

This is a repository copy of *The impact of waiting time on patient outcomes : Evidence from Early Intervention in Psychosis services in England*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/132381/>

Version: Accepted Version

Article:

Reichert, A. and Jacobs, R. orcid.org/0000-0001-5225-6321 (2018) The impact of waiting time on patient outcomes : Evidence from Early Intervention in Psychosis services in England. *Health Economics*. pp. 1772-1787. ISSN 1057-9230

<https://doi.org/10.1002/hec.3800>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

The impact of waiting time on patient outcomes: Evidence from Early Intervention in Psychosis services in England

Anika Reichert^{a*}, Rowena Jacobs^a

^aCentre for Health Economics, University of York, YO105DD, UK

* Address for correspondence: A. Reichert, Centre for Health Economics, University of York,
Alcuin College Block A, York, YO105DD, UK (Email: ar1314@york.ac.uk)

Keywords: waiting times, mental health, psychosis, routine outcome measures, treatment
intensity

ABSTRACT

Recently, new emphasis was put on reducing waiting times in mental health services as there is an ongoing concern that longer waiting time for treatment leads to poorer health outcomes. However, little is known about delays within the mental health service system and its impact on patients. We explore the impact of waiting times on patient outcomes in the context of Early Intervention in Psychosis (EIP) services in England from April 2012 to March 2015. We use the Mental Health Services Data Set and the routine outcome measure the Health of the Nation Outcome Scale. In a generalised linear regression model, we control for baseline outcomes, previous service use and treatment intensity to account for possible endogeneity in waiting time. We find that longer waiting time is significantly associated with a deterioration in patient outcomes 12 months after acceptance for treatment for patients that are still in EIP care. Effects are strongest for waiting times longer than 3 months and effect sizes are small to moderate. Patients with shorter treatment periods are not affected. The results suggest that policies should aim to reduce excessively long waits in order to improve outcomes for patients waiting for treatment for psychosis.

1. INTRODUCTION

Waiting times are a notable phenomenon in publicly funded health care systems such as the English National Health Service. Waiting lists can serve to stock available demand and optimise utilisation of the scarce supply of resources such as skilled staff and medical equipment (Culyer *et al.*, 1976). However, concerns arise when cases are affected in which waiting time may impede the patient's utility gain from the treatment. In the case of psychosis, timely access to care is considered a key priority in successful treatment. It has significant implications for the prevention of impairments and disabilities, functional and symptomatic recovery, as well as the level of treatment engagement of patients (Doyle *et al.*, 2014; Penttilä *et al.*, 2014). This is why recently, new emphasis has been placed on reducing waiting times in mental health services with the introduction of maximum waiting time targets for Early Intervention in Psychosis (EIP) services in England (NHS England *et al.*, 2015). But to date little is known about delays within the mental health service system and their impact on patients.

This paper seeks to improve the understanding of the relationship between waiting times and patient outcomes in the context of EIP services in England. We investigate whether the time from acceptance onto the EIP caseload to start of treatment, not only leads to a deterioration in the patient's condition while waiting but also impedes the patient's ability to benefit from treatment up to 12 months after the start of treatment. The distinct feature of EIP services is that treatment is delivered over several months or years and treatment intensity can vary from patient to patient. Further, recovery in psychosis is a long lasting process where keeping patients in a stable condition is considered a good outcome (Revier *et al.*, 2015). Rather than looking at the outcomes immediately after a single treatment event as in previous literature, we look at patient outcomes after 12 months, incorporating treatment intensity during this time period. Our outcome measure, the Health of the Nation Outcome Scale (HoNOS), comes with a number of advantages for our analysis. Being clinician-reported, it provides a measure of patient outcome, independent of the

patient's subjectivity, which on the one hand is a desired dimension in patient-reported measures (Fitzpatrick et al., 1998) but may be challenging for people with severe mental illness (McCabe et al., 2007; Reininghaus et al., 2012). Previous work on waiting times using other outcome measures consistently found low to moderate effect sizes – it is however questionable whether effects that are statistically significant but small are also clinically relevant. We advance the analysis by estimating the impact of waiting time on a clinically reliable and significant change in HoNOS. While HoNOS is not specific to psychosis, it is routinely collected in administrative data which offers the potential to expand future analysis to other samples and mental health conditions in a comparable manner. As such, our work contributes to the literature discussing the feasibility and usefulness of routine outcome measures in general (Boswell et al., 2015) and for mental health conditions in particular (Gilbody et al., 2003; Tasma et al., 2017).

2. BACKGROUND

2.1. Theoretical framework

Waiting for treatment on a waiting list does not require patients to queue in person. Hence, there are no opportunity costs in terms of time spent waiting in order to clear markets. But still, waiting times impose costs, as introduced in the model of queuing by list by Lindsay *et al.* (1984). Patients join the waiting list in order to obtain the right to receive treatment at some point in the future. The value of obtaining this right depends on the price to be paid when receiving the treatment (which is zero in the case of no co-payments) and the delay (waiting time) to receive of the good. The delay effects are composed of two factors: First, a positive interest rate will lead to a discounted value of the good consumed in the future relative to its present value. Second, the time of delivery affects the present value of consumption due to, for example, pain, uncertainty, and disability. In other words, the treatment received tomorrow is worth less since the patient (and caring relatives) have to experience suffering and inconvenience of living with

a disease. Koopmanschap *et al.* (2005) further demonstrated theoretically that the negative impact of waiting time can be long-term. The deteriorated condition of the patient while waiting may take longer to recover or will not be reversed at all after a critical waiting time has passed. In the context of psychosis, the suffering can be significant. Psychoses summarise a group of serious mental health conditions in which a person's perception, thoughts, mood and behaviour are significantly altered (NICE, 2014). Patients experience a high degree of impairment and the often unusual or bizarre behaviour leads to difficulties in managing their own life up to the point of social exclusion (Huxley *et al.*, 2003). Although psychosis affects people in the most productive period of their working lives (mid- to late-twenties), they face lower rates of employment, lower payment and less secure jobs (Marwaha *et al.*, 2004; Revier *et al.*, 2015). All of the above suggests that waiting for psychotic treatment will contribute to a deterioration in treatment outcomes as symptoms and impairments will worsen over time and patients may disengage with treatment (Doyle *et al.*, 2014; Penttilä *et al.*, 2014).

2.2. Early intervention in psychosis services and treatment

EIP services provide a full range of pharmacological, psychological, social, occupational and educational interventions (NICE, 2014). The main treatment options are antipsychotic medication in conjunction with psychological interventions such as cognitive behavioural therapy and family interventions. The psychological interventions are delivered in a number of planned sessions over a period of at least 3 to 12 months. Given the multidisciplinary nature of EIP services, the care coordinator plays a key role in the effective delivery of EIP care (NHS England *et al.*, 2015). They not only bring together all different professionals involved in the care of the patient, such as therapists, social workers, and psychiatrists. They are also responsible for engaging patients in treatment and supporting them across the total spectrum of their needs. Each EIP service is governed by the responsible mental health trust as the provider, which in

turn negotiates the overall budget with a number of payers and distributes it to the individual services in charge.

2.3. Related literature

Two strands of literature shall be distinguished in the discussion of waiting times and outcomes. The first strand focuses on physical health conditions with most studies in the area of non-urgent surgical procedures such as hip and knee replacement (Braybrooke *et al.*, 2007; Hamilton *et al.*, 1999; Hamilton *et al.*, 1996; Hirvonen *et al.*, 2007; Hirvonen *et al.*, 2009; Ho *et al.*, 2000; Nikolova *et al.*, 2016; Quintana *et al.*, 2011; Tuominen *et al.*, 2009, 2010), or more urgent surgical procedures such as organ transplantation (Meier-Kriesche *et al.*, 2000; Rauchfuss *et al.*, 2013), and coronary artery bypass surgery (Manji *et al.*, 2013; Moscelli *et al.*, 2016; Sari *et al.*, 2007). Fewer studies investigate the relationship of waiting time with non-surgical treatments such as rehabilitation (Collins *et al.*, 2015; Pedersen *et al.*, 2017), radiotherapy (Gupta *et al.*, 2016; Noel *et al.*, 2012; Seidlitz *et al.*, 2015), or HIV treatment (Su *et al.*, 2016). Results are inconsistent as to whether longer waiting causes worse chances of functional remission, recurrence, treatment adherence, quality of life, and mortality. Most of these studies use field data which are limited in sample size, number of providers, and covariates to control for confounders. More recently, studies have used administrative data to overcome some of these limitations. Moscelli *et al.* (2016) found that waiting for coronary bypass surgery did increase the number of emergency readmissions but not in-hospital mortality. Nikolova *et al.* (2016) analysed the impact of waiting for elective surgery on patient-reported outcomes. They found that a longer waiting time reduces health-related quality of life for hip and knee replacement, but not for varicose veins and inguinal hernia.

The second strand of literature focuses on the impact of treatment delays on outcomes regarding first episode of psychosis patients. The key measure of waiting time in this context is the duration

of untreated psychosis (DUP). DUP measures the time from the onset of the psychosis to the start of treatment and is mostly defined using patient interviews (Norman *et al.*, 2001). Penttilä *et al.* (2014) recently published a comprehensive review of 33 studies. Longer DUP was associated with more severe symptomatic outcomes and reduced remission rates with small to moderate effect sizes. Also, longer DUP correlated with poorer social functioning but not with employment or quality of life. Some recent studies looked at long-term effects of DUP on outcomes. In a 20-year follow-up, Cechnicki *et al.* (2014) found significantly deteriorated outcomes for the long DUP group (> 6 months) in terms of symptom recovery, social functioning, and employment. Tang *et al.* (2014) reported significantly higher symptom remission rates for the shorter DUP group after accounting for confounding factors in a 13-year follow-up period. Despite the quantity of studies, evidence remains limited, since studies tend to be small-scale with sample sizes between 23 and 776 patients using only a single or a few providers. Attrition rates ranged from 4–71% which could be a source of significant selection bias. Most studies are based on purely correlational methods or do not account adequately for the typically skewed nature of DUP (Marshall *et al.*, 2005; Norman *et al.*, 2001).

Our work aims to bridge the gap between these two distinct strands of literature. We advance the literature on psychotic patients by using well established methods from physical health care and a large, nationally representative sample. Our waiting time measure moves beyond the traditional concept of DUP to overcome some of its limitations. At the same time, we advance the literature in the physical health context by looking at a different treatment regime characterised by multiple treatment events over a period of several months. This stresses the importance of treatment intensity which we include in the analysis.

3. DATA AND KEY MEASURES

We use the Mental Health Services Data Set which is a national administrative database of mental health related treatment in hospitals and community settings within the English NHS. Patients were included if they had a first EIP episode within the study period April 2012 to March 2014 and followed up for a period of 12 months. Figure 1 summarises the study timeline and measurement points.

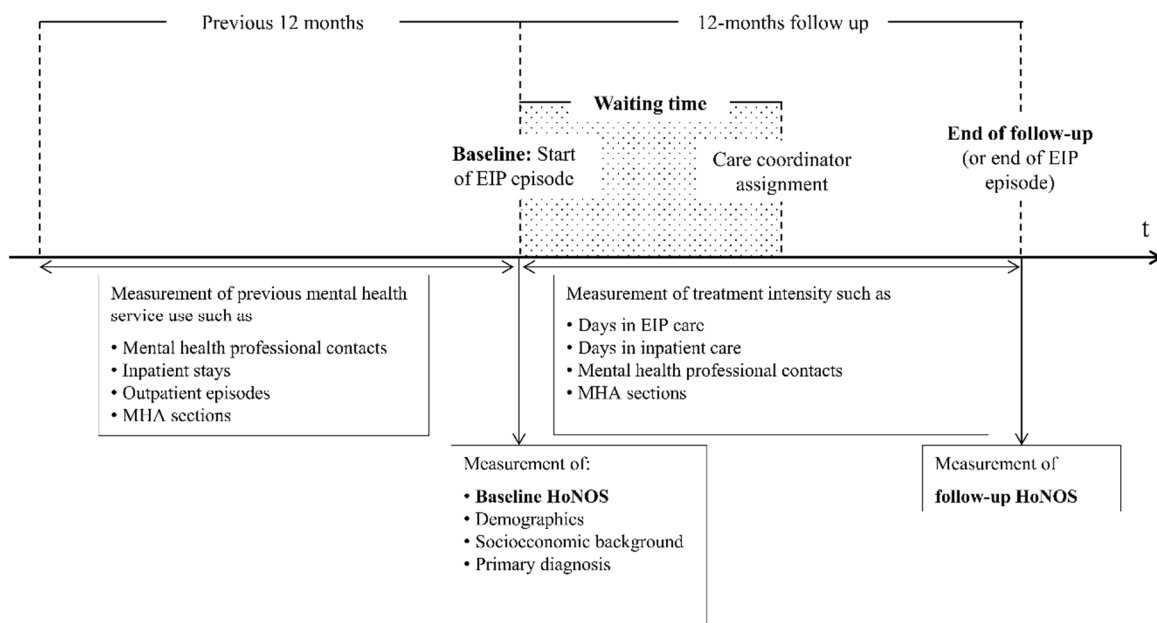


Figure 1 Study timeline with measurement points

Our outcome measure, HoNOS, is routinely collected by providers in our dataset (Wing *et al.*, 1998; Wing *et al.*, 1999). HoNOS is composed of 12 items covering the four sub-domains behaviour, impairment, symptoms, and social functioning (see Appendix A). Each item is evaluated by a trained clinician on a scale from 0 (no problem) to 4 (severe to very severe problems) and contributes equally to the total score ranging from 0 to 48. HoNOS measurements are conducted at treatment start and during the course of treatment. This allows us to observe a baseline HoNOS score at the start of the EIP episode and a score at follow-up at the end of the 12 months (or at the end of the EIP episode if treatment ended before the follow-up). We use the

baseline measurement to condition on pre-treatment severity. Further, we determine whether patients improved reliably and in a clinically meaningful way using the concept of reliable and clinically significant change introduced by Jacobson *et al.* (1991) and applied to HoNOS by Parabiaghi *et al.* (2005).

Inpatient waiting time as commonly used in physical health papers measures the time from the specialist's decision to treat until the start of the inpatient treatment (Siciliani *et al.*, 2014). We translate this concept to the context of psychosis by measuring the time from the patient's acceptance onto the EIP caseload (decision to treat) to the assignment of a care coordinator (start of treatment). The care coordinator is the key requirement for effective treatment to be initiated (NHS England *et al.*, 2015). Previous papers found the relationship between waiting time and outcomes to be non-linear with outcomes deteriorating significantly at a waiting time longer than 1 month (Tang *et al.*, 2014) or 3 months (Cechnicki *et al.*, 2014). Therefore, we employ three different transformations of waiting time: (i) a log transformation of waiting time in days, (ii) waiting time quintiles with an equal number of patients in each group, (iii) waiting time intervals based on the thresholds typically used in the previous literature (0.5 to 3 months, 3 to 6 months, and 6 to 12 months).

4. METHODS

4.1. The model

We denote h_{ijkl} as the mental health status of the i th-patient, $i = 1, \dots, N$ who lives in small area j , $j = 1, \dots, J$ and receives treatment at provider k , $k = 1, \dots, K$ in the financial year l , $l = 1, \dots, L$. The health status is measured prior to treatment (h_{ijkl}^1) and 12 months after treatment start (h_{ijkl}^2) as the total HoNOS score. Formally, the model is specified as

$$h_{ijkl}^2 = \alpha W_{ijkl} + \beta h_{ijkl}^1 + \gamma T_{ijkl} + \delta X_{ijkl} + \tau S_{ijkl} + y_l + u_k + \varepsilon_{ijkl} \quad (1)$$

where W_{ijkl} represents the patient's waiting time. The patient's outcome prospects are likely to depend on the severity of the condition at baseline. We therefore condition on the baseline HoNOS score h_{ijkl}^1 . T_{ijkl} encompasses measures of treatment intensity. Over the 12-month follow-up period, treatment intensity will vary between patients but may also impact on the patient's outcomes. We approximate treatment intensity by the following variables: (i) the number of days in EIP care, (ii) the number of days in inpatient care, and (iii) the number of mental health professional contacts until the end of follow-up (or end of EIP if earlier than follow-up). We further control for whether a patient was being detained under the Mental Health Care Act in that period since additional legislative requirements impose a higher level of treatment intensity. Since the degree to which each of the variables contributes to the patient's recovery process is unknown, we include each of them with equal weight into the model.

Patient characteristics that could impact both waiting time and outcomes are captured in X_{ijkl} . Alongside a range of demographic characteristics, we consider the patient's socioeconomic background. At patient-level, we include accommodation and employment status. Further, we used socioeconomic deprivation based on the Index of Multiple Deprivation measured at small-area level (McLennan *et al.*, 2011). Previous mental health service use, represented by S_{ijkl} may

be indicative of the patient's ability to navigate through the system and take advantage of treatment options (and thus impact waiting times as well as outcomes). The vector includes the number of inpatient stays (in intervals 0, 1-2, >2), outpatient episodes (in intervals 0, 1-2, >2), mental health professional contacts (in intervals 0, 1-10, >10), and primary as well as secondary diagnoses within the 12 months prior to the EIP start. There are L unobservable year effects y_l and K unobservable provider-level effects u_k for the 48 mental health trusts in our sample. The term ε_{ijkl} represents the idiosyncratic error.

Our main coefficient of interest is α which measures the effect of waiting time on follow-up HoNOS outcomes conditional on the included covariates. We expect follow-up outcomes to deteriorate if waiting time increases both because the waiting itself causes a worsening in the patient's condition and because the waiting impedes the patient's ability to benefit from treatment. Therefore, we expect a positive α indicating an increased (worse) follow-up HoNOS score. By the application of provider and time fixed effects, any variation has to be interpreted as within provider variation for a given year.

Both the Shapiro-Wilk test (Shapiro *et al.*, 1965) and the Shapiro-Francia test (Shapiro *et al.*, 1972) strongly rejected the null hypothesis of h_{ijkl}^2 being normally distributed. We accounted for the skewness by using generalised linear regression methods (Nelder *et al.*, 1972) which were shown to be an adequate choice in typically skewed data (Jones *et al.*, 2016; Sinko *et al.*, 2016). The modified Park test confirmed the Poisson distribution to fit the data best. Both the Pregibon link test (Pregibon, 1980) and the modified Hosmer-Lemeshow goodness-of-fit test (Hosmer *et al.*, 2005) accepted the square root link function. The RESET test (Ramsey, 1969) further confirmed the model specification. We used cluster robust standard errors for the 48 mental health trusts (see Appendix C).

4.2. Robustness checks

We applied the same model from equation (1) to each sub-domain of HoNOS resulting in four separate models for behaviour, impairment, symptoms and social outcomes. We estimated this system of linear equations as a Seemingly Unrelated Regression model without constraints to account for cross-model covariance which was supported by the Breusch-Pagan Lagrange multiplier test for error independence (Zellner, 1962). Further, we use concept of clinically significant and reliable change (see Appendix B for more details) to test whether the effect size we measure is of clinical relevance. We employed an ordered probit model to predict the impact of waiting time on the probability of a clinically significant and reliable change in the HoNOS score conditional on the same set of covariates as introduced above.

5. RESULTS

5.1. Descriptive statistics

We identified 14,912 patients (full sample) having a first EIP episode and a care coordinator within the study period. We excluded 5,874 patients (39.4%) for which we could not observe two complete HoNOS records. Another 89 patients (0.01%) were excluded which were from providers treating fewer than 30 patients in the sample. The remaining study sample included 8,949 patients being treated within 48 mental health trusts. Table I compares key characteristics of the study sample with those from the full and the excluded sample. Our study sample was on average 25.8 years old, predominantly male, of White ethnicity, single, and diagnosed with schizophrenia. Most lived in mainstream housing within the most deprived neighbourhoods, and were unemployed.

Table I Descriptive statistics

	Full sample	Study sample	Excluded sample
Number of patients	14,912	8,949	5,963
Number of providers	55	48	55
Patient demographics			
Patient age (mean)	24.9	25.8	23.6
Male (%)	63.7	63.6	64.0
White ethnicity (%)	73.4	72.1	75.6
Marital status: Single (%)	88.5	87.1	90.9
Schizophrenia diagnosis (%)	53.6	52.9	55.2
Socioeconomic background			
Mainstream housing (%)	83.5	83.6	83.3
Unemployed (%)	45.2	47.4	41.1
Least deprived quintile (%)	9.9	9.3	10.8
Most deprived quintile (%)	37.8	38.1	37.3
Mental health service use (before start of EIP care)			
Zero health professional contacts (%)	31.7	18.3	52.0
Zero outpatient episodes (%)	73.7	68.0	82.3
Zero inpatient admissions (%)	69.2	60.2	82.8
Zero Mental Health Care Act section (%)	77.5	71.6	86.5
HoNOS score at baseline (mean)			
Total (min 0, max 48)	14.0	14.1	13.5
Behaviour score (min 0, max 12)	2.8	2.8	2.9
Impairment score (min 0, max 8)	1.4	1.2	1.0
Symptoms score (min 0, max 12)	5.7	5.7	5.5
Social score (min 0, max 16)	4.4	4.4	4.1
Treatment intensity (during EIP care)			
Days in EIP care (mean)	291.5	306.6	268.9
Days in inpatient care (mean)	15.2	18.8	9.9
Mental health professional contacts (mean)	36.6	42.5	27.7
Mental Health Care Act sectioned (%)	20.5	23.9	15.3

Note: HoNOS observations are reported for the total study sample and for n = 10,012 in the full sample and n = 1,063 in the excluded sample.

The mean HoNOS score at baseline was 14.1. Figure 2 visualises the distributional shift of HoNOS scores towards zero from baseline to follow-up. During the 12 months follow-up, patients in the study sample spent on average 18.8 days in inpatient care and experienced 42.5 contacts with any kind of mental health professional. 23.9% were sectioned under the Mental Health Care Act at least once during the time of follow-up. Our study sample was on average two years older than the excluded patients and more likely unemployed. Most evident is that patients in the study sample were more likely to have been in contact with mental health services

in the previous 12 months. Also, treatment intensity during the EIP care was higher for the study sample. Mean HoNOS scores at baseline were, however, very similar on all dimensions.

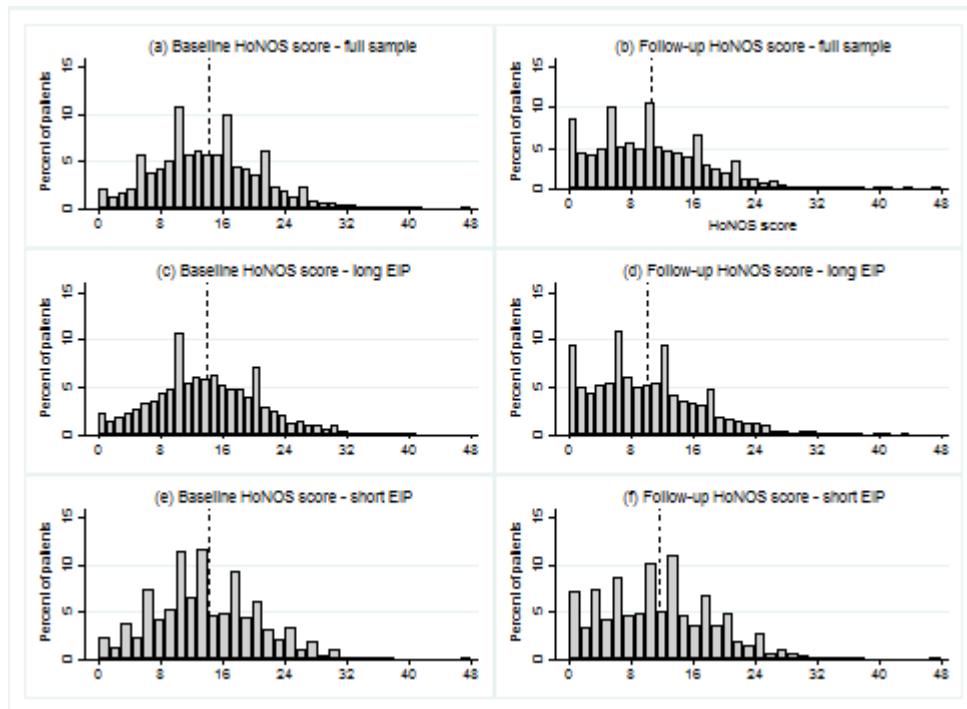


Figure 2 Histogram of HoNOS scores at baseline and follow-up

We note that not all patients spent the total follow-up time in EIP care. We therefore stratified the study sample by whether a patient finished EIP care before the end of follow-up (“short EIP” group, 31.4%) or not (“long EIP” group, 68.6%) and run analyses for the two subsamples separately. Table II shows that in all three samples, HoNOS decreased (improved) from baseline to follow-up by about 2 to 4 points. The short EIP group improved less in HoNOS but waited almost 15 days longer than the long EIP group.

Table II Summary statistics of waiting time and HoNOS

	Study sample		Long EIP		Short EIP	
	Mean	SD	Mean	SD	Mean	SD
Baseline HoNOS	14.1	6.8	14.1	6.8	14.0	6.8
Follow-up HoNOS	10.6	7.0	10.1	6.9	11.7	7.2
Waiting time	50.1	74.1	42.1	64.1	67.4	89.8
Observations	8,949		6,135		2,814	

Figure 3 summarises several descriptive statistics of our main explanatory variable, waiting time. As expected, we find waiting time to be heavily left skewed with a median of 20 and a mean of 50 days. Consequently, the largest proportion of patients was allocated to the waiting time interval of less than 0.5 months (panel (b)). We also see that taking the logarithm of waiting time helps to reduce a large amount of the skewness (panels (c) and (d)).

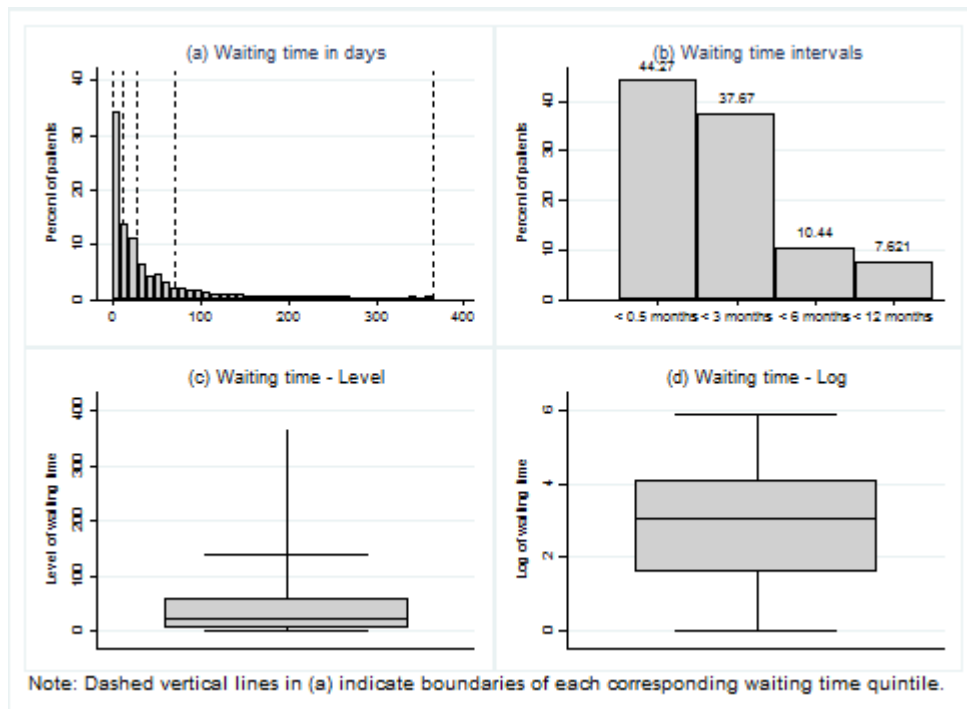


Figure 3 Descriptive statistics on waiting time

Figure 4 visualises the distribution of HoNOS scores across the different waiting time intervals. Baseline HoNOS scores in panel (a) vary very little across intervals of waiting particularly for the first three intervals. While patients from the longest waiting interval had the lowest median HoNOS at baseline, they improved least at follow-up. Panel (b) shows that median follow-up scores decreased (improved) most at follow-up for the shorter waiting time intervals.

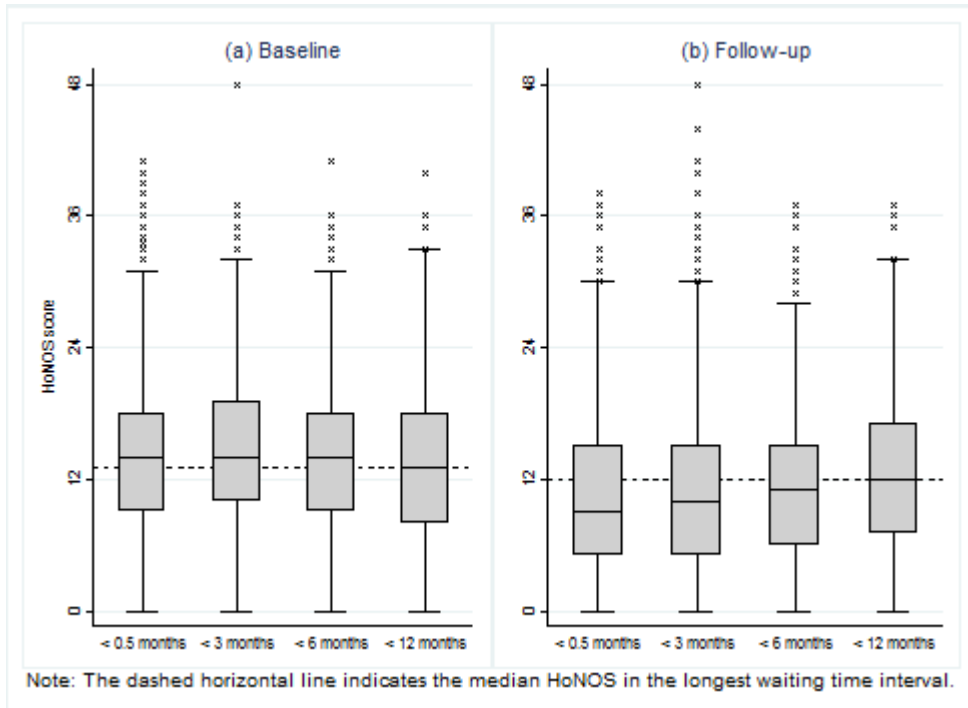


Figure 4 Box plots for HoNOS scores by waiting time intervals

5.2. Estimation results

Table III displays the estimation results from the regression of equation (1) including marginal effects (dy/dx). The estimates for the three different waiting time measures result from three independent regressions. Model (1) includes the whole study sample whereas models (2) and (3) look at long and short EIP patients respectively. We observe a significant but small effect of log waiting time on the HoNOS score 12 months after the EIP start for the total sample and the long EIP group. A 1% longer waiting time translates into an increase (worsening) in HoNOS by 0.20 to 0.27 points. The association between longer waiting and worse outcomes is only significant for the longest waiting quintile – however with a larger effect than the overall. Being in the longest waiting quintile is associated with a 0.78 to 1.27 points higher (worse) HoNOS compared to the shortest waiting time quintile. For long EIP patients, we observe a clear gradient looking at the waiting time intervals. Patients waiting between 0.5 and 3 months (3 to 6 months; 6 to 12 months) had a 0.34 (1.15; 1.61) higher HoNOS score than patients waiting less than 0.5 months.

Patients with an EIP episode shorter than the follow-up time seem to be not significantly affected by the length of waiting.

Table III Generalised Linear Model results with follow-up HoNOS as dependent variable

	(1) Study sample			(2) Long EIP			(3) Short EIP		
	Coeff.	SE	dy/dx	Coeff.	SE	dy/dx	Coeff.	SE	dy/dx
Log waiting time (continuous)	0.032***	(0.009)	0.20	0.043***	(0.010)	0.27	0.022	(0.015)	0.15
Waiting time quintiles (ref.cat.: shortest quintile)									
2nd shortest quintile	-0.050	(0.036)	-0.32	-0.015	(0.049)	-0.09	-0.109	(0.064)	-0.73
3rd shortest quintile	-0.057	(0.041)	-0.36	0.004	(0.051)	0.02	-0.186**	(0.060)	-1.24
4th shortest quintile	0.031	(0.033)	0.20	0.050	(0.039)	0.31	0.036	(0.081)	0.25
Longest quintile	0.119**	(0.045)	0.78	0.199***	(0.053)	1.27	0.021	(0.072)	0.14
Waiting time intervals (ref.cat.: less than 0.5 months)									
Waiting time 0.5 to 3 months	0.040	(0.023)	0.25	0.054*	(0.027)	0.34	0.034	(0.045)	0.23
Waiting time 3 to 6 months	0.120***	(0.034)	0.78	0.181***	(0.048)	1.15	0.076	(0.052)	0.52
Waiting time 6 to 12 months	0.215***	(0.059)	1.41	0.250***	(0.065)	1.61	0.189	(0.098)	1.30
Observations	8,949			6,135			2,814		
Provider and year fixed effects	yes			yes			yes		
Covariates	yes			yes			yes		

Note: * p<0.05, ** p<0.01, *** p<0.001. Model (1) includes the complete study sample. Model (2) includes only patients with an EIP episode longer than the follow-up. Model (3) includes only patients with an EIP episode shorter than the follow-up. "dy/dx" represents average marginal effects in days. For factor levels they present the discrete change from reference category. All models use cluster robust standard errors for 48 provider clusters.

Table IV reports the estimated coefficients and marginal changes of the baseline HoNOS as well as the treatment intensity variables based on equation (1). As expected, we observe a strong positive relationship between baseline and follow-up HoNOS scores. A higher or worse baseline condition strongly predicts worse outcomes 12 months after treatment start. Most severely affected patients had an up to 5 points worse outcome at follow-up. Overall, treatment intensity does not seem to impact outcomes much. Although significant, effect sizes are small. More days spent in EIP care seem to improve outcome prospects. Interestingly, more days of inpatient care and more mental health professional contacts are associated with a deterioration in follow-up outcomes. This may be explained by the fact that the two variables also capture some level of baseline severity of the patient that is not captured in the other control variables. In this case, more severe patients would need more inpatient care and service contacts but at the same time

have worse outcome prospects regardless of treatment intensity. For the short EIP group, we observe patients who were sectioned under the Mental Health Care Act to have significantly worse outcomes. This again may be explained by the variable capturing some different dimension of baseline severity, but it may also be an indication that involuntary treatment worsens outcome prospects.

Table IV Generalised Linear Model results of baseline HoNOS and treatment intensity on follow-up HoNOS

	(1) Study sample		(2) Long EIP		(3) Short EIP	
	Coeff.	dy/dx	Coeff.	dy/dx	Coeff.	dy/dx
Baseline HoNOS (ref.cat.: least severe)						
2nd least severe quintile	0.220***	1.34	0.154***	0.92	0.347***	2.18
3rd least severe quintile	0.310***	1.92	0.268***	1.63	0.404***	2.57
4th least severe quintile	0.400***	2.51	0.297***	1.82	0.620***	4.08
Most severe quintile	0.544***	3.49	0.449***	2.82	0.759***	5.09
Treatment intensity						
Number days in EIP care	-0.001***	-0.01	-	-	-	-
Number of days in inpatient care	0.001*	0.00	0.001**	0.01	0.000	0.00
Number of mental health professional contacts	0.005***	0.03	0.005***	0.03	0.004***	0.03
Sectioned under MHA within follow-up	0.023	0.15	-0.023	-0.15	0.153**	1.04
Observations	8,949		6,135		2,814	
Provider and year fixed effects	yes		yes		yes	
Covariates	yes		yes		yes	

Note: * p<0.05, ** p<0.01, *** p<0.001. All models include log waiting time as regressor.

5.3. Robustness of results

Results from the effects of waiting time on the different HoNOS sub dimensions are provided in Appendix D. In line with previous findings, we find the strongest negative impact of waiting time on the symptoms dimension. But also, all other sub dimensions are negatively affected by a longer waiting time. As before, patients waiting longer than 3 months are affected most by a deterioration in outcomes on each sub-domain. Although effect sizes are moderate, we find evidence of a significant increase in the probability of a reliable and clinically significant deterioration for the study sample and the long EIP group. The likelihood of a clinically relevant deterioration is again highest for the longest waiting patients (see Appendix E).

6. DISCUSSION

Waiting times for mental health services in general and for EIP services in particular have recently gained considerable policy interest. But little is known about the detrimental effect of delays within the care system on outcomes for patients with psychosis. We document a moderate decline in patient outcomes 12 months after treatment acceptance for additional days of waiting. Despite moderate effect sizes, the risk of a clinically significant and reliable deterioration is elevated by longer waiting time. Effects are significant in the waiting time range from 3 to 12 months which supports the threshold theory discussed in previous papers. Also consistent with previous literature, all outcome dimensions are affected with the largest impact on symptomatic and social outcomes.

Our study contributes a number of aspects to existing evidence. First, we developed a strategy to measure a system-related waiting time measure in contrast to the commonly used DUP. DUP has been criticised in its suitability to measure service effectiveness as definitions vary considerably across studies and are prone to a self-report bias by patients (Norman *et al.*, 2001; Register-Brown *et al.*, 2014; Singh, 2007). Our waiting time measure allows us to investigate the impact of delays within the care system rather than the help-seeking behaviour of patients (Gronholm *et al.*, 2017). Second, we consider treatment intensity during the time of follow-up. It allows us to reflect recovery in psychosis as a long lasting process and patient outcomes as a result of repeated service contacts over a period of several months. Finally, we are the first to study a routine outcome measure (HoNOS) to look at psychosis outcomes. HoNOS has been found to have adequate or good validity, reliability, sensitivity to change, and feasibility (Amin *et al.*, 1999; McClelland *et al.*, 2000; Pirkis *et al.*, 2005; Wing *et al.*, 1998). Given its generic nature, it may lack clinical precision. But our findings are consistent with studies that use specific but heterogeneous outcome measures.

We note some limitations of our work. First, we may have underestimated waiting time as we excluded any waiting time that occurred between the first service contact (e.g. general practitioner) or self-referral and the specialist's decision to treat. If longer waiting time does indeed have negative effects on outcomes, we would have estimated a lower bound of the effect. Second, we restricted our follow-up period and thus treatment intensity to 12 months given the boundaries of data availability. Longer follow-up has, however, been shown to increase the impact of waiting time on outcomes (Penttilä *et al.*, 2014). If this is the case, then again, our results are a lower bound estimation. Third, our outcome measure demonstrates the clinician's judgement of the patient's condition which may not necessarily match the patient's perception (Kramer *et al.*, 2003). Fourth, this work is limited by the relatively high number of missing HoNOS records which is common when working with clinician-reported measures (Jacobs, 2009). The remaining study sample had substantially higher proportions of mental health service contacts prior to the EIP treatment than excluded patients. This would have limited the external validity of our results if the relationship between waiting for treatment and outcomes of the same treatment would be different dependent on past service experience. On the one hand, patients may have learned coping strategies during previous service contacts which help them to deteriorate less during the time of waiting. On the other hand, patients with more service contacts in the past may be in a more severe condition overall which will worsen even more during waiting. Whereas in the first case we would have underestimated the negative impact of waiting time, we would have overestimated it if the latter case is true. Without further knowledge about the role of previous service use in the interplay of waiting time and outcomes, our results have to be interpreted as representative for a patient cohort with relatively high mental health service use in the past. If there were systematic differences in HoNOS coding quality between providers which in turn may be associated with the provider's performance regarding patient waiting times and outcomes, we have controlled for these through the use of provider fixed effects.

Finally, the estimated effect is based on the assumption that the baseline health outcome conditional on other individual characteristics, including previous service use and treatment intensity, is sufficient to account for the individual's unobserved pre-treatment severity. We find the baseline outcome to be a strong predictor for the follow-up outcome. Also, accounting for previous service use and treatment intensity may have captured some remaining severity not observed by the baseline HoNOS. However, there may still have remained unobserved severity that explains both longer waiting times and worse outcomes. Future research should aim to consider either a valid instrument or a suitable comparison group to deal with this challenge.

Our results have direct implications for the recently introduced waiting time target policy for EIP services. As has been the case in many previous target policies in other health areas, the 14-day target appears to have been chosen arbitrarily rather than based on evidence. A comprehensive discussion on the optimal targeted waiting time needs to consider the effects on patient outcomes but also implications for the supply side. Our paper sheds some light on the demand dimension. According to our results, the target policy can only be effective in improving patient outcomes if it leads to a reduction in excessive waits longer than 3 months.

REFERENCES

- Amin, S, Singh, SP, Croudace, T, Jones, P, Medley, I, & Harrison, G. (1999). Evaluating the Health of the Nation Outcome Scales. *The British Journal of Psychiatry*, 174.
- Boswell, JF, Kraus, DR, Miller, SD, & Lambert, MJ. (2015). Implementing routine outcome monitoring in clinical practice: benefits, challenges, and solutions. *Psychotherapy research*, 25(1), 6-19.
- Braybrooke, J, Ahn, H, Gallant, A, Ford, M, Bronstein, Y, Finkelstein, J, & Yee, A. (2007). The impact of surgical wait time on patient-based outcomes in posterior lumbar spinal surgery. *European Spine Journal*, 16(11), 1832-1839.
- Cechnicki, A, Cichocki, L, Kalisz, A, Bładzinski, P, Adamczyk, P, & Franczyk-Glita, J. (2014). Duration of untreated psychosis (DUP) and the course of schizophrenia in a 20-year follow-up study. *Psychiatry Research*, 219(3), 420-425.
- Collins, CL, Suskin, N, Aggarwal, S, & Grace, SL. (2015). Cardiac rehabilitation wait times and relation to patient outcomes. *European Journal of Physical and Rehabilitation Medicine*, 51(3), 301-309.
- Culyer, AJ, & Cullis, JG. (1976). Some Economics of hospital waiting lists in the NHS. *Journal of Social Policy*, 5(3), 239-264.

- Doyle, R, Turner, N, Fanning, F, Brennan, D, Renwick, L, Lawlor, E, & Clarke, M. (2014). First-episode psychosis and disengagement from treatment: a systematic review. *Psychiatric Services*, 65(5), 603-611.
- Fitzpatrick, R, Davey, C, Buxton, M, & Jones, D. (1998). Evaluating patient-based outcome measures for use in clinical trials: a review. *Health Technology Assessment*, 2(14).
- Gilbody, SM, House, AO, & Sheldon, TA. (2003). Outcome measures and needs assessment tools for schizophrenia and related disorders. *Cochrane Database Syst Rev*(1), Cd003081.
- Gronholm, PC, Thornicroft, G, Laurens, KR, & Evans-Lacko, S. (2017). Mental health-related stigma and pathways to care for people at risk of psychotic disorders or experiencing first-episode psychosis: a systematic review. *Psychological Medicine*, 15, 1-13.
- Gupta, S, King, WD, Korzeniowski, M, Wallace, DL, & Mackillop, WJ. (2016). The Effect of waiting times for postoperative radiotherapy on outcomes for women receiving partial mastectomy for breast cancer: a systematic review and meta-analysis. *Clinical Oncology*, 28(12), 739-749.
- Hamilton, BH, & Bramley-Harker, RE. (1999). The impact of the NHS reforms on queues and surgical outcomes in England: Evidence from hip fracture patients. *Economic Journal*, 109(457), 437-462.
- Hamilton, BH, Hamilton, VH, & Mayo, NE. (1996). What are the costs of queuing for hip fracture surgery in Canada? *Journal of Health Economics*, 15(2), 161-185.
- Hirvonen, J, Blom, M, Tuominen, U, Seitsalo, S, Lehto, M, Paavolainen, P, . . . Sintonen, H. (2007). Evaluating waiting time effect on health outcomes at admission: A prospective randomized study on patients with osteoarthritis of the knee joint. *Journal of Evaluation in Clinical Practice*, 13(5), 728-733.
- Hirvonen, J, Tuominen, U, Seitsalo, S, Lehto, M, Paavolainen, P, Hietaniemi, K, . . . Blom, M. (2009). The effect of waiting time on health-related quality of life, pain, and physical function in patients awaiting primary total hip replacement: a randomized controlled trial. *Value in Health*, 12(6), 942-947.
- Ho, V, Hamilton, BH, & Roos, LL. (2000). Multiple approaches to assessing the effects of delays for hip fracture patients in the United States and Canada. *Health Services Research*, 34(7), 1499-1518.
- Hosmer, DW, & Lemeshow, S. (2005). Assessing the fit of the model. In D. W. Hosmer, S. Lemeshow, & R. X. Sturdivant (Eds.), *Applied Logistic Regression* (pp. 143-202). New Jersey: John Wiley & Sons, Inc.
- Huxley, P, & Thornicroft, G. (2003). Social inclusion, social quality and mental illness. *Br J Psychiatry*, 182, 289-290.
- Jacobs, R. (2009). *Investigating Patient Outcome Measures in Mental Health*. University of York, Centre for Health Economics: CHE Research Paper 48.
- Jacobson, NS, & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of consulting and clinical psychology*, 59(1), 12-19.
- Jones, AM, Lomas, J, Moore, PT, & Rice, N. (2016). A quasi-Monte-Carlo comparison of parametric and semiparametric regression methods for heavy-tailed and non-normal data: an application to healthcare costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(4), 951-974.
- Koopmanschap, MA, Brouwer, WBF, Hakkaart-van Roijen, L, & van Exel, NJA. (2005). Influence of waiting time on cost-effectiveness. *Social Science & Medicine*, 60(11), 2501-2504.
- Kramer, TL, Owen, RR, Wilson, C, & Thrush, CR. (2003). Relationship Between Self-Report and Clinician-Rated Impairment in Depressed Outpatients. *Community Mental Health Journal*, 39(4), 299-307.
- Lindsay, CM, & Feigenbaum, B. (1984). Rationing by waiting lists. *American Economic Review*, 74(3), 404-417.
- Manji, RA, Jacobsohn, E, Grocott, HP, & Menkis, AH. (2013). Longer in-hospital wait times do not result in worse outcomes for patients requiring urgent coronary artery bypass graft surgery. *Hospital practice (1995)*, 41(3), 15-22.

- Marshall, M, Lewis, S, Lockwood, A, Drake, R, Jones, P, & Croudace, T. (2005). Association between duration of untreated psychosis and outcome in cohorts of first-episode patients: a systematic review. *Archives of General Psychiatry*, 62(9), 975-983.
- Marwaha, S, & Johnson, S. (2004). Schizophrenia and employment - a review. *Soc Psychiatry Psychiatr Epidemiol*, 39(5), 337-349.
- McCabe, R, Saidi, M, & Priebe, S. (2007). Patient-reported outcomes in schizophrenia. *The British Journal of Psychiatry*, 191(50), s21-s28.
- McClelland, R, Trimble, P, Fox, ML, Bell, B, & Stevenson, MR. (2000). Validation of an outcome scale for use in adult psychiatric practice. *Quality in Health Care*, 9(2), 98-105.
- McLennan, D, Barnes, H, Noble, M, Davies, J, Garratt, E, & Dibben, C. (2011). The English indices of deprivation 2010. Retrieved from <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010>
- Meier-Kriesche, HU, Port, FK, Ojo, AO, Rudich, SM, Hanson, JA, Cibrik, DM, . . . Kaplan, B. (2000). Effect of waiting time on renal transplant outcome. *Kidney International*, 58(3), 1311-1317.
- Moscelli, G, Siciliani, L, & Tonei, V. (2016). Do waiting times affect health outcomes? Evidence from coronary bypass. *Social Science & Medicine*, 161, 151-159.
- Nelder, JA, & Wedderburn, RWM. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 135(3), 370-384.
- NHS England, & NICE. (2015). *Implementing the early intervention in psychosis access and waiting time standard: guidance*. London, Manchester: NHS England, NICE.
- NICE. (2014). *Psychosis and schizophrenia in adults: prevention and management. Clinical guideline*. Manchester: National Institute for Health and Care Excellence (NICE).
- Nikolova, S, Harrison, M, & Sutton, M. (2016). The Impact of Waiting Time on Health Gains from Surgery: Evidence from a National Patient-reported Outcome Dataset. *Health Economics*, 25(8), 955-968.
- Noel, G, Huchet, A, Feuvret, L, Maire, JP, Verrelle, P, Le Rhun, E, . . . Taillandier, L. (2012). Waiting times before initiation of radiotherapy might not affect outcomes for patients with glioblastoma: A French retrospective analysis of patients treated in the era of concomitant temozolomide and radiotherapy. *Journal of Neuro-Oncology*, 109(1), 167-175.
- Norman, RM, & Malla, AK. (2001). Duration of untreated psychosis: a critical examination of the concept and its importance. *Psychological Medicine*, 31(3), 381-400.
- Parabiaghi, A, Barbato, A, D'Avanzo, B, Erlicher, A, & Lora, A. (2005). Assessing reliable and clinically significant change on Health of the Nation Outcome Scales: Method for displaying longitudinal data. *Australian and New Zealand Journal of Psychiatry*, 39(8), 719-724.
- Pedersen, TJ, Bogh, LNB, & Lauritsen, JM. (2017). Improved functional outcome after hip fracture is associated with duration of rehabilitation, but not with waiting time for rehabilitation. *Danish Medical Journal*, 64 (4) (no pagination)(A5348).
- Penttilä, M, Jääskeläinen, E, Hirvonen, N, Isohanni, M, & Miettunen, J. (2014). Duration of untreated psychosis as predictor of long-term outcome in schizophrenia: systematic review and meta-analysis. *The British Journal of Psychiatry*, 205(2), 88-94.
- Pirkis, JE, Burgess, PM, Kirk, PK, Dodson, S, Coombs, TJ, & Williamson, MK. (2005). A review of the psychometric properties of the Health of the Nation Outcome Scales (HoNOS) family of measures. *Health and quality of life outcomes*, 3(76).
- Pregibon, D. (1980). Goodness of link tests for generalized linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(1), 15-14.
- Quintana, JM, Garcia, S, Bilbao, A, Navarro, G, Perea, E, De Larrea, NF, & Begiristain, JM. (2011). Waiting time for cataract extraction: Predictive factors and influence on outcomes. *Journal of Cataract and Refractive Surgery*, 37(1), 19-26.
- Ramsey, JB. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 31(2), 350-371.

- Rauchfuss, F, Zidan, A, Scheuerlein, H, Dittmar, Y, Bauschke, A, & Settmacher, U. (2013). Waiting time, not donor-risk-index, is a major determinant for beneficial outcome after liver transplantation in high-MELD patients. *Annals of Transplantation*, 18(1), 243-247.
- Register-Brown, K, & Hong, LE. (2014). Reliability and validity of methods for measuring the duration of untreated psychosis: a quantitative review and meta-analysis. *Schizophrenia Research*, 160(1-3), 20-26.
- Reininghaus, U, & Priebe, S. (2012). Measuring patient-reported outcomes in psychosis: conceptual and methodological review. *The British Journal of Psychiatry*, 201(4), 262-267.
- Revier, CJ, Reininghaus, U, Dutta, R, Fearon, P, Murray, RM, Doody, GA, . . . Jones, PB. (2015). Ten-year outcomes of first-episode psychoses in the MRC ÆSOP-10 study. *The Journal of Nervous and Mental Disease*, 203(5), 379-386.
- Sari, I, Acar, Z, Nurkalem, Z, Uslu, N, Davutoglu, V, Ates, M, . . . Aksoy, M. (2007). Preoperative clinical status but not waiting time predicts in-hospital outcomes of surgery in patients with left main coronary artery stenosis. *Tohoku Journal of Experimental Medicine*, 213(2), 173-180.
- Seidlitz, A, Siepmann, T, Lock, S, Juratli, T, Baumann, M, & Krause, M. (2015). Impact of waiting time after surgery and overall time of postoperative radiochemotherapy on treatment outcome in glioblastoma multiforme. *Radiation Oncology*, 10 (1) (no pagination)(172).
- Shapiro, SS, & Francia, RS. (1972). An approximate analysis of variance test for normality. *Journal of American Statistical Association*, 67, 215-216.
- Shapiro, SS, & Wilk, MB. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591-611.
- Siciliani, L, Moran, V, & Borowitz, M. (2014). Measuring and comparing health care waiting times in OECD countries. *Health Policy*, 118(3), 292-303.
- Singh, SP. (2007). Outcome measures in early psychosis: Relevance of duration of untreated psychosis. *The British Journal of Psychiatry*, 191(50), s58-s63.
- Sinko, A, Turner, A, Nikolova, S, & Sutton, M. (2016). *Models for waiting times in healthcare: comparative study using Scottish administrative data*. Working paper. University of Manchester: Institute of Health Sciences.
- Su, S, Li, S, Gao, L, Cai, Y, Fu, J, Guo, C, . . . Zhang, L. (2016). Gaps in the Continuum of HIV Care: Long Pretreatment Waiting Time between HIV Diagnosis and Antiretroviral Therapy Initiation Leads to Poor Treatment Adherence and Outcomes. *BioMed Research International*, 2016 (no pagination)(2648923).
- Tang, JY, Chang, WC, Hui, CL, Wong, GH, Chan, SK, Lee, EH, . . . Chen, EY. (2014). Prospective relationship between duration of untreated psychosis and 13-year clinical outcome: a first-episode psychosis study. *Schizophrenia Research*, 153(1-3), 1-8.
- Tasma, M, Liemburg, EJ, Knegtering, H, Delespaul, PAEG, Boonstra, A, & Castelein, S. (2017). Exploring the use of Routine Outcome Monitoring in the treatment of patients with a psychotic disorder. *European Psychiatry*, 42, 89-94.
- Tuominen, U, Sintonen, H, Hirvonen, J, Seitsalo, S, Paavolainen, P, Lehto, M, . . . Blom, M. (2009). The effect of waiting time on health and quality of life outcomes and costs of medication in hip replacement patients: a randomized clinical trial. *Osteoarthritis and Cartilage*, 17(9), 1144-1150.
- Tuominen, U, Sintonen, H, Hirvonen, J, Seitsalo, S, Paavolainen, P, Lehto, M, . . . Blom, M. (2010). Is longer waiting time for total knee replacement associated with health outcomes and medication costs? Randomized clinical trial. *Value in Health*, 13(8), 998-1004.
- Wing, JK, Beevor, AS, Curtis, RH, Park, SB, Hadden, S, & Burns, A. (1998). Health of the Nation Outcome Scales (HoNOS). Research and development. *The British Journal of Psychiatry*, 172(1), 11-18.
- Wing, JK, Curtis, RH, & Beevor, A. (1999). Health of the Nation Outcome Scales (HoNOS). Glossary for HoNOS score sheet. *The British Journal of Psychiatry*, 174(5), 432-434.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association*, 57(298), 348-368.

APPENDIX

Appendix A Health of the Nation Outcome Scale (HoNOS) items, sub-domains and total score

HoNOS item	HoNOS sub score	Total HoNOS score
1. Aggression	A: Behaviour	Total HoNOS score
2. Self-harm	(max. score = 12)	(max. score = 48)
3. Drug & alcohol use		
4. Cognitive problems	B: Impairment	
5. Physical illness & disability	(max. score = 8)	
6. Hallucinations & delusions	C: Symptoms	
7. Depression	(max. score = 12)	
8. Other symptoms		
9. Relationships	D: Social	
10. Activities in daily living	(max. score = 16)	
11. Residential environment		
12. Day-time activities		

Note: Based on Wing *et al.* (1998); Wing *et al.* (1999).

Appendix B Reliable and clinically significant change

For more details, we refer the interested reader to Jacobson *et al.* (1991) and Parabiaghi *et al.* (2005). A reliable and clinically significant change satisfies two criteria: (1) a clinically significant change would move a person from a score typical of the “dysfunctional” population to a score typical of the “functional” population, and (2) a reliable change is beyond what could be attributed to measurement error or chance.

(1) Clinically significant change:

Patients were defined as dysfunctional if they had a score of ≥ 3 in at least two of the 12 items. All others made up the functional population. The cut-off point where the chance of belonging to either population (dysfunctional or functional) is equal is calculated as follows:

$$cut-off = \frac{(mean_{dysfunc} \times SD_{func}) + (mean_{func} \times SD_{dysfunc})}{(SD_{func} \times SD_{dysfunc})} = \frac{(17.9 \times 4.5) + (8.9 \times 5.5)}{(4.5 + 5.5)} = 13.0$$

A change of at least 13 score points in HoNOS was considered as clinically significant.

(2) Reliable change

We calculated a reliable change index (RC_{index}) based on the baseline HoNOS score:

$$RC_{index} = 1.96 \times SE_{diff} = 1.96 \times 5.2 = 10.2$$

where SE_{diff} is the standard error of measurement of a difference:

$$SE_{diff} = SD_1 \times \sqrt{2} \times \sqrt{1-\alpha} = 6.8 \times \sqrt{2} \times \sqrt{1-0.71} = 5.2$$

SD_1 is the standard deviation of the baseline score and α is Cronbach's coefficient. A change of at least 10 score points in HoNOS was considered as reliable.

Based on this, we classified patients with a given HoNOS at baseline ($score_1$) and follow-up ($score_2$) as improved if $score_2 \geq RC_{improv}$, where $RC_{improv} = score_1 + RC_{index}$ and deteriorated if $score_2 \leq RC_{deter}$, where $RC_{deter} = score_1 - RC_{index}$. In the study sample, 77.4% were classified as stable, 18.7% improved, and 3.9% deteriorated.

Appendix C Generalised Linear Model diagnostics - dependent variable: total HoNOS score at follow-up

Test for normality of follow-up HoNOS					
	Obs	W	V	z	Prob>z
Shapiro-Wilk test	8,949	0.98	103.33	12.38	0.000
Shapiro-Francia test	8,949	0.98	109.33	12.32	0.000

Within/between provider variance in follow-up HoNOS					
	Mean	SD	Min	Max	Obs
Overall variance	50.01	74.1	0	365	N = 8,949
between provider variance		29.1	0	133.0	n = 48
within provider variance		69.5	-82.9	380.7	T-bar = 186.44

GLM model diagnostics		based on model with independent variable: log waiting time			
Log pseudolikelihood	-36006				
Squared correlation btw. y and yhat	0.183				
Linktest yhat	P> t = 0.000				
Linktest yhat squared	P> t = 0.051				
Hosmer-Lemeshow test	F(10, 8939) = 0.42; Prob>F = 0.9385				
Ramsey RESET test	chi2(1) = 5.78; Prob>chi2 = 0.0162				
Park test	Gaussian	chi2(1) = 26.12; Prob>chi2 = 0.0000			
	Poisson	chi2(1) = 2.20; Prob>chi2 = 0.1376			
	Gamma	chi2(1) = 4.59; Prob>chi2 = 0.0322			
	Inverse Gaussian	chi2(1) = 33.26; Prob>chi2 = 0.0000			

Appendix D Seemingly Unrelated Regression results - dependent variable: HoNOS sub score at follow-up

	(1) Behaviour		(2) Impairment		(3) Symptoms		(4) Social	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Log waiting time (continuous)	0.040**	(0.014)	0.025**	(0.009)	0.078***	(0.018)	0.061**	(0.021)
Waiting time quintiles (ref.cat.: shortest quintile)								
2nd shortest quintile	-0.084	(0.076)	-0.084	(0.049)	-0.044	(0.100)	-0.157	(0.117)
3rd shortest quintile	-0.111	(0.076)	-0.009	(0.048)	-0.060	(0.099)	-0.218	(0.117)
4th shortest quintile	0.055	(0.077)	0.076	(0.049)	0.068	(0.100)	0.004	(0.118)
Longest quintile	0.145	(0.077)	0.053	(0.049)	0.341***	(0.101)	0.198	(0.118)
Waiting time intervals (ref.cat.: less than 0.5 months)								
Waiting time 0.5 to 3 months	0.066	(0.048)	0.089**	(0.031)	0.074	(0.063)	0.045	(0.074)
Waiting time 3 to 6 months	0.146*	(0.074)	0.078	(0.047)	0.295**	(0.096)	0.259*	(0.114)
Waiting time 6 to 12 months	0.314***	(0.085)	0.154**	(0.054)	0.516***	(0.111)	0.419**	(0.130)
Observations	8,949		8,949		8,949		8,949	
Provider and year fixed effects	yes		yes		yes		yes	
Covariates	yes		yes		yes		yes	

Note: * p<0.05, ** p<0.01, *** p<0.001. The dependent variable is in model (1) the HoNOS behaviour sub score (items 1-3), in model (2) the HoNOS impairment sub score (items 4-5), in model (3) the HoNOS symptoms sub score (items 6-8), in model (4) the HoNOS social sub score (items 9-12).

Appendix E Ordered probit estimation results - dependent variable: reliable and clinically significant change in total HoNOS score

	(1) Study sample		(2) Long EIP		(3) Short EIP	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
Log waiting time (continuous)	0.033**	(0.012)	0.042**	(0.014)	0.029	(0.020)
Waiting time quintiles (ref.cat.: shortest quintile)						
2nd shortest quintile	-0.066	(0.061)	-0.052	(0.080)	-0.083	(0.109)
3rd shortest quintile	-0.117*	(0.058)	-0.042	(0.075)	-0.312**	(0.108)
4th shortest quintile	-0.012	(0.054)	-0.020	(0.063)	0.080	(0.127)
Longest quintile	0.137*	(0.064)	0.220**	(0.083)	0.042	(0.101)
Waiting time intervals (ref.cat.: less than 0.5 months)						
Waiting time 0.5 to 3 months	-0.023	(0.040)	-0.024	(0.046)	0.017	(0.069)
Waiting time 3 to 6 months	0.122	(0.063)	0.216*	(0.084)	0.048	(0.091)
Waiting time 6 to 12 months	0.253***	(0.077)	0.290**	(0.096)	0.239*	(0.116)
Observations	8,949		6,135		2,814	
Provider and year fixed effects	yes		yes		yes	
Covariates	yes		yes		yes	

Note: * p<0.05, ** p<0.01, *** p<0.001. Models (1) - (3) are ordered probit models with the dependent variable defined as: 1 = clinical improvement, 2 = clinical stability, 3 = clinical deterioration. Model (1) includes the complete study sample. Model (2) includes only patients that were in EIP care for the whole duration of follow-up. Model (3) includes only patients with an EIP episode shorter than the 12-months follow-up. All models use cluster robust standard errors for 48 provider clusters.