that health care providers respond to financial incentives (for excellent overviews see Chandra, Cutler and Song, 2012 and McGuire, 2000). Most empirical evidence concerns the US and shows that fee-for-service payment provides incentives for overtreatment. Some of the first papers on this topic are Epstein et al. (1986), Hickson et al. (1987) and Stearns (1992). Recently, similar behavioral responses have also been reported since the introduction of regulated competition in the Dutch hospital market (Douven, Mosca, Mocking, 2013) and market for general practitioners (Van Dijk et al., 2013). Less research has been done on casemix-based funding in the mental health care market (Mason and Goddard, 2009). In the US, Jennison and Ellis (1987) found an 18 percent increase in the rate of visits per mental health provider per month when they shifted from a salaried basis to a fee-for-service basis. Rosenthal (2000) has examined the effects of risk sharing with mental health care providers. She found that providers that received a salary reduced their number of visits by 20 to 25 percent compared to providers who were still paid for each visit. Bellows and Halpin (2008) studied the impact of Medicaid reimbursement on mental health quality indicators and found evidence of upcoding of quality indicators to increase reimbursement.

This is the first study to evaluate the introduction of a new reimbursement schedule in mental health care. The reimbursement function follows a discontinuous discrete step function —once the provider has passed a treatment duration threshold the fee does not increase until a next threshold is reached.

[1] Managed competition in the Dutch curative care sector was introduced in 2006 (Van de Ven and Schut, 2008).
[2] The casemix refers to the mix of different types of patients that are treated by the provider.

We look at two effects: efficiency and unintended effects. Our study shows that the unintended effects, i.e. providers treat patients longer to reach a next threshold and obtain a higher fee outweigh the efficiency effect, i.e. on the flat part of the fee schedule providers prolong treatment only if marginal benefits to patients outweigh marginal costs. We separate out these two effects by using regression discontinuity design type of ideas (see e.g. Lee and Lamieux, 2010). Providers' behavior around discontinuous fee thresholds are most likely be explained by the change in fee, and not by other contemporary factors such as medical quality, treatment outcome, location or other unobserved factors. We use a quasi-experimental design in which 10 percent of all mental health care providers are paid according to the new reimbursement schedule, while 90 percent of providers were not subject to the reform. This latter group serves as a control group. We find an efficiency effect: we estimate a reduction in treatment duration by 2 to 6% and lower costs by 3 to 5% compared to a control group. However, we also find unintended effects: in total, about 11 to 13% of treatments are shifted to over a next threshold, resulting in a cost increase of approximately 7 to 8%.

The outline of our paper is as follows. Section 2 provides a concise overview of the Dutch mental health care system. Section 3 describes the economic theory relating to the new reimbursement schedule. Section 4 describes the data and section 5 presents the estimation methods. Section 6 presents the results and section 7 concludes.

2. **The Dutch mental health care system**

Although the mental health status of the Dutch population has been roughly stable since 1975, the number of people that use professional mental health services has increased with about 10 percent per year from 535,000 patients in 2001 to about 1 million patients in 2009 (GGZ Nederland, 2010).

Dutch mental health care distinguishes between primary and secondary care. Patients with mild mental disorders usually go to primary care, which is provided by a general practitioner, psychologist, psychotherapist or psychiatrist.[3] Patients with a more serious condition need specialized care are referred to secondary care. Secondary care is split into curative care and long-term care. Long-term care patients usually remain in an institution such as a residence or other kind of mental health facility for longer than a year. Our study focuses on the patients that receive curative care. They can receive care in an inpatient or outpatient setting and their treatment does not last longer than a year.

The reform to regulated competition in 2008 required many changes for providers, health insurers and regulators. The government decided upon a transition period between 2008 and 2010, in which health insurers did not incur financial risk on providing mental health care.[4] Since 2008, providers are reimbursed on their casemix, called a DBC (Diagnosis Treatment Combination). A DBC refers to the complete treatment episode of a patient. It starts with the initial consultation and continues until the provider ends the treatment. Consider for example a patient with mild depression that for ten months receives each month an individual therapy for sixty minutes by a psychotherapist (and no other form of medication or treatment). This patient's treatment can be coded with the following DBC: "Depression, 250 to 800 minutes, no medication" (DBC Onderhoud, 2013). If a treatment episode lasts longer than one year, the DBC is closed automatically. After that year a new DBC is opened. With the closed DBC a provider can receive reimbursement from his patient's health care insurer. The fee covers all labor and capital costs related to the treatment episode. The reimbursement fee for a DBC was fixed during our period of study and set prospectively by the Dutch Healthcare Authority (NZa). Patients' out-of pocket payments were small.[5]

Most mental health providers work in large regional institutions. These institutions can be a regional facility for ambulatory care, but also a specialized psychiatric hospital. Often, many different types of mental health care specialists work together. Their payment was before (and after) 2008 still based on annual budgets. These budgets were based on expected casemix and several regional budget parameters (such as inflation, wages, capital costs etc.). Mental health care specialists who work at a budgeted institution received a fixed salary[6]. These mostly large mental health care institutions account

---

[3] As of 2008, groups of practice nurses, social workers and psychologists (named POH-GGZ) entered the market to support general practitioners.

[4] Health insurers had therefore no financial incentives to control costs. The policy was that first a proper risk adjustment system should be implemented before health insurers could bear more financial risks. In 2013, DBC-fees became subject to negotiation between insurers and providers. To stimulate efficiency, the government started programs to develop quality indicators in mental health care. In 2013, a critical report (Rekenkamer, 2013) concluded that the stability and quality of most indicators is poor and needs to be improved.

[5] There was a mandatory annual deductible of 150 euros (in 2008) to 165 euros (in 2010) for all curative services (including mental health services) except general practitioner care and obstetrics.

[6] The government made agreement with labor unions about these salaries.

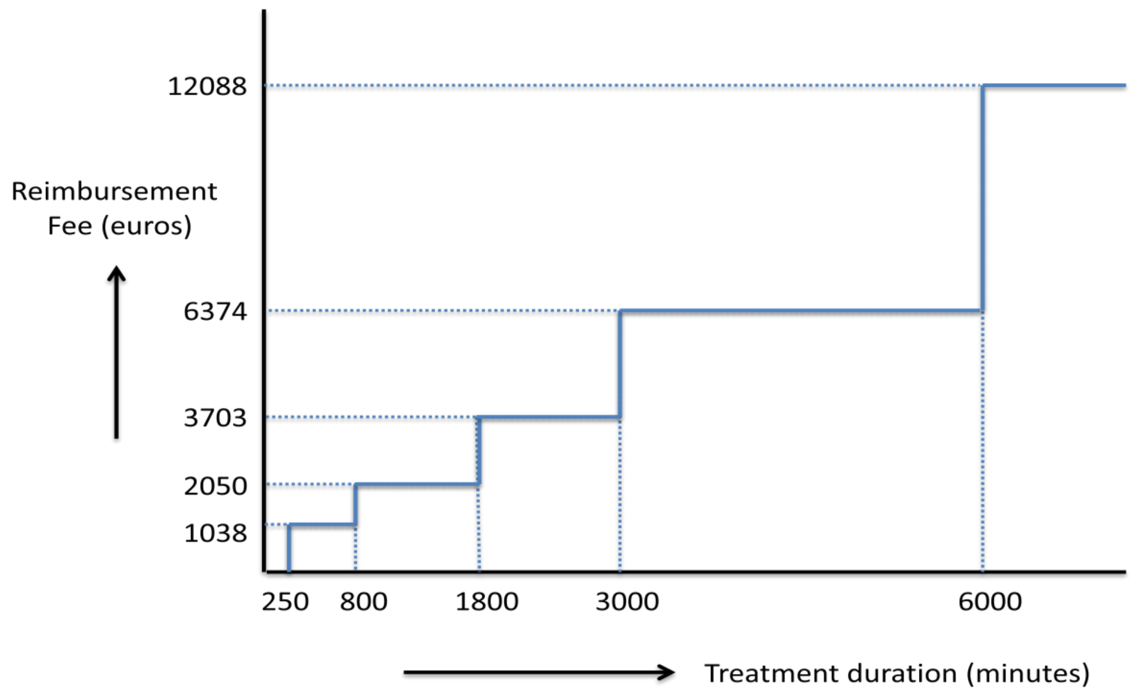
for about 90 percent of the sector (NZa, 2012). Henceforth, we will use these ‘budgeted’ or B providers in our study as a comparison (or control) group because their individual salaries during 2008-2010 were not related to the new reimbursement schedule.

About ten percent of the mental health care providers choose to work independently in for example private practices. Only this group of self-employed providers, and new providers that entered the market after January 1st of 2008, received their income according to the new reimbursement schedule. This group of self employed providers will be our treatment group and, henceforth, we will call these providers ‘non-budgeted’ or NB providers.

3. The economic theory of bunching of treatment durations

In this section we will explain in more detail the new reimbursement schedule and how we separate out efficiency and unintended effects. Treatment duration forms the basis of the size of the fee in the new schedule, and is calculated as a weighted sum of several components. Time spent by the provider *with* the patient receives the highest weight and can a consult, intake or therapy session. Lower weighted components are the time that a patient spends on other organized activities, such as group therapy session, and the number of days that a patient stays overnight in an institution. If for example, a patient receives eight therapy sessions of one hour (duration is 480 minutes), ten hours of group sessions (weighted as a total duration of 150 minutes) and three days in a residence (weighted as 180 minutes), it accounts for total treatment duration of 810 minutes. Figure 1 shows the reimbursement schedule for DBC category 'depression'. The X-axis shows the different classes of treatment duration. All DBC categories, in all specialties of mental health care, have the same treatment duration thresholds: at 250, 800, 1800, 3000, 6000 minutes.[7] The Y-axis shows the corresponding fees. They are unweighted averages for the years 2008 – 2010. Figure 1 shows that the reimbursement schedule is a discrete step function, in which fees are flat and only increase after a threshold is reached. The fees at each duration threshold slightly differ across specialties (e.g. depression, anxiety disorders etc.). Only the specialty 'other childhood disorders' has higher fees for treatments with more than 3000 minutes (NZa, 2007; NZa, 2008; NZa, 2009).[8]

Figure 1. Reimbursement schedule for the specialty 'depression'.



[7] Thresholds occur also at 12000, 18000 and 24000 minutes but we capped the duration time at 7000 minutes because such long treatment durations were rare.
[8] No major changes occurred in the reimbursement schedule throughout the studied period.

The financial incentives at thresholds in this system are obvious. The reimbursement fee for a treatment duration of 2900 minutes is €3703 euro, while a prolongation of the treatment with 100 minutes yields €6374 euro. This is a small difference in terms of treatment duration but a large difference in financial reward. The reimbursement schedule in Figure 1 may result in “bunching” of treatments at thresholds.

In line with Ellis and McGuire (1986, 1990), referred to as E&M from here on, we formulate a utility function of provider j for providing patient i with health severity θ_i a treatment duration x_i. The function is composed of two parts: benefits B_i to the patient and profits π_i for the provider.

$$U_{ij} = B_i(x_i, \theta_i) + \alpha_j \pi_i(x_i) \quad (1)$$

As in E&M, the agency parameter α_j describes the extent to which a provider weights the benefits to the patient relative to its own profits. For example, an entrepreneurial provider may attribute a higher α_j to profits. For the benefits to the patient $B_i(x_i, \theta_i)$ we make the standard assumptions $\frac{\partial B_i(x_i,\theta_i)}{\partial x_i} > 0$ at $x_i = 0$ and $\frac{\partial^2 B_i(x_i,\theta_i)}{\partial^2 x_i} < 0$, indicating that at the start of the treatment there is a positive benefit to the patient and the marginal benefit to the patient declines as treatment duration increases.

We model the profit function $\pi_i(x_i)$ for a NB provider in (1) as follows:

$$\pi_i(x_i) = P(x_i) - cx_i \quad \text{with } P(x_i) = P_l \text{ for } k_l \leq x_i < k_{l+1} \quad (2)$$

k_l represents the treatment duration thresholds, with $l = 1, \ldots, 5$, and $k_1 = 250, k_2 = 800, k_3 = 1800, k_4 = 3000, k_5 = 6000$ minutes. $P(x_i)$ is the flat fee rate for a treatment duration x_i. For example, in figure 1, $P(350) = 1038$ and $P(1000) = 2050$ euros.[9] Provider costs are represented by a simple linear cost function cx_i and indicate production costs as well as indirect costs such as foregone leisure time. Note that the profit function $\pi_i(x_i)$ is discontinuous at a threshold $x_i = k_l$. In line with E&M we assume that a provider maximizing its utility solves the problem:

$$\max_{x_i} B_i(x_i, \theta_i) + \alpha_j (P(x_i) - cx_i) \quad (3)$$

Thus, given a patient’s severity θ_i, and provider agency type α_j, a provider will choose a treatment duration x_i that solves the maximization problem in (3). Solving (3) returns that marginal benefits equal marginal costs:

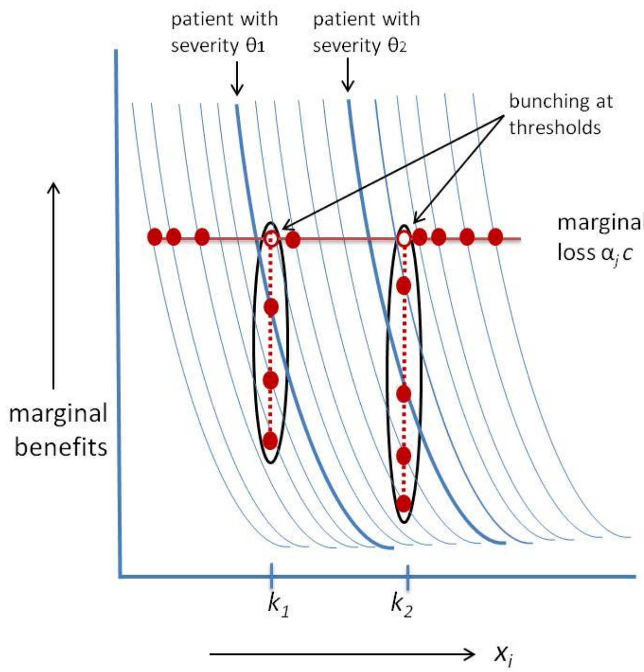
$$\frac{\partial B_i(x_i,\theta_i)}{\partial x_i} = \alpha_j c, \text{ with discontinuities at } x_i = k_l. \quad (4)$$

In Figure 2, we illustrate that solving this optimization problem results in bunching at treatment duration thresholds k_l. We plot the various marginal benefit functions $\frac{\partial B_i(x_i,\theta_i)}{\partial x_i}$ and the marginal loss line $\alpha_j c$, which is discontinuous at thresholds k_1 and k_2. We observe a spike at both treatment duration thresholds because reaching such a threshold implies that the provider receives a higher reimbursement

[9] These are the reimbursement fees for depression.

fee (or bonus). The size of both spikes depends on the fee difference before and after the threshold.[10] When the marginal benefit function of the patient with severity θ_1 crosses the marginal loss line $\alpha_j c$ in Figure 2 the provider will not end its treatment but prolong treatment until k_1 because its utility is maximized at the threshold k_1. A similar reasoning applies to the marginal benefit function θ_2, the provider prolongs treatment until k_2.

Figure 2. Bunching at treatment duration thresholds



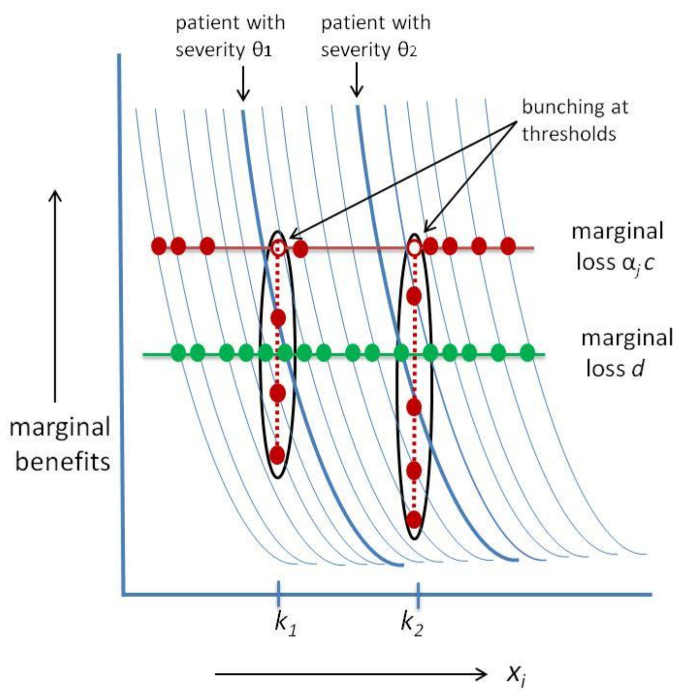
The result is bunching. The distribution of treatment durations will exhibit gaps before treatment duration thresholds. These gaps are expected to be larger for treatment durations closer to k_1 and k_2.[11] The reimbursement schedule provides also incentives for efficiency. For example, if $\alpha_j = 1$ then $\frac{\partial B_i(x_i,\theta_i)}{\partial x_i} = c$, and all dots on the marginal loss line $\alpha_j c$ in Figure 2 correspond with socially optimal treatment durations, where marginal benefit to the patient equals marginal costs (McGuire, 2000). In the case of $\alpha_j = 1$, bunching implies overtreatment. If $\alpha_j > 1$, all dots on the marginal loss line $\alpha_j c$ correspond with under treatment. Bunching implies that some treatment durations are prolonged and become closer to the cost efficient duration (although some overshooting may also happen). Similarly, if $\alpha_j < 1$, there is overtreatment and bunching implies even more overtreatment.

[10] The size of the spike has to be determined empirically. Around k_l, locally holds $\frac{\partial P_i(x_i)}{\partial x_i} = -\infty$, implying an infinite spike. However, in practice the decision to prolong treatment is more discrete in nature. For substantial shorter treatment durations than at thresholds k_l, the provider has to trade off the costs associated with treating the patient longer versus the size of the fee difference equal to $P_{l+1} - P_l$.

[11] Suppose treatment duration is at a local optimum. The farther away this treatment duration is from a threshold duration k the more costly it will be for a provider to move to the threshold k.

Important for our estimation procedure is the notion that there is only a financial incentive to prolong, and not to shorten, treatment durations. For example, a provider that hits a treatment duration threshold will not end the treatment but will prolong treatment as long as marginal benefits to the patient outweigh marginal costs.[12] Now, consider our comparison group, the B providers who receive a fixed salary. Compared to NB providers, we expect no bunching at treatment duration thresholds k_l because B providers face no particular financial consequences around these thresholds. We make a general assumption about the behavior of B providers, namely that $\frac{\partial B_i(x_i,\theta_i)}{\partial x_i} = d$, where d is a constant. The incentive structure may differ between B and NB providers. First of all, NB providers might have a stronger production incentive because production is directly related to their income. Keeping treatment durations short allows NB providers to treat more patients in a given time frame. Moreover, B providers may have weaker incentives to control costs than NB providers because their institution covers partly these costs. In the extreme case, $d = 0$, and B providers care only about patient benefits and not about costs. In practice B providers face some costs, but we hypothesize to a lesser extent than NB providers. If that hypothesis holds then we have $0 < d < \alpha_j c$. In Figure 3, we plotted the marginal loss line d of the budgeted providers below the marginal loss line $\alpha_j c$.

Figure 3: Marginal profit line of budgeted providers lower than $\alpha_j c$.



[12] An exception could be a provider with (too) many patients in his practice. Such a provider may have a financial incentive to end a treatment after hitting a threshold because treating a new patient may be more rewarding (in terms of profits and total patient benefits). Vice versa a provider with a shortage of patients may have an incentive to prolong treatment duration securing his financial income. In our analysis we assume that these are second order effects.

If this marginal loss line d is located below the largest vertical spikes of the NB providers then B providers would treat all patients longer than NB providers. The situation in Figure 3 represents a mix. Optimal treatment durations of B providers can be shorter (for example for the patient with severity θ_2) and longer (for example for patients with severity θ_1) than for NB providers. Figure 3 also shows the trade-off that we test empirically in this paper. Introducing the new reimbursement schedule to our control group, the B providers, may generate two effects. First, there may be an "efficiency" effect. This efficiency effect is measured by the vertical distance between the marginal loss line d and the marginal loss line $\alpha_j c$. Second, there are unintended effects of bunching around thresholds (vertical lines in Figure 3). In the next sections we will estimate these two effects separately.

4. Descriptive statistics

We obtained our dataset from an administrative database maintained by the NZa and contains all registered DBCs from providers in the secondary curative mental health care in the Netherlands for the period 2008 to 2010. All treatments had a minimum duration of 250 minutes and there were only few DBCs with a very long treatment duration, therefore we restricted our sample to DBCs with a maximum treatment duration of 4,000 minutes.[13] Table 1 summarizes the data. It contains approximately 1,4 million observations in fifteen specialties.

Table 1. Description of data*

	Number of DBCs			
Specialty	2008	2009	2010	*Total*
Depression	80.444	78.944	76.975	236.363
Anxiety disorders	57.829	60.262	60.706	178.797
Other mental disorders and problems	49.282	50.901	49.602	149.785
Adjustment disorders	46.693	49.865	49.080	145.638
Hyperkinetic disorders	35.271	41.463	43.442	120.176
Personality disorders	39.077	39.122	39.127	117.326
Other diagnoses	26.797	28.362	30.611	85.770
Schizophrenia	24.832	27.053	28.234	80.119
Pervasive disorders	27.425	25.820	24.968	78.213
Delirium. dementia and other disorders	17.796	17.680	17.617	53.093
Other substance use disorders	14.544	15.004	15.353	44.901
Alcohol use disorders	14.170	14.071	13.796	42.037
Other childhood disorders	9.427	13.398	18.576	41.401
Bipolar disorders	13.228	12.423	12.349	38.000
Total	456.815	474.368	480.436	1.411.619

*The numbers in the table correspond to DBCs with treatment durations smaller than 4000 minutes.

Table 2 distinguishes between B and NB providers. B providers produce the most DBCs for all categories. and some mental disorders are almost exclusively treated by B providers, for example this holds for the categories ‘delirium, dementia and other disorders’, ’alcohol use disorders’ and to a lesser extent ‘schizophrenia’. For NB providers we observe in many cases the profession of the therapist, we have 1302 psychologists, 431 psychiatrists and 74 providers working in institutions. For B providers we do not observe the profession because they are all grouped together in a large regional institution. The data contains for each DBC information on the type of therapy (for example adult, forensic, crisis or child care) and whether this is individual therapy, or (also) group therapy or a overnight stay. Other variables are the reason for closing a DBC (for example closed on a regular basis, or duration exceeding a year, or patient dissatisfied with treatment), and whether providers have prescribed drugs during a treatment. Another important variable are the global assessment of functioning (GAF) scores. The GAF-score is a quality measure for the severity of a patient’s mental illness. GAF-scores range between 0 (very severe symptoms) and 100 (no symptoms). Providers report these GAF-scores at the beginning of a treatment.

[13] Treatment durations below 250 minutes belong to primary mental health care. About 8% of the treatments had treatment duration longer than 4000 minutes. Note that 4000 minutes is well below 6000 minutes, so estimation errors that occur because providers prolong treatment duration to 6000 minutes are likely to be small.

Table 2. Type of provider and number of DBCs (years 2008-2010)*

Specialty	Budgeted providers	Non-budgeted providers	Total
Depression	181.487	54.876	236.363
Anxiety disorders	142.747	36.050	178.797
Adjustment disorders	115.416	30.222	145.638
Personality disorders	107.545	12.631	120.176
Hyperkinetic disorders	91.126	26.200	117.326
Other diagnoses	73.216	12.554	85.770
Schizophrenia	75.096	5.023	80.119
Pervasive disorders	76.633	1.580	78.213
Delirium. dementia and other disorders	52.891	202	53.093
Other substance use disorders	43.958	943	44.901
Alcohol use disorders	40.717	1.320	42.037
Other childhood disorders	27.969	13.432	41.401
Bipolar disorders	34.557	3.443	38.000
Other mental disorders and problems	112.309	37.476	149.785
Total	1.175.667	235.952	1.411.619

*The numbers in the table correspond to DBCs with treatments duration smaller than 4000 minutes.

Table 2 shows that patients are unevenly distributed across providers. To obtain enough power for our tests we narrowed down our patient sample and considered only patients within the following specialties: depression. anxiety disorders, adjustment disorders, and personality disorders.[14] To obtain similar patient characteristics for comparing our treatment and control group we only selected patients in the category "adults" that received individual therapy sessions. Also, DBCs were closed on a regular basis and patients received no prescribed medication. Furthermore, we corrected the subsamples for the severity of the diseases. Based on the GAF scores four subsamples per specialty were created. The first subsample considers all patients that received as initial assessment a GAF score between 41-70. The other three subsamples are selected from this subsample, each containing only patients with one of the following GAF scores: 41-50, 51-60, or 61-70.[15] For these subsamples patients treated by B and NB providers have exactly the same characteristics and, thus, can be compared.[16] Table 3 summarizes and shows the number of observations for each subsample.[17] Note we also included the total sample in our estimations. Patient characteristics of the total sample are very likely to differ between B and NB providers but it provides an estimate of the total effect of prolonging treatment durations due to the existence of various thresholds.

[14] We choose for these four categories because they are most prevalent treated mental illnesses with a clear diagnosis (see Table 2).

[15] The patient has some mild symptoms (e.g., depressed mood and mild insomnia) [GAF scale 61-70], moderate symptoms (e.g., flat affect and circumlocutory speech, occasional panic attacks) [GAF-scale 51-60] or serious symptoms (e.g., suicidal ideation, severe obsessional rituals, frequent shoplifting) [GAF-scale 41-50].

[16] The age and sex distributions are very similar across subsamples.

[17] The number of observations shrinks the more narrowly the subsample is defined. Important is also that many records were not filled in completely, and therefore had to be excluded from our subsample.

Table 3. Number of observations in various subsamples (years 2008-2010)**

		Budgeted providers	Non-budgeted providers	Total
1. Total Sample		1.175.667	235.952	1.411.619
2. Sample Depression		181.487	54.876	236.363
2a.	GAF: 41-70*	57.740	30.508	88.248
2b.	GAF: 41-50*	12.132	4.963	17.095
2c.	GAF: 51-60*	32.730	17.395	50.125
2d.	GAF: 61-70*	12.878	8.150	21.028
3. Sample Anxiety Disorders		142.747	36.050	178.797
3a.	GAF: 41-70*	55.505	21.581	77.086
3b.	GAF: 41-50*	10.360	2.934	13.294
3c.	GAF: 51-60*	31.051	12.086	43.137
3d.	GAF: 61-70*	14.094	6.561	20.655
4. Sample Adjustment Disorder		115.416	30.222	145.638
4a.	GAF: 41-70*	55.545	21.607	77.152
4b.	GAF: 41-50*	5.985	1.934	7.919
4c.	GAF: 51-60*	30.571	12.067	42.638
4d.	GAF: 61-70*	18.989	7.606	26.595
5. Sample Personality Disorder		107.545	12.631	120.176
5a.	GAF: 41-70*	39.571	17.977	57.548
5b.	GAF: 41-50*	8.467	2.457	10.924
5c.	GAF: 51-60*	22.102	10.056	32.158
5d.	GAF: 61-70*	9.002	5.464	14.466

**In these samples we only consider individual adult therapies without medical prescriptions that were closed on a regular basis.

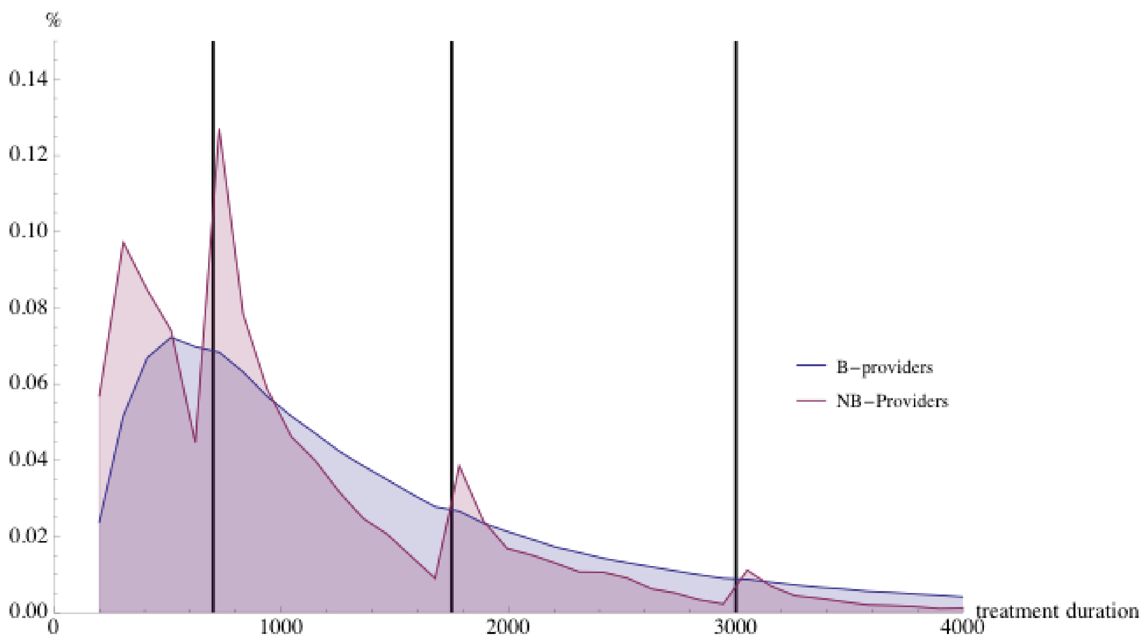
**The numbers in the table correspond to DBCs with treatments duration smaller than 4000 minutes.

5. Estimation Method

Figure 4 shows the distribution of treatment durations in the total sample (1. Total Sample in Table 3) for both types of providers. The three vertical black lines correspond to three treatment duration thresholds at 800, 1800 and 3000 minutes. The distribution function clearly differs between the B and NB providers. The treatment distribution for the budgeted providers is smooth for all treatment durations. However, in stark contrast with the B providers, for NB providers we observe large gaps and spikes at thresholds. Similar figures are obtained if we plot subsamples of our dataset.

To estimate the efficiency and unintended effects we use ideas from regression discontinuity design (RDD).[18] However, while RDD-studies use local linear smoothing around single thresholds to determine non-linear responses, we have reasonably large bunches and gaps of several thresholds that may be connected.[19] Therefore, we use a global estimation approach which allows us to estimate in one step the distribution functions for both types of providers.

Figure 4. Distribution of treatment duration for B and NB providers (all categories)



[18] RDD studies related to health care include Card, Dobkin and Maestas (2008, 2009) who study the discontinuity of health care utilization around age 65 when US citizens become eligible for Medicare. Sojourner et al. (2012) use RDD to study the effects of unionization of nursing homes. Shi (2013) finds evidence of income manipulation when studying labor supply responses to income cutoffs of a subsidized health insurance program in Massachusetts. Einav, Finkelstein and Schrimpf (2013) study the response of drug expenditure to non-linear contracts in Medicare part D. These studies are all related to consumer responses. Our study is about provider responses and more related to Bajari et al. (2011) who study hospital's responses to discontinuities in linear reimbursement schedules. Their identification strategy is more complicated than in our paper because reimbursement schedules are only discontinuous in the first derivative, and thresholds are not fixed but may differ across hospitals.

[19] For example, combining several separate local linear estimation procedures to one distribution function may not necessarily result in a smooth function.

We fit the non-linear regression equation (2) for each mental disorder category i, and provider type j (in what follows we omit i, j):

$$Y_t = f(\beta) + \eta_t \text{ with } \eta_t = B_t - G_t + \varepsilon_t \quad (2)$$

where Y_t , t =3,…,39 is the distribution function of treatment durations defined in treatment duration classes of 100 minutes.[20] Alike Lee and Lemieux (2010) we assume that all factors evolve "smoothly". If there are no discontinuities ($G_t = 0, B_t = 0$) in the reimbursement schedule, $f(\beta)$ would be a reasonable guess for explaining Y_t. This assumption is confirmed by estimates of $f(\beta)$ for the distribution function of B providers.

In standard RDD applications, sudden shifts in the outcome variable result from an exogenous change. In this study we have the same. Bunches and gaps in treatment durations of the NB providers are caused by exogenous changes in the fee structure, and not by medical outcome or other unobserved factors of individual patients. This implies that a conscious prolongation of treatment duration by NB providers, introduces systematic errors in Y_t. In (2), η_t represents both systematic and random errors. We distinguish systematic positive errors or "bunches" after a threshold ($B_t \geq 0$) and systematic negative errors or "gaps" before a threshold ($G_t \geq 0$). Lastly, ε_t represents the random error term in (2).

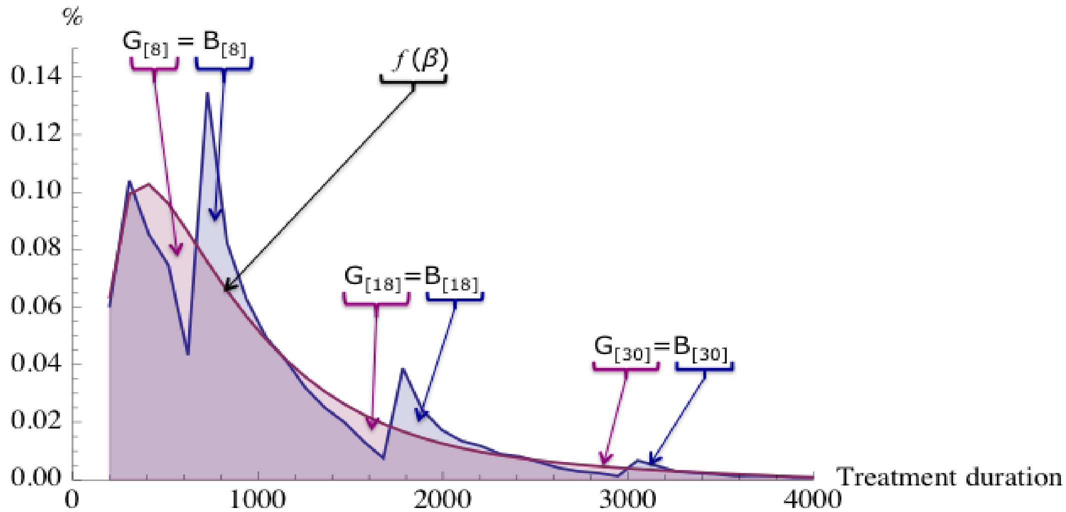
To estimate the smooth function $f(\beta)$ we constructed a class of smoothing functions that are able to describes similar shapes as the B providers in Figure 4. A property of this function is that it must increase at t = 300, has a top somewhere between t = 300 and 800 minutes, and monotonically declines thereafter. Furthermore, the function must be flexible enough to capture various shapes. Exponential function (3) satisfies these criteria:

$$f(\beta) = \beta_1 + \beta_2 t + \beta_3/t + \beta_4 e^{-\beta_5 t} \quad (3)$$

We have to estimate the five parameters β_j, $j = 1, \dots, 5$ in function (5). First, we substitute (3) in (2). Then we estimate (2). The size of the gap before each threshold $[k] \in \{[8], [18], [30]\}$ should equal the size of the bunch after this threshold (see Figure 5). This restriction reflects our theory in section 2: bunching after a threshold occurs through a *shift* of treatment durations from before to after a threshold.

[20] Thus, Y_3 represents all treatment durations in the 300-400 minutes time interval and Y_{39} in the 3900-4000 minutes time interval. The size of the surface of all distributions is normalized to 1. Note that we could also choose for smaller than 100 minutes time intervals but this does not change the nature of our story, but it would require a slightly more complex exponential specification as specified in (3).

Figure 5. Estimation of distribution function and bunches and gaps



To estimate β, we follow a weighted non-linear least squares minimization problem with four restrictions.

$$\min_{\beta} \sum_{t=3}^{40} w_t[Y_t - f(\beta)]^2 \text{ with restrictions:}$$

$$\textstyle\sum_{t=5}^{10}[Y_t - f(\beta)] = 0, \quad \sum_{t=11}^{20}[Y_t - f(\beta)] = 0, \quad \sum_{t=21}^{32}[Y_t - f(\beta)] = 0, \quad \sum_{t=3}^{39}[Y_t - f(\beta)] = 0 \qquad (4)$$

The first three restrictions correspond to the shift of treatment durations: $B_{[k]} - G_{[k]} = 0$, for $k = 8, 18, 30$.[21] We observed in the data that bunching occurs up to 300 minutes after a threshold. Therefore we fixed possible bunching to the first 300 minutes after a threshold in our restrictions.

To obtain smooth convergence of our non-linear estimations, we added a fourth restriction: the total sum of the errors is zero.[22] Weights w_t were also introduced.[23] Our global estimation strategy with restrictions is quite powerful compared to three separate local RDD-estimations at each individual threshold. The global approach allows us to connect the “bunches” and “gaps” estimates at individual thresholds making our identification strategy more reliable.

Minimization procedure in (4) generates $\hat{\beta}_1, \dots, \hat{\beta}_5$. This allows us to compute $\hat{\eta}_t = Y_t - f(\hat{\beta})$. Next, we can compute our estimates for the gaps and bunches: $\hat{G}_{[8]} = -\sum_5^7 \hat{\eta}_t$, $\hat{B}_{[8]} = \sum_8^{10} \hat{\eta}_t$, $\hat{G}_{[18]} = -\sum_{11}^{17} \hat{\eta}_t$, $\hat{B}_{[18]} = \sum_{18}^{20} \hat{\eta}_t$, $\hat{G}_{[30]} = -\sum_{21}^{29} \hat{\eta}_t$, $\hat{B}_{[30]} = \sum_{30}^{32} \hat{\eta}_t$.

[21] $G_{[8]} = -\sum_5^7 \eta_t$, $B_{[8]} = \sum_8^{10} \eta_t$, $G_{[18]} = -\sum_{11}^{17} \eta_t$, $B_{[18]} = \sum_{18}^{20} \eta_t$, $G_{[30]} = -\sum_{21}^{29} \eta_t$, $B_{[30]} = \sum_{30}^{32} \eta_t$.

[22] This implies all systematic shifts are explained by the three previous restrictions, and that no treatments with duration between 300-500 minutes are shifted to over 800 minutes threshold, and between 3300-4000 minutes are shifted to over the 6000 minutes threshold.

[23] In most cases we used $w_t = 1$, however sometimes we experimented with somewhat higher weights to obtain smooth convergence. We performed our optimizations with the numerical non-linear global optimization function “NMinimize” of the software program Mathematica. To obtain convergence we sometimes had to alter the minimization method in Mathematica (gradient-based and direct search methods), weights and starting values.

In order to present the significance of our estimates for bunches and gaps we need an estimate for our error term ε_t in (2). Because our computation does not allow us to compute for each t, $\hat{B}_t, \hat{G}_t$ in (2) separately, we cannot properly estimate the random error term ε_t. Therefore we assume $\hat{\varepsilon}_t = \hat{\eta}_t^B$ where $\hat{\eta}_t^B$ are the estimated errors of the budgeted providers after estimating (2). Thus, we assume the standard error of the non-budgeted providers s^{NB} in (2) equals the standard error of the budgeted providers s^B:[24]

$$s^{NB} = s^B = \sqrt{\frac{1}{(37-5)}\sum_t (\hat{\eta}_t^B)^2} \tag{5}$$

We use a 32 degrees of freedom correction (see e.g. Verbeek, 2004) , 37 minus 5 (parameters β to estimate in (3)). After obtaining these statistics we can derive additional statistics such as an estimate of the average treatment duration, prolongation time as a result of shifting treatments and associated costs.

[24] We make the reasonable assumption that the random errors and corresponding standard deviations s^B and s^{NB} are of the same order of magnitude. If there are small systematic errors in $\hat{\eta}_t^B$ we will overstate s^{NB}. Note that we calculate s^B from a Y_t distribution that has the same number of observations as the corresponding Y_t distribution of the NB providers.

6. Estimation results

In this section we present our estimation results for all the samples described in Table 3. We first show our results graphically in Figure 6 for the three samples 1, 2 and 2a in Table 3: "total sample", "depression" and the subsample "depression with similar patient characteristics (GAF-scores 41-70)". Figure 6 contains for each sample three panels. The first panels, Figures 6a,d,g, show Y_t and the corresponding estimate $f(\hat{\beta})$ of the B provider, from which we will derive an estimate for our standard error. The estimates indicate that our exponential identification in (3) can fit $f(\beta)$ to Y_t very well. The middle panels, Figures 6b,e,h, indicate the unintended effects. Bunches and gaps are present in all three samples. The size of bunches and gaps are remarkably stable across subsamples. Bunches and gaps are largest (and significant) at the first two thresholds of 800 and 1800 minutes and positive (but insignificant) at the threshold of 3000 minutes in all cases.[25] The efficiency effects are presented in the right three panels, Figures 6c,f,i. For the total sample and depression sample (panels 6c,f) we observe large efficiency effects; on average NB providers treat patients much shorter than B providers. However, the efficiency effect almost disappears in the case of patients with similar characteristics (panel 6i). Controlling for patient characteristics is therefore crucial to identify possible efficiency effects between B and NB providers.

The estimation results of the three subsamples are summarized in Table 4. The first column presents the unintended effects: the percentage of treatments that are shifted over each of the three thresholds. In total about 11-13% of treatments are shifted to over a next threshold. The second column in Table 4 presents average treatment duration. The difference between $f(\hat{\beta})$ and Y_t for B providers is small, confirming the good fit and resulting in small standard errors s^B.[26] For NB providers the average treatment duration corresponding to $f(\hat{\beta})$ is 20-26 minutes lower than Y_t, indicating that the increase in average treatment duration as a result of bunching is relatively small.[27] Important is the large difference in average treatment duration between B and NB providers in the "Total Sample", 22.0%, and "Total Sample Depression", 23.7%, indicating that B providers treat on average more sick patients. After controlling for patient characteristics ("Total Sample Depression, GAF scores 41-70") the efficiency effect shrinks to 2.2%. In the third column of Table 4 we present average treatment costs. The unintended effects increase average costs per treatment by 137 to 157 euros or a cost increase of 7.1 to 7.9%. The efficiency effect for the "Total Sample Depression, GAF scores 41-70" yields that on average treatments are 2.2% (or 37 euros) more expensive for B than NB providers. The efficiency effect is however more than offset by the unintended effects; summing up the unintended and efficiency effect yields that NB providers treat on average patients 157-37=120 euros more expensive than B providers.[28]

---

[25] There are two reasons. One there are few patients treated around 3000 minutes. Second, we have only 39 observations, which is rather small. We would obtain significant results if we would increase our number of observations by narrowing the bins of the distribution function Y_t.

[26] For smaller subsamples the graph Y_t is less smooth increasing the size of the standard error s^B.

[27] The average prolongation of treatment duration for treatments that are shifted over to a next threshold is about 200 minutes.

[28] We have tested the significance of the efficiency effect with the non-parametric Kolmogorov-Smirnov test. It rejected the hypothesis of similar $f(\hat{\beta})$ distribution functions for B and NB providers in the first two samples in Table 4. However, it does not reject this hypothesis for the subsample with the same GAF-scores. One simple way to improve the power of the

Figure 6a-i. Y_t and $f(\widehat{\beta})$ for NB and B providers for various samples

Kolmogorov-Smirnov test is to increase the number of observations by increasing the number of bins. In one analyses we used bins of 50 minutes instead of 100 minutes (not shown here). Using more bins does not change our main results but it increases the power and resulted in a rejection of the equal distribution hypothesis.

Table 4. Estimation results for "Total Sample", Total Sample Depression" and "Total Sample Depression (GAF scores 41-70).[a]

	Bunches and Gaps (%) Unintended effect				*Average Treatment Duration (minutes)*			*Average Treatment Costs (euros)*		
	$\hat{B}_{[8]}$	$\hat{B}_{[18]}$	$\hat{B}_{[30]}$	Total	Y_t	$f(\hat{\beta})$	*dif*	Y_t	$f(\hat{\beta})$	*Unintended effect*
Total Sample (all specialties)										
B Providers					1382	1380		2308	2299	
NB Providers	7.4**	3.2**	0.7	11.3**	1147	1131	26	2060	1923	137 (7.1%)
Efficiency effect						22.0%			19.5%	
Total Sample Depression										
B Providers					1321	1316		2251	2238	
NB Providers	8.2**	3.6**	0.4	12.2**	1084	1064	20	2012	1867	145 (7.8%)
Efficiency effect						23.7%			19.9%	
Total Sample Depression (GAF scores 41-70)										
B Providers					1199	1185		2060	2034	
NB Providers	8.2**	4.1**	0.6	12.9**	1181	1159	22	2154	1997	157 (7.9%)
Efficiency effect						2.2%			1.9%	

[a] The *,** in the Table indicate significance levels of 0.05, respectively 0.01. Average treatment costs for B providers are calculated on the premise that they are paid according to the reimbursement schedule for NB providers.

In addition to the three subsamples, we have also looked into other mental illnesses (see Table 3 for the subsamples and the number of observations in each subsample). We performed the same estimations for these sixteen subsamples. The results are are reported in Tabel 5. Columns (1)-(3) present the volume effects. Column (1) represents the size of the unintended effects: the percentage treatments that are shifted to over a next threshold. Column (2) shows the average treatment duration for the actual distribution Y_t , and estimated distribution $f(\hat{\beta})$, and column (3) shows the efficiency effect; the percent change in treatment duration between NB and B providers. Columns (4)-(6) show the same effects but now for fees. Column (4) shows the average fee of a treatment for Y_t and $f(\hat{\beta})$. Column (5) presents the unintended cost effects; the percent difference between the two variables. Finally, column (6) represents the cost difference related to the efficiency effect between NB and B providers.

The results in Table 5 confirm our previous findings. First of all, we observe that the unintended effects (column (1)) are present in all subsamples. The effects are fairly stable across all our subsamples and vary roughly between ±11-13%, with some outliers.[29] This correspond with a cost increase that varies between ±7-8% (column 5). The efficiency effect in column (3) shows that B providers treat patients approximately ±2-6% longer than NB providers with corresponding cost increases of approximately ±3-5% (column (6)).[30] Thus, for almost all cases we find that the marginal loss line d is situated somewhat below the line $\alpha_j c$ (see Figure 3 in Section 3). The unintended financial effects in column (5) are in all cases larger than the "efficiency" effects (column (6)).

[29] The estimation results for the unintended effects are all significant on a 0.01 level.

[30] Only for the subsample depression GAF: 41-50 and adjustment disorders, GAF:61-70, we find a 0.1%, respectively 1.5%, higher average treatment duration for NB providers.

Table 5. Estimation results for subsamples 2a-2d, 3a-3d, 4a-4d, 5a-5d (see Table 3)

Subsample	**(1) bunches, gaps (%)**	***(2) avg. treatment duration (mins)***		***(3) efficiency effect***	***(4) avg. treatment costs (euros)***		***(5) unintended effect***	***(6) efficiency effect***
Type of Provider	NB:	NB:	NB:	(NB-B)/B:	NB:	NB:	%	(NB-B)/B:
Distribution	Y_t	Y_t	$f(\hat{\beta})$	$f(\hat{\beta})$	Y_t	$f(\hat{\beta})$	Change	$f(\hat{\beta})$
Depression								
2a. GAF: 41-70	12.9%	1181	1159	−3.0%	2154	1997	7.9%	−2.6%
2b. GAF: 41-50	13.9%	1307	1284	0.1%	2359	2183	8.1%	0.1%
2c. GAF: 51-60	13.0%	1192	1167	−3.8%	2168	2005	8.2%	−3.2%
2d. GAF: 61-70	12.1%	1081	1001	−8.5%	1999	1759	13.6%	−7.8%
Anxiety disorders								
3a. GAF: 41-70	12.0%	1131	1108	−5.6%	2047	1893	8.1%	−4.9%
3b. GAF: 41-50	11.2%	1302	1268	−7.1%	2340	2150	8.8%	−6.1%
3c. GAF: 51-60	12.6%	1177	1148	−5.1%	2122	1950	8.9%	−4.9%
3d. GAF: 61-70	11.3%	1073	1056	−6.3%	1951	1819	7.2%	−5.1%
Adjustment disorders								
4a. GAF: 41-70	10.1%	1030	1010	−2.6%	1764	1645	7.2%	−2.4%
4b. GAF: 41-50	10.4%	1192	1180	−0.9%	2019	1895	6.6%	−0.4%
4c. GAF: 51-60	10.5%	1038	1015	−4.9%	1773	1647	7.7%	−4.6%
4d. GAF: 61-70	9.3%	977	962	1.5%	1683	1582	6.4%	1.3%
Personality disorders								
5a. GAF: 41-70	11.4%	1339	1320	−5.3%	2397	2234	7.3%	−4.8%
5b. GAF: 41-50	11.7%	1473	1438	−6.4%	2617	2414	8.4%	−6.1%
5c. GAF: 51-60	12.3%	1399	1372	−5.4%	2497	2306	8.3%	−5.1%
5d. GAF: 61-70	10.2%	1264	1287	−2.7%	2271	2193	3.5%	−1.9%

To conclude, the unintended effects appear very clear in the data and are very stable across all subsamples. The efficiency effects are smaller and less certain because these effects are estimated by comparing B and NB providers. A limitation of our measure for the efficiency effects could be that there is still unobserved variation in the treatment and control group that we do not capture adequately. In future research we may be able to address this point by adding more socioeconomic information on individual patients.

7. Discussion

We evaluate the implementation of a new reimbursement schedule in Dutch mental health care. The reimbursement schedule follows a discontinuous discrete step function —once the provider has passed a treatment duration threshold the fee is flat until a next threshold is reached. We find an efficiency effect: on the flat part of the fee schedule providers prolong treatment only if marginal benefits to patients outweigh marginal costs. We estimate a reduction in treatment duration by 2 to 6% and lower costs by 3 to 5% compared to a control group. However, we also find unintended effects: providers treat patients longer to reach a next threshold and obtain a higher fee. The data shows gaps and bunches in the distribution function of treatment durations, just before and after a threshold. In total, about 11 to 13% of treatments are shifted to over a next threshold, resulting in a cost increase of approximately 7 to 8%.

An important message of our study is that the unintended effects clearly demonstrate that mental health care providers react to financial incentives. Monitoring providers' behavior is therefore an important element for the system to function properly. In the Dutch system of managed competition health insurers have the role to discipline providers. However, until 2014 health insurers lacked information about the exact treatment duration of health care providers. They received only global information on treatment duration of individual providers, i.e. they received only information between which two treatment duration thresholds the provider performed the treatment, and not the exact treatment time. Thus, insurers had no possibility to perform the same analysis as we carried out in this paper. This is now gradually changing; since 2014 health insurers obtain exact information about treatment durations and are also becoming more financially responsible for mental health care cost containment.

We measure a small efficiency effect. However, we cannot be certain that we measure genuine efficiency since we cannot rule out the possibility that patients may also have received too little care. Our efficiency arguments do hold if we assume $\alpha_j = 1$ in our utility function (1), which is a fairly standard assumption (McGuire, 2000). In that case NB providers produce cost efficient on the flat part of the reimbursement schedule and bunching corresponds to overtreatment. Efficiency differences between B and NB providers could also be related to differences in practice styles or quality of treatments (see e.g. Chandra, Cutler, Song, 2012). To address these issues more properly quality information about treatments would be necessary. Quality information would also allow us to make statements about welfare effects.

In 2014, the Dutch government decided to pay B providers also according to the new reimbursement schedule. Our findings suggest that, all else equal, this policy will lead to higher costs since the higher costs associated with the unintended effects outweigh the lower costs of the efficiency effect. However, the caveat of this statement is the "all else equal" assumption. There are still many external dynamic demand and supply factors that are difficult to assess. For example, budgeted providers may put a lower weight on profits (lower agency parameter α_j in (1) for B providers) than NB providers because the latter category of providers is of a more entrepreneurial type. In that case the unintended effects may turn out to be lower than we report in this paper. Also, insurers may be better equipped to monitor providers' treatment duration and, in the longer run, quality. Another important difference is that the Dutch government changed the flat reimbursement fees to maximum fees. Thus, health insurers can bargain with providers lower reimbursement fees, if providers' performances turn out to be inadequate.

Also the conclusion that the introduction of the new reimbursement schedule for NB providers in 2008 led to higher costs is premature. Before 2008, NB providers received a fixed fee for each visit. A fee for each visit is similar to the reimbursement schedule in our study but now there are thresholds after each visit of sixty minutes. A fee for each visit is closer to a fee-for-service type of payment and may also result in overtreatment. Unfortunately, we have no data for the period before 2008 available, making a comparison between the two regimes not possible.

An important policy question is how an optimal reimbursement schedule for mental health care providers should look like. One could think of several ways to improve the reimbursement schedule in our study. A first option would be to change the positioning of the thresholds. Ideally, thresholds should be placed where the mass of the distribution function $f(\beta)$ is small. If the mass before a threshold is small, unintended effects will diminish because there are only few treatments to shift over to a next threshold. Unfortunately, the threshold of 800 minutes is placed just after the top of the distribution function (see Figure 4), thus exacerbating the unintended effects. Moving the 800 minutes threshold to 500 minutes, just before the top of the distribution function, would diminish the unintended effects. A second option would be to increase or decrease the number of thresholds. A reduction in the number of thresholds to a single fee for each treatment could remove the unintended effects. However, if patients' characteristics across providers differ substantially, it could also result in a larger income variation across providers. A single fee might also increase the incentives for selecting more favorable patients, i.e. patients that need short treatment duration, and/or stinting of mental health services (McGuire, 2000). However, adding more thresholds may create more incentives for providers to prolong treatment duration. A third option would be to get rid of the discontinuities in the reimbursement schedule. A mixed payment system of a prospective fee and a linear reimbursement schedule as advocated by Ellis and McGuire (1990) would be continuous and may diminish the unintended effects as well.

In this study we rely on providers that record their own DBCs. We assume that providers record their treatment duration correctly in their administration. However, literature indicates that fraudulent behavior may also occur in payment systems based on DRGs in the US, or DBCs in the Netherlands. This fraudulent behavior is often referred to as 'upcoding' (Steinbusch et al., 2007). The Dutch reimbursement system may be vulnerable to this 'upcoding' because Dutch providers code DBCs themselves. They could tamper with the data. Especially, in mental health care the risk for fraud may even be greater than for less discretionary treatments, as hip or knee replacements. Third parties, such as health insurers, also might find it particularly difficult to verify and dispute mental health diagnoses.

References

Bajari, P., H. Hong, M. Park, R. Town (2011) Regression Discontinuity Designs with an Endogenous Forcing Variable and an Application to Contracting in Health Care, NBER Working Paper No. 17643

Bellows, N.M. and H.A. Halpin (2008), Impact of Medicaid Reimbursement on Mental Health Quality Indicators, *Health Services Research* 43, pp. 582-97.

Card, D., C. Dobkin and N. Maestas (2008), The impact of Nearly Universal Insurance Coverage on Health Care Utilization: Evidence from Medicare, *American Economic Review* 98 (5), pp. 597-636.

Card, D., C. Dobkin and N. Maestas (2009), Does Medicare Saves Lives?, *Quarterly Journal of Economics* 123 (1) pp. 597-636.

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

Douven, R., R. Mocking and I. Mosca (2012), The Effect of Physician Fees and Density Differences on Regional Variation in Hospital Treatments, CPB Discussion Paper 208, CPB, The Hague, and iBMG Working Paper W2012.01, Erasmus University Rotterdam.

van Dijk, C.E., B. van den Berg, R. A. Verheij, P. Spreeuwenberg, P. P. Groenewegen, D. H. de Bakker (2013), Moral Hazard and Supplier-Induced Demand: Empirical Evidence in General Practice, *Health Economics* 22 (3), pp. 340–352.

DBC Onderhoud (2013), Spelregels, DBC-registratie GGZ, versie RG13a, DBC Onderhoud, Utrecht.

Einav, L., A. Finkelstein and P. Schrimpf (2013), The Response of Drug Expenditure to Non-Linear Contract Design: Evidence from Medicare Part D, MIT Working Paper.

Ellis R.P. and T.G. McGuire (1986), Provider behavior under prospective reimbursement. Cost sharing and supply, *Journal of Health Economics* 5 (1986), pp. 129-151.

Ellis R.P. and T.G. McGuire (1990), Optimal payment systems for health services, *Journal of Health Economics* 9(4), pp. 375-396.

Epstein, A.M. et al. (1986), The use of ambulatory testing in prepaid and fee-for-service group practices: relation to perceived profitability, *New England Journal of Medicine* 314, pp. 1089-1093.

Frank R.G. and T.G. McGuire (2000), Economics and Mental Health, in A.J. Culyer and J.P. Newhouse (eds.), Handbook of Health Economics, Vol. 1B, pp. 893-954. Amsterdam Elsevier.

GGZ Nederland (2010), Zorg op waarde geschat, update, Sectorrapport ggz 2010, Amersfoort (in Dutch).

Hickson G.B. et al. (1987), Physician reimbursement by salary or fee-for-service: effect on a physician's practice behavior in a randomized prospective study, *Pediatrics* 80, pp.744-750.

Jennison K. and R.P. Ellis (1987), Comparison of psychiatric service utilization in a single group practice, in: McGuire and Scheffler (eds.), The Economics of Mental Health Services: Advances in Health Economics and Health Services Research, Vol. 8, JAI Press, Greenwich, USA, pp. 175-194.

Lee D.S. and T. Lemieux (2010), Regression Discontinuity Designs in Economics, Journal of Economic Literature 48, pp. 281–355.

Mason, A. and M. Goddard (2009), Payment by Results in Mental Health: A Review of the International Literature and an Economic Assessment of the Approach in the English NHS, Research Paper 50, Centre for Health Economics, The University of York.

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.

NZa (2007), Tariefbeschikking DBC GGZ 2008. Nederlandse Zorgautoriteit, Utrecht (in Dutch).

NZa (2008), Tariefbeschikking DBC GGZ 2009. Nederlandse Zorgautoriteit, Utrecht (in Dutch).

NZa (2009), Tariefbeschikking DBC GGZ 2010. Nederlandse Zorgautoriteit, Utrecht (in Dutch).NZa (2010), Invoering prestatiebekostiging curatieve GGZ: Advies op hoofdlijnen. Duth Healthcare Authority, Utrecht (in Dutch).

NZa (2010), De curatieve GGZ in 2009: Ontwikkelingen in aanbod en volume, Monitor, Nederlandse Zorgautoriteit, Utrecht (in Dutch).

NZa (2012), Marktscan Geestelijke Gezondheidszorg. Weergave van de markt 2008-2011. Nederlandse Zorgautoriteit, Utrecht (in Dutch).

Rekenkamer (2013), Indicatoren voor kwaliteit in de zorg, Algemene Rekenkamer, March 28, The Hague, The Netherlands.

Rosenthal, M.B. (2000), Risk Sharing and the Supply of Mental Health services, *Journal of Health Economics* 19(6), pp. 1047-1065.

Shi, J. (2013), Labor Supply Response to Income Cutoffs of Health Insurance in the Massachusetts Reform, Working Paper, Boston University.

Sojourner A.J., D.C. Grabowski, R.J. Town, M.C. Chen, B.R. Frandsen (2013), Impacts of Unionization on Quality and Productivity: Regression Discontinuity Evidence From Nursing Homes, Working Paper, https://economics.byu.edu/frandsen/Documents/Nursing_Home_Unions.pdf.

Stearns, S., B. Wolfe and D. Kindig (1992), Physician responses to fee-for-service and capitation payment, *Inquiry* 29, pp. 416-425

Steinbusch, P.J.M., J.B. Oostenbrink, J.J. Zuurbier, F.J.M. Schaepkens (2007), The risk of upcoding in casemix systems: A comparative study, *Health Policy* 81(2-3), pp. 289-299.

Van de Ven, W.P.M.M. and F.T. Schut (2008), Universal Mandatory Health Insurance In The Netherlands: A Model For The United States?, *Health Affairs* 27(3), pp. 771-781.

Verbeek, M. (2004), A Guide To Modern Econometrics, Wiley, New York, 2nd edition.

VWS (2010), Interdepartementaal beleidsonderzoek curatieve GGZ, Attachment by the Report: Heroverweging curatieve zorg, Ministry of Health, Welfare and Sport, The Hague (in Dutch).

Publisher:

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

November 2014 | ISBN 978-90-5833-664-4