CPB Netherlands Bureau for
Economic Policy Analysis

Treatment responses of mental health care providers after a demand shock

This paper investigates how two different groups of Dutch mental health care providers responded to an exogenous 20% drop in the number of patients in 2012. Providers that operated under a budget increase treatment duration on average by 8%.
We find suggestive evidence for over-treatment as the longer treatments did not result in better patient outcomes.

For the group of self-employed providers, paid by a stepwise fee-for-service function, we find only a small significant increase in treatment duration for the least altruistic providers, which we relate to an income effect.

CPB Discussion Paper
Rudy Douven,
Minke Remmerswaal
and Tobias Vervliet
September 2019

Treatment responses of mental health care providers after a demand shock*

Rudy Douven [†1, 2], Minke Remmerswaal[1, 3], and Tobias Vervliet [‡4]

[1]CPB Netherlands Bureau for Economic Policy Analysis, the Hague, the Netherlands
[2]Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, the Netherlands
[3]Department of Economics, Tilburg University, Tilburg, the Netherlands
[4]SEO Amsterdam Economics, University of Amsterdam, Amsterdam, the Netherlands

September 3, 2019

This paper investigates how two different groups of Dutch mental health care providers responded to an exogenous 20% drop in the number of patients in 2012. Providers that operated under a budget increase treatment duration on average by 8%. We find suggestive evidence for over-treatment as the longer treatments did not result in better patient outcomes (i.e. general assessment of functioning scores). Both professional uncertainty and income effects may explain the results. For the group of self-employed providers, paid by a stepwise fee-for-service function, we find only a small significant increase in treatment duration for the least altruistic providers, which we relate to an income effect.

Keywords: physician incentives, mental health care, treatment outcomes, payment system

JEL Classification: H51, I11, I12, J22, J31 and J33

*We are very grateful to Robin Zoutenbier for his work at the start of this research. We gratefully acknowledge comments received during presentations at the CPB, Dutch Healthcare Authority (Utrecht), Dutch Ministry of Health, Welfare and Sports (Den Haag), ESHPM and ESE (Rotterdam), euHEA (Maastricht), GGZ-Nederland (Utrecht), iHEA (Boston), lolaHESG (Rotterdam) and Parnassia (Den Haag). We also thank our colleagues at the CPB, Ministry of Health and Dutch Healthcare Authority for comments. Moreover thanks to Jan Boone, Raf van Gestel, Frank Hoogendijk, Lucy Kok, Tom McGuire, Joe Newhouse, Erik Schut, Joyce van der Staaij, Paula Terra and Gert-Jan Verhoeven for comments on earlier versions of this paper. Lastly, special thanks to the Dutch Healthcare Authority for providing the data.

†Corresponding author: R.Douven@cpb.nl.

‡At the time of writing this paper, Tobias Vervliet was employed by the CPB Netherlands Bureau for Economic Policy Analysis.

1. Introduction

Smart payment models in health care are increasingly seen as a promising way to curb health care expenditures while maintaining good quality of care (McClellan, 2011). There is a growing empirical literature which addresses how payments systems in health care influence provider behavior (Ellis and McGuire, 1986; McGuire, 2000; Chandra et al., 2012; Chandra and Skinner, 2012; Christianson and Conrad, 2012; Johnson, 2014). Each payment model may provoke different responses by health care providers, and to find the best payment system, these provider responses and corresponding patient outcomes must be studied. Designing a payment system is especially challenging for health services that are supply sensitive with heterogenous, unknown or marginal treatment benefits (Skinner, 2012). In this respect mental health care is a particularly interesting sector to study as uncertainty and variation in treatments are greater than for other health services and responses to financial incentives are often exacerbated (Frank and McGuire, 2000).

In this paper, we study treatment behavior of mental health care providers in the Netherlands. We compare two groups of providers each paid by a different payment scheme, and thus both groups face different incentives. On the one hand, there are large institutions with salaried employees that operate under a budget constraint. On the other hand, there is a group of self-employed providers that are reimbursed per treatment episode by a stepwise fee-for-service function. Both types of providers faced a large, sudden demand shock in 2012 and 2013 because the government reduced insurance coverage and increased the level of the deductibles. The policy led to a plausible exogenous drop in the number of patients of about 20%.[1] We study empirically to what extent both types of providers changed their treatment behavior in response to this demand shock. Our approach is thereby similar to Gruber and Owings (1996) who use as demand shock a decline in fertility over a long period. They find that the declining fertility reduced the income of obstetricians/gynaecologists which led them to substitute from normal childbirth toward a more highly reimbursed alternative, Caesarean delivery.

The starting point of our analyses is a standard imperfect agency model that describes treatment behavior for the two types of providers. With this model we develop several hypotheses about how both types of providers might respond to the demand shock. Our paper contributes to the literature by including three additional mechanisms to the standard model: professional uncertainty, income effects and rationing.

Professional uncertainty is the fact that differences in beliefs, decision making and motivation of providers are important drivers of supply side variation. There is a growing body of

[1] Lambregts and van Vliet (2018) and Ravesteijn et al. (2017) discuss more extensively the demand side effects of this policy shock.

literature on professional uncertainty. For example, Cutler et al. (2019) find that cardiologists' responsiveness to financial factors and patient demand play a relatively small role in explaining equilibrium variations in utilization patterns in Medicare. They argue that different beliefs of physicians about the effectiveness of treatments and specific procedures, often unsupported by clinical evidence, are more important. Currie et al. (2016) and Currie and McLeod (2017) show also that there is a great deal of variation in both responsiveness and treatment methods across doctors and that these characteristics of doctors are fairly stable over time. Currie and McLeod (2018) argue that treatment choice depends on a physicians diagnostic skill, so that the optimal treatment can vary even for identical patients. Abaluck et al. (2016) document enormous across-doctor heterogeneity in imaging tests for pulmonary embolism. Douven et al. (2019) show that self-employed mental health care providers differ in their degree of altruism, or professionalism, and find that altruistic providers report better treatment outcomes.

Second, we incorporate rationing for budgeted providers, as budgets may restrict capacity or time that is available to treat patients optimally. Note that modelling budgets in healthcare is by its very nature complex (Christianson and Conrad, 2012). We take an agnostic approach and assume that tight budgets may surpress provider responses, which may influence treatment duration and quality. Important for our paper is the link with professional uncertainty. After the demand shock providers responses become less restricted which may reveal insight in professional uncertainty.

Third, income effects may be important. A drop in the number of patients may reduce (future) income and providers may try to recoup some of this income loss by changing their treatment behavior. There is ample evidence that physicians may treat patients differently when financial incentives are involved. For example, one of the first to find evidence of income effects were Gruber and Owings (1996). More recently, van Dijk et al. (2013) used an exogenous change from capitation to fee-for-service payments and showed that Dutch general practitioners increased their services after the change. Clemens and Gottlieb (2014) used an exogenous payment shock in Medicare to show that providers in areas with higher payment shocks experienced significant increases in health care supply.

To test our hypotheses empirically we use a large administrative data set which contains all treatment episodes for all patients in the secondary curative mental health care in the Netherlands. Our sample period covers the years before (2008-2011) and after the demand shock (2012-2013).

For the budgeted institutions with salaried employees we find, after controlling for changes in case-mix, an 8% increase in treatment duration after the demand shock. This increase in treatment duration does not result in better treatment outcomes, which suggests over-treatment.[2] Both professional uncertainty and income effects may explain the results. Professional uncertainty suggests that before the demand providers perceived implicitly or explicitly some form of rationing. After the demand shock more capacity became available and provider treated patients longer because they expected that it would benefit patients, but these expectations did not materialize ex-post. Income effect may occur because the demand shock implied a potential loss in current and future income of providers, and longer treatments may be a mechanism to secure their income. At the employee level also "shirking" may have played a role (i.e. employees became less productive per hour).

We find almost no changes in treatment duration for the group of self-employed providers that are reimbursed by a discontinuous fee-for-service function. The discontinuities in the payment function seem to have prevented an increase in treatment duration after the demand shock. Only for the least altruistic self-employed providers we find a small significant increase in treatment duration, which we relate to an income effect.

Taking into account how payment systems affect patient health outcomes is important for performing (partial) welfare analysis. A common problem with empirical studies is often gathering and comparing patient outcomes, as this is often complicated by patient heterogeneity and endogenous provider choices. In this paper we circumvent this problem because we can compare for each treatment a patients health status before and after treatment.[3]

This paper complements our previous work on the supply side of the Dutch mental health care sector. Douven et al. (2015) show that self-employed providers who were paid according to the discontinuous payment scheme showed different treatment behavior than budgeted providers between 2008 and 2010. Moreover, altruistic providers treated mental health patient shorter and reported better patient outcomes than financially motivated providers (Douven et al., 2019). In this paper, we gathered three years of additional data which allows us to study treatment responses of providers after a large demand shock in 2012.

Lastly, we contribute to the literature by showing that for supply sensitive treatments, such as mental health care services, financial mechanisms play an important role. It is well-known that tight budget policies may result in lower quality of care, and this paper provides suggestive

[2]Changes in patient benefits were measured in terms of changes in GAF scores, which is a crude outcome measure in mental health care. See also section 4.

[3]Previous researchers have used (implicit) outcome measures, such as mortality rates (e.g. Clemens and Gottlieb, 2014), survival rates (Jacobson et al., 2017), treatment choices (Gruber and Owings, 1996) or a variety of hospital conditions (Currie et al., 2016, Doyle et al., 2015, 2017).Brosig-Koch et al. (2018) use a controlled laboratory setting to define quality and show that a fee-for-service payment system may lead to overprovision of services and capitation type of systems to underprovision of services.

evidence of the opposite: loose budgets may result in over-treatment. Designing optimal budgets for supply sensitive treatments is extremely complicated and can result in large inefficiencies. This problem is of course not restricted to health care but widespread present in all parts of the economy where budgets play a role. From the group of self-employed providers we learn that provider responses do not only depend on the characteristics of the payment system but also on the characteristics of providers, i.e. their degree of altruism and sensitivity to exogenous policy shocks and income effects. Ideally, these different aspects should all be taken into account when designing an optimal payment system.

The structure of our paper is as follows. Section 2 describes the institutional setting of the Dutch mental health care sector and the demand shock. Section 3 explains our theoretical framework. Section 4 describes the data and provides a descriptive analysis. Section 5 and 6 explain the empirical strategy and a discussion of the results. In section 7 we conclude.

2. Institutional setting

This study focuses on the secondary curative mental health care sector in the Netherlands. Curative mental health care is specialized care for patients with a relatively serious mental health condition. Unlike long term mental health care, these patients do not remain in a residence or other mental health facility for a long period. In the Netherlands, curative mental health care costs four billion euros per year, which accounts for roughly 65% of total mental health care expenditure (Dutch Healthcare Authority, 2013) and about 10% of total expenditure on curative care. In 2008, the Dutch government placed secondary mental health care under a regime of regulated competition.[4]

Secondary curative mental health care is part of the basic benefit package and therefore covered by the mandatory insurance scheme for all inhabitants of the Netherlands.[5] Patients need a referral from their general practitioner to have access to secondary curative mental health care, but with this referral they are free to choose any mental health care provider. However, in practice most patients tend to follow the advice of their general practitioner. Patients face out-of-pocket payments for mental health care services: a mandatory generic deductible which applies to most of the services in the basic insurance package.[6] In 2008, this deductible was 150 euros and raised annually by 5 to 10 euros.

We distinguish two types of providers for mental health services: budgeted providers and self-employed providers. Henceforth, we will refer to budgeted providers as B-providers, and to self-employed providers or non-budgeted as NB-providers.

Roughly 10% of all treatments in curative mental health care are provided by NB-providers who often operate in small practices. NB-providers are compensated by health insurers according to their production and case-mix, which is defined in a DBC or Diagnosis Treatment Combination.[7] Every DBC is a treatment episode which refers to a specific diagnosis and a specific treatment.[8] For example, one DBC may encompass an intake plus multiple therapy sessions that take place over the course of one year.[9] The treatment episode is closed once a

[4]Regulated competition in Dutch curative health care was introduced in 2006. The goal of this policy was to improve efficiency in the sector by letting insurers buy care on behalf of their enrollees.

[5]The basic benefit package covers most types of curative care, such as pharmaceutical care, hospital care, GP care, physiotherapy, et cetera.

[6]GP care and care related to pregnancy and child birth are exempted from the deductible. Also, the deductible does not apply to persons below 18 years old.

[7]In Dutch: Diagnose Behandel Combinatie.

[8]The DBC has strong similarities with a Diagnostic Related Group (DRG) that is used in many other countries. See Westerdijk et al. (2012).

[9]Consider for example a patient with mild depression who has an individual therapy session of 60 minutes each month for a period of ten months. The patient does not receive any medication or other types of treatment. These

treatment is completed or when one year has passed since the start of the treatment episode (then a new treatment episode may be started for the next year). NB-providers negotiate a tariff for each DBC with insurers, meant to cover average estimated labor and capital costs for a treatment. The maximum tariff for each DBC is determined by the Dutch Healthcare Authority (NZa). Figure 9 in Appendix B shows that the tariff structure of a DBC follows a stepwise fee-for-service function with thresholds at 250, 800, 3000, 6000 and 12000 minutes of treatment. Between these thresholds, the tariffs are flat. For example, the tariff for DBC "Depression, 250 to 800 minutes" was 956 euros in 2010.[10]

The majority of the treatments, about 90%, are provided by B-providers. B-providers are large institutions such as regional facilities for ambulatory care or specialized psychiatric hospitals. Since B-providers are more specialized they also attract more severe patients than NB-providers (see also section 4). Importantly, until 2014, these institutions were not compensated according to their DBC production, but based on annual budgets. The budgets were determined by (expected) production and regional parameters such as labor and capital costs, and were negotiated with the largest health insurer in the geographical region.[11] Also, budgeted providers recorded all their treatment episodes as a DBC. The differences of the payment systems and financial incentives for B-providers and NB-providers will be discussed in more detail in the next section.

Market developments in 2012 and 2013

In 2012 and 2013, the Dutch government implemented several reforms with the intention to reduce public spending on curative mental health care. Insurance coverage for mental health services was reduced, cost-sharing for mental health use was increased, and the regulated maximum prices for treatment episodes were lowered.

First, the government excluded treatments with a diagnosis "Adjustment disorder", about 10% of all curative mental health services, from the basic insurance package in 2012. In 2013, treatments with diagnoses "V-codes" were also excluded, which covered about 7% of all treat-

therapy sessions are provided by a psychotherapist. This patient's treatment episode is classified as: "Depression, 250 to 800 minutes, no medication" (DBC Onderhoud, 2013).

[10]In general NB-providers negotiate with insurers a percentage of the maximum tariff. We have no information about these negotiated percentages but most of these percentages are between 75% and 100% of the maximum tariff.(Dutch Healthcare Authority, 2013).

[11]The Netherlands was divided in 32 regions and in each region a dominant health insurer was appointed by the government. This dominant insurer received a regional budget from the government for all mental health services in the region. In 2014, this concept was abolished and B-providers had to negotiate with each individual health insurer separately. These developments fall outside of the sample period of this research.

ments.[12] These treatments were no longer covered by the basic benefit package and patients therefore had to pay the entire treatment out of their own pocket. Important for our analyses is that the number of patients with these disorders vanished almost completely in the administrative database in 2012 and 2013. This will be shown in Section 4.

Second, in 2012, the government raised co-payments from 10 to 20 euros per visit in primary mental health care care and introduced a deductible of 200 euros for secondary mental health care specifically.[13] Already in 2013, the government abolished this deductible, but simultaneously increased the mandatory general deductible to 350 euros.[14] Lambregts and van Vliet (2018) and Ravesteijn et al. (2017) show that the deductible for mental health care has prevented many patients to visit a mental health care provider. If a patient decides to visit a mental health care provider in 2012 then the patient has to pay the full 200 euros deductible. Thus, the demand side effect of the reduction the deductible is mainly a yes/no decision to visit a provider. Once the initial decision to visit a provider is taken by the patient then any follow-up decision is without any monetary costs for the patient. Therefore, we will assume in the rest of the paper that follow-up decisions by patients to visit a mental health care provider are not influenced by the deductible. This allows us to relate changes in treatment duration responses mainly to the supply side and not to the demand side.

Lastly, the government lowered maximum price tariffs for all treatment episodes with 5.5% in 2012 (Dutch Healthcare Authority, 2014). Furthermore, in 2013 the government, insurers and mental health care sector agreed to limit the future growth of curative mental health care spending.[15]

As shown in previous studies by Lambregts and van Vliet (2018) and Ravesteijn et al. (2017) the reforms resulted in a large drop in the number of patients in 2012 and 2013. In our data we find a reduction in the number of patients of roughly 20%. Note that this reduction is the net result of the aforementioned policy changes and subsequent responses by mental health providers to secure patients.[16] The strong decline in patient demand was unexpected for the government,

[12] Adjustment disorders are conditions related to stressful events and 'V-codes' to relational or occupational problems.

[13] This was on top of a mandatory deductible for all curative care services. Costs of emergency treatments were excluded from the deductible.

[14] The reason for the abolishment of the deductible in 2013 were related to budgetary windfalls in 2013 and, presumably, the unexpected strong response by patients to the deductible which attracted a lot of media attention.

[15] In this agreement, for example, was laid out that substitution from secondary to primary mental health care should be stimulated.

[16] Health care providers have only limited control about whether a patient seeks (or does not seek) mental health care. However, anecdotical evidence suggests that providers have used various channels to attract more patients. For example, patients with a diagnoses of "adjustment disorder" or "V-codes" may have been recoded to for example a "mood disorder". Some providers may have offered to pay the co-payments and deductible. We

insurers and providers, and therefore plausibly exogenous.

While the number of patients declined substantially in 2012, total spending on Dutch secondary mental health care stayed relatively stable, see Table 1.[17] This implies that cuts in individual provider budgets were limited. For example, the 0.2 billion euro difference between 2011 and 2012 roughly corresponds to a 5% budget cut. Dutch Healthcare Authority (2013) shows that insurers negotiated lower price tariffs for DBC's with NB-providers between 2011 and 2012 although the exact prices are unknown.

Table 1 indicates that the number of workers in the total mental health care sector increased during 2009-2011 up to 89,000 but remained relatively stable in 2012 and 2013. Only in 2014 there was a drop in employment.[18]

Table 1: Total spending and employees in Dutch secondary mental health care sector

Year	2009	2010	2011	2012	2013	2014
Spending (x 1 billion euros)	-	4.0	4.3	4.1	4.0	4.0
Number of employees (x 1,000)	87	87	89	89	89	86

Notes: The total spending figures are retrieved from Dutch Healthcare Authority (2014, 2016) and the number of employees from www.azwinfo.nl (retrieved at April, 23, 2019).

do not find evidence in our data that providers have increased the number of treatment episodes per patient (see Table 7, 8). Also, providers may have obtained more patients by reducing waiting lists. However, we do not observe lower average annual waiting times after the shock (Dutch Healthcare Authority, 2014), which remains a puzzle.

[17] From a budgetary perspective the policy measures turned out to be successful for the government as the growth of secondary mental health care spending had stopped in the year 2012 and 2013.

[18] This holds also for full-time employees which increased during 2009-2011 from 71,000 to 73,000 but remained relatively stable with 72,000 and 73,000 in 2012 and 2013. In 2014, it dropped to 71,000.

3. Theoretical Framework

In the setup of our model we follow the literature on physician agency models, see for example McGuire (2000); Chandra and Skinner (2012); Cutler et al. (2019); Douven et al. (2019). We extend the standard models in several ways.

First, we incorporate income effects in the model so that providers can respond to a reduction in (future) income by changing their treatment behavior. We model that income becomes more important in the utility function if total income declines. Second, we include professional uncertainty to allow for supply side variation as a result of differences among providers in beliefs, decision making, and motivation (Chandra and Skinner, 2012). Third, for both type of providers we explicitly model the key characteristics of their payment function. For example, for providers that operate under a budget we consider the possibility of rationing.

3.1. General model

In our model, a provider j decides on a patient's treatment episode i. The key instrument in this decision is the duration of the treatment episode x_i, as measured by the total number of treatment minutes.

The demand side of the model is an indirect patient's utility function which is a function of health, patient's out-of-pocket payments, preferences and treatment duration. This function reflects the demand of a fully informed patient (Cutler et al., 2019). In this function we denote "true" patient benefit or quality from treatment as Q_i. We assume ${Q_i}''(x_i) \leq 0$, for $x_i \geq 0$. That is, marginal patient's benefits from treatment decline as the treatment duration increases. Solving the demand function for optimal treatment duration yields x_i^{D}, which we assume is the fully informed patient's demand. We assume that "over-treatment" occurs if for a treatment duration x_i holds that $x_i > x_i^{\mathrm{D}}$ and $Q_i(x_i) - Q_i(x_i^{\mathrm{D}}) \leq 0$, i.e. for additional care provided at the margin a patient gains no improvements in health. This corresponds to the "flat of the curve" hypothesis, see e.g. Fuchs (1986).

On the supply side, the provider's utility from treating patients depends on two components: the utility a provider perceives from expected patient's benefits, denoted by S_{ij} and the utility from net financial benefits for all treatments, denoted by Π_j. We assume that the utility of provider j when performing q treatment episodes, with $i = 1, ...q$, is given by:

$$U_j(x_i, \theta_i, q) = \sum_{i=1}^{q} S_{ij}(x_i, \theta_i) + \alpha_j \Pi_j^{\frac{1}{\gamma_j}} \tag{1}$$

The provider's utility is the sum of the provider's assessment of expected patient benefit $S_{ij}(x_i, \theta_i)$, which depends on treatment duration x_i and the patient's health status θ_i. We allow S_{ij} to differ across providers because of professional uncertainty, hence the subscript j. Providers may have different beliefs about expected patient's benefits, because uncertainty and variations in mental health services are great (Frank and McGuire, 2000). Because of professional uncertainty, some providers may decide on a treatment duration $x_i > x_i^{\mathrm{D}}$ because $S_{ij}(x_i) - S_{ij}(x_i^{\mathrm{D}}) > 0$ while in fact $Q_i(x_i) - Q_i(x_i^{\mathrm{D}}) \leq 0$. These providers "overtreat" because they believe it is in the best interest of the patient to do so. We assume $S_{ij}''(x_i) \leq 0$, for $x_i \geq 0$.[19] Moreover, we assume that provider utility is additive in the number of treatment episodes.

Provider j attributes weight α_j to the utility it receives from its net financial benefits Π_j. We allow α_j to be provider-specific and time-invariant. Douven et al. (2019) show that there is a large variation in α_j's between self-employed mental health care providers. With $\gamma_j \geq 1$ we allow for a concave relationship of Π_j in the utility function. This reflects the income effect. If $\gamma_j > 1$ there is more pressure on financial benefits when these benefits are low. We will define Π_j more precisely in the next two subsections as both groups of providers have different payments systems.

We assume that provider j maximizes provider utility (1) for given θ_i's and q:

$$\max_{x_i} U_j(x_i, \theta_i, q) \tag{2}$$

In the next two subsection we describe for both provider how they optimize their utility function and how the optimization problem alters after the demand shock in 2012 and 2013, i.e. a drop in q. The most important difference between both providers is the payment model, i.e. they differ in financial benefits Π_j (as mentioned in section 2). In the next two subsections we model the payment models explicitly in the utility function. We use superscripts NB and B to distinguish between both types.

3.2. NB-providers

NB-providers receive a financial compensation per treatment episode i which looks like a stepwise (or staircase) fee-for-service function, which is given by:

$$p_i^{\mathrm{NB}}(x_i, \theta_i) = P_i(k^l, \theta_i) \quad \text{for} \quad k^l \leq x_i < k^{l+1} \tag{3}$$

where k^l represents the treatment duration threshold with $l = 1, ..., 5$. See Figure 9 in Appendix B.[20] NB-providers are single specialists or a few cooperating self-employed specialists

[19] Owen et al. (2016) provide some evidence for the assumption that marginal benefits to patients decline in mental health care.

[20] The treatment duration thresholds are the same for all treatments: $k^1 = 250, k^2 = 800, k^3 = 1800, k^4 =$

who work in small private practices with much lower investment costs and flexible labor contracts. We make the simplifying assumption that costs for each individual treatment episode are given by $c_j^{\mathrm{NB}}(x_i, \theta_i) = c_j^{\mathrm{NB}} x_i$, thereby only considering variable costs and ignoring fixed costs. Thus, net financial benefits for provider j are

$$\Pi_j^{\mathrm{NB}} = \sum_{i=1}^{q} \left(p_i^{\mathrm{NB}}(x_i, \theta_i) - c_j^{\mathrm{NB}} x_i \right) \tag{4}$$

To obtain the optimal treatment durations x_i^* for each treatment episode we substitute (4) in (1) and optimize with respect to x_i:[21]

$$\frac{\partial S_{ij}^{\mathrm{NB}}}{\partial x_i^{*\mathrm{NB}}} = \overline{\alpha_j}^{\mathrm{NB}} c_j^{\mathrm{NB}} \qquad \text{or} \qquad x_i^{*,\mathrm{NB}} = k^l \tag{5}$$

with

$$\overline{\alpha_j}^{\mathrm{NB}} = \frac{\alpha_j^{\mathrm{NB}}}{\gamma_j} \Pi_j^{\mathrm{NB}\,\frac{1-\gamma_j}{\gamma_j}}$$

The first term in (5) is the interior solution and the second term is the corner solution. Note that we obtain these two solutions because the derivative of the stepwise fee-for-service function p_i^{NB} is zero or doesn't exist (at thresholds k^l). For each interior solution $x_i^{*,\mathrm{NB}}$ there is also a nearest threshold $k^l > x_i^{*,\mathrm{NB}}$ for which treatment duration is longer and the reimbursement is higher. A provider may jump to a corner solution k^l if its utility is higher at k^l than at the interior solution. This will be more often the case the closer $x_i^{*,\mathrm{NB}}$ is located to k^l. Whether a NB-provider will decide to jump to the corner solution k^l will depend on several factors: the size of γ_j, the weight α_j^{NB} and the variable costs c_j^{NB}. Thus, in the case of jumping to a next threshold, the step-function p_i^{NB} may result in overprovision of services as the provider may prolong treatment duration for own financial reasons. For a more thorough discussion on this stepwise fee-for-service system, see Douven et al. (2015, 2019).

Our main question is how providers respond to the policy reforms and the concomitant demand shock in 2012 and 2013, i.e. an exogenous drop in the number of patients q, and a tariff cut, i.e. a drop in p_i^{NB} (by around 5.5%).

At the demand side, we argue that the introduction of deductibles for mental health care does not alter fully informed patient's demand x_i^{D}. The reason is that the only argument that changes in the patient's utility function is the out-of-pocket payment due to changes in deductibles or the basic benefit package. However, as explained before, the deductible is mainly important for

$3000, k^5 = 6000$. For example, in figure 9, the fees for schizophrenia in 2011 are given by $P(350) = 1,070$ euro and $P(1000) = 2,020$ euro. Fees $P_i(k^l, \theta_i)$ might differ slightly across diagnoses.

[21] We assume that providers treat patients independently from each other, which allows us to optimize each x_i independently.

a patient to make the yes/no decision to visit a psychologist or psychiatrist. As soon as a patient decides to go to see a psychiatrist or psychologist, he or she will exhaust the deductible. Patients have to pay the full amount for a treatment related to adjustment or relational disorders as these were excluded from the basic benefit package. To conclude, we assume that before and after the shock treatment duration was not influenced by different patient behavior.

At the supply side the fall in the number of patients q and tariff cuts are likely to lower net financial benefits. Thus, a provider who is sensitive to income effects may alter treatment duration. We distinguish two cases.

(1) If $\gamma_j = 1$, providers do not react to changes in their income, i.e. $\overline{\alpha_j}^{\text{NB}} = \alpha_j{}^{\text{NB}}$, and the optimal treatment duration in (5) will be the same before and after the shock.

(2) If $\gamma_j > 1$, providers are sensitive to changes in their total income, i.e. $\overline{\alpha_j}^{\text{NB}}$ decreases for a given α_j^{NB}. The effect on provider responses may be twofold. First, consider the case of an internal solution before the shock, i.e. say $x_i^{*\text{NB}} \neq k^l$. Optimal treatment duration may decline after the shock as treating a patient beyond a threshold becomes more costly. In this case treatment duration may decrease at the margin. However, as NB-providers are sensitive to changes in income, they may also decide to jump to a corner solution, i.e. the next threshold k^{l+1}. The jump $k^{l+1} - x_i^{*,\text{NB}}$ can be large, it will only occur when the increase in fee is larger than the increase in costs. Second, consider the case of a corner solution before the shock, i.e. $x_i^{*\text{NB}} = k^l$. Now, NB-providers only have an incentive to increase treatment duration to a next fee threshold k^{l+1}. Again, this implies a large jump as the distances between two consecutive thresholds k^l and k^{l+1} is substantial. Whether we will observe such a jump depends on the size of γ_j. We can deduce the following testable hypothesis for NB-providers.

Hypothesis NB-providers. NB-providers will prolong treatment duration after the shock if they care (strongly) about their income ($\alpha_j^{\text{NB}} > 0$) and if they are sensitive to changes in income ($\gamma_j > 1$).

Note that in this case professional uncertainty does not reveal itself as NB-provider are unconstrained in their treatment responses before and after the demand shock. As we will see in the next subsection this may be different in the case for B-providers. Due to budgets B-providers may have experienced rationing before the shock which may have have restricted their treatment responses. After the demand shock relative capacity increased which allows them to reveal, or to get closer to, their unconstrained treatment reponses.

3.3. B-providers

An important difference between B- and NB-providers is that B-providers face budgets. Budgets complicate the analysis since we have to take into account that B-providers may be constrained to treat patients optimally. We assume that B-providers operate under a budget constraint Y_j^{B} that they determine ex-ante after a negotiation process with the most dominant health insurer in the region. In practice, the budget negotiations are private and unobserved. For example, if B-providers negotiate with insurers about their budget then several factors play a role such as the previous years' budget, fixed costs, the number of patients, the severity and treatment duration of patients etc. As there is, opposed to NB-providers, no directly observable price tag for an individual treatment episode i, we approximate it by $p_i^{\mathrm{B}}(x_i, \theta_i)$ which is some monotone non-decreasing function of x_i. We approximate the total level of production for B-provider j who produces q treatment episodes at the end of a year therefore by:

$$R_j^{\mathrm{B}}(x_i, \theta_i, q) = \sum_{i=1}^{q} p_i^{\mathrm{B}}(x_i, \theta_i) \tag{6}$$

B-providers will only receive the ex-ante budget Y_j^{B} if $R_j^{\mathrm{B}} \geq Y_j^{\mathrm{B}}$, otherwise they will receive the realized total production R_j^{B}. If B-providers produce less, i.e. $R_j^{\mathrm{B}} < Y_j^{B}$, then they also run the potential risk of budget cuts in subsequent years. Moreover, we assume that B-providers have relatively high fixed costs c_j^{B}. This reflects that B-providers are organized in large regional institutions, such as a regional facility for ambulatory care or a specialized psychiatric hospital, who have often invested in large buildings or facilities that have to be paid off over time. In addition, they have hired many salaried employees who typically have long-term employment contracts. We model net financial benefits for B-provider j as follows:

$$\Pi_j^{\mathrm{B}} = \min\left\{R_j^{\mathrm{B}}, Y_j^{\mathrm{B}}\right\} - c_j^{\mathrm{B}} \tag{7}$$

The first term in this equation demonstrates that a B-provider either receives its budget Y_j^{B} or an amount $R_j^{\mathrm{B}} < Y_j^{\mathrm{B}}$, if production falls below the predetermined budget. To determine the optimal treatment duration $x_i^{*\mathrm{B}}$ for each individual patient we first maximize the utility of B-provider j for the unrestricted case:

$$\sum_{i=1}^{q} S_{ij}^{B}(x_i, \theta_i) + \alpha_j^{\mathrm{B}} \left\{ \sum_{i=1}^{q} p_i^{B}(x_i, \theta_i) - c_j^{\mathrm{B}} \right\}^{\frac{1}{\gamma_j}} \tag{8}$$

where we assume that each psychologist or psychiatrist who works for a B-provider can, in the unrestricted case, freely choose their optimal treatment duration for each individual patient, the solution $x_i^{*\mathrm{B}}$ satisfies:

$$\frac{\partial S_{ij}^{\mathrm{B}}}{\partial x_i^{*\mathrm{B}}} = \overline{\alpha_j}^{\mathrm{B}} \frac{\partial p_i^{\mathrm{B}}(x_i^{*\mathrm{B}}, \theta_i)}{\partial x_i^{*\mathrm{B}}} \tag{9}$$

with

$$\overline{\alpha_j}^{\mathrm{B}} = \frac{\alpha_j^{\mathrm{B}}}{\gamma_j} \left\{ \sum_{i=1}^{q} p_i^B(x_i^{*\mathrm{B}}, \theta_i) - c_j^{\mathrm{B}} \right\}^{\frac{\gamma_j - 1}{\gamma_j}}$$

The solution $x_i^{*\mathrm{B}}$ holds for the case where the budget is non-binding, i.e. $R_j^{\mathrm{B}}(x_i^{*\mathrm{B}}, \theta_i, q) = \sum_{i=1}^{q} p_i^{\mathrm{B}}(x_i^{*\mathrm{B}}, \theta_i) \leq Y_j^{\mathrm{B}}$, and states that the marginal benefit to the patient of an additional unit treatment duration is equal to its marginal financial benefits. When the budget constraint is binding, i.e. $R_j^{\mathrm{B}}(x_i^{*\mathrm{B}}, \theta_i, q) > Y_j^{\mathrm{B}}$, then the budget is too tight for B-providers and they have not enough resources, in terms of money, employees or facilities, to reach the level of production that matches with their internal optimum $x_i^{*\mathrm{B}}$. In that case optimal treatment duration will be lower than $x_i^{*\mathrm{B}}$. This corresponds to rationing of health care. Note that without any further information we do not know which of both solutions is closer to x_i^{D}, but the demand shock allows us to shed more light on this problem.

Our main research question is how B-providers will respond to price cuts and a fall in the number of patients. Again, as argued in the NB-case, we assume that patient demand x_i^{D} does not change for those patients that decide to visit a psychologist or psychiatrist, i.e. the policy shock affects the demand side at the extensive margin, i.e. a fall in the number of treatment episodes q, but not at the intensive margin, i.e. the number of treatment minutes.

As the supply side we have to take several aspects into account. Although we have no information about budgets Y_j^{B} at the individual provider level, we showed in section 2 that the total expenditure (and thus corresponding budgets) for secondary curative mental health care fell with about 5% in 2012. Individual budget cuts were on average likely of the same size. The tariff cuts of about 5.5% were anticipated in the budget negotiations and the prices $p_i(x_i, \theta_i)$ for individual treatments as this was announced in 2011 by the government.[22] However, the large fall in the number of patients of about 20% in 2012 was presumably not anticipated at the end of 2011 when budget negotiations for 2012 were concluded. Considering these factors and assume that providers would treat patients in a similar way as before the demand shock, then this would result in a considerable reduction of total production R_j^{B}.[23] Moreover, we showed

[22]Note that there were few incentives for further budget cuts as negotiations with providers were performed by the most dominant insurer in the region who run few financial risks.

[23]If treatment durations x_i do not alter before and after the shock for a similar patient, then $Y_j^{\mathrm{B}} - \sum_{i=1}^{q} p_i^B(x_i, \theta_i)$ before the shock will be considerably smaller than $0.95Y_j^{\mathrm{B}} - \sum_{i=1}^{0.8q} 0.945 p_i^B(x_i, \theta_i)$ after the shock, where the latter term corresponds to a 5% lower negotiated budget, a 80% subset of q treatment episodes before the shock, and a 5.5% cut in price per treatment episode.

in section 2 that the number of (full) employees, mostly paid on a salary basis with long term contracts, in the total mental health care sector remained relatively stable during 2010-2012, which suggests that capacity, and fixed costs c_j^{B}, did not change much in 2012 and 2013. As a result, we argue that relative capacity of B-providers has most likely increased for an individual patient in 2012.

The increase in relative capacity after the demand shock creates the possibility for B-providers to treat patients longer than before the shock. Whether B-providers react to this shock depends not only whether they respond to an income effect, as was the case for NB-providers, but also whether they implicitly or explicitly experienced some form of rationing in the years before the shock. This leads to the following testable hypothesis:

Hypothesis B-providers: Denote optimal treatment duration for a similar patient before the shock with $x_i^{*\mathrm{B},1}$ and after the shock with $x_i^{*\mathrm{B},2}$.

First, we consider the case that $x_i^{*\mathrm{B},1}$ was an interior solution, i.e. $R_j^{\mathrm{B}}(x_i^{*\mathrm{B},1}, \theta_i) \leq Y_j^{\mathrm{B}}$:
(1) $\gamma = 1$ (no income effect). The optimization problem does not change: $x_i^{*\mathrm{B},2}{=}x_i^{*\mathrm{B},1}{=}x_i^{*\mathrm{B}}$.
(2) $\gamma > 1$ (an income effect). We expect $x_i^{*\mathrm{B},2}{>}x_i^{*\mathrm{B},1}$ as providers respond to a loss in income.[24]

Contrary to the case of NB-providers, we have to take the possibility of rationing into account. Therefore, we consider the second case that $x_i^{*\mathrm{B},1}$ is a corner solution, i.e. $R_j^{\mathrm{B}}(x_i^{*\mathrm{B},1}, \theta_i) > Y_j^{\mathrm{B}}$:
(3) $\gamma = 1$ (no income effect). We expect $x_i^{*\mathrm{B},2}{>}x_i^{*\mathrm{B},1}$, as providers receive more capacity per patient to reach the interior treatment duration optimum $x_i^{*\mathrm{B}}$. This may be a case of good agency or professional uncertainty (see subsection below).[25]
(4) $\gamma > 1$ (an income effect). We expect $x_i^{*\mathrm{B},2}{>}x_i^{*\mathrm{B},1}$, as this is a combination of (2) and (3).

3.4. Testing for over-treatment

Let d =B,NB. In our empirical analyses we will test whether or not treatment duration increases after the shock, i.e whether $x_i^{*\mathrm{d},2}$ is larger, equal or smaller than $x_i^{*\mathrm{d},1}$ and whether outcomes for similar patients have changed after the shock, i.e. whether $Q(x_i^{*\mathrm{d},2})$ differs from $Q(x_i^{*\mathrm{d},1})$.[26]

Over-treatment: We have suggestive evidence for over-treatment of mental health services if

[24] There is an income effect in equation (9) because after the shock the number of patients q and prices p_i^B fall. Treating patients in a similar way as before the shock results in lower net benefits $\sum_{i=1}^{q} p_i^B(x_i^{*\mathrm{B},1}, \theta_i) - c_j^{\mathrm{B}}$. Hence, $\overline{\alpha_j}^{\mathrm{B}}$ decreases after the shock.

[25] Good agency is a special case of professional uncertainty. Good agency implies that providers ex-ante expect that longer treatments will be on average beneficial for patients, and these expectations turn out to be correct ex-post.

[26] Note, that we use that we use Q instead of S because in the data we observe the "true" patient benefits, while S reflects the expected patient benefits which may differ from realized patient benefits.

providers prolong treatment duration after the demand shock, i.e. $x_i^{*\mathrm{d},2} > x_i^{*\mathrm{d},1}$, and $Q(x_i^{*\mathrm{d},2}) - Q(x_i^{*\mathrm{d},1}) \leq 0$.

The identification of the mechanisms is straightforward for NB-providers; if we find suggestive evidence for over-treatment we can attribute this to an income effect.

For B-providers we may not be able to distinguish between the four different cases that we discussed in the hypotheses for B-providers, however, we can distinguish three situations:

- $x_i^{*\mathrm{B},2} = x_i^{*\mathrm{B},1}$ we are in case (1). Rationing and income effects do not play a role.
- $x_i^{*\mathrm{B},2} > x_i^{*\mathrm{B},1}$ and $Q_i(x_i^{*\mathrm{B},2}) - Q_i(x_i^{*\mathrm{B},1}) > 0$ we are in case (3) or (4). We label this situation as good agency since B-providers increase treatment duration to improve patient benefits. We reject the income effect hypothesis from case (2), because when starting from an internal optimum we do not expect significant increases in patient benefits when longer treatments are driven only by financial motivations.
- $x_i^{*\mathrm{B},2} > x_i^{*\mathrm{B},1}$ and $Q_i(x_i^{*\mathrm{B},2}) - Q_i(x_i^{*\mathrm{B},1}) \leq 0$ we are in case (2), (3) or (4). We identify over-treatment. But, without additional information, we cannot distinguish between the mechanism, whether it is driven by an income effect, professional uncertainty, or by both. Note that in the case of professional uncertainty, B-providers may act in the best interest of the patient, as they perceive it, and not be aware of any over-treatment.

4. Data and descriptive statistics

For this research we use a large administrative data set provided by the Dutch Healthcare Authority. It contains all treatment episodes for all patients in the secondary curative mental health care in the Netherlands. Our sample period covers the years before (2008-2011) and after the demand shock (2012-2013).

4.1. Description of the data

The data set contains detailed information on all treatments in the Dutch mental health care sector. The data can be grouped into patient characteristics, provider characteristics, treatment characteristics and treatment outcomes.

For each patient age and gender is available. The patient's diagnosis, consisting of a main and sub-diagnosis, is also registered by the provider. To illustrate, we can observe that a patient has a "Mood" disorder with sub-diagnosis "Depression", and for example not a "Bipolar" disorder. There are 19 main diagnoses and over a hundred sub-diagnoses. At the beginning of a treatment episode, each practitioner assesses the mental health status of a patient by means of the Global Assessment of Functioning (GAF). This GAF score is measured on a ten point scale, where lower GAF scores indicate more severe mental health conditions and higher GAF scores imply less severe conditions.[27]

Providers are grouped into B- and NB-providers (see also Section 2). Using a unique provider ID, we can follow each provider over the six year period. We also know which treatment episodes were performed by which provider.

For each treatment episode several characteristics are recorded. The exact day of the start and end of a treatment episode. It is possible that a treatment episode is finished, but the treatment is not. In that case, a provider starts a new treatment episode labeled "continued treatment". The prior treatment episode is then labeled "regular treatment". Regular and continued treatments comprise over 90% of all treatments episodes. Providers record each minute they spend on a patient in the treatment episode. As a result, we observe treatment duration per treatment episode, which is one of the main dependent variables in our empirical analysis. Furthermore, providers distinguish between direct treatment time, when a provider is treating

[27] The GAF score ranges from 1 to 100, but is measured in ten categories: 1-10, 11-20, ... , 91-100. A GAF score of 1-10 means: "Persistent danger of severely hurting self or others (e.g., recurrent violence), persistent inability to maintain minimal personal hygiene or serious suicidal act with a clear expectation of death.", whereas as a GAF score of 91-100 means "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." For a detailed description of the GAF score, see the DSM-V handbook of American Psychiatric Association (2000)

the patient in the actual presence of the patient, and indirect treatment time, when the provider is doing preparation or administrative work for the patient.

The data also offer treatment outcomes: the improvement in mental health during the treatment episode. This is the difference between the GAF score at the start and end of a treatment episode, which we will henceforth refer to as *DIFGAF*, another key outcome variable in the empirical analysis.

4.2. Data cleaning and sample selection

The total data set contains roughly six million treatment episodes over the period 2008-2013. However, for the analyses the data were cleaned in several steps. Firstly, treatment episodes which included missing values and outliers were removed. Next, we used only the majority of treatment episodes that were labelled as a "regular" or "continued" treatment and removed other less common treatment episodes.[28] Treatment episodes with a very long treatment duration, i.e. over 8000 minutes, were also removed from the data as they are uncommon, maybe outliers, and may refer to very specialized treatments. As we want to study treatment duration responses of incumbent providers that were on the market before and after the shock, we constructed a balanced data set and selected only providers that were all years on the market.[29] The final sample consists of 357 B-providers with 3,893,294 treatment episodes and 740 NB-providers with 253,261 treatment episodes. Thus, B-providers account for 94% of all treatment episodes. The number of treatment episodes per B-provider is roughly 11,000 compared to 342 per NB provider (30 times difference). Because both differ in size, type of payment system and type of patients we analyze them separately. To obtain this final sample roughly 30% of the treatment episodes in the raw data were removed. An overview of the data cleaning steps is provided in Table 6 in Appendix A.

4.3. Descriptive statistics

The two panels in Figure 1 show how the number of treatment episodes for B and NB-providers developed between 2008 and 2013. Up to 2011 the number of treatment episodes do not change much for both provider types. Then, in 2012 we clearly observe the demand shock. The precise numbers are presented in Tables 7 and 8 of Appendix A for B-providers and NB-providers

[28] These treatment episodes were mainly short and less common such as one-time (urgent) consults, intercollegial consultations or second opinions, and acute admissions of patients for intensive treatment. We excluded these rare and short treatments to keep our sample as homogenous as possible. Moreover, for short treatments follow-up decisions by patients may play a role making it more difficult to relate responses to the supply side.

[29] We balanced on yearly not monthly level. It is therefore possible that a provider does not record treatment episodes in a particular month between 2008 and 2013.

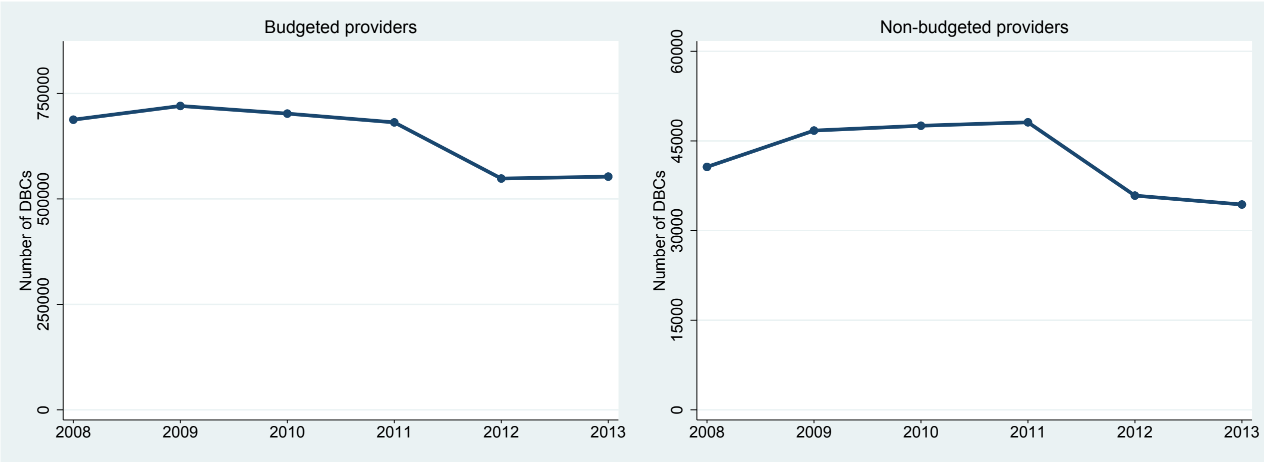


Figure 1: Evolution of the total number of treatment episodes per year

respectively. For B-providers, the number of treatment episodes decreases from 681,507 in 2011 to 548,372 in 2012. This drop is roughly 20%. The number of treatment episodes of NB-providers decreases from 48,117 in 2011 to 35,851 in 2012; a reduction of 22%. The lower number of treatment episodes of 2012 persists in 2013.

Tables 7 and 8 present also information for five (of the nineteen) largest diagnosis groups. We find mixed results. For some diagnoses, there is a drop in the number of treatment episodes in 2012 while for others there is no decrease or even a slight increase. Diagnosis groups "Adjustment" disorders and "V-codes" exhibit remarkable evolutions: the drop in 2012 is so large that there are almost no treatment episodes left in 2012.[30] This is caused by the removal of both diagnoses from the basic benefit package as we explained in Section 2. For prevalent diagnoses, such as "Mood", "Personality" and "Anxiety" disorders, we observe almost no drop in the number of treatment episodes in 2012. Anecdotical evidence suggests that recoding may have taken place: patients who previously have been diagnosed with a "Adjustment" disorder or "V-codes" were instead diagnosed differently as to be covered by insurance.[31]

Figure 2 shows how the distribution of mental health status of patients at the start of a treatment episode changed before and after the demand shock. Both distributions of B-providers are more skewed to the left than for NB-providers, which indicates that B-providers treat more

[30]The policy reforms were announcement well before 2012. This may explain that we observe anticipation effects in our data: the number of treatment episodes for "Adjustment" disorders already start dropping at the end of 2011.

[31]There is the possibility of recoding. Proving or thoroughly analyzing this potential provider response falls outside the scope of this paper. In Appendix D we show that recoding does not affect our estimation results for treatment duration.

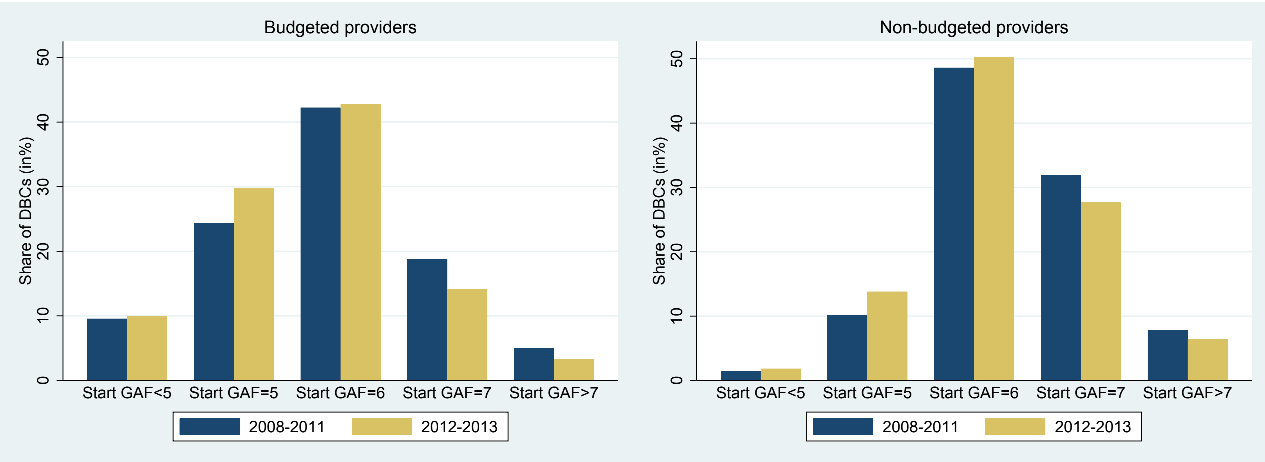


Figure 2: Relative change in case-mix

patients with a severe mental health condition. After the shock in 2012, there are relatively fewer patients with a "mild" mental health condition and relatively more patients with a more severe condition. This is the case for both B- and NB-providers. Apparently, more patients with a relatively mild mental health condition dropped out. Hence, the case-mix changed in 2012 (and 2013).

For both types of providers, average treatment duration per treatment episode increased in 2012. This increase is clearly visible in Figure 3. For B-providers, average treatment duration increased from 1,251 minutes in 2011 to 1,427 minutes in 2012. This increase is roughly 13%.

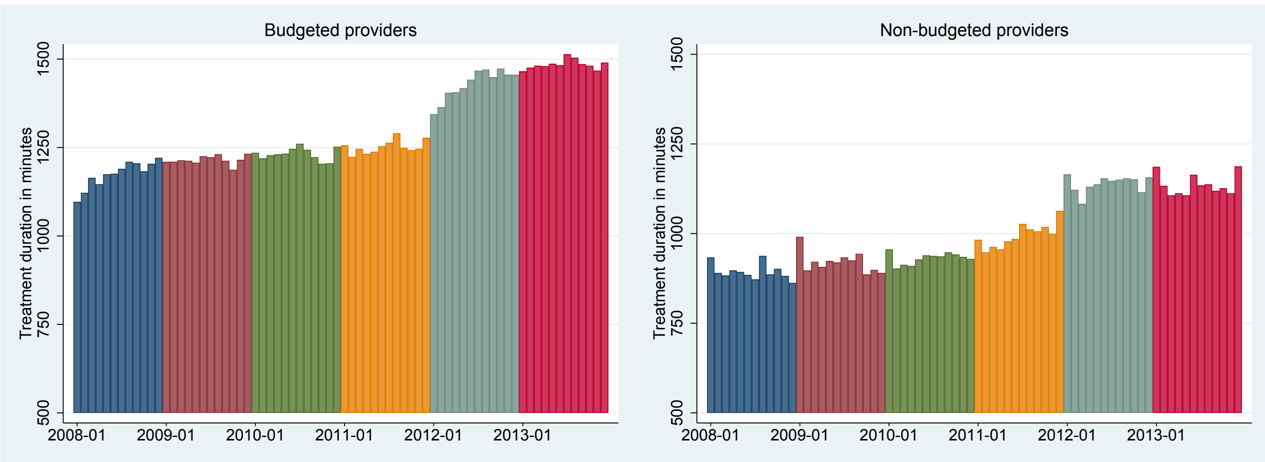


Figure 3: Average treatment duration per treatment episode per month (months are plotted on the horizontal axis with 2008-01 as January 2008.)

In 2013, average treatment duration for B-providers increased even further to 1,483 minutes. Average treatment duration of NB-providers increased from 995 minutes in 2011 to 1,138 minutes in 2012; an increase of 12%. In 2013, average treatment duration remained at 1,138 minutes. Tables 7 and 8 show that the increase in 2012 is also present for all start GAF categories. Note that the absolute increase in treatment duration is considerably larger for B-providers as their average treatment duration is about 30% higher as for NB-providers.

An increase in treatment duration in 2012 and 2013 may be the result of a change in case-mix. Tables 7 and 8 show that indeed persons with a more severe mental condition, i.e. a lower GAF-score, are treated longer. For example, patients with a start GAF<5 are treated on average about three times longer than patients with a start GAF>7. As there are relatively more severe patients after the demand shock (see Figure 2) and because these patients are treated longer on average, it follows that after the demand shock average treatment duration increases due to case-mix changes. Moreover, Tables 7 and 8 show that the average treatment duration for all start-GAF categories increases in 2012 and 2013 compared to 2008-2011.

Tables 7 and 8 also show that a treatment results on average in better patient outcomes. For B-providers average GAF scores at the beginning of a treatment are 6.30 (averaged over all years) and these scores improve on average after treatment (average *DIFGAFs* over all years is 0.24). For NB-providers average start GAF-score are 6.31 and average *DIFGAF* are 0.86. This shows that on average NB-providers have larger GAF improvements.[32] The total number of *DIFGAFs* produced in a year can be interpreted as an output measure. For both types of providers the total number of *DIFGAFs* declines substantially in 2012 and 2013 which is mainly related to the decline in the number of patients.

Finally, the distribution of treatment duration for B- and NB-providers differs greatly. Figure 4 shows that B-providers have relatively smooth treatment duration distributions. The mass of treatment durations is between 200 and 2000 minutes. In contrast, the distribution of NB-providers shows gaps before and bunches just after treatment duration thresholds (indicated in Figure 4 by the vertical lines).[33] Figure 4 shows that the distributions of both B- and NB-providers after the policy reforms are more skewed to the right, which reflects an increase in average treatment duration.

[32] For more information about GAF-scores of B and NB-providers, see (R. Zoutenbier, 2016).

[33] We refer to Douven et al. (2015, 2019) for an extensive exposition about the treatment responses of NB-providers around treatment duration thresholds.

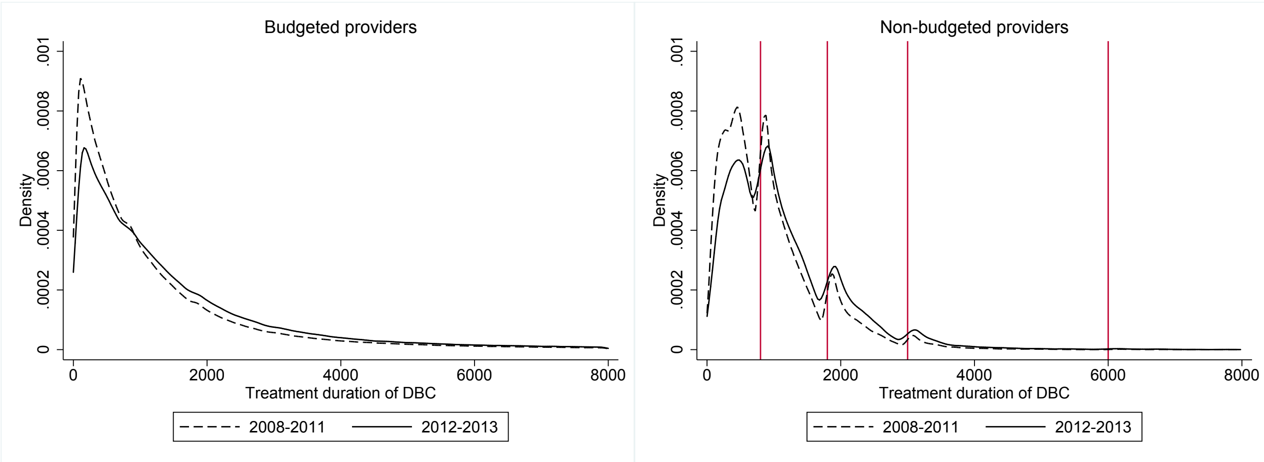


Figure 4: Distribution of treatment duration before and after the demand shock for B- and NB-providers

5. Estimation methodology

The goal of our empirical approach is to analyze how providers changed their treatment behavior in 2012 and 2013. We therefore measure the drop in the number of treatment episodes, the effect on treatment duration and the effect on patient outcomes (respectively, q, x_i and Q_i in the theoretical framework).

For our identification there are three key variables which form the basis of all our analyses. The first variable is a linear trend which describes the development of the market before the policy reform and demand shock, i.e. in the period from 2008 to 2011. If we extend this linear trend in our outcome variable to 2012 and 2013, we obtain a baseline or counterfactual trend, which represents how the market would have developed if there were no policy reforms or demand shocks. The variable is denoted by *baseline*.

The second variable measures the effect of the demand shock in 2012 on our outcome variable. It is given by a dummy variable for the year 2012. This variable, *response2012*, describes the short-term provider responses relative to the baseline.

The third variable, *response2013*, is a dummy for the year 2013 and measures provider responses in 2013 relative to the baseline.

Our empirical consists of two parts. First, we analyze the major developments in the mental health care sector at an aggregated provider level. Secondly, we go into more detail by studying the provider behavior at treatment episode level. Both analyses will be done for B-providers as well as NB-providers. We study four dependent variables. At the provider level the number of treatment episodes and total treatment duration and at the treatment episode level treatment duration and *DIFGAF*.

Provider level

For the analyses at the provider level we use a fixed effects panel model. Using this model we are able to describe the evolution of our outcome variables over the years and, specifically, estimate *baseline*, *response2012*, and *response2013*, as described above. We aggregate our data for each mental health care provider at a monthly level. The model is defined as follows:

$$\begin{aligned} Y_{jym} = c &+ \beta_1 baseline_y + \beta_2 response2012_y + \beta_3 response2013_y \\ &+ \delta_1 D_m + \delta_2 D_{2008,m} + \alpha_j + u_{jym} \end{aligned} \tag{10}$$

where the dependent variable Y_{jym} is the outcome variable for provider j in year $y = 2008, ..., 2013$ and month $m = 1, ..., 12$. At the provider level Y_{jym} is either the number of treatment episodes or the total production, i.e. total treatment duration (in minutes), per provider per month. The *baseline* trend is estimated by β_1 which describes the average yearly change in our outcome variable Y over the years 2008-2011. The *response2012* is estimated by β_2 and measures the deviation of the outcome variable from the baseline trend in 2012 as a result of the policy reforms and subsequent demand shock. Similarly, β_3 is the estimate for the *response2013* variable and measures the deviation of the outcome variable from the baseline trend for the year 2013. This provides insight in how providers accommodate to the policy reforms and demand shock in 2013.

We insert month dummies D_m to control for within year variation in the outcome variable. $D_{2008,m}$ are dummies for the first six months of 2008, which control for the fact that 2008 was the first year that mental health care was reformed from a public to a regulated competition system and providers were still adjusting to this new system.[34] The error term in the fixed effect panel model is composed of a time-invariant provider specific effect α_j and an idiosyncratic error term u_{jym}. The standard errors are clustered at the provider level.

Model (10) provides a description of the general developments in the sector, as we do not control for case-mix differences. We will control for those factors in the analyses on treatment episode level below.

Treatment episode level

We zoom in and study provider responses at the treatment episode level to see how the duration for an individual treatment episode has changed. We will estimate the effect on treatment

[34] We tested this empirically. The adjustment effects of the 2008 reforms disappeared after about six months, and adding more monthly dummies did not alter our results.

duration without and with controlling for changes in case-mix. We do this by including patient characteristics as well as attributes specific to the treatment episode. The model with controls is formulated as follows:

$$\begin{aligned} Y_{ijym} = c + \beta_1 baseline_y + \beta_2 response2012_y + \beta_3 response2013_y \\ + \delta_1 D_m + \delta_2 D_{2008,m} + \delta_3 D_j + \gamma X_{iym} + \epsilon_{iym} \end{aligned} \tag{11}$$

The dependent variable Y_{ijym} represents the treatment duration or the difference in our patient outcome measure between the start and end of a treatment episode i ($DIFGAF$). Treatment episode i is opened by provider j in month m of year y. In this way we analyze to what extent the treatment duration of a treatment episode on average changes over time and, specifically, after a demand shock. As in model (10) we have the same three variables of interest *baseline*, *response2012* and *response2013*. Also, dummies D_m and $D_{2008,m}$ are the same as in model (10).

A difference between models (10) and model (11) is the set of explanatory variables X which are specific to treatment episode iym and captures the case-mix of treatment episode i. X includes patient characteristics gender and age, as well as treatment episode characteristics such as the main diagnosis and sub-diagnosis, the GAF score at the start of the treatment episode, whether a patient stays over night at a mental health institution, and the type of a treatment episode.[35] As there are great differences between providers we also include a dummy for each provider, D_j. As treatment episodes within a provider are likely to be related, we cluster the standard errors on provider level.

To simplify the interpretation of our estimates, we scale the regression coefficients β_1, β_2 and β_3 relative to our baseline trend for both models. By using the predicted values we can construct a baseline trend as a counterfactual for 2012 and 2013 as there would be no policy reforms. The scaled regression coefficients can be interpreted as the percentage change in the outcome variable relative to this baseline trend.

Figure 5 provides a visual interpretation of an example of a positive β_1 (*baseline*) and a positive β_2 (*response2012*) and β_3 (*response2013*). The baseline trend is expanded throughout 2012-2013 as a counterfactual scenario in which we assume that the market would have developed similar to the pre-period without policy reforms. β_2 and β_3 capture the effect of the policy reform in 2012 and 2013, respectively, relative to the baseline.

[35]We include one dummy for gender, 100 age dummies (one for each age-year), over 100 dummies for each main and sub-diagnosis group, 10 dummies for each start-GAF category and 2 dummies whether it is a regular or continued treatment.

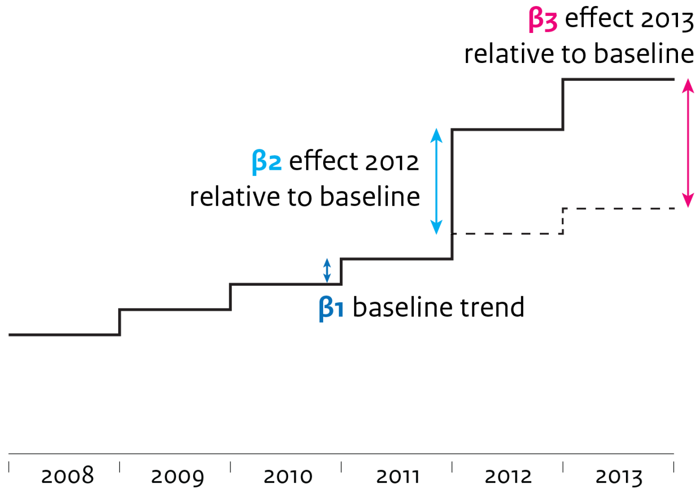


Figure 5: Illustration of key variables in empirical strategy

6. Results

In this section we present the estimation results of panel model (10) and treatment episode model (11). First, the results for B-providers are discussed, then the results for NB-providers. At the end of the section, in subsection 6.3, we link the results to the hypotheses from our theoretical framework. In this section we only present the interpretation of the estimation results (see

Table 2: Results for B-providers

	Provider level		*Treatment episode level*		*DIFGAF*
	#Treatment episodes	Total production	Average treatment duration no case-mix	Average treatment duration case-mix	case-mix
Baseline (β_1)	-0.9%	0.6%	1.7%***	1.2%***	0.01
Response 2012 (β_2)	-19.5%***	-9.6%***	12.2%***	7.8%***	-0.004
Response 2013 (β_3)	-18.7%***	-6.1%	14.8%***	8.6%***	-0.01
Number of observations	23,635	23,635	3,893,294	3,893,294	3,893,294

Notes: For the complete estimation results, see Appendix C. In this table the interpretation of the results are presented (see section 5. The levels of significance refer to the significance of the coefficients underlying these percentages. *, **, and *** indicate significance based on a two-sided test at the .10, .05, and .01 levels, respectively. No case-mix refers to the estimations without case-mix controls. The estimates in the *DIFGAF* column represent absolute changes and refers to the regression with case-mix controls included. In Appendix C we also show the results for *DIFGAF* without controls.

section 5), the full estimation outputs are reported in Appendix C.

6.1. B-providers

The first column in Table 2 presents the results of the panel model for B-providers. The baseline trend of the number of treatment episodes is fairly constant between 2008 and 2011, as β_1 is small and insignificant. This is in line with Figure 1. In 2012, the number of treatment episodes drops significantly by 19.5%. The number of treatment episodes remain low in 2013, and is 18.7% lower than the baseline.

The second column in Table 2 shows that total production, measured in total treatment minutes, drops by 9.6% in 2012 and in 2013 6.1% lower than baseline. The standard errors in Appendix C (Table 10, column *Total production*) are relatively large for β_2 in 2012 and, especially, β_3 in 2013, which indicates that the shock provoked large differences in responses across B-providers. The estimated percentage change in total production is in all years larger (or less negative) than the change in the number of treatment episodes, which indicates an increase in average treatment duration, especially in 2012 and 2013.

The third and fourth column show the results for average treatment duration, estimated for treatment episode level model (11). Without controlling for case-mix, average duration per treatment episode increases annually with 1.7%, i.e. the baseline trend. The policy reform and demand shock result in a significant increase in average treatment duration of 12.2% in 2012 and 14.8% in 2013 compared to baseline. Changes in case-mix are an important driver of the increase in average treatment duration. Controlling for case-mix changes lowers the change in average treatment duration after the demand shock substantially. However, apart from the changes in case-mix, treatment duration still increases significantly by 7.8% in 2012, and 8.6% in 2013, compared to baseline (see also figure 8 in the conclusion section). Case-mix differences do not strongly affect the baseline trend.

The last column of Table 2, show that our outcome measure $DIFGAF$ did not change before and after the demand shock. β_2 and β_3 are both negative and insignificant, indicating that longer treatment durations have on average not improved patient outcomes, as measured by GAF-scores.

The results above do not reveal the variation in effects for different diagnosis groups or start GAF levels. Therefore, we run 95 regressions (for every combination of 19 diagnosis groups and 5 groups of start GAF scores) of model (11). Figure 6 shows the relationship between the change in treatment duration and outcome improvements for these 95 regressions for 2012 and 2013. Treatment duration (on the horizontal axis) increases for the vast majority of the treatment episodes in Figure 6 while the changes in GAF-scores (on the vertical axis) all lie around zero.

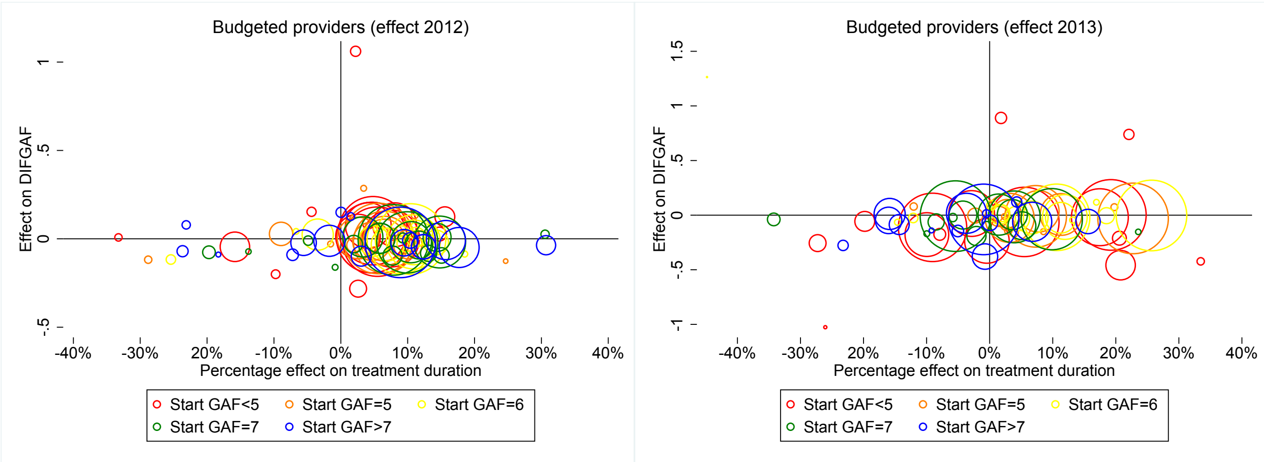


Figure 6: Treatment duration (x-axis) and *DIFGAF* (y-axis) responses for 2012 and 2013 per diagnosis group and start GAF level. Each diagnosis and start GAF combination is represented by the center in a circle, and the size of the circle represents the relative size of the number of treatment episodes in this GAF-diagnosis group.

Furthermore, there is much more heterogeneity in treatment duration responses in 2013 than in 2012, compared to baseline, suggesting that the degree of accomodation to the shock over time differs across GAF-diagnose groups.

6.2. NB-providers

Table 3 shows the results for NB-providers. Compared to B-providers, we see a significant and positive baseline trend between 2008 and 2011: each year the number of treatment episodes increased by 4.4% and total annual treatment duration by 6.6%. This means that production of NB-providers has increased over the years both at the extensive as intensive margin. NB-providers faced a significant decline in the number of treatment episodes after the shock of 28.9% in 2012 and 32.4% in 2013 compared to (the increasing) baseline. Moreover, total production decreased by 19.5% in 2012 and 25.3% in 2013. These results indicate that NB-providers adjusted their total production more than B-providers in response to the demand shock. Average treatment duration increased by 9.9% in 2012 and 6.7% in 2013 without controlling for case-mix. However, after controlling for case-mix, we find that treatment duration in 2013 returns almost to the level of before the shock in 2011 (see also figure 8 in the conclusion section). We find a relatively small increase, compared to the large negative shock, in treatment duration of 3.5% in 2012 but in 2013 we observe a decline of 4.3% relative to baseline. Thus, compared to B-providers, we find for NB-providers relatively small increases in treatment durations in 2012 and even a decline in 2013.[36]

Similar to B-providers, we find no significant effect on treatment outcomes in 2012 and 2013 for NB-providers. Figure 7 shows again the relationship in 2012 and 2013 between treatment duration responses and *DIFGAF* score for all 95 combinations of diagnoses and start GAF scores. We see that most circles are centered around the origin, both in terms of change in treatment duration and in terms of *DIFGAF* outcomes.

Table 3: Results for NB-providers

	Provider level		*Treatment episode level*		*DIFGAF*
	#Treatment episodes	Total production	Average treatment duration		
			no case-mix	case-mix	case-mix
Baseline (β_1)	4.4%***	6.6%***	2.7%***	2.9%***	-0.006
Response 2012 (β_2)	-28.9%***	-19.5%***	9.9%***	3.5%***	.000
Response 2013 (β_3)	-32.4%***	-25.3%***	6.7%***	-4.3%***	.029
Observations	44,908	44,908	252,776	252,776	252,776

Notes: For the complete estimation results, see Appendix C. In this table the interpretation of the results are presented (see section 5. The levels of significance refer to the significance of the coefficients underlying these percentages. *, **, and *** indicate significance based on a two-sided test at the .10, .05, and .01 levels, respectively. No case-mix refers to the estimations without case-mix controls. The estimates in the *DIFGAF* column represent absolute changes and refers to the regression with case-mix controls included. In Appendix C we also show the results for *DIFGAF* without controls.

6.3. Further explanation of the results

In this subsection we link our empirical results to the theoretical framework of Section 3.

NB-providers

Douven et al. (2015, 2019) and the right panel in Figure 4 show that NB-providers already responded strongly to the discontinuities in the payment system before the demand shock. NB-providers are on average sensitive to financial incentives (i.e. $\alpha_j^{\mathrm{NB}} > 0$). After the demand shock however, we observe after controlling for casemix hardly any change in average treatment duration for NB-providers. The theoretical model and Figure 4 provide a possible explanation for these findings. Marginal changes in treatment duration yield almost no financial gain for NB-providers, because there are not many treatments with a treatment duration distribution just

[36] Note that the relatively large baseline trend (β_1=2.9%) for NB-providers, after controlling for case mix, makes the results more difficult to interpret as these results may signal "overtreatment" in the years before the shock.

Figure 7: Treatment duration (x-axis) and *DIFGAF* (y-axis) responses for 2012 and 2013 per main diagnosis and start GAF level.

before a treatment duration threshold. Providers can only increase their income by prolonging treatment duration to a next treatment duration threshold. In our model this is only possible if γ_j is large enough.

To study the importance of income effects for NB-providers, we compare the results of NB-providers for different levels of altruism, i.e. different levels of α_j^{NB}. Douven et al. (2019) estimate this degree of "altruism" or "professionalism" for each NB-provider by measuring how often a NB-provide ends a treatment episode just after reaching a treatment duration threshold. Altruistic providers (who do not respond to thresholds) have a small α_j^{NB} and more financially motivated providers (who do respond to thresholds) have a large α_j^{NB}. Note that if $\gamma_j > 1$ then $\overline{\alpha_j}^{\mathrm{NB}}$ is increasing in α_j^{NB}, which implies that more financially motivated providers will be more sensitive to the income effect than altruistically motivated providers.[37]

In line with Douven et al. (2019), we grouped NB-providers into four quartiles using the average reimbursement per hour for the years 2008-2011 as a proxy for whether a provider is altruistically or financially motivated. The first quartile contains altruistic providers who did not respond to the treatment duration thresholds between 2008 and 2011 and who earned on average 111 euros per hour or less. The fourth quartile contains the most financially motivated providers, who strongly responded to the treatment duration thresholds and earned on average 122 euros per hour or more. Figure 10 in Appendix B shows for each quartile how the distribution of treatment duration changed after 2012. Average reimbursement per hour remained about the

[37]Moreover, γ_j and α_j^{NB} are likely to be positively correlated which further enlarges the income effect between the most and least altruistically motivated providers.

same for altruistic providers in quartile 1 after 2012, but declined for more financially motivated providers in quartile 2-4 (see Table 9 in Appendix B). This suggests that it became more difficult for these providers to end a treatment just after a threshold in 2012 and 2013. The increasing severity of patients after the shock in 2012 may have made longer treatments necessary making it more difficult to end a treatment just after a treatment duration threshold.

We estimate the treatment episode model (11) for NB-providers in each quartile where we control for case-mix. The results are presented in Table 4.[38]

Table 4: Treatment duration responses of altruistic and financially motivated NB-providers

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Baseline (β_1)	2.4%***	4.2%***	3.1%***	2.1%**
Response 2012 (β_2)	1.2%	1.9%	4.8%***	8.2%***
Response 2013 (β_3)	-3.3%	-8.3%***	-5.9%**	0.9%
Baseline fixed on the level of treatment duration in 2011				
Response 2012 (β_1 + β_2)	3.6%	6.1%	7.9%	10.3%
Response 2013 ($2\beta_1$ +β_3)	1.5%	0.1%	0.4%	5.1%
Observations	63,276	63,485	62,920	63,095

Notes: Quartile 1 corresponds to altruistic NB-providers and quartile 4 to the most financially motivated NB-providers. The results were estimated with regression equation (11) with average treatment duration as the dependent variable. The levels of significance refer to the significance of the coefficients underlying these percentages. *, **, and *** indicate significance based on a two-sided test at the .10, .05, and .01 levels, respectively. The estimations are with case-mix controls. The full regression output can be retrieved by the authors upon request.

Comparing the results among the quartiles is difficult, because each quartile has a different baseline trend β_1. Therefore, we also computed the responses in 2012 and 2013 compared to a flat baseline equal to the level of treatment duration in year 2011. The results indicate that the effects in 2012 increase with the quartiles, which reflects that the size of the response is positively related to α_j^{NB}, which is in line with the income effect hypothesis. Moreover, for 2013 we also find the biggest effect for financially motivated providers in quartile 4.

To conclude, we cannot reject the income effect hypothesis, i.e. $\gamma_j>1$, at least not for the most financially motivated NB-providers. The discontinuities in the stepwise fee-for-service function and their behavior before the demand shock where they already exploited the discontinuities, may have prevented financially motivated NB-providers with a $\gamma_j>1$ to prolong

[38]We run the four regressions also for *DIFGAF* but this did not yield significant changes in GAF-scores before and after the shock. The results from these regression can be obtained by the authors upon request.

treatment duration further after the demand shock.

B-providers

B-providers respond to the demand shock by increasing their average treatment duration. Moreover, we find that on average GAF-scores did not improve which we interpret as suggestive evidence for over-treatment. Our theoretical model yields two possible explanations for these findings: professional uncertainty and income effects. We will not be able to discriminate between the two hypotheses in a similar, clean way like for NB-providers because we lack exogenous variation between providers, such as the degree of altruism. However, we can follow the strategy of Yip (1998), who estimated provider responses after a large reduction in Medicare prices after a reform. She found evidence of the income effect hypothesis by showing that providers with the largest losses in their income had also the largest volume responses to recoup this income.

We group B-providers in four quartiles, where the first quartile are providers that faced the largest reduction (29% or more) in treatment episodes in 2012 and the fourth quartile are providers with the smallest reduction (18% or less). If only the income effect would play a role then we would expect a positive correlation between prolonging treatment duration and the magnitude of the reduction in the number of patients after the policy reform. More random responses would suggest that professional uncertainty plays a role as providers beliefs and prior rationing may differ across providers and therefore may be more unrelated to the size of the shock.

Table 5: Treatment duration responses of B-providers ranked according to the size of the shock

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
Baseline (β_1)	2.3%***	0.3%	1.2%*	0.8%
Response 2012 (β_2)	6.7%***	7.2%***	11.5%***	6.5%***
Response 2013 (β_3)	3.8%**	6.0%*	16.2%***	9.3%***
Baseline fixed on treatment duration in 2011				
Response 2012 ($\beta_1 + \beta_2$)	9.0%	7.5%	12.7%	7.3%
Response 2013 ($2\beta_1 + \beta_3$)	8.4%	6.6%	18.6%	10.9%
Observations	939,064	850,081	892,243	844,494

Notes: Quartile 1 corresponds to B-providers that have suffered the largest demand shock and quartile 4 to the smallest. The results were estimated with equation (11) with average treatment duration as the dependent variable. In this table the interpretation of the regression results are presented (see section 5. The levels of significance refer to the significance of the coefficients underlying these percentages. *, **, and *** indicate significance based on a two-sided test at the .10, .05, and .01 levels, respectively. All regression included case-mix controls. The full regression output can be retrieved by the authors upon request.

Table 5 shows that responses in 2012 and 2013 are not positively correlated with the size of the demand shock.[39] Thus, these results do not support the income effect hypothesis. Hence, professional uncertainty seems to be important as well. Note that we cannot interpret these estimates as causal. First, as we already showed for NB-providers, also B-providers may differ in their degree of altruism. For example, if altruistic providers received larger shocks than financially motivated providers, then we also may not find a positive correlation. Also, unobserved heterogeneity may be large since we do not observe other possible activities of providers. After the shock B-providers might differ in the way they have layed off personnel or employed other activities that we do not observe. We conclude that both, professional uncertainty and income effects, have played a role but we cannot disentangle the impact of the two mechanisms.

6.4. Robustness analyses

To test if our results are not driven by other mechanisms, we have performed several additional analyses. The results are listed in Appendix D. Here we repeat the main findings.

First, we show that our results are not driven by indirect treatment duration, i.e. providers have mainly increased their administrative tasks after the demand shock instead of increasing their face-to-face contacts with patients. Our regression results do not change when we use direct treatment duration, i.e. face-to-face contacts with patients, instead of total treatment duration.

Second, we test whether announcement effects play role. The policy reform was legally announced by the government in June 2011. Providers may have anticipated to this announcement. We test this leaving out all observations from July 2011 until December 2011 in one regression and leaving out all observations from July 2011 until June 2012 in another regression. The results are again comparable with the main results in our paper, albeit the provider responses are somewhat stronger for B-providers in the second regression.

Third, we test whether our estimates may be biased because of recoding in 2012 and 2013. Some providers may have recoded patients with "Adjustment" and "V-codes" to other diagnoses, such as "Mood", in order to prevent that these patients need to pay the full fee for the treatment. To test this possibility we ran a new regression were we coded all diagnoses "Adjustment" and "V-codes" for all years in our dataset as "Mood" (and we left out all subdiagnoses for "Mood" in our controls). We found almost similar results as in our main regression which indicates that a possible bias in our results due to recoding is neglegible.

Fourth, we test whether providers treated patient longer when measured in the number of days, instead of measured in treatment minutes. We measure the number of days as the

[39] We also performed the same regression without controls with similar results.

difference between the first day and the last day that a patient visits a provider (as recorded in a DBC). When providers receive fewer patients than the remaining patients can, theoretically, be treated in fewer days. However, since we find longer treatment durations in minutes after the shock, we might also find an increase in the number of days after the shock. The results are shown in the column "Number of days". The results indicate that, compared to baseline, we find an increase in the number of days after the shock but the percentages are smaller than compared to the percentages of the main results in the column "All".

7. Conclusion

We study if and how Dutch mental health care providers respond to a sudden drop in the number of patients of 20% and compare providers who receive a budget (B-providers) with self-employed providers that are paid according to a discontinuous fee-for-service scheme (NB-providers). Our main results are summarized visually in Figure 8 below.

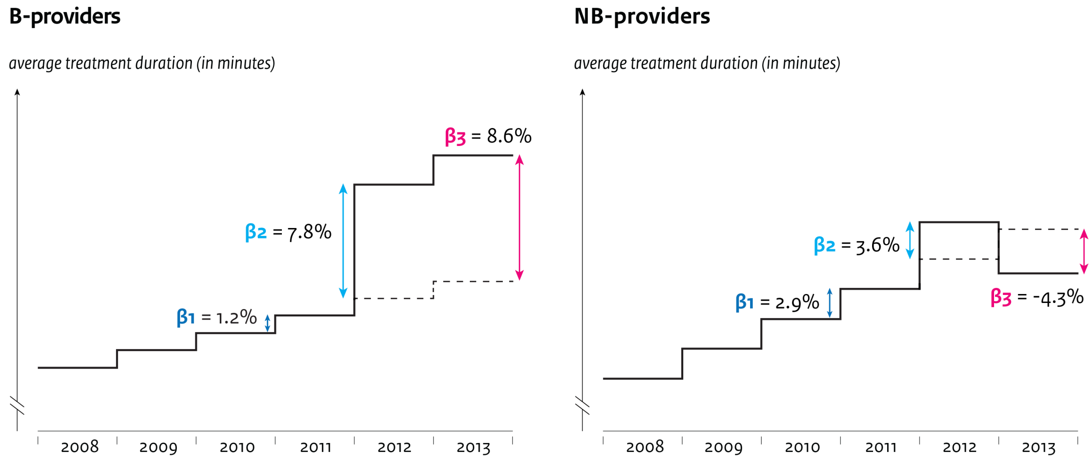


Figure 8: Treatment duration responses of B and NB-providers

The left panel in Figure 8 shows that B-providers respond more strongly to the demand shock than NB-providers. They increase treatment duration on average by approximately 8% for both years after the demand shock (β_2 and β_3). This increase in treatment duration does however not result in better treatment outcomes.

We distinguish two explanations why B-providers prolong patients' treatments: professional uncertainty and income effects. The professional uncertainty explanation means that B-providers have experienced some form of implicit or explicit rationing before the shock. After the shock, when more capacity became available, they treated patient longer because they believed it would benefit patients. Professional uncertainty plays an important role in the case of supply sensitive treatments, such as treatments in mental health care.

Another explanation is income effects. A large reduction in the number of patients may affect provider income. At the management level, managers may have stimulated their employees to treat patients longer as to recoup some their income loss. However, it could also be that at employee level, most of them having a salary and fixed working hours, employees acted indepently from their manager. When more time became available, treating patients longer could provide a signal to their manager that they are productive. Thus, shirking may have played a role. It is clear that the professional uncertainty and income effect mechanisms are extremely difficult to disentangle in practice. This remains a challenge for future research.

The right panel in Figure 8, shows the result for NB-providers who are reimbursed by a stepwise fee-for-service payment scheme. NB-providers face a larger demand shock but the changes in treatment duration are much smaller than for B-providers. We find some evidence that only the most financially motivated NB-providers increased treatment duration in both years after the shock. Previous research showed that financially motivated NB-providers react strongly to the treatment duration threshold in their payment scheme (Douven et al., 2015, 2019). The theoretical exposition in our paper suggests that the treatment duration thresholds may have prevented a further increase in treatment duration after the shock. Note also the difference between B- and NB-providers. The treatment duration responses of B-providers may be larger because of prior rationing which does not play a role for NB-providers. This allows us to attribute all treatment duration effects for NB-providers to an income effect.

A limitation of our study is that our conclusions are based on only one outcome indicator (GAF change) that we use 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 Vatnaland et al. (2007), but there are also a number of studies that show the GAF is a reliable indicator Hilsenroth et al. (2000), especially when comparing groups of patients (Jones et al., 1995, Soderberg et al., 2005). Unfortunately, alternative outcome measures, such as Routine Outcome Monitoring (ROM), were not available for the years that we investigated. Another limitation is that we have very limited information about provider characteristics, such as information about their location, the number of employees, annual financial reports, et cetera.

Our study shows that differences in provider and payment scheme characteristics can have a big impact on treatment behavior by providers and patient outcomes. A stepwise fee-for-service scheme induces self-employed providers to prolong treatments, but may also prevents the prolonging of treatments when faced with a large sudden drop in the number of patients. A budget payment scheme does not cause bunching of treatments just after the treatment duration threshold. However, it is difficult to determine the "right" budget. In our case, looser budgets (or more capacity available for individual patients) are likely to have resulted in over-treatment. These findings show that designing an optimal payment system is complicated. The ideal payment system should not only take into account that individual providers differ in a variety of ways, such as their differences in medical decision-making and their preferences for monetary versus other rewards, but also be robust to external shocks, such as policy reforms.

One way forward seems to be designing payment systems that reward and monitor quality and efficient provision of care, such as population-based payment mechanisms with suitable quality indicators. However, not all quality measures can be used to reward providers. For example, GAF-scores are not suitable because they are scored by providers themselves, an thus prone to manipulation.

References

Abaluck, J., Agha, L., Kabrhel, C., Raja, A. and Venkatesh, A. (2016), 'The determinants of productivity in medical testing: intensity and allocation of care', *American Economic Review* **106**(12), 3730–3764.

American Psychiatric Association (2000), *Diagnostic and statistical manual of mental disorders*, Vol. 4, American Psychiatric Association, Washington D.C.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N. and Wiesen, D. (2018), 'The effects of introducing mixed payment systems for physicians: Experimental evidence', *Health Economics* **26**(2), 243–262.

Chandra, A., Cutler, D. and Song, Z. (2012), *Handbook of Health Economics*, Vol. 2, Elsevier, Amsterdam, chapter Who ordered that? The economics of treatment choices in medical care, pp. 397–432.

Chandra, A. and Skinner, J. (2012), 'Technology growth and expenditure growth in health care', *Journal of Economic Literature* **50**(3), 645–680.

Christianson, J. and Conrad, D. (2012), *The Oxford Handbook of Health Economics*, Oxford University Press, chapter Provider Payment and Incentives.

Clemens, J. and Gottlieb, J. (2014), 'Do physicians' financial incentives affect medical treatment and patient health?', *American Economic Review* **104**(4), 1320–1349.

Currie, J. and McLeod, W. (2017), 'Diagnosing expertise: Human capital, decision making and performance among physicians', *Journal of Labor Economics* **35**(1), 1–34.

Currie, J. and McLeod, W. (2018), Understanding physician decision making: The case of depression, Technical report, NBER Working Paper 24955.

Currie, J., McLeod, W. and Van Parys, J. (2016), 'Physicians practice style and patient health outcomes', *Journal of Health Economics* **47**, 64–80.

Cutler, D., Skinner, J., Stern, A. and Wennberg, D. (2019), 'Physician beliefs and patient preferences: A new look at regional variation in health care spending', *American Economic Journal: Economic Policy* .

Douven, R., Remmerswaal, M. and Mosca, I. (2015), 'Unintended effects of reimbursement schedules in mental health care', *Journal of Health Economics* **42**, 139–150.

Douven, R., Remmerswaal, M. and Zoutenbier, R. (2019), 'Do altruistic providers have better treatment outcomes?', *Journal of Human Resources* **54**(2), 310–341.

Doyle, J., Graves, J. and Gruber, J. (2017), 'Uncovering waste in us healthcare', *Journal of Health Economics* **54**, 25–39.

Doyle, J., Graves, J., Gruber, J. and Kleiner, S. (2015), 'Measuring returns to hospital care: Evidence from ambulance referral patterns', *Journal of Political Economy* **123**(1), 170–214.

Dutch Healthcare Authority (2013), Marktscan en beleidsbrief geestelijke gezondheidszorg: weergave van de markt 2008-2012, Technical report, Dutch Healthcare Authority.

Dutch Healthcare Authority (2014), Marktscan en beleidsbrief geestelijke gezondheidszorg: Betaalbaarheid en wachttijden 2009-2014, Technical report, Dutch Healthcare Authority.

Dutch Healthcare Authority (2016), Marktscan ggz 2016, Technical report, Dutch Healthcare Authority.

Ellis, R. and McGuire, T. (1986), 'Provider behavior under prospective reimbursement: cost-sharing and supply', *Journal of Health Economics* **5**(2), 129–151.

Frank, R. and McGuire, T. (2000), *Handbook of Health Economics*, Vol. 1B, Elsevier, Amsterdam, chapter Economics and mental health, pp. 893–954.

Fuchs, V. (1986), 'More variation in use of care, more flat-of-the-curve medicine', *Health Affairs* **23**(6), 104–107.

Gruber, J. and Owings, M. (1996), 'Physician financial incentives and cesarean delivery', *RAND Journal of Economics* **27**(1), 99–123.

Hilsenroth, M. J., Ackerman, S., Blagys, M. D., Baumann, B. D., Matthew, R., Smith, S., Price, J., Smith, C., Heindselman, T., Mount, M. and Holdwick, D. (2000), 'Reliability and validity of dsm-ivaxis v', *American Journal of Psychiatry* **157**(11), 1858–1863.

Jacobson, M., Chang, T., Earle, C. and Newhouse, J. (2017), 'Physician agency and patient survival', *Journal of Economic Behavior and Organization* **134**, 27–47.

Johnson, E. (2014), *Encyclopedia of Health Economics*, Vol. 3, Elsevier, Amsterdam, chapter Physician-Induced Demand, pp. 77–82.

Jones, S., Thornicroft, G., Coffey, N. and Dunn, G. (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* **166**, 654–659.

Lambregts, T. and van Vliet, R. (2018), 'The impact of copayments on mental healthcare utilization: a natural experiment', *The European Journal of Health Economics* **19**(6), 775–784.

McClellan, M. (2011), 'Reforming payments to healthcare providers: the key to slowing healthcare cost growth while improving quality?', *Journal of Economic Perspectives* **25**(2), 69–92.

McGuire, T. (2000), *Handbook of Health Economics*, Vol. 1B, Elsevier, Amsterdam, chapter Physician Agency, pp. 461–536.

Owen, J., Adelson, J., Budge, S., Kopta, S. and Reese, R. (2016), 'Good-enough level and dose-effect models: Variation among outcomes and therapists', *Psychotherapy Research* **26**(1), 22–30.

R. Zoutenbier (2016), Relatie behandelduur en gaf-scores in de ggz, Technical report, CPB Netherlands Bureau for Economic Policy Analysis.

Ravesteijn, B., Schachar, E., Beekman, A., Janssen, R. and Jeurissen, P. (2017), 'Association of cost sharing with mental health care use, involuntary commitment, and acute care', *JAMA Psychiatry* **74**(9), 932–939.

Skinner, J. (2012), *Handbook of Health Economics*, Vol. 2, Elsevier, Amsterdam, chapter Causes and consequences of Regional Variations in Health Care, pp. 45–93.

Soderberg, P., Tungstrom, S. and Armelium, B. A. (2005), 'Special section on the gaf: Reliability of global assessment of functioning ratings made by clinical psychiatric staff', *Psychiatric Services* **56**(4), 434–438.

van Dijk, C., van den Berg, B., Verheij, R. A., Spreeuwenberg, P., Groenewegen, P. and de Bakker, D. (2013), 'Moral hazard and supplier-induced demand: Empirical evidence in general practice', *Health Economics* **22**(3), 340–352.

Vatnaland, T., Vatnaland, J., Friis, S. and Opjordsmoe, S. (2007), 'Are gaf scores reliable in routine clinical use?', *Acta Psychiatrica Scandinavica* **115**(4), 326–330.

Westerdijk, M., Zuurbier, J., Ludwig, M. and Prins, S. (2012), 'Defining care products to finance health care in the netherlands', *The European Journal of Health Economics* **13**(2), 203–221.

Yip, W. (1998), 'Physician responses to medical fee reductions: Changes in the volume and intensity of supply of coronary artery bypass graft (cabg) surgeries in the medicare and private sectors', *Journal of Health Economics* **17**(6), 675–700.

Appendix A: Descriptives

Table 6: Overview of data cleaning and sample selections

	Raw sample	Selection 1 (cleaning)	Selection 2 (selecting type of treatment episode)	Selection 3 (selecting durations)	Final sample (balancing)
B-providers					
- number of providers	902	877	876	876	357
- number of treatment episodes	5,515,997	4,905,074	4,403,711	4,290,083	3,893,294
NB-providers					
- number of providers	1,738	1,716	1,716	1,716	740
- number of treatment episodes	432,860	392,275	383,910	383,348	253,261
Total					
- number of providers	2,640	2,593	2,592	2,592	1,097
- number of treatment episodes	5,948,857	5,297,349	4,787,621	4,673,431	4,146,555

Notes: The table shows the various steps in our data cleaning and sample selection process. The first column refers to all observations in our raw sample for the years 2008-2013. Column Selection 1 refers to removing providers and treatment episodes with missing values and outliers. Column Selection 2 refers to the selection of all "regular" and "continued" treatments. Column Selection 3 refers to the selection of all treatments with a duration less than 8000 minutes. Column Selection 4 refers to the selection of all incumbent providers that were on the market during all years 2008-2013.

Table 7: Descriptive statistics B-providers

	2008	2009	2010	2011	2012	2013
Number of budgeted providers	357	357	357	357	357	357
Number of treatment episodes	687,854	720,420	702,249	681,507	548,372	552,892
- Mood	136,533	134,075	129,286	137,379	125,265	132,486
- Anxiety	7,4846	78,578	76,967	82,299	74,288	80,983
- Personality	49,541	51,205	51,578	55,349	51,577	58,956
- Adjustment	67,621	73,024	69,291	44,755	2,558	832
- V-codes	71,779	72,621	66,555	58,767	26,773	1,986
- Start GAF < 5	72,748	67,769	65,961	60,850	52,901	56,948
- Start GAF $= 5$	155,817	174,607	173,890	176,044	157,972	170,640
- Start GAF $= 6$	279,689	304,823	300,321	294,811	236,632	234,802
- Start GAF $= 7$	137,691	136,788	128,814	120,031	81,730	73,861
- Start GAF > 7	41,909	36,433	33,263	29,771	19,137	16,641
Number of treatment episodes per patient	1.08	1.06	1.07	1.08	1.06	1.07
Average treatment duration	1,169	1,215	1,232	1,251	1,427	1,483
- Start GAF < 5	1,524	1,704	1,744	1,726	1,868	1,919
- Start GAF $= 5$	1,433	1,481	1,514	1,536	1,675	1,742
- Start GAF $= 6$	1,173	1,183	1,181	1,202	1,367	1,399
- Start GAF $= 7$	850	865	863	869	1,007	997
- Start GAF > 7	587	617	633	629	713	694
Average direct time	660	664	671	693	785	796
Average indirect time	471	512	521	511	560	569
Average Start GAF	5.77	5.79	5.77	5.76	5.67	5.62
Average DIFGAF	0.2330	0.2157	0.2164	0.2709	0.2639	0.2613
Total DIFGAF	160,288	155,419	151,970	184,634	144,713	144,475
Average age patient	38.98	38.51	38.31	38.16	37.10	37.96
Share females	51.86%	51.41%	51.24%	51.55%	50.11%	50.53%
Share of overnight stays	6.94%	6.83%	6.77%	6.38%	6.96%	7.01%
Type of care						
- Share regular treatment	53.60%	54.67%	54.03%	53.70%	51.85%	54.63%
- Share continued treatment	46.40%	45.33%	45.97%	46.30%	48.15%	45.37%

Notes: Treatment duration, average direct and indirect time are specified in minutes. Total DIFGAF represents the sum of all End GAF-Start GAF scores for all treatment episodes. Average DIFGAF is Total DIFGAF divided by the total number of treatment episodes in a given year.

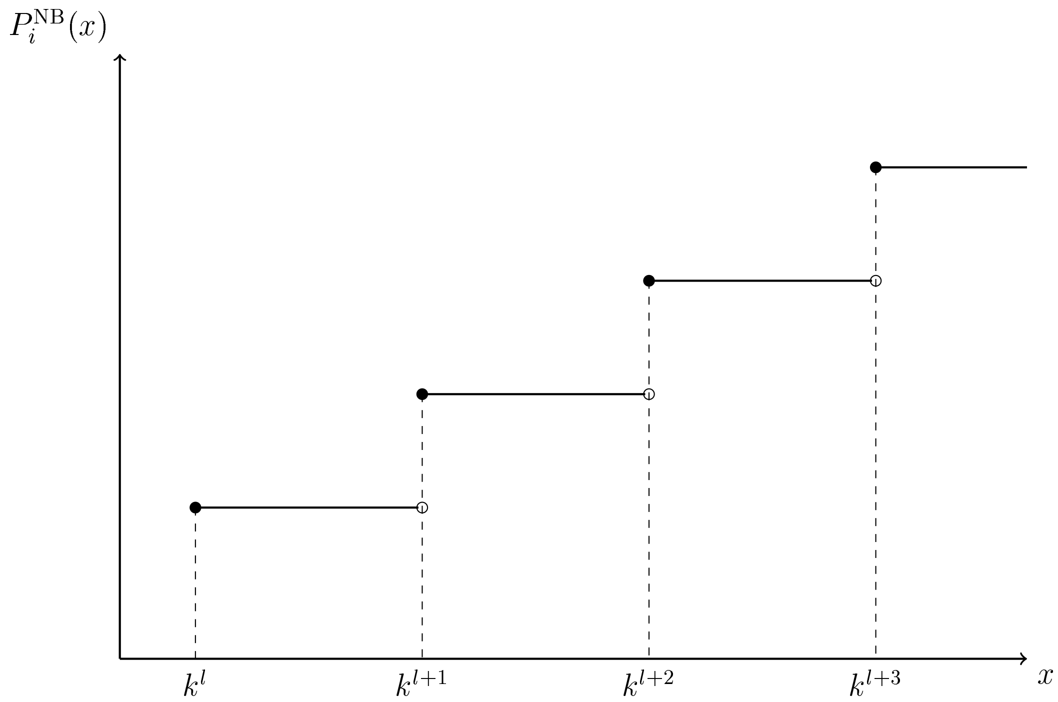
Table 8: Descriptive statistics non-budgeted providers

	2008	2009	2010	2011	2012	2013
Number of non-budgeted providers	740	740	740	740	740	740
Number of treatment episodes	40,655	46,733	47,545	48,117	35,851	34,360
- Mood	8,565	9,645	9,504	10,659	10,159	12,254
- Anxiety	6,455	7,465	7,772	9,273	8,682	9,962
- Personality	5,086	5,651	5,607	5,921	5,870	6,140
- Adjustment	6,614	7,596	7,657	4,529	199	20
- V-codes	10,037	11,547	11,736	11,468	5,145	116
- Start GAF $<$ 5	611	568	698	811	701	596
- Start GAF = 5	4,302	4,768	4,418	5,036	4,529	5,174
- Start GAF = 6	20,550	22,701	22,560	23,179	17,788	17,457
- Start GAF = 7	12,617	15,125	15,723	15,016	10,413	9,073
- Start GAF $>$ 7	2,575	3,571	4,146	4,075	2,420	2,060
Number of treatment episodes per patient	1.06	1.08	1.12	1.14	1.10	1.09
Average treatment duration	902	927	933	995	1,138	1,138
- Start GAF $<$ 5	1,125	1,100	946	1,062	1,158	1,227
- Start GAF = 5	1,092	1,090	1,169	1,264	1,408	1,408
- Start GAF = 6	958	999	1,011	1,067	1,199	1,174
- Start GAF = 7	798	839	837	881	1,002	986
- Start GAF $>$ 7	585	597	619	657	765	790
Average direct time	632	642	645	691	793	791
Average indirect time	269	283	286	300	340	335
Average Start GAF	6.29	6.35	6.38	6.34	6.25	6.19
Average DIFGAF	0.887	0.856	0.818	0.878	0.846	0.891
Total DIFGAF	36,067	40,002	38,873	42,227	30,318	30,625
Average age patient	38.78	38.72	39.07	39.05	38.25	38.76
Share females	64.30%	64.35%	64.59%	64.38%	63.09%	64.72%
Share of overnight stays	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Type of care						
- Share regular treatment	79.22%	71.93 %	70.23%	70.35%	69.37%	70.08%
- Share continued treatment	20.78%	28.07%	29.77%	29.65%	30.63%	29.92%

Notes: Treatment duration, average direct and indirect time are specified in minutes. Total DIFGAF represents the sum of all End GAF-Start GAF scores for all treatment episodes. Average DIFGAF is Total DIFGAF divided by the total number of treatment episodes in a given year.

Appendix B: Figures

Figure 9: Step-wise increasing payment scheme for NB-providers



Note: The treatment duration thresholds k are the same for all diagnoses, with $k^1 = 250$, $k^2 = 800$, $k^3 = 1800$, $k^4 = 3000$ and $k^5 = 6000$ minutes. For example, the tariff in 2011 for the diagnosis schizophrenia with treatment duration x is $P_i^{\mathrm{NB}}(x) = 1,070$ euros for $200 < x < 800$ minutes, $P_i^{\mathrm{NB}}(x) = 2,020$ euros for $800 \leq x < 1800$ minutes, $P_i^{\mathrm{NB}}(x) = 3,700$ euros for $1800 \leq x < 3000$ minutes, $P_i^{\mathrm{NB}}(x) = 6,100$ euros for $3000 \leq x < 6000$ minutes. $P_i^{\mathrm{NB}}(x) = 11,300$ euros for $6000 \leq x < 12000$ minutes. The tariffs are rounded off and may differ slightly across diagnoses.

Figure 10: Distributions of treatment duration by quartiles of NB-providers' relative altruism

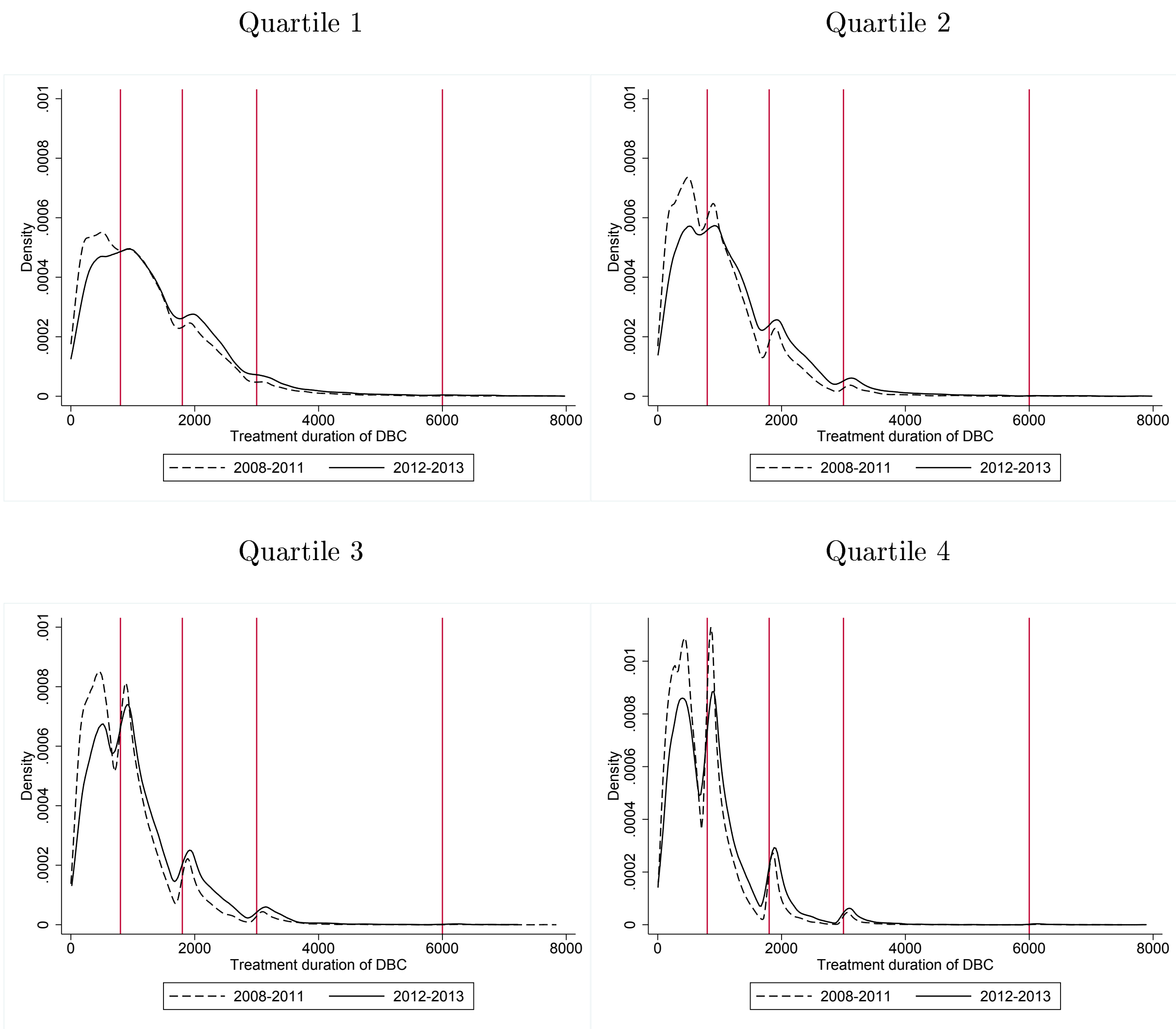


Table 9: Average reimbursement in euros per hour

	Quartile 1	Quartile 2	Quartile 3	Quartile 4
2008-2011	108,19	113,58	118,79	128,21
2012	108,43	111,92	115,44	124,16
2013	108,82	111,66	114,91	123,72

Notes: All average reimbursements are calculated with the tariffs set by the regulator in 2011.

Appendix C: Estimation output

Table 10: Estimation results at B-provider level

	Number of treatment episodes Coefficient (std.err.)	*Total production* Coefficient (std.err.)
Baseline (β_1)	-1.62 (2.03)	1336.02 (2309.93)
Response 2012 (β_2)	-33.77*** (6.88)	-21206.33*** (5686.75)
Response 2013 (β_3)	-32.07*** (8.52)	-13476.04 (11504.62)
Intercept	211.95 (8.20)	267147.60 (10481.75)
Dummies		
- Month	Yes	Yes
- Start 2008	Yes	Yes
- Individual	Yes	Yes
Observations	23,635	23,635
Number of providers	357	357
R^2	0.004	0.002
Fraction of variance due to u_i	0.911	0.908

Notes: *, **, and *** indicate significance based on a two-sided test at the .10, .05, and .01 levels, respectively. The robust standard errors are presented between parentheses.

Table 11: Estimation results at treatment episode level for B-providers

	No case-mix controls		With case-mix controls	
	Treatment duration	*DIFGAF*	*Treatment duration*	*DIFGAF*
	Coefficient (std.err.)	Coefficient (std.err.)	Coefficient (std.err.)	Coefficient (std.err.)
Baseline (β_1)	21.05*** (6.05)	0.007 (0.009)	14.89*** (5.17)	0.010 (0.007)
Response 2012 (β_2)	154.92*** (12.86)	0.002 (0.012)	98.45*** (11.59)	-0.004 (0.008)
Response 2013 (β_3)	191.06*** (23.76)	0.004 (0.027)	109.80*** (23.76)	-0.014 (0.020)
Intercept	594.18*** (14.55)	1.573*** (0.017)	813.82*** (164.94)	4.353 (0.445)
Dummies				
- Month	Yes	Yes	Yes	Yes
- Start 2008	Yes	Yes	Yes	Yes
- Main diagnosis	No	No	Yes	Yes
- Sub-diagnosis	No	No	Yes	Yes
- Type of care	No	No	Yes	Yes
- Overnight stays	No	No	Yes	Yes
- Age	No	No	Yes	Yes
- Gender	No	No	Yes	Yes
- GAFopen	No	No	Yes	Yes
- Provider	Yes	Yes	Yes	Yes
Observations	3,892,093	3,892,093	3,892,093	3,892,093
R^2	0.055	0.104	0.227	0.189

Notes: *, **, and *** indicate significance based on a two-sided test at the .10, .05, and .01 levels, respectively. The robust standard errors are presented between parentheses.

Table 12: Estimation results at NB-provider level

	Number of treatment episodes	*Total production*
	Coefficient (std.err.)	Coefficient (std.err.)
Baseline (β_1)	0.29*** (0.09)	439.04*** (99.02)
Response 2012 (β_2)	-1.92*** (0.24)	-1297.69*** (195.29)
Response 2013 (β_3)	-2.25**** (0.35)	-1796.76*** (304.03)
Intercept	9.91*** (0.28)	10122.12*** (284.81)
Dummies		
- Month	Yes	Yes
- Start 2008	Yes	Yes
- Individual	Yes	Yes
Observations	44,908	44,908
Number of providers	740	740
R^2	0.061	0.045
Fraction of variance due to u_i	0.467	0.423

Notes: *, **, and *** indicate significance based on a two-sided test at the .10, .05, and .01 levels, respectively. The robust standard errors are presented between parentheses.

Table 13: Estimation results at treatment episode level for NB-providers

	No case-mix controls		With case-mix controls	
	Treatment duration	*DIFGAF*	*Treatment duration*	*DIFGAF*
	Coefficient (std.err.)	Coefficient (std.err.)	Coefficient (std.err.)	Coefficient (std.err.)
Baseline (β_1)	27.19*** (3.96)	-0.015 (0.009)	30.10*** (4.03)	-0.006 (0.006)
Response 2012 (β_2)	100.62*** (9.14)	0.031* (0.017)	36.52*** (8.29)	-0.000 (0.014)
Response 2013 (β_3)	69.68*** (14.27)	0.102*** (0.025)	-45.47*** (13.99)	0.029 (0.019)
Intercept	1460.44*** (8.50)	1.004*** (0.018)	909.27*** (185.01)	1.777*** (0.502)
Dummies				
- Month	Yes	Yes	Yes	Yes
- Start 2008	Yes	Yes	Yes	Yes
- Main diagnosis	No	No	Yes	Yes
- Sub-diagnosis	No	No	Yes	Yes
- Type of care	No	No	Yes	Yes
- Overnight stays	No	No	Yes	Yes
- Age	No	No	Yes	Yes
- Gender	No	No	Yes	Yes
- GAFopen	No	No	Yes	Yes
- Provider	Yes	Yes	Yes	Yes
Observations	252,776	252,776	252,776	252,776
R^2	0.173	0.217	0.248	0.304

Notes: *, **, and *** indicate significance based on a two-sided test at the .10, .05, and .01 levels, respectively. The robust standard errors are presented between parentheses.

Appendix D: Robustness analyses

In this section, we perform four aditional analyses and test whether our results are not driven by other mechanisms.

First, we perform our estimations of model (11) for direct treatment duration only. Total treatment duration is the sum of direct, i.e. face-to-face time with patients, and indirect treatment duration, i.e. administrative tasks. Table 14 shows that the regression results for both B- and NB-providers in the column "Direct treatment duration". The differences between the results for total, i.e. column "All", and direct treatment duration are about the same, which indicates that the increase in treatment duration is driven by face-to-face treatment minutes of the health care provider with the patient, and not mainly driven by longer (or shorter) administrative tasks of the provider.

Second, we test whether announcement effects are important. The policy reform was legally announced by the government in june 2011. Providers may have anticipated to this announcement. For example, they may have started new treatments already at the end of 2011, for example to prevent patients from paying a deductible in 2012. Also Table 3 suggests an ad-

Table 14: Robustness analyses

Dependent variable: Treatment duration						
	All	Direct treatment duration	Announcement effect (1)	Announcement effect (2)	Recoding effect	Number of days
B-providers						
Baseline (β_1)	1.2%***	0.7%	1.2%**	1.0%	1.2%***	5.7%**
Response 2012 (β_2)	7.8%***	8.3%***	7.7%***	10.8%***	7.7%***	3.8%***
Response 2013 (β_3)	8.6%***	6.8% **	8.5%***	9.5%***	8.6%***	1.8%
NB-providers						
Baseline (β_1)	2.9%***	2.9%**	2.6%***	2.7%***	2.9%***	11.5%***
Response 2012 (β_2)	3.5%***	4.9%***	5.1%***	4.0%***	3.4%***	2.2%***
Response 2013 (β_3)	-4.3%***	-3.2%**	-2.6%	-2.6%	-4.2%***	-3.9%***

Notes: In this table the interpretation of the regression results are presented (see section 5. The levels of significance refer to the significance of the coefficients underlying these percentages. *, **, and *** indicate significance based on a two-sided test at the .10, .05, and .01 levels, respectively. All regressions are with case-mix controls. The full regression output can be retrieved by the authors.

justment effect, for especially B-providers, as average treatment duration is lower in the first than second half of 2012. It is outside the scope of this paper to identify all possible provider effects of the announcement but here we are mainly interested whether these responses may have influenced our main results. Therefore we run our regression model (11) twice. In one regression, outcomes shown in column "Announcement effect (1)", we leave out all observations from July 2011 until December 2011 and the other regression, outcomes shown in column "Announcement effect (2)", we leave out all observations from July 2011 until June 2012. Table 14 shows that for both regressions the results are comparable to the results in the column "All", albeit the provider responses are somewhat stronger for B-providers when we also leave out the observations for January 2012 until June 2012.

Third, we test whether our estimates may be biased because of recoding in 2012 and 2013. As we discussed in the paper diagnoses "Adjustment" and "V-codes" were excluded from the basic benefit package. Table 7 and 8 show that for both types of providers the number of treatment episodes drops drastically for both diagnoses in 2012 and 2013. Hence, some providers may have recoded patients with "Adjustment" and "V-codes" to other diagnoses, such as "Mood", in order to prevent that the patient needs to pay the full fee for the treatment. To test this possibility we ran a new regression were we coded all diagnoses "Adjustment" and "V-codes" for all years in our dataset as "Mood" (and we left out all subdiagnoses for "Mood" in our controls). The results in the column "Recoding effect" in Table 14 show that a possible bias in our results due to recoding is neglegible.

Fourth, we test whether providers treated patient longer when measured in the number of days, instead of measured in treatment minutes. We measure the number of days as the difference between the first day and the last day that a patient visits a provider (as recorded in the DBC). When providers receive fewer patients the remaining patients can, theoretically, be treated in fewer days. However, since we find longer treatment durations after the shock, we might also find an increase in the number of days after the shock. The results are shown in the column "Number of days". The results indicate, compared to baseline, an increase in the number of days after the shock but the percentages are smaller than compared to the percentages of the main results in the column "All".