The economics of waiting time in mental health

Anika Kreutzberg

Ph.D.

University of York
Economics and Related Studies

November 2018

Abstract

This thesis empirically explores aspects of the economics of waiting times in the mental health context. Waiting times are of persistent policy concern as they risk poorer treatment outcomes and threaten the desired principles of timely and equitable access. The empirical applications focus on first-episode psychosis patients and early intervention in psychosis (EIP) services in the English National Health Service where policymakers recently placed new emphasis on reducing waiting times.

Analyses are based on the nationally representative Mental Health and Learning Disabilities Dataset 2011 to 2015. We develop procedures to measure various dimensions of waiting times at the patient level – such as duration of untreated psychosis, inpatient waiting time, and referral-to-treatment waiting time. We apply generalised linear modelling to accommodate the heavy-tailed distribution of waiting time and use duration analysis to overcome the challenge of right-censoring. We further make use of difference-in-difference and matching techniques to evaluate the impact of the newly introduced EIP waiting time target.

We found significant socioeconomic inequalities in duration of untreated psychosis. Also, hallucinations and delusions, as well as previous mental health service use, were influencing factors for patients to access services. Waiting for a care coordinator was associated with a clinically significant deterioration in patient outcomes independent of treatment intensity. The implementation of the EIP waiting time target led to an increased probability of waiting below target whereas waiting times along the distribution did not improve. However, waiting times improved already in anticipation of the policy change with little evidence of unintended effects such as re-prioritisation of patients or gaming behaviour of providers. Results of this thesis can help to inform the development of strategies to reduce inequalities in access to EIP services, and the prospective implementation of waiting time targets in other mental health service areas as well as its adaptation to other countries.

Contents

List of figures

List of tables

Acknowledgements

First of all, I thank my supervisor, Professor Rowena Jacobs, for her time, dedication and belief in myself and my research. She let me grow as an independent researcher offering advice and guidance when needed. She made me feel valued when self-doubts were dreading my motivation. Thank you, Rowena!

Next, I would like to thank Professor Luigi Siciliani and Professor Nigel Rice as members of my Thesis Advisory Panel especially for providing methodological expertise in fruitful discussions. They continually challenged me to think beyond my methodological comfort zone which made me grow more than I would have without. Thank you, Nigel and Luigi!

Further, I am grateful not only for the financial support I was given in the form of a studentship but also for the opportunity and resources I received from the Centre for Health Economics (CHE) in York. CHE provided an excellent work and social environment to pursue my PhD. I always found a welcoming open door when seeking advice from colleagues. I want to particularly point out the valuable support I received from Maria Jose Aragon and Rita Santos.

My Ph.D. fellows made life in- and outside the office enjoyable!

Most importantly, I thank my husband Leo for his love, support, and encouragement. I would neither have started, nor continued, not to mention successfully finished this study without his reinforcement and patience.

Author's declaration

I declare that this thesis is my own original work and has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. I am the sole author of chapters 1 and 6. Chapters 2 to 5 are co-authored with Professor Rowena Jacobs from the Centre for Health Economics, University of York. In all four co-authored chapters, I am the lead author as I substantially contributed by developing the research idea, conceiving and designing the methodological approaches, cleaning and analysing the data, interpreting the results, and drafting early and final versions.

An earlier version of the research contained in chapter 2 was presented at the Health Economics Study Group meeting at the University of Birmingham in January 2017. The chapter has been published as a peer-reviewed paper entitled 'Socioeconomic inequalities in duration of untreated psychosis: Evidence from administrative data in England' in *Psychological Medicine,* 48(5), 822-833 in 2018. Previous versions of the research of chapter 3 were presented at the Health Econometrics and Data Group seminar at the University of York in March 2017 and at the RGS PhD conference in Economics at the Universität Duisburg-Essen in February 2018. Chapter 3 further has been published as a peer-reviewed paper entitled 'The impact of waiting time on patient outcomes: Evidence from Early Intervention in Psychosis services in England' in *Health Economics,* 2018(27), 1772-1787 in 2018. Research based on chapter 4 was presented at the Health Economics Study Group meeting at City University in London in January 2018. Previous versions of chapter 5 were presented at the Workshop on the Economics of Mental Health Policy in Rome organised by Monash University as well as at the workshop of the German Health Economics Study Group on "Access and Accessibility in Health Care Services" at the Technische Universität Berlin both in September 2018. A revised chapter version was accepted for a poster presentation at the Health Economics Study Group meeting at the University of York in January 2019.

The main dataset used in chapters 2 to 5 is the Mental Health and Learning Disabilities Dataset (MHLDDS).[1] Under a Data Sharing Agreement with NHS Digital the MHLDDS is released on condition that it is not shared with any third party. Copyright © 2011/2012 - 2015/16, re-used with the permission of NHS Digital. All rights reserved. As we are using secondary data no ethical approval was required. All data analyses have been conducted using Stata 14.0 unless stated differently.

[1] Formerly known as Mental Health Minimum Dataset (2000-2013) – we consistently use the term MHLDDS to refer to either of the versions.

"Far too many people of all ages wait too long to get the mental health services they need. The longer they wait for support, the more likely it is their condition gets worse. This has to change. […]. Simply making services available is not enough. We are also looking at ways to overcome inequalities around service usage – and around the outcomes those services achieve."

Department of Health (2014a), pp.12-13

1. Chapter: Introduction

1.1 Research objective

Providing access to services for people in need of care is a key perspective for health systems around the world (Willcox et al., 2007). Hence, waiting times are of persistent policy concern in countries with publicly funded health care systems and universal access such as the United Kingdom, Canada, New Zealand, or Australia (Siciliani, Moran and Borowitz, 2014; Cullis, Jones and Propper, 2000). Waiting lists can be used to stock available demand and optimise utilisation of the scarce supply of resources such as skilled staff and medical equipment (Culyer and Cullis, 1976). However, excessively long waiting times do not only risk poorer patient outcomes they also create anxiety and disability during waiting (Propper, 1995, 1990; Lindsay and Feigenbaum, 1984) and threaten the desired principles of timely and equitable access to care (Oudhoff et al., 2007).

This thesis empirically explores aspects on the economics of waiting times in the mental health context. Recently, new policy emphasis has been placed on reducing waiting times in mental health services in England. With its strategy "No health without mental health" published in February 2011, the Department of Health acknowledged the often neglected importance of mental health for individuals, society and the economy (Department of Health, 2012, 2011). With its aim to ensure mental health having parity of esteem with physical health, some of the key objectives of this strategy were to improve access to specialist services, promote early intervention for people with severe mental illness, and to reduce inequalities in accessing services. The key objectives were later specified by the announcement of the first ever waiting time targets in mental health services in January 2014 (Department of Health, 2014a).

This thesis aims to investigate to what extent the governmental objectives regarding shorter waiting times and more equal access to specialist mental health services have been met after the first four years of its announcement. The empirical applications focus on first-episode psychosis patients within the English National Health Service (NHS) which is one of the key areas the waiting time target policy focused on. On the demand side, we will analyse inequalities in accessing specialist services and explore the impact of waiting on treatment outcomes. On the supply side, we will provide answers to whether the increased policy focus on reducing waiting times led to any

intended or unintended changes in provider behaviour. We look into provider responses to enforced waiting time targets and analyse changes in prioritisation patterns, and gaming behaviour.

1.2 Background

1.2.1 Clinical and institutional background

Psychotic disorders, including schizophrenia, encompass a group of conditions that are considered to be among the most serious mental illnesses (Iyer et al., 2015). A psychosis is characterised by disturbances in thinking and perception which manifests in positive symptoms such as delusions, hallucinations, and disorganised thoughts and behaviour as well as negative symptoms such as monotone speech, lowered levels of motivation, lack of interest in social interaction, or inability to feel pleasure (NICE, 2014; Andreasen, 1984, 1983). In 2016, there were 20.9 million prevalent cases of schizophrenia worldwide (Vos et al., 2017). The estimated lifetime median prevalence of psychotic disorders is 4.0 per 1,000 persons (McGrath et al., 2008). The pooled incidence of all psychosis in England is 31.7 per 100,000 person-years (15.2 for schizophrenia) (Kirkbride et al., 2012a). Incidence varies markedly by demographic as well as neighbourhood related factors. For example, men have an almost twofold greater risk than woman (McGrath et al., 2008; Kirkbride et al., 2006). Also, migrants were found to have a more than four-fold higher incidence compared to native-born people (McGrath et al., 2008) and incidence is two- to three-fold higher in most deprived or socially fragmented neighbourhoods (O'Donoghue et al., 2016).

The economic and social consequences of psychosis are tremendous. Once a psychosis has emerged, the majority of patients suffers from repeating psychotic episodes over their lifetime. Only 44% will symptomatically recover within eight years (Revier et al., 2015). Psychosis is closely linked to poorer physical health and in consequence shorter life expectancy (Naylor et al., 2012; Thornicroft, 2011; McGrath et al., 2008; Henderson et al., 2000). It affects people during late adolescent and early adult years when they are just about to pursue further education or employment. At the same time, patients face lower rates of employment, poorer pay and less secure jobs throughout their lifetime (Revier et al., 2015; Marwaha and Johnson, 2004). The often unusual or

bizarre behaviour leads to difficulties in managing their own life and social exclusion (Revier et al., 2015; Huxley and Thornicroft, 2003). Fear of stigma and reluctance prevents patients from seeking help (Dawson, Jordan and Attard, 2013). Worldwide, the annual national costs for schizophrenia are estimated to range from US$94 to US$102 billion (Chong et al., 2016). In England, schizophrenia causes total cost of £8.8 billion per year with service costs contributing 40%, and informal care 13% (Kirkbride et al., 2012b).

The first two to five years following the onset of psychosis are referred to as "first-episode psychosis" (Breitborde, Srihari and Woods, 2009). Here is where the majority of the decline in functioning emerges and treatment response is highest. There is compelling scientific evidence that the early phase of a psychosis is critical and delays in first treatment are robustly linked to poorer outcomes (Birchwood, Todd and Jackson, 1998). Based on this idea, specialised early intervention in psychosis (EIP) services have been developed in many countries (Marshall and Rathbone, 2006; Joseph and Birchwood, 2005). EIP services provide multidisciplinary care including pharmacological, psychological, social, occupational and educational interventions (Cheng and Schepp, 2016; NICE, 2014). Treatment is delivered by stand-alone services within the community in planned sessions over a period of two to five years (Malla et al., 2017). Key components of EIP care should comprise easy and rapid access to services, integrated biopsychosocial care plans, a multidisciplinary team including a psychiatrist, treatment of comorbidities, and formal evaluation of care quality (Addington et al., 2013). Given the multidisciplinary nature, the care coordinator plays a key role in the effective delivery of care (Iyer et al., 2015). She brings together all involved professionals and is responsible for engaging and supporting patients in treatment. Treatment engagement is a particular challenge in first-episode psychosis as about 30% of patients disengage from treatment over time (Doyle et al., 2014).

Specialised EIP care during first-episode psychosis effectively improves outcome prospects through various channels (Fusar-Poli, McGorry and Kane, 2017): (1) reduced time from onset of symptoms until first treatment; (2) improved treatment response; (3) improved well-being including caring relatives; (4) treatment of comorbidities; (5) secondary prevention of illness progression. Furthermore, the cost-

effectiveness of EIP care has been demonstrated (Behan et al., 2015; Valmaggia et al., 2015; Hastrup et al., 2013). In England, specialist mental health services such as EIP care are provided by mental health trusts. There are just over 50 mental health care trusts within the English NHS. Each trust covers a certain geographical area with a number of inpatient wards as well as community-based service teams. In the following, we will refer to mental health trusts as providers.

1.2.2 Policy context

Specialised EIP services have experienced great policy interest in the past two decades. But although they are well-established in many countries, availability of services and implementation standards are very heterogeneous between countries (Nordentoft and Albert, 2017) and access to services is falling behind its expectations (Anderson et al., 2018). Surveys in countries such as Australia (Catts et al., 2010), Canada (Nolin et al., 2016), the United States (Breitborde and Moe, 2017), and Italy (Cocchi et al., 2018) report insufficient progress in the nationwide implementation of EIP care. Service availability is mainly impeded by the lack of governmental funding, and explicit standards EIP services will be operating from. EIP provision is highly influenced by the local EIP network, including clinicians and front-line staff in mental health services that support the EIP idea (Cheng, Dewa and Goering, 2011).

England is one of the few countries which has had a nationwide EIP implementation strategy from the early 2000s onwards. In 1999, the English government issued a National Service Framework for Mental Health which initiated the implementation of functional community psychiatric services including EIP teams (Joseph and Birchwood, 2005). By December 2004, fifty EIP services were expected to be implemented nationwide with the aim of providing access to specialist care for every young person with first-episode psychosis for the first three years of their illness. The strategy was accompanied with a considerable amount of funding and a detailed guidance about how EIP services were expected to be specified and operated. However, after initial funding, EIP provision began to decline and waiting times were increasing (Kirkbride et al., 2017; Marwaha et al., 2016). Against this backdrop, the English government introduced the first waiting time target in mental health history (Department of Health, 2014a). From April 2015, 50% of patients being referred to an

EIP service were expected to wait no longer than 14 days from referral to treatment (NHS England, 2015). Referrals may come from any internal or external source (e.g. other mental health service, inpatient ward, prison, general practitioner, school, or self-referral). A £40 million funding package was provided to support its implementation. The target is planned to be raised to 60% by 2020/21 and within the following years, all mental health services shall be affected by a similar target. Since 2000, England continuously focused on setting maximum waiting time targets in order to improve access to various areas of health care such as first outpatient appointments, elective inpatient treatment or routine cardiac surgery (Willcox et al., 2007). Often combined with substantial penalties for failing providers, they have been shown to be effective in reducing waiting times (Besley, Bevan and Burchardi, 2009; Propper et al., 2008; Bevan and Hood, 2006). To date it is however unknown whether a comparable target within the mental health context can be similarly effective.

1.2.3 Theoretical framework

Queuing theory provides an overall theoretical framework of why waiting times are prevalent in publicly funded health systems with zero prices. Queues for health care emerge for three main reasons (Culyer and Cullis, 1976): (1) the price of health care is below the market clearing level and thus causes an excess in demand; (2) demand for health care is stochastic; (3) patients have preferences about when to be treated. Queues can serve as a rationing instrument as the price for health care becomes the opportunity cost of time spent waiting (Martin and Smith, 1999). However, waiting time as discussed in this thesis does not require patients to queue in person. Once referred to a specialist service, one is free to use the time of waiting for anything else without the costs of wasted time. Lindsay and Feigenbaum (1984) developed the theory of queuing by list which explains why waiting times impose costs to patients even without wasted time. The model introduces an exponential demand decay rate which reflects the fact that a good or service received later is worth less today. When waiting for health care, patients may experience pain, anxiety, disability, and restrictions in their daily activities. A patient's health status may deteriorate during the time of waiting and will make the awaited treatment less likely to be successful - particularly if waiting affects the patient's ability to benefit from treatment. The

negative impact of waiting time can be also long-term as the deteriorated condition of the patient due to waiting may take longer to recover or will not be reversed at all after a critical waiting time has passed (Koopmanschap et al., 2005). Given the fact that waiting time imposes deadweight losses, it can be welfare improving to impose different waiting times for different patient groups (Gravelle and Siciliani, 2009; Gravelle and Siciliani, 2008). Which groups to prioritise depends on where the total cost of waiting will be reduced through the prioritisation.

A second theoretical foundation for this thesis builds on the principal-agent economic model in the presence of asymmetric information. It motivates the analysis of provider responses to performance targets. The policymaker (principal) wishes to maximise some welfare function that depends on an unobserved health outcome which can be influenced by the provider (agent)'s level of effort. Due to asymmetric information, the policymaker can only imperfectly observe the provider (agent)'s effort to achieve the unknown health outcome (Goddard, Mannion and Smith, 2000). The waiting time target serves as a quantifiable measure to approximate the provider's performance. Target performance is linked to some kind of financial or non-financial reward (or penalty) which incentivises the provider to achieve a good target performance. However, providers act in a complex environment with multiple stakeholders, budgets, and objectives that need to be served (Besley and Ghatak, 2003). Being under pressure to meet different objectives with limited resources, the provider may also take unintended actions to improve target performance at the expense of worse performance in non-targeted areas (Smith, 1995).

1.3 Thesis structure

The remainder of this thesis is structured into four main chapters and a concluding chapter. Chapter 2 investigates whether there exist inequalities in duration of untreated psychosis associated with socioeconomic deprivation in a national cohort of first-episode psychosis patients in England. Chapter 3 contributes to the still ongoing debate whether longer waiting time for treatment leads to poorer health outcomes. We explore the impact of waiting time for specialist treatment on HoNOS outcomes for EIP patients in England. Chapters 4 and 5 focus on provider responses to the newly introduced EIP waiting time target. In chapter 4, we explore providers' changes in

behaviour in anticipation of the EIP target policy. Looking at the years prior to the policy, we examine whether the public announcement of the policy was associated with changes in waiting time as well as unintended changes in prioritisation patterns. We further explore whether providers tried to game the target by making performance look better than it actually was. Research in chapter 5 aims to measure the causal effect of the EIP target policy in reducing referral-to-treatment waiting time after the first six months of its implementation using a difference-in-difference design. To validate our approach, we also test whether there have been any spill-over effects from the targeted EIP services towards the non-targeted standard community mental health teams.

1.4 Data sources, main measures, and methods

1.4.1 Mental Health and Learning Disabilities Dataset (MHLDDS)

All research presented in this thesis uses the administrative Mental Health and Learning Disabilities Dataset (MHLDDS) as the main data source. The MHLDDS contains patient-level data on any mental health related treatment in hospitals and community settings within the English NHS (NHS Digital, 2017). Its first version, named Mental Health Minimum Dataset, was introduced in 2000 and became mandatory in 2003 for all providers of specialist mental health services funded by the NHS. From September 2014, the dataset was renamed Mental Health and Learning Disabilities Dataset following the inclusion of people in contact with learning disability services. Data collection paused in November 2015 in order to introduce a new version, the Mental Health Services Dataset, from April 2016. For the purpose of this thesis, data were available from April 2011 to November 2015. In the following we consistently use the name MHLDDS to refer to any of the corresponding versions.

A main measure of disease severity within the MHLDDS is the Health of the Nations Outcomes Scale (HoNOS) (Wing, Curtis and Beevor, 1999; Wing et al., 1998). We will discuss HoNOS in more detail in chapter 3 where we use it as our main outcome measure to estimate the effect of waiting time on treatment outcomes. Next to the total HoNOS score, the HoNOS item 6 is of particular interest for our analyses as it measures problems with hallucinations and delusions. As such it allows us to measure symptom severity for psychotic patients. Moreover, the MHLDDS provides rich

information on patient demographics which we use to control for patient case mix in all chapters. To enrich information on the patient's socioeconomic background we use the lower super output area (LSOA) of the patient's place of residence in our data and link it to the Index of Multiple Deprivation (IMD) (McLennan et al., 2011). More detail will be provided in chapter 2 where the IMD is our main explanatory variable to explain socioeconomic inequalities in waiting time. In chapter 5, we further use the patient's LSOA to calculate travel distances from the patient's place of residence to the nearest EIP site postcode. Provider information within the MHLDDS is at the trust-level with no information about the local community service site the patient was receiving care from. Therefore, we create a novel dataset including postcode information of all EIP sites within the English NHS. We will present the procedures of the data collection in chapter 5 where this information is used.

1.4.2 Measuring and analysing waiting time

Waiting time can be measured in various ways (Siciliani, Moran and Borowitz, 2014; Godden and Pollock, 2009). First, waiting time needs to be distinguished from the concept of waiting lists often published in official waiting time statistics. Waiting lists represent the number of patients waiting at a given point in time. As such they reflect a cross-sectional snapshot rather than the patient's full waiting experience. Waiting time, in contrast, is longitudinal in nature and measures the time the patient spends on the list from being added until the start of treatment. It can only be generated retrospectively once a patient has finished waiting.

Second, patients may face several waiting times during their care pathway which is illustrated for a first-episode psychosis patient in Figure 1.1. The most comprehensive waiting time concept specific to first-episode psychosis is the duration of untreated psychosis (DUP). DUP measures the time from the first onset of symptoms (1) to the start of treatment (4) thus it includes the period of help-seeking where no service contact has happened yet (Norman and Malla, 2001). A large body of literature mainly located in psychiatric outlets exists on measuring DUP, its predictors and its impact on patient outcomes (Penttilä et al., 2014). DUP is considered to be one of the key parameters in managing patients with psychosis. But is has also been criticised as the operative measurement varies widely across studies (Register-Brown and Hong,

2014). The date of onset needs to be reconstructed retrospectively and relies on the self-report of the patients or carers which may in turn be related to the patient's severity of illness.

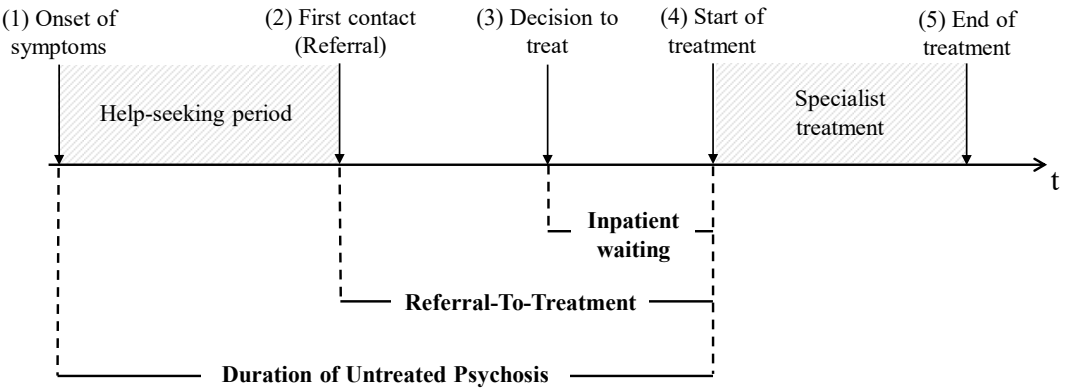


Figure 1.1: Different waiting times along the patient pathway

In the health economic literature, inpatient waiting time is most commonly used which measures the time from the specialist's decision to treat (3) to the start of (inpatient) treatment (4) (Siciliani, Moran and Borowitz, 2014). It has been particularly applied in the investigation of the relationship between waiting time and treatment outcomes (Moscelli, Siciliani and Tonei, 2016; Nikolova, Harrison and Sutton, 2016). The measure's advantage is that it is quite easy to derive when there are administered waiting lists that contain information on when the patient was put on the list and when treatment started. However, it may miss important parts of the patient's waiting experience – particularly if the inpatient treatment only follows a range of other primary care and outpatient treatments in the community as in the case of first-episode psychosis. Referral-to-treatment waiting time, therefore, considers the additional waiting time from referral (2) to the first specialist assessment (3) and adds it to the inpatient wait. This concept has gained increased attention by policy makers since the perception that an emphasis on inpatient waits may come at the cost of longer waits for a specialist assessment (Marques et al., 2014; Kelman and Friedman, 2009). Its measurement requires, however, to relate a referral to the treatment the referral is directed to. This challenges most existing data structures as referrals will be likely to happen in a different care sector (e.g. primary care or social care agency) then the treatment (e.g. secondary mental health service).

To date there exists no waiting time concept that is superior to all others. Each measure comes with its strengths and limitations and has to be chosen in accordance with the purpose of the analysis. This research aims to investigate a broad range of different waiting time concepts to overcome each single measure's limitations and provide different perspectives of the concept. None of these measures can be observed directly in our data. However, we develop methods to estimate waiting times based on the procedures outlined below. Table 1.1 provides an overview of how we defined the relevant measurement points along the patient pathway based on Figure 1.1 in our data.

The MHLDDS contains information on a patient's first-episode psychosis which we make use of to calculate DUP in chapter 2. We use the emergent date of the first-episode psychosis as start point and the start of antipsychotic medication as end point of our DUP measure. In chapter 3, we use a variation of inpatient waiting time. We refer to it as specialist waiting time as the service to be waited for is a specialist treatment (EIP care) within the community rather than inpatient treatment. We measure the time from the start of EIP care to the assignment of a care coordinator. As outlined before, the care coordinator plays a key role in the delivery of EIP care and is responsible for implementing a care plan concordant with the treatment guidelines of the National Institute for Health and Care Excellence (NICE). Hence, we assume that once a care coordinator is assigned the patient will receive a relevant care package. This same endpoint is also used in chapters 4 and 5 were we estimate referral-to-treatment waiting times in accordance with the EIP waiting time target policy. The difference is the start point which requires one to identify the relevant referral. Referrals within the MHLDDS cannot be directly linked to the service they were directed to. We used a number of measures to identify the referral directed to the relevant EIP episode. The MHLDDS defines care spells which are overarching and continuous periods of time a patient spent in the care of a single or multiple healthcare providers (Monitor, 2015). We considered all care spells that started within the study period and where the patient's first team episode was with an EIP service. We identified referrals that initiated the care spell (i.e. happened before the start of the spell). Referrals could have been received from multiple sources, including primary and secondary care providers, other tertiary mental health or social care providers, agencies within the justice system and self-referrals. We considered only referrals that

were accepted by the receiving provider. If there were multiple accepted referrals before the start of a care spell, we used the referral closest to the start of the care spell. We defined acceptance onto the EIP caseload as the start of the EIP team episode which initiated the care spell. The first care coordinator the patient was assigned to following the EIP acceptance stopped the waiting time clock.

Table 1.1: Measurement point definitions to estimate waiting times in the thesis

Measurement points	Definition in thesis	MHLDDS variable used
(1) Onset of symptoms	emergent date of psychosis	EMERPSYCHDATE
(2) Referral	date of referral	REFRECDATE
(3) Decision to treat	start of EIP care episode	STARTDATE if CLINTEAM = "A14"
(4) Start of treatment	date of care coordinator assignment	STARTDATE if EPITYPE = "CCASS"

Note: A14 is the code for the early intervention in psychosis team. CCASS refers to an episode of an assigned care coordinator.

Econometric analysis is challenged when using waiting time as variable of interest for two reasons. First, waiting time has typical count data properties. It takes only positive integer values and is usually skewed heavily to the right as it contains a large proportion of zero or very short waiting patients but an extensive right tail of a few very long waiters (Jones et al., 2013). One way to solve this is to bin waiting time in several categories such as short (< 6 months) and long waiters (> 6 months). Thresholds for these categories are, however, often chosen arbitrarily. Another common approach is to use transformations of waiting time (e.g. logarithm). In consequence, interpretation of results can no longer be made on the original scale. More recently, count data models have been used in the economic literature. For example, generalised linear models (GLM) have gained increased attention as they offer a high degree of flexibility in modelling mean and variance functions (Deb and Norton, 2018; Jones et al., 2016; Sinko et al., 2016). At the same time, they allow the interpretation of the dependent variable on its natural scale. We use GLM in chapters 2 and 3 to exploit its advantages for our analysis. Second, waiting time can only be measured retrospectively, i.e. once a patient finished waiting (or started treatment). This means that patients still waiting at the end of the study period have to be excluded

in conventional regression analysis. Duration analysis can help to overcome this challenge as it accounts for the time in the study and prevents the need to truncate the sample (Rabe-Hesketh and Skrondal, 2012). We make use of duration analysis techniques in chapter 4 for these reasons.

1.4.3 Limitations of the available data sources

There are two limitations of the MHLDDS which shall be discussed here as they are relevant to the analyses in all of the following chapters. The first limitation is the identification of first-episode psychosis patients relevant to the study. Typically, when working with administrative data, relevant patients can be identified by their primary diagnosis which is based on the International Statistical Classification of Diseases and Related Health Problems code in its 10th version (ICD-10). In the mental health context however, diagnosis information is less informative. First, diagnosing a psychosis is complex and requires a long process of investigation and mutual exclusion of competing conditions (American Psychiatric Association, 2013; Williams and Doessel, 2001). Hence, a confirmed diagnosis may only be recorded once treatment commenced already. Second, a severe mental health diagnosis is still attached with considerable stigmatisation which is why clinicians tend to not diagnose patients to prevent them from "being labelled as insane" (Wykes and Callard, 2010). Third, the concept of first-episode psychosis is still lacking a precise definition and studies find their own ways to operationalise the concept (Breitborde, Srihari and Woods, 2009). In consequence, we apply a number of selection criteria used in previous studies to identify relevant patients (Kirkbride et al., 2017; Tsiachristas et al., 2016): (1) schizophrenia diagnosis (ICD-10 F.20-F.29); (2) received treatment by an EIP team; (3) reported problems with hallucinations and delusions (HoNOS item 6 > 0), or (4) allocated to the first-episode psychosis mental health care cluster[2]. The detailed methods of the sample selection will be presented in each chapter respectively.

[2] Mental healthcare clusters are reference groups used to group service users with similar needs and problem severities related to their mental health within the English NHS. There are 20 clusters each of which describes a particular type, combination and severity of needs. Similarly, to the idea of health resource groups (HRGs) in the acute physical health context, mental health care clusters can be combined with a fixed price to form a system of prospective payment for mental health services (Department of Health, 2013). However, only a handful of providers have adopted this payment system to date (Jacobs et al., 2018).

The second limitation to be highlighted is related to data completeness and data quality in the MHLDDS. Data quality has improved considerably over the first years of data collection. However, not all variables of interest for us are mandatory for providers to report and hence data missingness is an issue. For categorical explanatory variables in our regression models, we used the missing-group method suggested by Cohen and Cohen (1975) to deal with missing information. That is, we created a separate group for patients with missing values. This approach allows us to use all available information and avoids the risk of selection bias by dropping subjects for which missing information is not random as well as lower power by the considerable reduction in sample size. In cases where information was missing on the main explanatory variable of interest, we followed the complete-case approach by dropping patients with missing values from the analysis: In chapter 2, we excluded patients with missing information regarding the onset of psychosis date variables which were required to measure DUP. In chapter 3, patients with a missing HoNOS rating at the end of the 12-month follow-up had to be excluded. We discuss the potential impact of these procedures on our results in the corresponding chapters. In general, missing data may be indicative of the quality of provider coding practices which in turn may be associated with the provider's performance regarding patient waiting times and outcomes. Waiting times in return are also known to vary across providers due to supply factors such as the availability of staff, infrastructure, and other resources as well as regional demand factors such as overall health of the population. We therefore control for provider heterogeneity in all our models using either provider fixed effects (chapter 2, 3, and 5) or stratification by provider (chapter 4).

1.5 Contribution to research

This thesis makes substantial contributions to the existing evidence base in a number of ways. It provides the first comprehensive and rigorous analysis of waiting times for nationally representative first-episode psychosis patient samples in England using large administrative data. The policy maker's neglect of the importance of mental health over many years is also mirrored in a lack of empirical evidence in the health economic literature. To date the majority of studies analysing aspects of access to care for people with mental illnesses have a more clinical or epidemiological focus and

often rely on simple correlations without controlling for any confounding. Waiting time is mostly just a side aspect rather than the main focus of research.

The lack of health economic studies in the area of mental health is closely linked to the lack of available data. Most studies in the psychotic literature are based on observational data derived from epidemiological field cohorts such as the National EDEN prospective cohort study (Marshall et al., 2014) or the Aetiology and Ethnicity of Schizophrenia and Other Psychoses (AESOP-10) study (Morgan et al., 2014). Cohort data rely on small sample sizes in a limited geographical catchment area and a limited number of providers. Most cohort studies are correlational only without aiming for causal inferences. This research overcomes this problem by using a unique, large, administrative dataset. It not only allows us to build a nationally representative patient and provider sample, but it also helps us to appropriately risk adjust in regression analyses thanks to the rich information of disease-related as well as demographic patient variables. Unlike most other administrative datasets, we are able to use information on both inpatient care as well as care taking place within the community. Since the main focus of EIP care is to keep patients within the community and prevent costly inpatient admissions, this is crucial to our analysis.

This thesis further presents the first attempts to measure waiting times for first-episode psychosis patients using the MHLDDS. The dataset in its current version is not designed to directly derive waiting times which is one more reason why evidence in this area is scarce. We develop procedures to estimate waiting times in a number of ways and critically appraise each measure's advantages and limitations. We move beyond the well-established psychosis-specific measure DUP and analyse waiting times that are mainly induced by delays within the health care system which is of relevance for policy makers. We are also the first to create evidence on the regional distribution of EIP care availability and distances patients have to travel to access EIP care.

Moreover, our work extends existing evidence with regard to methodological aspects. We use generalised linear modelling to appropriately account for the skewness in waiting time and quantile regression to analyse waiting time along the entire distribution. Further, we apply duration analysis to overcome the challenge of right

censoring in waiting time studies. To evaluate the effectiveness of the EIP waiting time target, we make use of quasi-experimental methods. Difference-in-difference designs are well established in the analysis of waiting time targets. However, most analyses rely on country-level comparisons without appropriately controlling for patient case mix and different institutional settings. We add to this evidence by conducting difference-in-difference analysis at the patient level.

The research presented spans across various areas of health policy analysis and is thus of potential benefit for policymakers from different angles. We provide insights into whether socioeconomic determinants contribute to the unequal distribution of waiting times. Policymakers can draw conclusions as to which patients are potentially disadvantaged when accessing specialist services and develop strategies to target these patient groups before others. We also add to the understanding of whether delayed access to mental health services is linked to patient outcomes. In doing this, we are one of the first to use the routinely collected HoNOS to measure outcomes. Policymakers can more likely justify investments for policies that aim to reduce waiting times if outcome prospects can be improved by shorter waiting times. We use HoNOS as a non-psychosis specific outcome measure which will further foster comparisons across studies of different mental health conditions in the future. Further, this research offers evidence on the responsiveness of mental health providers towards performance measures. As such it translates well established evidence and experience from the physical health area to the mental health context which entails a separate group of providers acting in a different incentive system. We provide the first evidence on the effectiveness of the newly introduced EIP waiting time target. We do not only uncover the intended effects in terms of reduced waiting times but also look at potential unintended consequences in non-monitored performance outcomes as well as non-targeted service areas. Results can aid the prospective development of the waiting time targets in other mental health service areas and its adaptation to other countries.

2. Chapter: Socioeconomic inequalities in duration of untreated psychosis

2.1 Introduction

Equitable access to care is one of the corner stones of modern health care systems with universal access. However, the existence of a socioeconomic gradient in accessing health care has been proven in many countries and various areas of physical health care (Cookson et al., 2016; Siciliani, 2016; Abasolo, Negrin-Hernandez and Pinilla, 2014; Kaarboe and Carlsen, 2014; Monstad, Engesæter and Espehaug, 2014; Johar et al., 2013; Laudicella, Siciliani and Cookson, 2012; Arnesen, Erikssen and Stavem, 2002). The relationship can differ depending on the type of health care. For example, low-income individuals and ethnic minorities have been shown to use secondary care to a lower but primary care to a higher extent (Morris, Sutton and Gravelle, 2005). Little is known about the extent to which socioeconomic status plays a role in accessing specialist mental health care. This chapter aims to empirically investigate inequalities in DUP by socioeconomic deprivation for a national first-episode psychosis sample in England.

In its mental health policy strategy the English Department of Health outlined its ambitions to not only make services available but also overcome inequalities around service usage (Department of Health, 2014a). DUP – measuring the time from the first onset of psychotic symptoms to the initiation of treatment (Norman and Malla, 2001) – is a key measure of access to care for first-episode psychosis patients. A growing body of evidence shows that a long DUP is associated with poorer symptomatic outcomes, reduced chances of recovery, poorer social functioning and less treatment engagement (Doyle et al., 2014; Penttilä et al., 2014; Tang et al., 2014; Marshall et al., 2005; Perkins et al., 2005). Since DUP is widely accepted as a modifiable risk factor in first-episode psychosis, plenty of studies have focused on investigating the determinants contributing to a longer DUP. Mostly studied were factors such as migration status, ethnicity, age at onset, gender, and history of substance abuse (Apeldoorn et al., 2014; Broussard et al., 2013; Cascio et al., 2012; Morgan et al., 2006b). Only few studies looked into variables of socioeconomic deprivation and its impact on DUP. There is, however, some evidence that suggests the existence of a socioeconomic gradient (O'Donoghue et al., 2016). DUP is composed of two components: (1) delays due to individual help-seeking; and (2) delays within care

services (Birchwood et al., 2013). Both aspects may influence DUP differentially depending on the level of socioeconomic deprivation. First, DUP is associated with fewer social interactions, limited coping skills, unemployment, and less family involvement (O'Donoghue et al., 2016; Poyraz et al., 2015; Morgan et al., 2006a; Drake et al., 2000). Relatives and friends play an important role in the help-seeking process and in engaging the patient to receive treatment (Fridgen et al., 2013; Hui et al., 2013). In more socioeconomically deprived areas this supportive social network may be less well established which contributes to longer DUP for those patients. Further, it has been shown that a lower degree of education contributes to a longer DUP (Hardy et al., 2018). Second, service provision in socioeconomically deprived areas may be less developed. Particularly in rural and remote areas, availability of EIP care is challenged which may further contribute to longer DUP (Cheng et al., 2014).

Theoretically, the notion of socioeconomic inequalities in health care is motivated by the Grossman model which characterises health care as one input factor into the production process of health (Grossman, 1972). Patients need to invest time and other inputs such as diet, physical activity, or non-smoking behaviour in order to produce better health in future. So even with health care at zero prices, socioeconomically disadvantaged people may invest less in health as they are more time constrained due to worse living and working conditions and have less income to spend on a healthy lifestyle. Second, the ability to produce health even with given inputs depends on the individual's level of education and the environment. People from socioeconomically disadvantaged backgrounds will more likely have to face less supportive social networks, poorer infrastructures, and reduced ability to navigate through a complex health care system in order to seek best quality of care.

Most previous work in this field is limited to small regionally restricted samples with a small number of providers. Also, studies are mainly looking at correlations without adequately controlling for confounding factors. The work presented in this chapter will contribute to the existing literature in a number of ways. First, this is the first study that focuses on the relationship between DUP and socioeconomic deprivation in England. We measure deprivation at small area level using the Index of Multiple Deprivation which is a widely used and accepted measure in the analysis of socioeconomic inequalities in health and in particular with regard to waiting times

(Gutacker, Siciliani and Cookson, 2016; Siciliani, 2016; Laudicella, Siciliani and Cookson, 2012). Second, this is the first study to use administrative data to measure DUP which allows us to analyse a national cohort of first-episode psychosis patients and a large number of mental health providers from different regions in England. It further allows us to control for a rich set of possible confounding factors. Particularly, we control for the severity of hallucinations and delusions, and previous mental health service use. Third, we explicitly model non-linearity to account for the skewed nature of DUP using generalised linear modelling (Jones et al., 2016). Fourth, we look at the effect of socioeconomic deprivation on the whole distribution of DUP using quantile regression recently suggested in the literature (Guloksuz et al., 2016).

2.2 Methods

2.2.1 Data and measures

This study uses secondary patient-level data from the MHLDDS (see section 1.4.1). We identified first-episode patients for our study if they had a record of the emergent or manifest date of the psychosis. Patients were included if both, their first-episode psychosis and their anti-psychotic treatment, started within the study period April 2012 to March 2015.

DUP is defined as the time from the onset of psychosis to the start of treatment (Norman and Malla, 2001). The majority of patients with psychotic disorders experiences a prodromal phase where the first noticeable change in behaviour takes place. First psychotic symptoms may occur during this phase but do not cross the diagnostic duration and severity thresholds of a psychosis. The time between the start of this prodromal phase and the start of anti-psychotic treatment is sometimes referred to as duration of untreated illness (DUI). The onset of psychosis is given once the psychotic symptom thresholds have been crossed. In this regard, the emergent date marks the start of first psychotic symptoms with required severity and the manifest date defines the time when symptoms have lasted for more than a week. These dates usually have to be assessed retrospectively by a trained clinician once the patient has first contact with a specialist mental health service. Assessment methods can vary between providers from clinical interviews, to generic psychosis assessment

instruments (e.g. the Positive And Negative Symptoms Scale) or instruments specifically used to assess DUP (e.g. the Interview for the Retrospective Assessment of the Onset of Schizophrenia). The interrater reliability of these instruments was found to be generally good with no substantial differences between the various instruments (Register-Brown and Hong, 2014).

The definition of treatment start varies in previous DUP studies. Most commonly used end points were the first psychiatric hospitalisation and the first prescription of anti-psychotic medication (mostly regardless of dose, duration, and compliance) (Register-Brown and Hong, 2014; Norman and Malla, 2001). The NICE treatment guideline for schizophrenia recommends oral anti-psychotic medication in conjunction with psychological interventions as the preferred treatment options for first-episode psychosis (NICE, 2014). If patients want to try psychological interventions alone, providers are asked to advise that these are more effective when delivered in conjunction with antipsychotic medication. Hence, we assume that the majority of patients will receive medication. In contrast, not every patient necessarily gets hospitalised during the course of the illness. In fact, a shorter DUP may contribute to a reduced likelihood of an inpatient admission which is why hospitalisation may not be a good measurement end point.

The MHLDDS provides the following information related to the onset of psychosis and treatment: the prodromal date, the emergent date, the manifest date, the date of anti-psychotic medication, and the treatment date (defined as medication taken for 75% of the next month). We have no information on the clinicians that reconstructed the dates and which assessment instruments they used. Figure 2.1 compares median and mean durations in our data when using the different start and end points as introduced above. For the following analysis, we used the emergent date as start and the date of anti-psychotic medication as the endpoint to measure DUP. To increase the number of patients with observed DUP, we used the manifest as alternative start and treatment date as alternative end point if the emergent or medication date were missing.[3]

[3] This is why our final study sample is larger than the ones displayed in Figure 2.1.

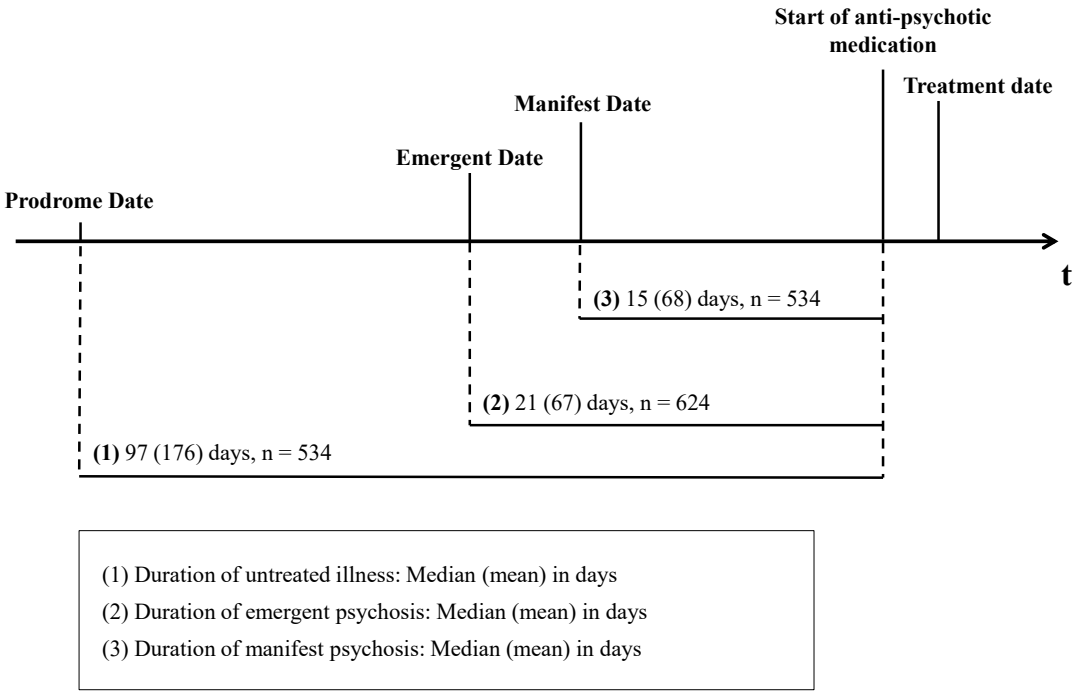


Figure 2.1: Median (mean) days for three different definitions of duration of untreated psychosis (DUP)

We measured socioeconomic status through the Index of Multiple Deprivation (IMD) 2010 which captures deprivation at lower super output area (LSOA) level (McLennan et al., 2011). The IMD includes seven domains of deprivation (income, employment, health and disability, education, barriers to housing, crime, and living environment) which are measured by 38 different indicators. Domains are each weighted according to their perceived importance to calculate the overall index. Each LSOA is ranked, where a rank of 1 equals the most deprived and a rank of 32,482 equals the least deprived area. We derived quintiles of the rank based on the distribution in the general population to indicate the 20% least deprived to the 20% most deprived small areas in England.

The severity of hallucinations and delusions is likely to impact a patient's DUP as patients may lack insight into their illness, fear of being stigmatised, or not be able to attend appointments due to their condition (Compton et al., 2011). We used the HoNOS item 6 which focuses on problems with hallucinations and delusions to approximate the patient's severity of condition (see sections 1.4.1 and 3.3). The item is evaluated on a scale between 0 (no problem) and 4 (severe to very severe problems).

Since symptom severity is an important confounder in the relationship between DUP and socioeconomic background, we excluded patients for which HoNOS information was missing. For patients with a valid HoNOS score, ratings have been conducted at different points in time which vary considerably for patients. Ideally, we are interested in the HoNOS score before treatment started as at later stages the severity of symptoms is likely to be influenced by the treatment itself. Hence, we used the score closest to the psychosis start but within a maximum window of 30 days after the treatment started. We allow for this time window in order to reduce the number of patients that would be excluded from the analysis due to the HoNOS rating having been conducted later. We do not believe that allowing for this time window affects our results for two reasons. First, by definition the clinician is supposed to rate the patient's condition over the past two weeks. If a score was taken 30 days after treatment start it gives a picture of the patient's severity two weeks after treatment start. Second, symptom remission in psychosis is a long-lasting process taking several months and years. Keeping a patient stable is a common outcome at the early stages of treatment. Hence, we assume that the symptom severity at 30 days after treatment start is still a good indicator for the symptom severity before treatment started. Moreover, the patients' ability to navigate themselves through the health care system might be influenced by previous experience of service contacts. Therefore, we considered additional variables of previous mental health service use not related to the first-episode psychosis. For each patient, we counted the number of mental health related professional contacts, outpatient episodes, and ward stays in the twelve months prior to the psychosis start.

We included a set of patient characteristics: age at onset, gender, ethnicity, marital status, accommodation status, employment status, number of physical comorbidities, and number of mental comorbidities. Comorbidities were counted as the number of ICD-10 codes recorded as secondary diagnoses for each patient. ICD-10 codes starting with an "F" where categorised as mental illness comorbidities, while all others as physical comorbidities. Each patient characteristic was measured at the time of the psychosis start. We additionally controlled for the primary diagnosis group measured at the start of the anti-psychotic treatment to distinguish between affective and non-affective psychoses.

2.2.2 Model and statistical methods

We define DUP as the number of days elapsed between the emergence of the patient's psychosis and the start of the first anti-psychotic prescription. Formally, the model is specified in Equation 2.1.

Equation 2.1: $$DUP_{ijk} = \beta_1' d_j + \beta_2' s_{ik} + \beta_3' p_{ijk} + y_i + \alpha_k + u_{ijk}$$

DUP_{ijk} is the DUP for patient $i = 1, \dots, n$ living in LSOA $j = 1, \dots, J$ and being treated at provider $k = 1, \dots, K$. The socioeconomic status is represented by the vector d_j which contains a factor variable for the quintiles of overall deprivation at LSOA-level. The vector s_{ik} contains factor variables to account for severity, namely the HoNOS subscale and the variables of previous service use. The vector p_{ijk} summarises the patient demographics. We included year dummies y_i to eliminate any effects due to changes over time not being captured in the control variables and used provider fixed effects α_k to control for differences in DUP between providers. Previous literature has shown the importance of controlling for provider related differences in waiting times (Sharma, Siciliani and Harris, 2013; Laudicella, Siciliani and Cookson, 2012). Controlling for variations between providers by introducing provider fixed effects allows us to control for the fact that wealthier and better educated people may choose providers with shorter waiting times. As a result, all observed variation needs to be interpreted as inequalities within providers rather than between. The term u_{ijk} represents the idiosyncratic error.

Both the Shapiro-Wilk test (Shapiro and Wilk, 1965) and the Shapiro-Francia test (Shapiro and Francia, 1972) strongly rejected the null hypothesis of DUP_{ijk} being normally distributed. We accounted for the skewness of DUP by using GLM methods (Nelder and Wedderburn, 1972). GLM have been shown to be an adequate choice in typically skewed data such as waiting times (Deb, Norton and Manning 2017; Jones et al., 2016; Sinko et al., 2016; Jones, 2007). GLM allow the expectation of the outcome variable to be a function of the linear index of covariates (link function). At the same time, heteroskedasticity can be modelled explicitly by choosing a distribution family that appropriately defines the functional relationship between the variance of the outcome and its predicted value. Further, GLM permit predictions of the outcome

on the natural scale which avoids the problem of re-transformation and simplifies interpretation of results.

We chose the link function and distribution family of the GLM estimations based on standard procedures suggested in the literature (Deb, Norton and Manning 2017; Jones, 2007). First, we compared the Akaike information criterion (AIC) (Akaike, 1970) and the Bayesian information criterion (BIC) (Schwarz, 1978) after jointly choosing the link function and the distribution family. Both criteria favoured the log link and gamma distribution. Additionally, we performed a modified Park test (Park, 1966) which confirmed the gamma distribution to fit the data best. Pregibon's link test (Pregibon, 1980), the modified Hosmer-Lemeshow goodness-of-fit test (Hosmer and Lemeshow, 2005) as well as Ramsey's Regression equation specification error test (RESET) (Ramsey, 1969) further confirmed the model specification.

We used cluster robust standard errors for 31 provider clusters. To extend our results, we analysed the heterogeneous effects of socioeconomic status at different quantiles of the DUP distribution. Especially in the presence of extreme outliers, it can provide more accurate estimates (Guloksuz et al., 2016). Due to small sample sizes we could only estimate the effect of socioeconomic deprivation and unemployment on DUP without including further covariates.

2.2.3 Sensitivity analyses

We conduct a number of sensitivity analyses to test the robustness of our results. First, we test to what extent the start and end point definitions of our DUP measure influenced the results: (1) we use only the emergent and the prescription date as start and end dates (no substitution of manifest and treatment date), (2) we use the same DUP definition but include only observations that have a valid treatment date, (3) we calculate DUP with the end point being the treatment date only and compare results with and without provider fixed effects. Second, we test the results for the impact of potential outliers: (1) we restrict the sample to the ages 14 to 35 as the main target group for early intervention services, (2) we exclude patients with a DUP of zero as this may be an artefact in the data recording, (3) we exclude patients with a DUP longer than 2 years and 1.5 years respectively. Third, we use marital, accommodation, and employment status as alternative measures of socioeconomic status at the patient-level

and look at the differences compared to using our small-area measure or a combination of the two.

2.3 Results

2.3.1 Descriptive statistics

We identified 1,368 patients with a valid psychosis start and treatment date within the study period (full sample). Six observations were dropped due to missing LSOA codes and 97 observations due to missing HoNOS scores. We further excluded 365 patients from the analysis if the HoNOS rating happened more than thirty days after the treatment start. to account for the level of severity at the early stages of the psychosis. 16 providers (22 corresponding patients) were dropped as they treated fewer than three patients. The final sample comprised 887 patients (65% of full sample) and 31 providers (60%) (see Appendix A1 for more details).

Table 2.1 summarises the demographic characteristics of the study sample and compares it to the full sample as well as to other recent first-episode psychosis studies. The cohort was on average 26 years old, predominantly male (65.6%), of White ethnicity (69.8%), and single (66.2%). Most patients lived in mainstream housing (70.9%), many were unemployed (31.3%) and diagnosed with schizophrenia (38%). There are no significant differences between demographic characteristics of the study sample and the full sample implying that there is no selection bias due to the exclusion of incomplete observations (see also Appendix A2). Further, our study sample appears to be comparable to other recent first-episode psychosis studies by Tsiachristas et al. (2016), O'Donoghue et al. (2016), Kirkbride, Stubbins and Jones (2012), Morgan et al. (2006a).

Table 2.1: Demographic characteristics of the study sample compared to the full sample and other first-episode psychosis (FEP) studies

	Study sample	Full sample	Other FEP studies	References
Total number of observations (n)	887	1,368	831, 292, 357, 495	1,2,3,4
Duration of untreated psychosis in days, median (mean, SD)	22 (73.8, 125.8)	21 (65.9, 115.5)	36 (406, 1036)	4
Duration of untreated psychosis = 0, n (%)	112 (12.6)	192 (14.0)	-	-
Total HoNOS score (range 0-48), mean (SD)	15.3 (6.7)	14.39 (7.1)	-	-
Hallucinations and delusions (HoNOS 6, range 0-4), mean (SD)	2.33 (1.3)	2.11 (1.4)	-	-
Patient age, mean (SD)	26.7 (10.09)	26.12 (10.54)	24.7 (4.62)*	1
Gender - Male (%)	65.6	65.4	65.5, 66.2, 57.8	1,3,4
Ethnicity - British White (%)	69.7	69.8	56.7, 79.1, 43.8	1,3,4
Marital status - Single (%)	66.2	66.7	68.5, 72.5	2,4
Employment				
Unemployed (%)	31.3	30.3	29.16, 50.0	1,3
Employed (%)	21.8	19.7	12.24, 25.0	1,3
Students (%)	17.3	18.7	9.96, 19.0	1,3
Not known (%)	14.2	14.8	48.62, 2.0	1,3
Accommodation				
Mainstream housing (%)	70.9	69.9	45.6	1
Homeless (%)	9.4	8.8	4.4	1
Institutionalised (%)	5.1	5.2	-	-
Not known (%)	13.6	15.0	42.3	1
Diagnosis				
Schizophrenia (%)	38.0	37.0	44.9	2
Affective disorders (%)	12.2	10.4	11.0	2
Not known (%)	34.2	37.9	-	-

Note: Full sample includes all patients with a valid psychosis start date and a valid prescription date in the financial year 2012/13 - 2014/15. The study sample is based on the full sample and excludes observations with missing LSOA, missing HoNOS score (or HoNOS more than 30 days after treatment start), and providers where fewer than 3 patients were treated. "Institutionalised" includes accommodation with mental health or other care support or criminal justice, acute or long-stay healthcare facility, or sheltered housing. References: 1 = Tsiachristas et al. (2016), 2 = O'Donoghue et al. (2016), 3 = Kirkbride, Stubbins and Jones (2012), 4 = Morgan et al. (2006a). * Study sample was restricted to 16 to 35-year-old patients.

Figure 2.2 demonstrates the typically skewed distribution of our DUP measure. Most patients had a DUP smaller than 200 days. For a small proportion of patients DUP was between 200 and 400 days and very few patients waited longer than 400 days from onset of symptoms until start of medication.

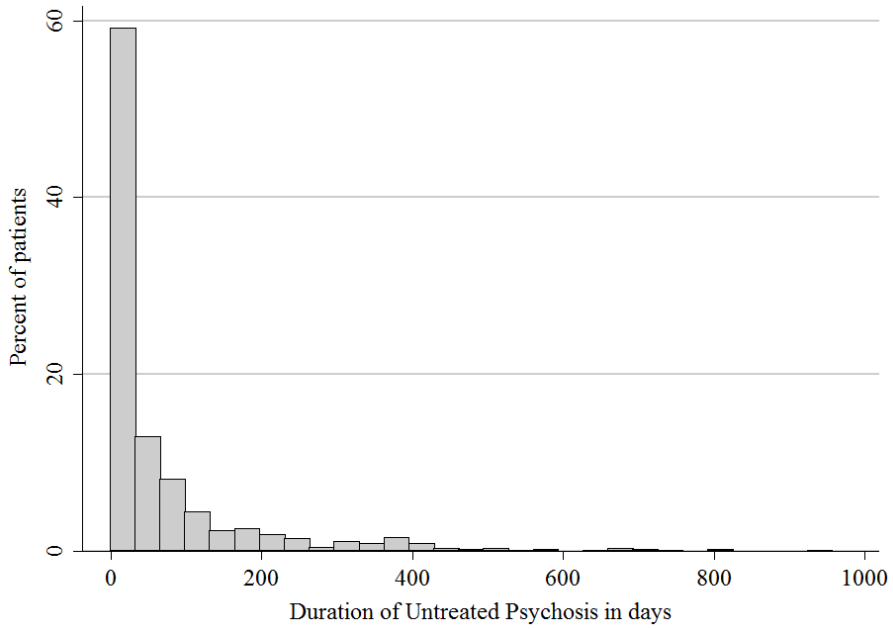


Figure 2.2: Distribution of duration of untreated psychosis

In Table 2.2, all covariates included in the model are presented for the study sample in total and by socioeconomic deprivation quintile. There is an increase in first-episode psychosis patients as the level of deprivation increases. At least 71% of all providers in our sample treated patients from all five socioeconomic quintiles. We note that providers are counted multiple times if they treated patients from more than one socioeconomic quintile. The median DUP overall was 22 days (mean = 73.8 days). Patients from the least deprived quintile waited shortest followed by a clear increase in DUP with every deprivation quintile – with the exception of the most deprived quintile. On average, patients had mild to moderately severe problems with hallucinations and delusions according to their HoNOS score (mean = 2.33). We note that patients from the most deprived quintiles differ in a number of characteristics from the rest of the sample. Compared to the study sample, they are more likely to be single, unemployed, homeless, and in contact with mental health services before the psychosis. The sample is distributed across all nine English regions with the largest

proportion of patients from the South East (25.7%) and the smallest proportion from the North East (0.3%) (see Appendix A3).[4]

[4] Please note that variations across regions may also be due to different incidence rates or age distributions in the general population.

Table 2.2: Distribution of patients, providers, and patient characteristics by socioeconomic status

	Study sample	Least deprived	2nd least deprived	3rd least deprived	4th least deprived	Most deprived
Number of patients, n (%)	887 (100)	145 (16.3)	142 (16.0)	180 (20.3)	191 (21.5)	229 (25.8)
Number of providers, n (%)	31 (100)	22 (71)	23 (74)	29 (94)	24 (77)	25 (81)
Duration of untreated psychosis, median (mean)	22 (73.8)	14 (46.3)	21 (75.2)	25.5 (80.5)	34 (100.3)	20 (62.8)
Disease severity, mean (SD):						
Total HoNOS score (range 0-48)	15.4 (6.7)	15.0 (6.5)	15.2 (6.9)	15.0 (6.5)	15.9 (6.8)	15.6 (6.8)
Hallucinations and delusions (HoNOS 6, range 0-4)	2.3 (1.3)	2.2 (1.4)	2.3 (1.3)	2.4 (1.3)	2.3 (1.3)	2.5 (1.3)
Mental health service use, mean (SD):						
Number of service contacts	2.8 (8.6)	2.8 (8.1)	2.3 (9.4)	2.3 (6.9)	2.8 (9.5)	3.3 (8.9)
Number of outpatient episodes	0.09 (0.4)	0.14 (0.5)	0.05 (0.3)	0.7 (0.3)	0.07 (0.3)	0.10 (0.6)
Number of ward stays	0.09 (0.4)	0.06 (0.3)	0.08 (0.5)	0.09 (0.4)	0.09 (0.5)	0.14 (0.5)
Number of physical comorbidities	0.01 (0.1)	0.01 (0.1)	0.02 (0.2)	0.01 (0.1)	0.01 (0.1)	0.2 (0.1)
Number of mental comorbidities	0.02 (0.2)	0.01 (0.1)	0.01 (0.1)	0.01 (0.1)	0.04 (0.3)	0.01 (0.1)
Patient characteristics:						
Patient age, mean (SD)	26.7 (10.1)	28.0 (14.1)	28.8 (13.4)	25.5 (9.0)	26.5 (11.1)	25.8 (7.4)
Male (%)	65.6	65.5	59.9	67.8	62.3	70.3
British White (%)	69.7	78.6	78.9	68.3	68.1	61.1
Single (%)	66.2	56.6	57.8	63.9	71.2	75.1
Employed (%)	21.8	26.9	30.3	20.6	23.0	13.1
Unemployed (%)	31.3	21.4	30.3	28.3	27.2	44.1
Students (%)	17.3	25.5	16.2	14.4	18.6	13.5
Long-term disabled (%)	8.1	6.2	4.9	8.9	8.9	10.0
Other employment (%)	7.3	9.0	5.6	8.9	7.3	6.1
Mainstream housing (%)	70.9	84.8	74.7	69.4	71.7	60.3
Homeless (%)	9.4	3.5	7.0	11.1	6.3	15.7
Institutionalised (%)	5.1	2.8	5.6	1.1	7.3	7.4

	Study sample	Least deprived	2nd least deprived	3rd least deprived	4th least deprived	Most deprived
Schizophrenia diagnosis (%)	38.0	31.1	26.8	41.7	41.4	43.7
Affective disorder diagnosis (%)	12.2	13.8	17.6	6.7	12.6	11.8
Substance abuse (%)	7.9	5.5	9.2	6.7	6.8	8.3
Financial year 2012/13 (%)	33.2	34.5	33.1	34.4	36.7	28.4
Financial year 2013/14 (%)	40.0	40.7	41.6	34.4	42.4	41.1
Financial year 2014/15 (%)	26.8	24.8	25.4	31.1	20.9	30.6

Note: Categorical variables may not sum up to 100% as categories of missing values are not presented. The number of providers refers to those that treated at least one patient from the given socioeconomic quintile, providers can be counted more than once if they treated patients from more than one socioeconomic quintile. Service contacts, outpatient episodes and ward stays refer to mental health related service use in the twelve months prior to the psychosis start.

2.3.2 Estimation results

Estimation results from Table 2.3 confirm significant variations in DUP by socioeconomic deprivation for the first four deprivation quintiles (least to fourth least deprived). Patients in the second least deprived quintile have a 35.5 day longer DUP than patients from the least deprived quintile. Patients from the third and fourth least deprived quintiles face a DUP that is 24 and 31 days longer than the patients' DUP from least deprived neighbourhoods. The most deprived quintile has a negative coefficient indicating a slightly shorter DUP for patients from most deprived areas compared to the least deprived quintile. However, the result is not statistically significant. Experiencing very severe problems with hallucinations and delusions has a significant impact on DUP. But different to what would be expected, patients suffering from severe hallucinations and delusions wait 21 days shorter than patients having no problems at all. Negative coefficients for moderately severe problems and minor problems indicate the same severity gradient in DUP, however the estimates are not statistically significant. This may indicate some effective prioritisation as patients being most severely affected receive treatment first. However, if this is the case then prioritisation is only effective for the very severely affected but not the patients with moderately and mildly severe problems. The observed effect may also be explained by the fact that the severe symptoms led to a quicker identification of the psychotic condition and motivated patients or carers to seek specialist help earlier. Mental health professional contacts in the twelve months prior to the psychosis start, significantly reduce DUP by 36 days for one to ten contacts, and by 53 days for more than ten contacts compared to no contact at all. Having had an outpatient mental health consultant episode before the psychosis, did not show a significant effect on DUP. However, for patients with more than three previous ward stays related to a mental health condition, DUP was 60 days shorter. Patient numbers in the latter case were low which might have affected their statistical significance. Regarding other patient characteristics, we find a small effect of age on DUP. Further, there is a strong relationship between employment status and DUP. Patients being unemployed have a 40 day longer DUP than employed patients. Also, students have a 30 day longer DUP compared to patients in employment. We could not find any significant inequalities in DUP with regard to gender, ethnicity, marital status, or accommodation status.

Table 2.3: Generalised linear model regression results

	Coeff.	Std. Err.	[95% Conf. Int.]		dy/dx
Socioeconomic status (ref.cat.: least deprived quintile)					
2nd least deprived quintile	0.459***	(0.131)	[0.204	0.715]	35.5
3rd least deprived quintile	0.331***	(0.099)	[0.137	0.525]	23.86
4th least deprived quintile	0.410*	(0.183)	[0.051	0.770]	30.89
Most deprived quintile	-0.012	(0.155)	[-0.315	0.291]	-0.73
Severity of hallucinations and delusions (ref.cat.: no problems)					
Minor problems	-0.147	(0.256)	[-0.648	0.354]	-11.95
Mild problems	0.011	(0.193)	[-0.367	0.389]	0.94
Moderately problems	-0.169	(0.200)	[-0.562	0.223]	-13.61
Severe problems	-0.271*	(0.118)	[-0.502	-0.040]	-20.77
Previous mental health service use (ref.cat.: zero service contacts, outpatient episodes, and ward stays)					
1-10 Service contacts	-0.526**	(0.163)	[-0.845	-0.206]	-35.79
>10 Service contacts	-0.929**	(0.297)	[-1.512	-0.347]	-52.96
1-3 Outpatient episodes	-0.713	(0.547)	[-1.786	0.359]	-40.10
> 3 Outpatient episodes	-0.211	(0.635)	[-1.456	1.034]	-14.95
1-3 Ward stays	-0.726	(0.435)	[-1.578	0.127]	-40.76
> 3 Ward stays	-1.417**	(0.519)	[-2.434	-0.399]	-59.83
Patient demographics					
Age	-0.017*	(0.009)	[-0.034	-0.000]	-1.34
Female	0.059	(0.139)	[-0.214	0.332]	4.63
Ethnicity (ref.cat.: White or White British)					
Mixed ethnic group	0.326	(0.291)	[-0.244	0.896]	28.71
Asian or Asian British	-0.474	(0.244)	[-0.952	0.004]	-28.11
Black or Black British	0.198	(0.201)	[-0.196	0.592]	16.32
Other ethnic group	-0.024	(0.277)	[-0.566	0.519]	-1.74
Marital status (ref.cat.: single)					
Married/civil partner	-0.060	(0.136)	[-0.326	0.207]	-4.32
Divorced/separated	-0.012	(0.314)	[-0.628	0.603]	-0.92
Accommodation (ref.cat.: mainstream housing)					
Homeless	0.011	(0.217)	[-0.414	0.435]	0.82
Institutionalised	0.077	(0.247)	[-0.408	0.562]	6.14
Other accommodation	-1.081***	(0.223)	[-1.518	-0.644]	-50.89
Employment (ref.cat.: employed)					
Unemployed	0.572***	(0.147)	[0.283	0.860]	39.98
Student	0.456*	(0.203)	[0.060	0.853]	29.98
Long-term disabled	0.270	(0.244)	[-0.209	0.749]	16.07
Other employment	0.193	(0.415)	[-0.621	1.006]	11.01
Diagnosis (ref.cat.: schizophrenia)					
Substance abuse	-0.096	(0.223)	[-0.5323	0.341]	13.12
Affective disorders	-0.431	(0.246)	[-0.913	0.050]	-10.72
Other diagnosis	-0.349	(0.276)	[-0.891	0.192]	-0.62
Number of physical comorbidities	-2.137	(1.179)	[-4.448	0.174]	-75.00
Number of mental comorbidities	-0.966	(0.517)	[-1.980	0.048]	-166.00

Note: * p<0.05, ** p<0.01, *** p<0.001. Dependent variable is duration of untreated psychosis in days (DUP). Included are year dummies for 3 financial years and provider dummies for 31 providers. Marginal effects (dy/dx) are average marginal effects in days. For factor levels they present the discrete change from the reference category. Cluster robust standard errors in parentheses.

The graphical analysis in Figure 2.3 confirms that the effect of socioeconomic deprivation increases towards the higher quantiles of the DUP distribution. The coefficients for all deprivation quintiles – except the most deprived quintile - are smaller for the lower quantiles and increase along the DUP distribution. However, confidence intervals increase towards the end of the DUP distribution. Also, the effect of unemployment on DUP increases along the DUP distribution.

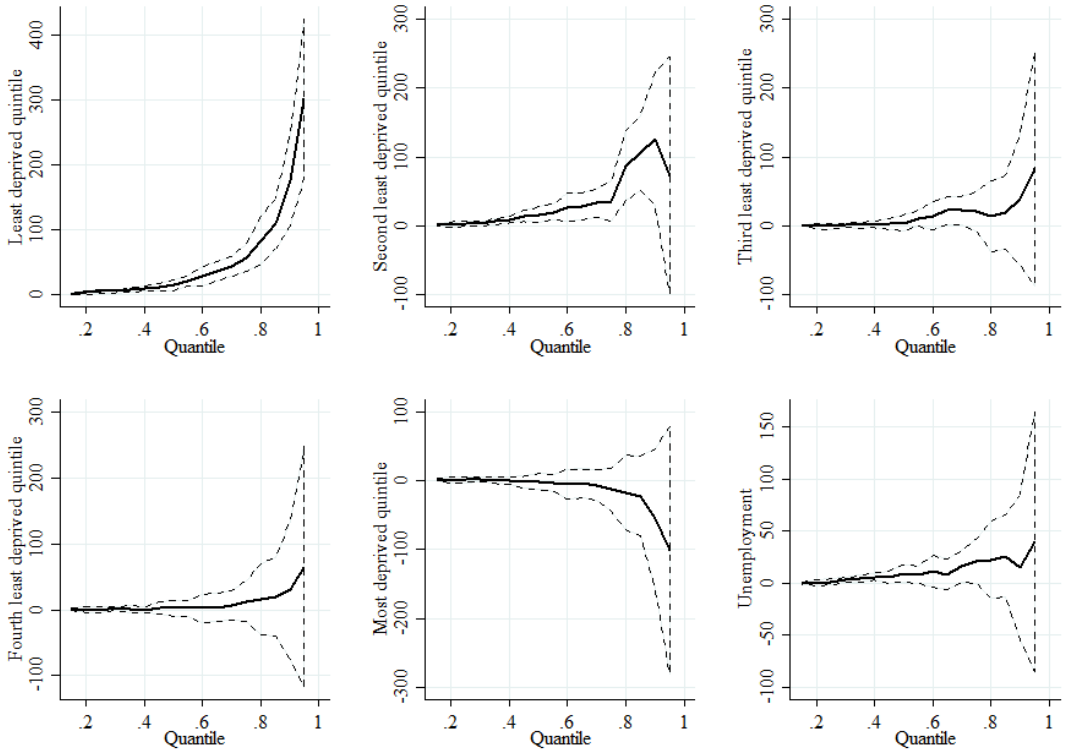


Figure 2.3: Differential effects of socioeconomic deprivation and unemployment by quantile

2.3.3 Sensitivity analyses

Estimation results for the sensitivity analyses can be found in Appendices A4 to A7. Results were shown to be robust against different definitions of the DUP measure (Appendix A4). Using only the emergent and prescription date as start and end point did not seem to influence the results in a significant way. Also, using only observations with a valid treatment date did not change the gradient we observed. Appendix A5 shows the results when using the treatment date as an alternative end point to calculate DUP. Again, the gradient remained similar with and without provider fixed effects. Restricting the sample to the ages 14 to 35 revealed an even stronger socioeconomic gradient compared to the full sample (Appendix A6). The socioeconomic gradient

decreased in magnitude and the second least deprived quintile lost significance as we exclude DUP that exceeds 1.5 years. This suggests that the socioeconomic gradient is stronger for patients with very long waits which was further confirmed when we excluded patients with a zero DUP. In the subsample of non-zero DUP patients, we observed a clear socioeconomic gradient. Excluding marital, accommodation, and employment status as patient-level measures of socioeconomic status from the regression, did not change the observed gradient (Appendix A7). Among the patient-level variables, only employment had a significant effect which was similar to what we observed in the main model. The IMD quintiles seem to capture aspects of deprivation which are not included as separate covariates in the model. The most deprived quintile remained insignificant regardless of the specification.

2.4 Discussion

Since the prevalence of first-episode psychosis in more deprived neighbourhoods is found to be higher compared to less deprived areas (O'Donoghue et al., 2016), we asked whether the level of socioeconomic deprivation determines the patient's help-seeking behaviour and access to care. As well as being the first to investigate the relationship between DUP and socioeconomic deprivation in England, we were able to use a large sample from administrative data including a large number of mental health providers. Compared to other literature in the field we controlled for a rich set of covariates and applied statistical methods that adequately account for non-linearity in DUP. The results were robust in a number of sensitivity analyses.

Our findings revealed significant inequalities regarding the level of socioeconomic deprivation. However, the effect was not linear across deprivation quintiles. Patients from the second least deprived quintiles had the longest DUP followed by patients from the fourth, and the third least deprived quintiles. For the most deprived quintile, differences were not statistically significant. Severe hallucinations and delusions and previous mental health service contacts not related to the psychosis significantly reduced the DUP. We did not find any significant inequalities in DUP with regard to age, gender, or ethnicity, which also confirms findings from previous studies (Ghali et al., 2013; Cascio et al., 2012; Large et al., 2008; Morgan et al., 2006a). We used a

comprehensive measure of socioeconomic status which captures various aspects of deprivation and is widely used in other literature on health inequalities. It should be noted that our measure is relative – not every person living in a highly deprived area will themselves be deprived and vice versa. At a patient level, marital, accommodation, and employment status could serve as proxies for the patient's socioeconomic status. Results consistently indicate that employment status plays an important role in the length of DUP as has been found by other studies (Morgan et al., 2006a). Marital and accommodation status, however, did not explain any differences in DUP. Since we control for provider fixed effects and patient-level socioeconomic status in our model, the observed socioeconomic gradient in DUP is independent of provider characteristics and of the patient's marital, accommodation, and employment status.

2.4.1 Limitations

It remains to be explained why the most deprived neighbourhoods have a shorter DUP than the other deprivation quintiles which contradicts findings of a clear socioeconomic gradient within the physical health literature (Siciliani, 2016). It may be that patients from most deprived neighbourhoods enter the system more often through the criminal justice system which may shorten their DUP, or they are more likely to be in contact with a general practitioner due to a poorer general health. Comparing most deprived patients with the rest of the sample revealed that they are more likely to be single, unemployed, and homeless. They were also more likely to have been in contact with mental health services before the psychosis which seems to support our theory. Further, this could represent a recall bias by the patient. The information on the emergence of the psychosis relies on self-report. Patients from more deprived neighbourhoods may systematically report their symptom history differently from others due to different educational levels or insight into the disease. Future work could aim to split DUP into two constituent parts in order to analyse the help-seeking period (from onset of symptoms to first referral) separately to see whether the effect of deprivation is different for this period. However, this requires the identification of the relevant referral which will lead to further restrictions in the selection of patients for the study with the given data. It is also unclear whether the interpretation of results would change considerably.

Our study focuses on DUP as one of the key parameters in managing first-episode psychosis patients. The importance of the DUP concept lies in its strong relationship to improved clinical outcomes while at the same time being a modifiable risk factor. The median DUP in our study was 22 days which is close to figures in some studies (Apeldoorn et al., 2014) but shorter compared to other studies reporting a median DUP of 50 to 120 days (O'Donoghue et al., 2016; Behan et al., 2015; Birchwood et al., 2013). On the one hand, differences may be caused by our study period being limited to three years. Thus, we possibly exclude a number of DUP observations exceeding the study period. As we can only observe patients that finished waiting by the end of the study period, long waiting patients were also more likely to be truncated at the end. This may explain the significant decline in DUP across the three years of study. However, this may as well reflect that the increasing international awareness of early intervention contributed to an overall reduction in DUP. In this case, our study provides a much more recent measure of DUP as previous studies used data from 1995 to 2011. If we are underestimating the DUP and it holds true that the socioeconomic gradient increases as the DUP increases, then we are likely to further underestimate socioeconomic inequalities. On the other hand, differences may be rooted in the measurement of DUP. Despite its strengths, the DUP concept has been criticised in the literature as its definition varies across studies (Register-Brown and Hong, 2014; Large et al., 2008; Singh, 2007). From our data, we are not able to provide information on the methods being applied to define the emergent date and what training the clinical teams received with regard to this. It is also very likely that methods varied between the providers in our sample. By applying provider fixed effects, we controlled for any measurement differences between providers. However, we were not able to capture any variation if clinicians within the same provider were recording dates differently. This would have influenced results if clinicians within a provider would record dates for patients from socioeconomically more deprived areas differently to those from less deprived areas. We defined the first antipsychotic prescription as the treatment start as it can be consistently defined within our dataset. But we appreciate that the prescription of medication does not necessarily imply that a patient has received effective treatment (Breitborde, Srihari and Woods, 2009). Using this approach introduces the problem of reliably defining effective treatment. To date, there is no agreed best way of measuring

DUP (Register-Brown and Hong, 2014). Assuming that effective treatment will be put in place from the first antipsychotic prescription, we are likely to underestimate the actual DUP and look at just a part of its full duration. We do, however, cover the period of help-seeking which is expected to be much more influenced by the patient's socioeconomic background than the aspect of receiving effective treatment after the first service contact. Our results were also robust against changing the DUP endpoint. Nevertheless, future research should aim to address this limitation by establishing a DUP measure that goes beyond the traditional definition using for example the acceptance onto the caseload of an EIP service as the endpoint. This approach will allow the inclusion of patients who never received any anti-psychotic medication.

Despite the policy relevance of DUP, the reporting of relevant data is not mandatory for providers. Hence, we cannot rule out that there is a bias in the composition of our sample as we may miss out first-episode psychosis patients not being reported by providers. We further had to exclude patients with missing HoNOS scores in a maximum 30-day window after treatment start in order to measure symptom severity as an important confounder. However, we see no reason to believe that patients with HoNOS ratings at later stages during treatment (which we excluded) are systematically different to our study sample in terms of their relationship between DUP and socioeconomic background. Overall, our sample proved to be comparable with first-episode psychosis patient cohorts from other recent studies.

Finally, any unobserved heterogeneity cannot be ruled out due to factors such as drug abuse, family history in psychosis, or patients' social network. For example, there is evidence of interactions between age, gender, and cannabis use (Donoghue et al., 2014; Broussard et al., 2013). Also stigma-related processes have been found to influence help-seeking and service contact at early stages of psychotic disorders (Gronholm et al., 2017). Although HoNOS is a validated tool in the application of psychoses it might not capture all aspects of disease related severity. This could lead to an over- as well as underestimation of the effects of socioeconomic deprivation on DUP depending on whether hallucinations or delusions are more prevalent in certain deprivation quintiles.

2.4.2 Implications for EIP services

DUP captures the complete waiting experience of the patient including time from first symptom to help-seeking, from referral to assessment, and from assessment to treatment. Therefore, we cannot distinguish between the patient's and the care system's contribution to the delay and factors are likely to interact with each other. However, socioeconomic deprivation is a contributing factor to a prolonged DUP independent of severity of hallucinations, previous service contacts, and patient demographics. Inequalities arise predominantly at the higher end of the DUP distribution. Policies to improve equitable access to care should therefore focus on preventing very long delays in treatment and target unemployed patients and students. Being known to mental health services for reasons other than psychosis seems to make it easier to access the system a second time regardless of the severity of the condition. Efforts aimed at shortening DUP should particularly target people that have not been in contact with any mental health professional in the past. For example, general practitioners or other health professional education campaigns could improve awareness of the signs of early psychosis and encourage them to refer patients promptly to specialist services (Lloyd-Evans et al., 2011). Also, information campaigns for young people and their families in schools or in mainstream media may contribute to a reduced stigmatising image of psychosis and will promote early help-seeking (Connor et al., 2016). The decrease in DUP over the past years indicates that the awareness of its importance has increased. However, significant variations within providers remain and should be addressed further to reduce inequalities.

3. Chapter: The impact of waiting time on patient outcomes

3.1 Introduction

Concerns about waiting times arise when cases are affected in which waiting time impedes the patient's utility gain from treatment. But to date, little is known about delays within the mental health service system and their impact on outcomes. This chapter 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 the assignment of a care coordinator, 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 twelve months after the start of treatment.

The theoretical foundation builds the model of queuing by list (Lindsay and Feigenbaum, 1984) introduced in section 1.2.3. As waiting for a care coordinator does not require patients to queue in person, there are no opportunity costs in terms of time spent waiting in order to clear markets. But still, waiting times impose costs. The treatment received tomorrow is worth less today since the patient (and caring relatives) have to experience suffering and inconvenience of living with a disease. This negative impact can also be long-term if the deteriorated condition of the patient due to waiting increases time of recovery or will not be reversed at all after a critical waiting time has passed (Koopmanschap et al., 2005). In case of first-episode psychosis, the suffering can be significant and intervening early is critical to successful treatment as has been demonstrated in section 1.2.1.

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 twelve months, incorporating treatment intensity during this time period. Our outcome measure, 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 (Reininghaus and Priebe, 2012; McCabe, Saidi and Priebe, 2007). 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 (Tasma et al., 2017; Gilbody, House and Sheldon, 2003).

3.2 Related literature

Two strands of literature can 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 (Nikolova, Harrison and Sutton, 2016; Quintana et al., 2011; Tuominen et al., 2010; Hirvonen et al., 2009; Tuominen et al., 2009; Braybrooke et al., 2007; Hirvonen et al., 2007; Ho, Hamilton and Roos, 2000; Hamilton and Bramley-Harker, 1999; Hamilton, Hamilton and Mayo, 1996), or more urgent surgical procedures such as organ transplantation (Rauchfuss et al., 2013; Meier-Kriesche et al., 2000), and coronary artery bypass surgery (Moscelli, Siciliani and Tonei, 2016; Manji et al., 2013; Sari et al., 2007). Fewer studies investigate the relationship of waiting time with non-surgical treatments such as rehabilitation (Pedersen, Bogh and Lauritsen, 2017; Collins et al., 2015), radiotherapy (Gupta et al., 2016; Seidlitz et al., 2015; Noel et al., 2012), 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, Siciliani and Tonei (2016) found that waiting for coronary bypass surgery did increase

the number of emergency readmissions but not in-hospital mortality. Nikolova, Harrison and Sutton (2016) analysed the impact of waiting for elective surgery on patient-reported outcomes. They found that a longer waiting time reduced 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 psychosis patients. The key measure of waiting time in this context is the DUP (see section 1.4.2). 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 to 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 and Malla, 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.

3.3 Data and key measures

We analyse a cohort of patients having a first EIP episode within the study period April 2012 to March 2014 from the MHLDDS (see section 1.4.1). Patients were followed up for a period of twelve months. Figure 3.1 summarises the study timeline and measurement points.

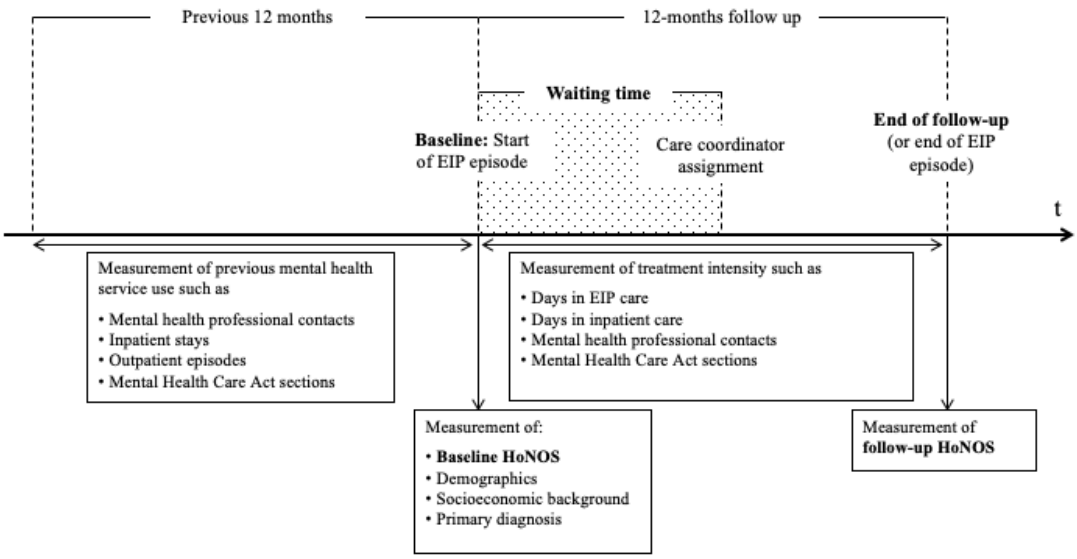


Figure 3.1: Study timeline with measurement points

We choose a twelve-month follow-up period to investigate both whether waiting itself is associated with a worsening of the patient's condition and whether the waiting impedes the patient's ability to benefit from treatment. In both cases, waiting time may be endogenous. First, a deteriorated patient condition after waiting may not only be caused by the waiting itself but will also depend on the patient's condition at the beginning of the waiting time. For example, if effective prioritisation is in place less severely affected patients may have longer waiting times but better outcomes as the initial severity level was lower. Second, the ability to benefit from treatment will not only depend on how long the patient waited for treatment but also on the treatment intensity during the twelve months of follow-up. Our strategy to encounter this endogeneity is to control for the initial level of severity at baseline as well as for treatment intensity during the follow-up period in our regression model.

Our outcome measure, HoNOS, is routinely collected by providers in our dataset (Wing, Curtis and Beevor, 1999; Wing et al., 1998). HoNOS is composed of twelve items covering the four sub-domains behaviour, impairment, symptoms, and social functioning (see Appendix A8). 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 twelve 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 and Truax (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, Moran and Borowitz, 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 and NICE, 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 one month (Tang et al., 2014) or three months (Cechnicki et al., 2014). Therefore, we employ three different transformations of waiting time: (1) a log transformation of waiting time in days, (2) waiting time quintiles with an equal number of patients in each group, (3) 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).

3.4 Methods

3.4.1 The model

We denote h_{ijkl} as the mental health status of the ith-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^1_{ijkl}) and twelve months after treatment start (h^2_{ijkl}) as the total HoNOS score. Formally, the model is specified in Equation 3.1.

Equation 3.1: $$h^2_{ijkl} = \alpha W_{ijkl} + \beta h^1_{ijkl} + \gamma T_{ijkl} + \delta X_{ijkl} + \tau S_{ijkl} + y_l + u_k + \varepsilon_{ijkl}$$

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^1_{ijkl}. 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: (1) the number of days in EIP care, (2) the number of days in inpatient care, and (3) 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 IMD measured at LSOA level (see section 1.4.1). 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 twelve 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 and Wilk, 1965) and the Shapiro-Francia test (Shapiro and Francia, 1972) strongly rejected the null hypothesis of h_{ijkl}^2 being normally distributed. We used GLM regression methods (Nelder and Wedderburn, 1972) to accommodate the skewness of the HoNOS distribution (see section 2.2.2 for a detailed discussion of GLM). GLM has been shown to be an adequate choice in typically skewed data. 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 and Lemeshow, 2005) accepted the square root link function. The Ramsey RESET test (Ramsey, 1969) further confirmed the model specification. We used cluster robust standard errors for the 48 mental health trusts.

3.4.2 Robustness checks

We applied the same model from Equation 3.1 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 used the concept of a clinically significant and reliable change (see Appendix A9 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.

3.5 Results

3.5.1 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 3.1 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. The mean HoNOS score at baseline was 14.1. During the twelve 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 twelve 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.

Table 3.1: 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 sections (%)	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.

Figure 3.2 visualises the distributional shift of HoNOS scores towards zero from baseline to follow-up.

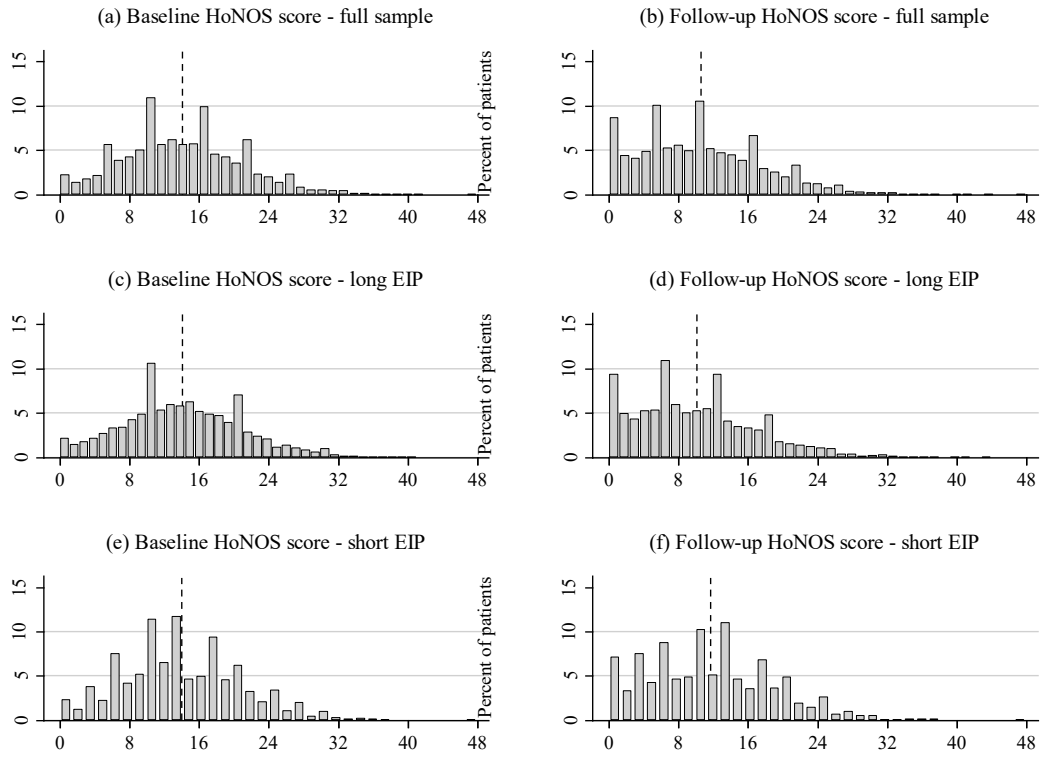


Figure 3.2: Histogram of HoNOS scores at baseline and follow-up

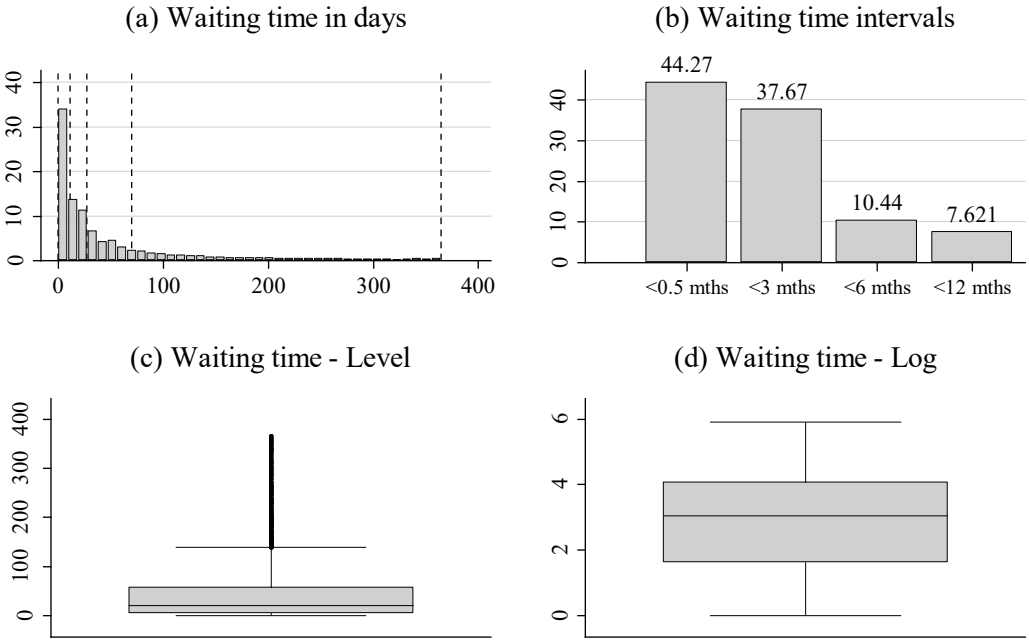
We note that not all patients spent the total follow-up time in EIP care. We therefore stratified the study sample by whether a patient finished EIP care before the end of follow-up ("short EIP" group, 31.4%) or not ("long EIP" group, 68.6%) and run analyses for the two subsamples separately. Table 3.2 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 3.2: 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 3.3 summarises several descriptive statistics of our main explanatory variable, waiting time. As expected, we find waiting time to be heavily left skewed with a

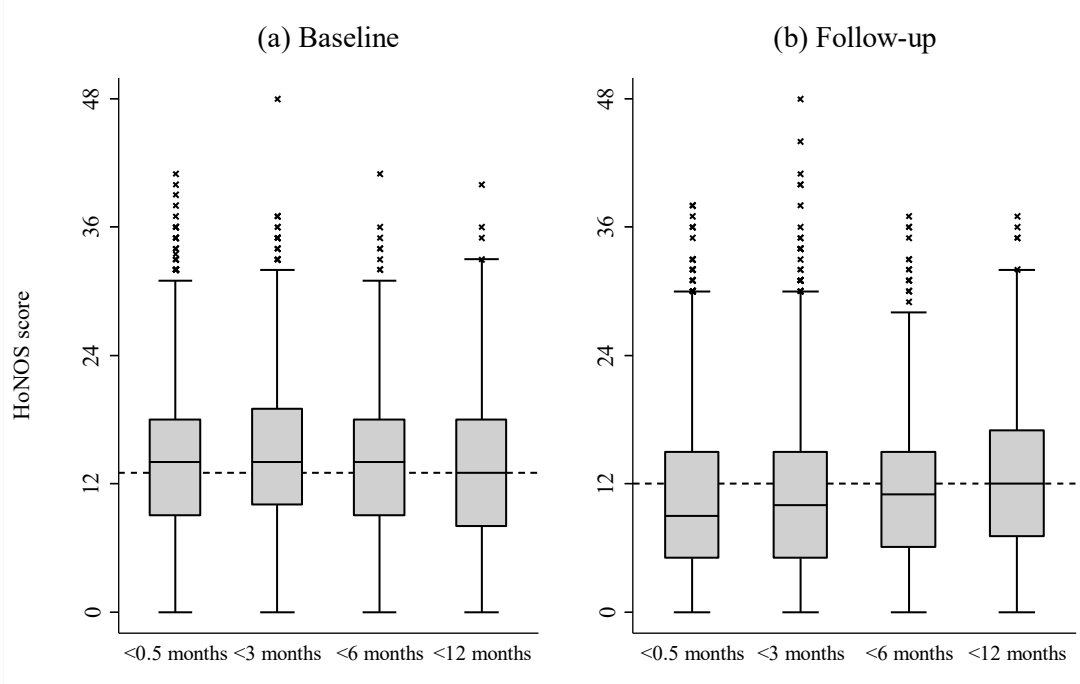
median of 20 and a mean of 50 days. Consequently, the largest proportion of patients was allocated to the waiting time interval of less than 0.5 months (panel (b)). We also see that taking the logarithm of waiting time helps to reduce a large amount of the skewness (panels (c) and (d)).



Note: Dashed vertical lines in (a) indicate boundaries of each corresponding waiting time quintile.

Figure 3.3: Descriptive statistics on waiting time

Figure 3.4 visualises the distribution of HoNOS scores across the different waiting time intervals. Baseline HoNOS scores in panel (a) varied 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.



Note: Dashed horizontal line indicates the median HoNOS in the longest waiting times interval.

Figure 3.4: Box plots for HoNOS scores by waiting time intervals

3.5.2 Estimation results

Table 3.3 displays the estimation results from the regression of Equation 3.1 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 twelve 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 3.3: Generalised linear model results with follow-up HoNOS as dependent variable

	(1) Study sample			(2) Long EIP			(3) Short EIP		
	Coeff.	Std. Err	dy/dx	Coeff.	Std. Err.	dy/dx	Coeff.	Std. Err.	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 (Std. Err.) for 48 provider clusters.

Table 3.4 reports the estimated coefficients and marginal changes of the baseline HoNOS as well as the treatment intensity variables based on Equation 3.1. As expected, we observe a strong positive relationship between baseline and follow-up HoNOS scores. A worse baseline condition strongly predicts worse outcomes twelve 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.

Table 3.4: 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
Mental Health Care Act sectioned 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.

Regarding treatment intensity, 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.

3.5.3 Robustness of results

Results from the effects of waiting time on the different HoNOS sub dimensions are provided in Appendix A11. 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 three months are affected most by a deterioration in outcomes on each sub-domain. We note that the marginal effects of waiting time on HoNOS scores are small (less than 2 score points). After applying the concept of clinically significant and reliable change to our study sample, a change of at least 10 score points would mean a reliable and a change of at least 13 score points a clinically meaningful change (see Appendix A9). Hence, our estimated effect of waiting time is likely to be clinically meaningless. However, 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 A12).

3.6 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 twelve months after treatment acceptance for additional days of waiting. However, we believe that this decline of less than two score points may not be of clinical relevance. Although a general agreed threshold on what defines a clinically meaningful change in HoNOS is missing, we find a change of at least 10 to 13 points for our study sample to be necessary to define a reliable and clinically relevant change. At the same time, we do find the risk of a clinically significant and reliable deterioration which is based on the above-mentioned thresholds to be elevated by longer waiting time. Effects are significant in the waiting time range from three to twelve 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 (Register-Brown and Hong, 2014; Singh, 2007; Norman and Malla, 2001). 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 (Pirkis et al., 2005; McClelland et al., 2000; Amin et al., 1999; 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 twelve 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 three months.

4. Chapter: Clinical priorities and gaming behaviour in the light of waiting time targets

4.1 Introduction

Waiting time targets are well established performance measures in systems with excess demand and rationing of care such as the NHS in England (Willcox et al., 2007). They guarantee patients' access to care within a defined window of time though the definition of this window varies widely across countries and areas of health care. In the market for physical health care, waiting time targets have been shown to be effective in reducing waiting time (Woodcock, Alan and Bell, 2013; Besley, Bevan and Burchardi, 2009; Dimakou et al., 2009; Propper et al., 2008). Whereas most of these targets in the past have been accompanied by substantial financial penalties with hospital managers in fear of losing their position (Propper et al., 2008), implicit mechanisms such as the public disclosure of performance information can have similar effects on provider behaviour (Marshall et al., 2000). Performance benchmarks may indirectly affect budgets, bonuses, job security, staff morale and recruitment (Goddard, Mannion and Smith, 2000), and the publishing of performance measures could result in reputational damage (Bevan and Hamblin, 2009; Hibbard, Stockard and Tusler, 2005). At the same time, there is an ongoing debate on whether the focus of providers on meeting arbitrary targets may lead to unintended consequences in non-targeted performance areas (Smith, 1995).

Since April 2015, the English NHS operates a waiting time target for EIP services as one of the first of its kind in the mental health context. Before the target came into effect, a number of policy initiatives contributed to a growing awareness about the importance of early access for patients seeking EIP care (see Figure 4.1). The strategy "No health without mental health" published in February 2011 initiated a sequence of activities aimed at improving access to evidence-based treatments at the early stages of a mental illness or crisis (Department of Health, 2012, 2011). In January 2014, the government announced the intention to introduce waiting time targets for mental health services from April 2015 onwards without specifying the services to be affected (Department of Health, 2014a). In October 2014, the EIP target was defined for the first time: "More than 50% of people experiencing a first episode of psychosis will be treated with a NICE approved care package within two weeks of referral" (Department of Health, 2014b). The target was supposed to be implemented from April 2015

onwards. However, providers were expected to reduce waiting times long before the target comes into effect. A detailed guideline on how the EIP target was to be implemented was published in February 2015 (NHS England and NICE, 2015). This guideline for the first time, introduced the assignment of a care coordinator as the key requirement to stop the waiting time clock.

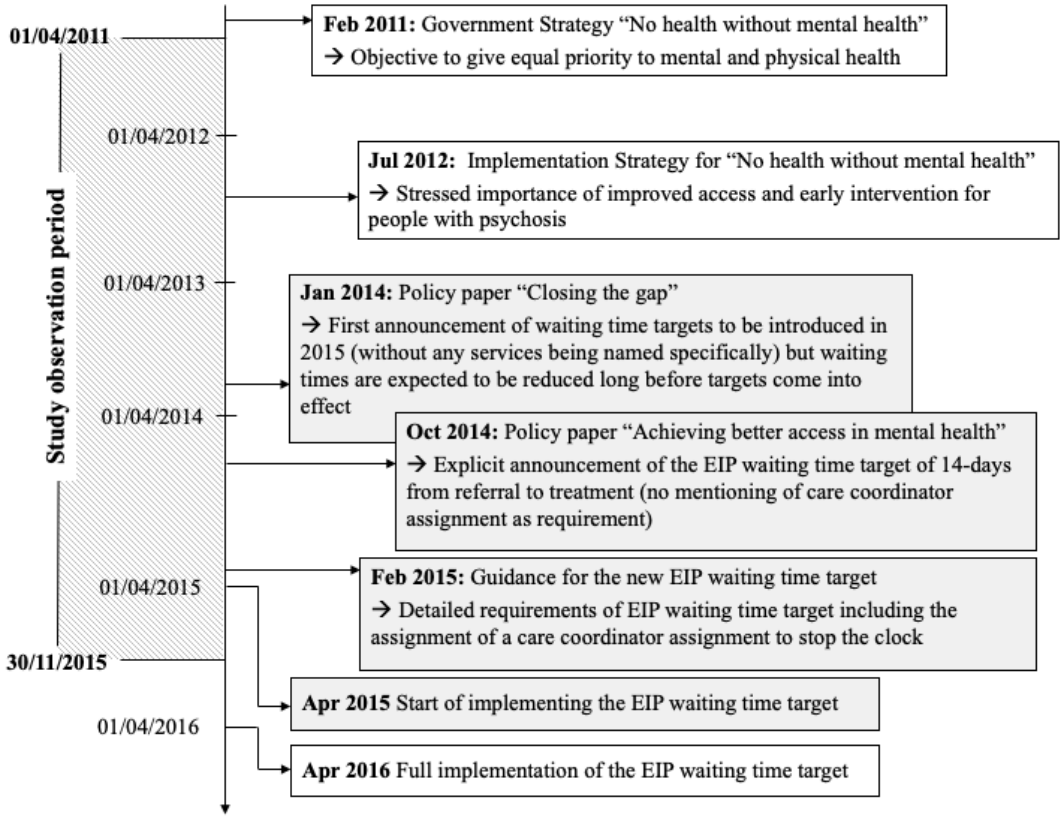


Figure 4.1: Timeline of announcements related to the mental health service reform in England 2011 to 2016

This chapter aims to explore whether referral-to-treatment waiting time changed over the years leading towards the EIP target implementation. If providers adapted behaviour in anticipation of the policy change, we expect referral-to-treatment waiting time for EIP patients to decrease from October 2014 after the first announcement the EIP target. We then focus on two different types of unintended effects that may have accompanied the change in waiting times: re-prioritisation and gaming. First, providers may change the order of treating patients and prioritise those that are most likely to breach the target regardless of the clinical urgency of treatment as a form of re-prioritisation (Appleby et al., 2005). Hence, we investigate whether changes in

waiting time varied by the patient's referral priority and disease severity as a form of re-prioritisation. Second, providers may focus on a reduction of the targeted part of the patient's total waiting time by prolonging non-targeted parts resulting in an unchanged or even longer waiting time in total (Kelman and Friedman, 2009). We investigate this form of gaming behaviour by differentiating two different waiting times: (1) time from referral to EIP caseload acceptance (time-to-EIP); and (2) time from EIP caseload acceptance to the assignment of a care coordinator (time-to-CCASS). We exploit the fact that in October 2014 little detail was given about what constitutes "NICE approved treatment" that is required to stop the waiting time clock. Assuming that providers expected the acceptance onto the EIP caseload to be sufficient to stop the waiting time clock, gaming behaviour would be present if time-to-EIP improved from October 2014 whereas time-to-CCASS increased at the same time. Only from February 2015 after detailed information revealed that time-to-CCASS will be monitored, we expect time-to-CCASS to have improved as well. Waiting time can only be measured retrospectively, i.e. once a patient finished waiting (or started treatment). We use duration analysis techniques to overcome the challenge of right-censoring as it allows us to account for a patient's time in the study (Rabe-Hesketh and Skrondal, 2012).

Our work is motivated by a principal-agent economic model in the presence of asymmetric information as introduced in section 1.2.3. As the policymaker can only observe the provider's effort by monitoring his target performance, providers – being under pressure to meet different objectives with limited resources – may take actions to make the target performance seem better than it actually is (Smith, 1995). These unintended consequences can occur in various ways and evidence is mixed. Appleby et al. (2005) found that additional admissions to meet the 15-month waiting time target in trauma and orthopaedics did not result in a distortion of clinical priorities. Januleviciute et al. (2013) compared the effects of waiting time targets with explicit prioritisation rules (Norway) to those without (Scotland). In both cases, waiting times did not change for high priority patients. However, explicit prioritisation led to a reduction in waiting times for low priority patients. Propper et al. (2010) could not identify any re-prioritisation of patients to meet the waiting time target for elective hospital admissions in England. In contrast, Nikolova, Sinko and Sutton (2015) found some evidence for re-ordering of patients as a consequence of the waiting time target

for elective surgery in Scotland. Fewer papers investigated gaming behaviour. Kelman and Friedman (2009) as well as Propper et al. (2010) provide evidence that improved target performance was neither associated with lower quality of care, nor with effort reduction in non-targeted activity such as waiting time or length of stay in other departments. Robinson et al. (2003) showed that time from referral to first hospital appointment improved as response to the 2-week waiting time target for women referred urgently with suspected breast cancer in England. However, waiting time from first appointment to treatment increased and consequently total waiting times changed very little. Similarly, Marques et al. (2014) showed that although the targeted inpatient waiting time in elective surgery in England decreased, total waiting time did not improve. Patients with shorter inpatient waits spent a longer time waiting prior to the inclusion in the waiting list.

We add to this evidence by making a number of important contributions. First, we extend existing evidence on the response of providers to waiting time targets to the mental health context where providers act in different market structures (e.g. payment systems) which may lead to different responses to targets. At the same time, we contribute to the still ongoing debate about the extent of unintended effects of enforced performance targets – particularly in the context where targets are not accompanied with direct penalties for providers. Second, we use duration analysis methods which allow us to overcome the challenge of right-censoring which typically occurs in waiting time studies. Finally, this chapter provides important evidence for the future evaluation of the EIP target policy itself. If providers adapted behaviour in anticipation of the policy change, the evaluation of the actual policy has to account for this anticipation.

4.2 Methods

4.2.1 Dataset and sampling

The analysis in this chapter uses secondary patient-level data from the MHLDDS (see section 1.4.1). We identified 10,744 patients aged 16 to 64 that had an EIP episode and a related referral between April 2011 and November 2015. We aimed to look at first-episode psychosis patients as treatment patterns may change for repeated episodes.

Hence, we excluded patients that were in contact with psychosis related services in the previous six months. Among those excluded, 336 (3%) had a previous EIP episode and 1,356 (13%) had a psychosis related care cluster episode. We further excluded 70 (1%) patients that were treated at independent providers as their care pathways may be different from NHS providers. To ensure that patient groups per provider are large enough, we excluded 16 (0.2%) patients that were with providers treating less than ten patients in our sample. Out of the remaining 8,966 patients, we excluded 2,800 (31%) with missing HoNOS information which we needed as a measure of priority. The final study sample consisted of 6,166 patients treated at 42 mental health providers.

4.2.2 Outcomes, covariates and empirical analysis

We use duration analysis to investigate changes in referral-to-treatment waiting time over the five-year study period. Referral-to-treatment time measures the time a patient waited from referral to the assignment of a care coordinator or censoring (see section 1.4.2 for more details). We included only referrals that have been accepted for action by the receiving provider. To analyse potential gaming behaviour, we differentiate two components of the referral-to-treatment time: (1) time from referral to acceptance onto EIP caseload (time-to-EIP); and (2) time from acceptance onto EIP caseload to the assignment of a care coordinator (time-to-CCASS). Censoring can occur because patients drop out of the study due to death (after acceptance of referral), because they are no longer in need of treatment, or because the care coordinator was assigned after the end of the study period. We interpret the number of days patients waited, as continuous time-to-event data. The start point, where analysis time $t = 0$, is defined by the patient's date of referral. We assume that the process by which patients entered the study is random at patient level.

First, we apply non-parametric methods to estimate survivor, hazard, and cumulative hazard functions of time-to-treatment (Aalen, 1978; Nelson, 1972; Kaplan and Meier, 1958). Second, we employ a stratified Cox regression model (Cox, 1972) to estimate the effect of the patient's referral year, referral priority and severity of condition on the probability (or hazard) of getting treated, conditional on a number of possible confounders. We parameterise the conditional hazard function, $h(t|x_{ij})$, for patient

$i = 1, \ldots, n$, living in small area $j = 1, \ldots, J$ as defined in Equation 4.1 and Equation 4.2.

Equation 4.1: $$h(t|x_{ij}) = h_{0s}(t)\exp\left(\beta_1 T_{ij}^1 + \beta_2 P_{ij} + \beta_3 (T_{ij}^1 \times P_{ij}) + \beta_4 X_{ij}\right)$$

Equation 4.2: $$h(t|x_{ij}) = h_{0s}(t)\exp\left(\delta_1 T_{ij}^2 + \delta_2 P_{ij} + \delta_3 (T_{ij}^2 \times P_{ij}) + \delta_4 X_{ij}\right)$$

Time to treatment is indicated by t for which we distinguish time-to-EIP, time-to-CCASS and referral-to-treatment. We define a variable T_{ij}^1 which equals 1 if a patient was referred to EIP after the first announcement of the EIP target policy and before the second announcement of the care coordinator assignment (October 2014 to January 2015). T_{ij}^2 equals 1 if the patient was referred to EIP after the second announcement (February 2015 to November 2015).

P_{ij} indicates the patient's priority status which we measure in two different ways. First, we use the patient's referral status which indicates the urgency of the referral. Patients are defined as high priority (P_{ij}=1) if the receiving mental health provider accepted the referral for immediate action and as low priority (P_{ij}=0) if the patient was placed on the appointment waiting list. Waiting time in this context may be related to both, the time until the acceptance onto the EIP caseload (i.e. start of the EIP episode) and the assignment of the care coordinator. Second, we use the HoNOS item 6 as a measure of severity of psychotic symptoms (see sections 1.4.1 and 3.3). We defined patients as high severity (P_{ij}=1) if the HoNOS 6 score was 3 or 4 and low severity (P_{ij}=0) if the HoNOS 6 score was below 3. HoNOS measurements must have taken place within a maximum window of 30 days before or after the start of the EIP treatment.

We control for patient characteristics X_{ij} that have been found to be related with patient waiting times (O'Donoghue et al., 2016; Apeldoorn et al., 2014). We consider general demographic factors such as age, gender, ethnicity, and marital status as well as socioeconomic variables. For the latter, we used accommodation and employment status at a patient-level and socioeconomic deprivation at LSOA level based on the IMD (see section 1.4.1). Furthermore, we controlled for the primary diagnosis during the EIP care and the source of referral (e.g. general practitioner, self-referral, or justice system).

We estimate Equation 4.1 and Equation 4.2 separately for time-to-EIP, time-to-CCASS and referral-to-treatment time. β_1 and δ_1 measure the estimated probability (hazard) of getting treated when being referred between October 2014 and January 2015 and between February and November 2015 respectively. We expect β_1 and δ_1 to be positive for referral-to-treatment if providers respond to the target policy by reducing waiting time. Potential gaming behaviour would be observed if β_1 for time-to-EIP i.e. the hazard of getting accepted onto the EIP caseload increased between October 2014 and January 2015 whereas β_1 for time-to-CCASS i.e. the hazard of getting assigned a care coordinator decreased at the same time. Only between February 2015 and November 2015, we expect time-to-CCASS to improve which would be indicated by a positive δ_1.

The interaction of T_{ij}^1 (and T_{ij}^2) and P_{ij} identifies high priority/severity patients being referred after the first (and second) policy announcement. Hence, the coefficients β_3 and δ_3 measure to what extent the hazard of getting treated changed for high priority patients after the first and second policy announcement respectively. We do not have an a priori assumption about the extent of prioritisation pre-policy. Any change in prioritisation over time ("good" or "bad") and hence a significant β_3 or δ_3 would indicate the presence of re-prioritisation.

The Cox model identifies the effect of each covariate on time to treatment in terms of hazard ratios $\widehat{HR} = \exp(\hat{\beta})$ which we will present in the result section alongside the estimated coefficients. The stratum-specific baseline hazard, $h_{0s}(t)$, indicates the probability of being treated when all covariates are zero. We assume there are $s = 1,2, \dots, S$ strata with $S = 42$ corresponding to the number of mental health providers in our sample. Stratification by provider allows us to control for any unobserved provider heterogeneity. At the same time, it is coherent with our study objective as we are not interested in provider effects per se but want to control for them. Estimated hazard ratios must be interpreted as the within-provider ratio of hazards (Rabe-Hesketh and Skrondal, 2012).

The strength of the Cox model is that is does not require the parameterisation of the baseline hazard function. However, this only allows estimating the probability of survivorship rather than absolute survival time in days. To predict adjusted variations

in median waiting times, we additionally model waiting time variation using accelerated failure time (ACF) models. ACF models use a parametric approach to estimate baseline survivorship over time which is assumed to follow a known distribution. We use the Akaike's information criterion (AIC) (Akaike, 1970) and the Bayesian information criterion (BIC) (Schwarz, 1978) to choose the best fitting distribution among the most commonly used exponential, Weibull, normal, logistic distribution.

4.2.3 Validation checks and assessing proportional hazards

We used Efron's method (Efron, 1977) to handle ties which was found to gain closer results to the exact partial likelihood (Hosmer, 2008). We evaluated the overall fit of the final models by plotting the Cox-Snell residuals against the Nelson-Aalen cumulative hazard function. The Cox model assumes that ratio of the hazards for any two individuals is constant over time (proportional). We assessed the proportional hazards assumption by performing overall as well as covariate specific score tests based on the (scaled) Schoenfeld residuals (Hosmer, 2008). We examined the magnitude of time dependencies by visually analysing scatterplots of the scaled Schoenfeld residuals as well as log-log plots. If the hazards are proportional, residuals should be scattered randomly around zero and log-log plots should be approximately parallel. Further, we investigated time varying interactions in an extended Cox model. Violations of the assumption are, however, only critical if (1) time dependencies are strong, or (2) there is a theoretical interest in analysing the time dependencies (Allison, 2010). Case (2) can be ignored given our study objective. In case of weak time dependencies, results can be interpreted as average effects over the range of times observed in the data (Allison, 2014).

4.3 Results

4.3.1 Descriptive statistics

Descriptive sample characteristics are summarised in Table 4.1 separately for the three time periods before, after first and after second policy announcement. For all 6,166 patients being referred to an EIP service within the study period, we observe the

complete time-to-EIP (100%).[5] The median time-to-EIP was one day in all three time periods with the mean decreasing from 29 to 8 days over time. 2,042 patients (33%) were right censored in their time-to-CCASS. The median time-to-CCASS was 52 days before the first announcement, increased to 64 days after the first announcement and decreased to 38 days after the second announcement. A similar pattern can be observed for the referral-to-treatment time.

Table 4.1: Summary statistics

Sample characteristics	Apr11-Sept14	Oct14-Jan15	Feb15-Nov15
n	5,178	321	667
Time-to-EIP, median (mean, SD)	1 (29, 98)	1 (12, 35)	1 (8, 21)
Time-to-CCASS, median (mean, SD)	52 (369, 501)	64 (161, 160)	38 (79, 88)
Referral-to-treatment, median (mean, SD)	71 (397, 514)	82 (172, 162)	44 (85, 89)
High priority, n (%)	3,184 (61.5)	178 (55.5)	381 (57.1)
High severity, n (%)	1,301 (25.1)	115 (35.8)	251 (37.6)
HoNOS score, range 0-48			
for low severity, mean (SD)	10.7 (5.4)	11.2 (5.5)	11.3 (5.4)
for high severity, mean (SD)	16.9 (6.0)	16.8 (5.4)	16.7 (5.9)
HoNOS 6 score, range 0-4			
for low severity, mean (SD)	0.9 (0.9)	1.0 (0.9)	1.1 (0.9)
for high severity, mean (SD)	3.2 (0.4)	3.2 (0.4)	3.2 (0.4)
Age 16 to 35, n (%)	5,023 (97.0)	315 (98.1)	649 (97.3)
Male, n (%)	3,275 (63.3)	192 (59.8)	423 (63.4)
Single, n (%)	3,883 (75.0)	205 (63.9)	383 (57.4)
White ethnicity, n (%)	3,628 (70.1)	216 (67.3)	409 (61.3)
Unemployed, n (%)	1,616 (31.2)	71 (22.1)	150 (22.5)
Least deprived quintile, n (%)	581 (11.2)	42 (13.1)	67 (10.0)
Most deprived quintile, n (%)	1,829 (35.3)	100 (31.2)	233 (34.9)
No fixed accommodation, n (%)	751 (14.5)	34 (10.6)	56 (8.4)

55.5 to 61.5% of patients were referred with high priority and 25.1 to 37.6% were classified as high severity. The average HoNOS score was 10.7 to 11.3 for low severity patients and 16.7 to 16.9 for high severity patients. The demographic and socioeconomic composition of our sample is comparable to other first-episode psychosis cohorts studied in the past (Kirkbride et al., 2017; Tsiachristas et al., 2016). The majority of our sample (97.0-98.1%) was aged 16 to 35 at time of referral, of male

[5] We were only able to identify patients relevant to our study if they had an EIP episode within the study period. Hence, right censoring of time-to-EIP was not possible by definition as for all patients we observed the start of the EIP episode.

gender (59.8-63.4%), and single (57.4-75.0%). Compared to the general population, we observe a relatively high proportion of people being unemployed (22.1-31.2%), living in the most deprived neighbourhoods (31.2-35.3%), and having no fixed accommodation (8.4-14.5%).

4.3.2 Graphical analysis of non-parametric functions

Figure 4.2 presents non-parametric functions for time-to-EIP, time-to-CCASS and referral-to-treatment time. The survival for time-to-EIP, panel 1(a), descends sharply for approximately the first 100 days and then slowly towards the end of the distribution. The descent is approximately constant from 400 days. The initial steep descent is a result of a relatively higher probability to be treated in the first days following referral. The treatment rate then decreases and remains at about the same level for the remainder of the follow-up period. The survival function based on time-to-CCASS in panel 2(a) shows a similar pattern. Survival diminishes fast, though not as fast as in the former case, within the first 100 days and then less than proportionally. The survival function finishes at about 46% reflecting the fact that 46% of the sample had a time-to-CCASS that was censored as they were still waiting for a care coordinator assignment at the end of the study period. The hazard of being treated is highest at the beginning of the duration and decreases sharply after. Both the hazard of EIP acceptance and of care coordinator assignment are at their minimum at around 400 days. After this point, the hazard of care coordinator assignment continues to decrease at a smaller rate whereas the hazard of EIP acceptance appears to rise again.

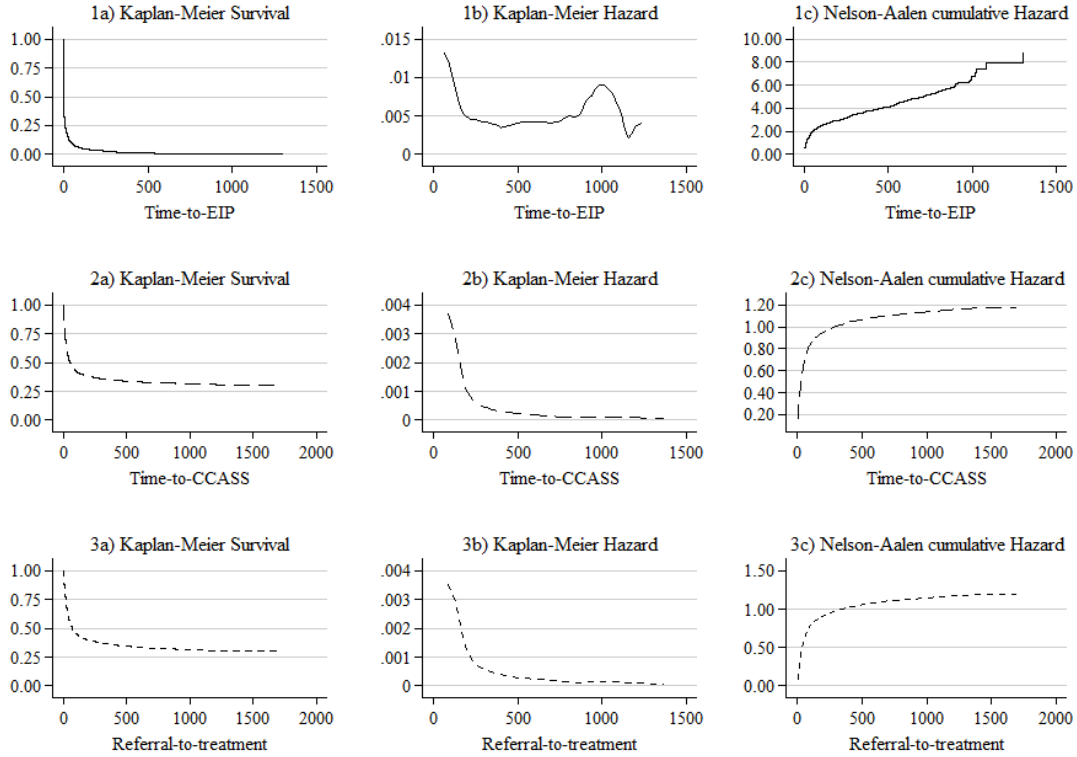
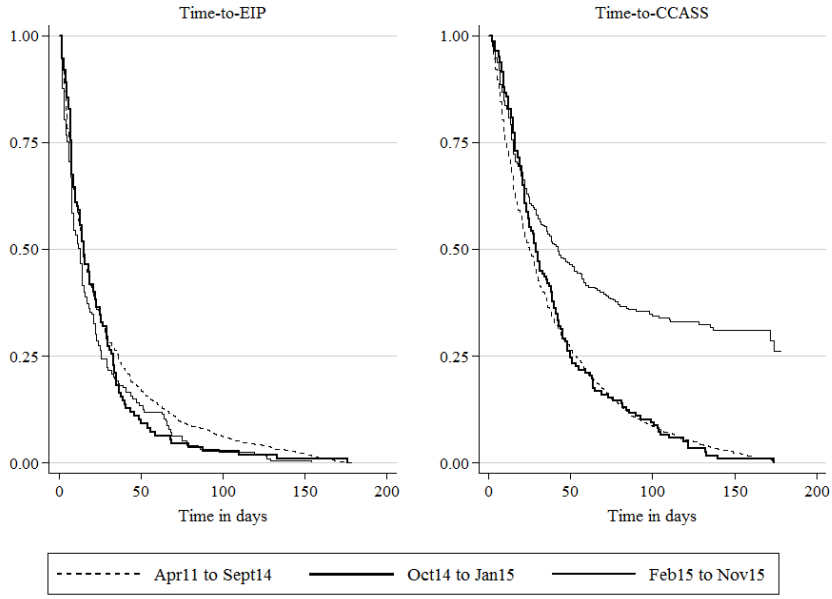


Figure 4.2: Non-parametric functions for time-to-EIP (1a-c), time-to-CCASS (2a-c) and referral-to-treatment time (3a-c)

Figure 4.3 compares the Kaplan-Meier survival curves for time-to-EIP and time-to-CCASS in the three different time periods. A steeper survival curve indicates a higher hazard of getting treated and thus a shorter time to treatment. For a time-to-EIP below 35 days, there were no apparent differences in survival between the pre-announcement period (Apr11-Sept14) and the period after the first announcement (Oct14-Jan15). However, for time-to-EIP above 35 days survival was steeper for the first announcement period and remained almost as steep in the second announcement period (Feb15-Nov15). Time-to-CCASS at the lower end of the distribution was shortest for the pre-announcement period (Apr11-Sept14). At the higher end of the distribution (50 days and above) time-to-CCASS for pre-announcement and first announcement period almost overlapped. However, there was a clear decrease in time-to-CCASS after the second policy announcement particularly for time-to-CCASS of 40 days and above.

The graphical analysis indicates that providers responded to both of the policy announcements. Time-to-EIP seems to have improved after the first announcement

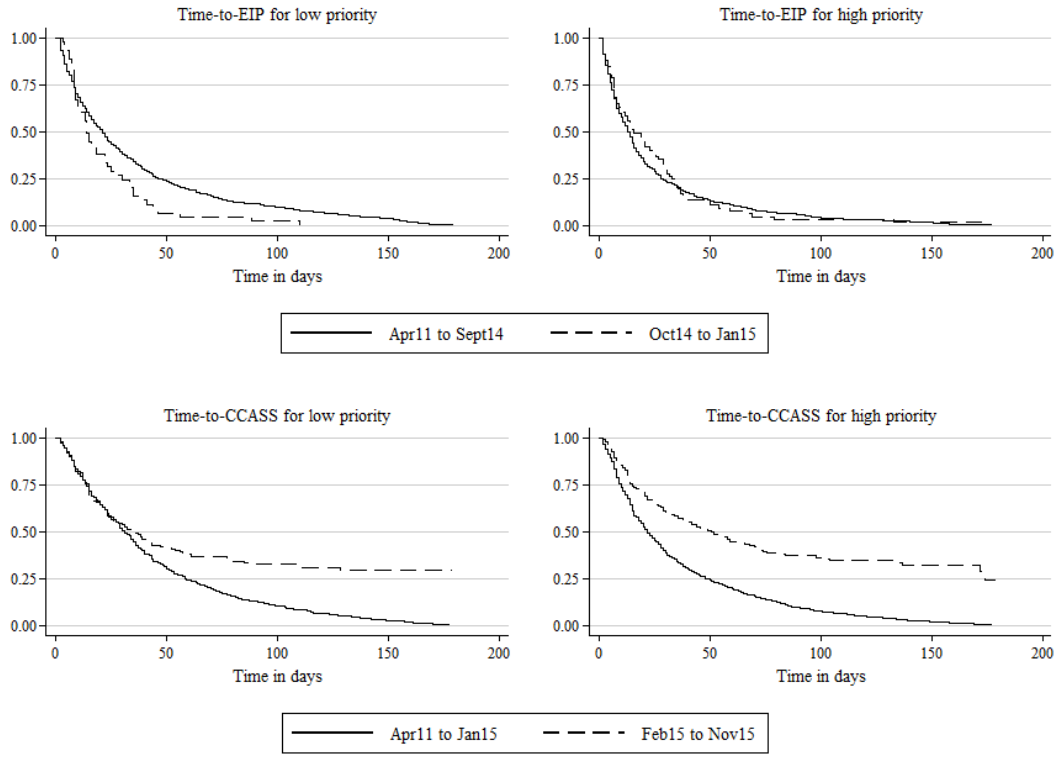
and time-to-CCASS improved after the second announcement. At the same time, we find no evidence of gaming. Time-to-CCASS did not seem to be affected negatively by the improvement in time-to-EIP after the first announcement.



Note: Graph truncated at time >1 and <180 days.

Figure 4.3: Kaplan-Meier survival curves for time-to-EIP and time-to-CCASS for different time periods

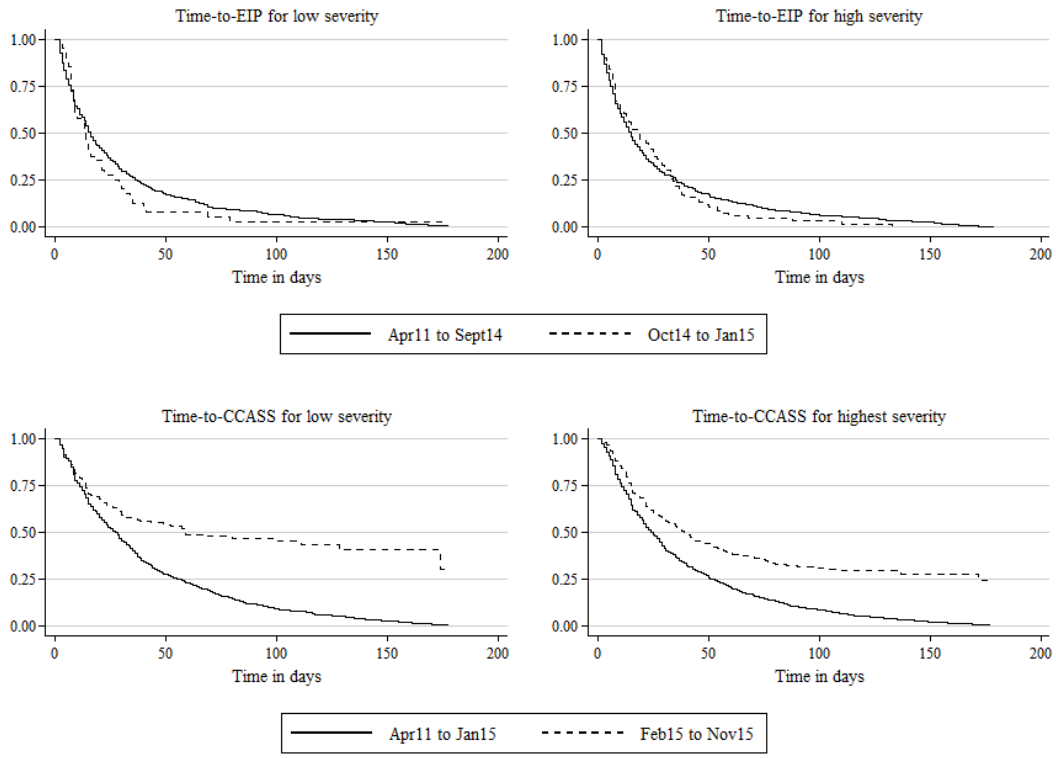
Figure 4.4 plots the survival curves for time-to-EIP and time-to-CCASS for low and high priority patients. Whereas the upper panel compares time-to-EIP pre-announcement (solid line) and after the first announcement (dashed line), the lower panel compares time-to-CCASS before and after the second announcement. We observe that time-to-EIP for low priority patients improved after the first announcement as the survival curve got steeper – particularly for time-to-EIP above 25 days. The same effect cannot be observed for high priority patients as both survival curves mostly overlap. Hence, providers seem to have improved time-to-EIP in response to the policy announcement but to the benefit of low priority rather than high priority patients. For time-to-CCASS we find the opposite. Time-to-CASS increased after the second policy-announcement for both low and high priority patients. The increase seems to have been larger for high priority patients.



Note: Graph truncated at time >1 and <180 days.

Figure 4.4: Kaplan-Meier survival curves for time-to-EIP and time-to-CCASS by priority

A very similar picture can be seen in Figure 4.5 here we compare survival curves for time-to-EIP and time-to-CCASS by severity. Again, we observe improvements in time-to-EIP for low severity patients after the first announcement but not for high severity ones. At the same time, time-to-CASS increased after the second policy-announcement for both low and high severity patients. But in contrast to before, this increase seems to have been larger for low severity patients compared to high severity ones.



Note: Graph truncated at time >1 and <180 days.

Figure 4.5: Kaplan-Meier survival curves for time-to-EIP and time-to-CCASS by severity

4.3.3 Regression results

Table 4.2 and Table 4.3 report the results from the stratified Cox models which control for patient characteristics and provider heterogeneity. The dependent variable is time-to-EIP for model (1), time-to-CCASS for model (2) and referral-to-treatment time for model (3). In panel (a) the policy announcement indicator was interacted with the referral priority indicator and in panel (b) with the severity dummy. Positive coefficients indicate an increase in the hazard of getting treated and hence a reduction in waiting time.

Results in Table 4.2 are based on Equation 4.1 which compares the pre-announcement period (Apr11-Sept14) with the period after the first announcement (Oct14-Jan15). Low priority patients significantly improved in time-to-EIP after the first announcement (hazard ratio (HR)=1.19). A similar but slightly smaller effect can be observed when using severity as measure (HR=1.15). At the same time, time-to-CCASS and referral-to-treatment time were not negatively affected by the first policy announcement which indicates no evidence of gaming. In fact, referral-to-treatment

time seemed to even have improved for high priority patients in the first post-announcement period (HR=1.35).

High priority patients had a longer time-to-EIP in the pre-announcement period (HR=0.87) compared to low priority patients. Improvements in time-to-EIP can only be observed for low priority and low severity patients whereas there was no significant change for high priority/severity ones. By severity, we observe some effective prioritisation in the pre-announcement period as high severity patients had a shorter time-to-CCASS (HR=1.17) and referral-to-treatment time (HR=1.19) than low severity ones. This prioritisation cannot be observed after the first policy announcement. Overall, there is no evidence of re-prioritisation associated with the first policy announcement as no patient group improved to the expenses of another patient group.

Table 4.3 presents results based on Equation 4.2 which compares the period before the second announcement (Apr11-Jan15) with the one after (Feb15-Nov15). Time -to-EIP continued to improve for low priority and low severity patients after the second announcement (HR=1.19 and 1.15 respectively). But this time, also time-to-CCASS and referral-to-treatment time improved significantly for low priority patients (HR=1.38). High priority patients did not improve in a similar way. When looking at prioritisation effects by severity the picture is the opposite around. High severity patients had a shorter time-to-CCASS (HR=1.18) and referral-to-treatment time (HR=1.21) before the second policy announcement compared to low severity patients. Both time measures decreased after the second announcement (HR=1.42/HR=1.41) without any effects on low severity patients.

Table 4.2: Stratified Cox regression results comparing Apr11-Sep14 and Oct14-Jan15

	(1) Time-to-EIP			(2) Time-to-CCASS			(3) Referral-to-treatment		
	Coeff.	Std. Err.	HR	Coeff.	Std. Err.	HR	Coeff.	Std. Err.	HR
a) Priority									
Low priority from Apr11-Sept14	ref. cat.		1.00	ref. cat.		1.00	ref. cat.		1.00
Low priority from Oct14-Jan15	0.178**	(0.065)	1.19	-0.139	(0.110)	0.86	-0.158	(0.114)	0.85
High priority from Apr11-Sept14	-0.141*	(0.063)	0.87	-0.046	(0.085)	0.95	-0.069	(0.083)	0.93
High priority from Oct14-Jan15	-0.073	(0.090)	0.93	0.234	(0.150)	1.27	0.300*	(0.152)	1.35
b) Severity									
Low severity from Apr11-Sept14	ref. cat.		1.00	ref. cat.		1.00	ref. cat.		1.00
Low severity from Oct14-Jan15	0.141*	(0.058)	1.15	-0.061	(0.099)	0.94	-0.059	(0.100)	0.94
High severity from Apr11-Sept14	0.029	(0.029)	1.03	0.156***	(0.037)	1.17	0.174***	(0.038)	1.19
High severity from Oct14-Jan15	-0.012	(0.092)	0.99	0.137	(0.150)	1.15	0.170	(0.151)	1.19

Note: * p<0.05, ** p<0.01, *** p<0.001. Regression based on Equation 4.1. Reported are standard errors (Std. Err.) and hazard ratios (HR).

Table 4.3: Stratified Cox regression results comparing Apr11-Jan15 and Feb15-Nov15

	(1) Time-to-EIP			(2) Time-to-CCASS			(3) Referral-to-treatment		
	Coeff.	Std. Err.	HR	Coeff.	Std. Err.	HR	Coeff.	Std. Err.	HR
a) Priority									
Low priority from Apr11-Jan15	ref. cat.		1.00	ref. cat.		1.00	ref. cat.		1.00
Low priority from Feb15-Nov15	0.156***	(0.044)	1.16	0.324***	(0.088)	1.38	0.327***	(0.087)	1.38
High priority from Apr11-Jan15	-0.154**	(0.058)	0.86	-0.001	(0.079)	1.00	-0.029	(0.078)	0.97
High priority from Feb15-Nov15	-0.045	(0.062)	1.05	-0.154	(0.114)	0.86	-0.109	(0.114)	0.90
b) Severity									
Low severity from Apr11-Sept14	ref. cat.		1.00	ref. cat.		1.00	ref. cat.		1.00
Low severity from Oct14-Jan15	0.218***	(0.041)	1.24	0.078	(0.081)	1.08	0.111	(0.080)	1.12
High severity from Apr11-Sept14	0.035	(0.028)	1.04	0.169***	(0.036)	1.18	0.190***	(0.036)	1.21
High severity from Oct14-Jan15	-0.096	(0.062)	0.91	0.347**	(0.110)	1.42	0.341**	(0.112)	1.41

Note: * p<0.05, ** p<0.01, *** p<0.001. Regression based on Equation 4.2. Reported are standard errors (Std. Err.) and hazard ratios (HR).

In Table 4.4, we report the predicted median waiting times when comparing the pre-announcement period with the period after the first announcement. The results are based on the ACF models for which we assumed a log-logistic distribution for time-to-EIP and a log-normal distribution for time-to-CCASS and referral-to-treatment time. Differences in median time-to-EIP are very small between the groups ranging between 2.5 and 2.8 days. The shortest median time-to-EIP of 2.5 days is observed for high severity patients from October 2014 to January 2015. Differences in predicted median time-to-CCASS are larger although we note that the predictions are generally quite large compared to the observed data. Median time-to-CCASS ranges between 129.0 days for high severity patients from October 2014 to January 2015 and 416.7 days for low priority patients from April 2011 to September 2014. Overall, median time to treatment for all three measures decreased after the first policy announcement. Hence, we find no evidence of gaming behaviour. Also, high priority/severity patients had consistently shorter time to treatment (except a slightly higher time-to-EIP for high priority patients) not only before but also after the first policy announcement which implies that there is no evidence of re-prioritisation.

We observe a very similar picture when looking at the difference in predicted median time to treatment before and after the second policy announcement as shown in Table 4.5. Median time-to-EIP decreased after the second policy announcement from 2.8 to 2.7 days, median time-to-CCASS from 345.6 to 216.9 days and referral-to-treatment time from 327.7 to 253.3 days when interacting priority with the announcement indicator. Overall, high priority as well as high severity had shorter time to treatment. However, we find some evidence of re-prioritisation by referral priority. Whereas time to treatment decreased after the second policy announcement for low priority patients it increased at the same time for high priority patients. We observe a similar re-prioritisation effect when using severity as priority measure for time-to-EIP but not for time-to-CCASS and referral-to-treatment time.

Table 4.4: Predicted median waiting times based on Accelerated Failure Models for Apr11-Sept14 compared to Oct14-Jan15

	(1) Time-to-EIP		(2) Time-to-CCASS		(3) Referral-to-treatment	
	Median	95% Conf. Int.	Median	95% Conf. Int.	Median	95% Conf. Int.
(a) Priority						
Apr11-Sept14	2.8	[2.6-2.9]	345.7	[267.3-424.0]	372.7	[301.1-444.5]
Oct14-Jan15	2.7	[2.3-3.2]	216.9	[124.5-309.4]	253.3	[159.3-347.3]
Low priority	2.7	[2.5-3.0]	401.4	[302.4-500.4]	424.7	[335.4-514.0]
High priority	2.8	[2.6-3.0]	280.7	[213.2-348.3]	311.7	[247.6-375.8]
Low priority from Apr11-Sept14	2.7	[2.5-3.0]	416.7	[311.8-521.6]	437.0	[342.9-531.0]
Low priority from Oct14-Jan15	2.8	[2.6-3.0]	291.7	[220.3-327.0]	322.4	[254.9-390.0]
High priority from Apr11-Sept14	2.6	[2.0-3.2]	263.1	[102.4-423.7]	308.6	[146.1-471.1]
High priority from Oct14-Jan15	2.8	[2.2-3.4]	181.8	[84.5-279.1]	210.0	[110.8-309.2]
(b) Severity						
Apr11-Sept14	2.8	[2.6-2.9]	346.2	[268.2-424.2]	374.1	[302.4-445.8]
Oct14-Jan15	2.7	[2.3-3.2]	255.7	[130.5-380.9]	306.2	[175.2-437.1]
Low severity	2.8	[2.6-3.0]	360.7	[280.3-441.1]	394.0	[319.0-469.1]
High severity	2.6	[2.4-2.9]	232.9	[171.7-294.1]	252.2	[194.9-309.5]
Low severity from Apr11-Sept14	2.8	[2.6-3.0]	369.3	[285.6-453.1]	399.4	[322.1-476.6]
Low severity from Oct14-Jan15	2.6	[2.4-2.9]	244.6	[178.8-310.4]	264.0	[202.5-325.5]
High severity from Apr11-Sept14	2.8	[2.2-3.4]	284.5	[133.8-435.2]	344.0	[185.8-502.3]
High severity from Oct14-Jan15	2.5	[1.9-3.1]	129.0	[51.4-206.7]	141.1	[66.3-215.9]

Table 4.5: Predicted median waiting times based on Accelerated Failure Models for Apr11-Jan15 compared to Feb15-Nov15

	(1) Time-to-EIP		(2) Time-to-CCASS		(3) Referral-to-treatment	
	Median	95% Conf. Int.	Median	95% Conf. Int.	Median	95% Conf. Int.
(a) Priority						
Apr11-Jan15	2.7	[2.5-2.9]	330.9	[253.1-408.7]	353.3	[288.7-417.8]
Feb15-Nov15	2.0	[1.8-2.2]	127.5	[86.8-168.2]	140.0	[103.5-176.5]
Low priority	2.6	[2.4-2.8]	339.5	[260.2-418.7]	359.1	[290.8-427.4]
High priority	2.6	[2.5-2.8]	246.7	[189.1-304.3]	273.6	[222.5-324.6]
Low priority from Apr11-Jan15	2.7	[2.5-2.9]	339.5	[295.1-512.3]	419.5	[332.2-506.9]
Low priority from Feb15-Nov15	2.7	[2.5-2.9]	246.7	[205.8-351.9]	303.2	[242.0-364.4]
High priority from Apr11-Jan15	1.8	[1.6-2.0]	116.6	[68.3-164.9]	118.9	[77.1-160.6]
High priority from Feb15-Nov15	2.6	[1.9-2.4]	135.3	[83.4-187.1]	156.0	[105.3-206.7]
(b) Severity						
Apr11-Jan15	2.7	[2.5-2.9]	326.5	[250.4-402.6]	349.9	[286.3-413.4]
Feb15-Nov15	2.0	[1.8-2.2]	161.6	[102.2-221.0]	174.0	[119.7-228.3]
Low severity	2.7	[2.5-2.8]	315.5	[249.6-381.4]	342.7	[284.7-400.7]
High severity	2.5	[2.3-2.8]	188.1	[143.0-233.2]	205.3	[163.8-246.8]
Low severity from Apr11-Jan15	2.7	[2.6-2.9]	352.1	[270.9-433.2]	378.3	[309.3-447.4]
Low severity from Feb15-Nov15	2.6	[2.3-2.8]	221.0	[161.9-280.1]	236.2	[184.5-287.8]
High severity from Apr11-Jan15	1.9	[1.7-2.2]	183.8	[112.7-254.9]	197.6	[131.3-263.9]
High severity from Feb15-Nov15	2.2	[1.8-2.5]	69.8	[44.1-95.6]	79.7	[54.3-105.0]

4.3.4 Validation of the approach

In the following we discuss the results from the validation checks detailed in section 4.2.3. In Appendix A13, we show that results are consistent independent of the choice of method to handle ties. The results of the global score tests – reported in Appendix A14 – indicate some evidence of non-proportional hazards. Particularly, in the models with time-to-CCASS as dependent variable the test showed significant p-values for all four functions of time. Hence, we visually assessed the magnitude of the non-proportionality in our two main explanatory variables, the post-announcement indicators. In Appendix A15, we plotted the scaled Schoenfeld residuals against each of the four functions of time. Residuals mostly appear to be randomly scattered around zero with the lowess smooth being roughly flat and horizontal. The log-log plots in Appendix A16 show that plots are mainly parallel and close to each other, except for some overlap for the second announcement period indicator based on the model of time-to-CCASS. The observed effect on both, time-to-EIP and time-to-CCASS, remained similar once we interact the main regressors with time in an extended Cox regression (see Appendix A17). Time interactions were partly significant but comparably small in magnitude. However, effects partly disappear after interacting the main regressors with log of time (see Appendix A18).

4.4 Discussion

The EIP waiting time target is the first of its kind in mental health service provision in England. We investigated whether waiting times changed already in anticipation of the policy change and to what extent this may have led to unintended consequences for patients.

Our findings showed that the first announcement of the EIP target in October 2014 was associated with a significant reduction in time-to-EIP. Furthermore, the second announcement of the care coordinator requirement in order to stop the waiting time clock in February 2015 was associated with a significant decrease in time-to-CCASS and also in total referral-to-treatment time. Hence, providers seem to have responded to the policy announcements and adapted behaviour in anticipation of the policy change. The absolute effect on median time-to-EIP was, however, very small (less than

a day). In contrast, the absolute effects on time-to-CCASS and referral-to-treatment time were quite substantial.

We found no evidence that this anticipation led to gaming behaviour as time-to-CCASS and referral-to-treatment time were not negatively affected by the improvements in time-to-EIP after the first announcement. We also find only little evidence of re-prioritisation and results from the non-parametric, semi-parametric and parametric analyses were partly contradictive and inconsistent. Although we observed differences in time to treatment by priority as well severity in the semi-parametric Cox regressions, there was no clear pattern apparent and no patient group benefitted at the expense of another. Overall, low priority/severity patients seemed to have benefitted from the improvements in time-to-EIP but without disadvantaging high priority/severity patients at the same time. For time-to-CCASS, we observed effective prioritisation by severity that did not change after the policy announcements. High severity patients had a shorter time-to-CCASS not only before but also after the second announcement. The parametric analyses indicated some re-prioritisation after the second policy-announcement as time-to-EIP decreased for low priority/severity patients but increased at the same time for high priority/severity patients. For time-to-CCASS and referral-to-treatment time, this re-prioritisation could only be observed for the priority but not the severity measure.

Three aspects may have contributed to the fact that patients seemed to be effectively prioritised by severity when assigning a care coordinator rather than accepting patients onto the caseload. First, time-to-CCASS was only focused on after the announcement of the care coordinator requirement in February 2015. Hence, a similar re-ordering of patients as for time-to-EIP may be observed in future data. Second, the patient's severity of condition is mostly unknown at the point of referral. Only after a period of assessment, clinicians can actually judge the patient's severity and hence prioritise patients accordingly. Consequently, we observe effective prioritisation by severity for time-to-CCASS that seems to not be affected by the efforts of providers to meet targets. In contrast, the observed discrimination of more severe patients facing a longer time-to-EIP may not only be due to provider's ambitions to meet targets but also an unintentional consequence of providers having insufficient information about the patient's severity at time of referral. Third, the patients themselves may cause the

longer time-to-EIP. We found that particularly problems with hallucinations and delusions are associated with longer time-to-EIP rather than any other dimensions of severity. The lack of insight into the illness or the fear of being stigmatised due to hallucinations and delusions may cause patients to not being able to take advantage of the referral and delaying follow-up appointments with EIP services. Hence, their time-to-EIP is longer than for other patients. It is, however, clearly stated in the EIP target guideline that it is the responsibility of the provider to actively encourage patients' attendance as this will not be taken into consideration when evaluating the providers performance against the EIP target.

Our methodological approach has a number of strengths in addressing our research question. A key issue in the analysis of waiting times is censoring. Since some patients are still waiting at the end of the study period, they would need to be excluded from conventional regression analyses. This is likely to produce a selection bias if the probability of waiting longer is not random. Duration analysis allows us to account for a patient's time in the study and it further offers the advantage of adequately modelling the skewed distribution of waiting time. As we use non- and semi-parametric survival methods, we do not impose any assumptions on the statistical distributions of the baseline hazard. We take care of systematic differences between providers by the stratification of the model.

We note some potential limitations to our results. First, the observed changes in waiting time over time may be caused by other factors rather than providers anticipating the policy change. We do however control for a large number of potential confounders. Second, our measures of priority and severity may be imperfect. Some unobserved component may also be captured in the other control variables. In this case, we would have underestimated the impact of priority and/or severity on time to treatment. Moreover, HoNOS measures for different patients were taken at different times during the patient's complete waiting experience. This may have overestimated the severity of patients for which we measure severity later if the patient's condition deteriorates while waiting and vice versa. However, this will have influenced our results only if there was a systematic difference in the timing of HoNOS measurements over the five years of study. Finally, our findings on the impact of severity on time to treatment may not be generalizable to patients with missing HoNOS. Although

differences in observed characteristics between the subsamples with and without HoNOS were very small, some unobserved factors may be related to both the fact of not having received a HoNOS measurement and the time-to-treatment.

4.5 Conclusions

Our study entails a number of important policy implications. First, the growing public awareness about the importance of early access to treatment for patients seeking EIP care was associated with improvements in patient waiting times. However, the growing political emphasis on waiting times may have put implicit clinical priorities at stake. Particularly at the early stages of the total waiting time, when providers are not yet able to judge the patient's severity of condition, more severely affected patients are in danger of being disadvantaged. Future work should aim to analyse other potential sources of undesired behavioural changes such as redistribution of resources across different services within a provider in order to meet targets.

5. Chapter: The effects of the EIP waiting time target

5.1 Introduction

As introduced in section 1.2.2, England has had a nationwide EIP implementation strategy from the early 2000s onwards (Joseph and Birchwood, 2005). However, EIP provision began to decline after initial funding (Marwaha et al., 2016) and first evidence of referral-to-treatment waiting times recently showed an increase in waiting between 2009 and 2013 (Kirkbride et al., 2017). In this context, the English government introduced the first waiting time target in mental health history (Department of Health, 2014a). From April 2015, 50% of patients being referred to an EIP service were expected to wait no longer than 14 days from referral to treatment. Waiting time targets are a common strategy to tackle excessive waiting times in a number of countries and areas of health care (Willcox et al., 2007). To date it is however unknown whether a comparable target within the mental health context can be similarly effective.

This chapter investigates the effects of the EIP target after the first six months of its implementation. We exploit the fact that patients with first-episode psychosis may receive care from two different service models: EIP care or standard community mental health care (standard care in the following). Whereas EIP patients are affected by the target policy, standard care patients are not and hence serve as our control group. Assuming that on average both groups would have common trends in the absence of the policy, the control group provides an estimate for the post-policy outcome of the treatment group had they not been affected by the target policy (Dimick and Ryan, 2014; Abadie, 2005). We use controls that had no access to EIP services within 15 kilometres travel distance. Travel time can be interpreted as a nonmonetary price for obtaining care and it has been found to be strongly associated with health care utilisation (Fortney et al., 2005). Psychotic care requires the patient to travel to services several times a week over a period of up to three years in order to receive various interventions. We assume that a patient who is actually eligible for EIP care but would have to face a long travel distance to receive it, would rather be treated by a comparable standard care service nearby. We assume this patient would not necessarily be different in terms of severity of condition and need of treatment. However, to ensure comparability between groups, we employ matching methods to control for observed characteristics (Heckman et al., 1998) with a difference-in-difference regression model

which further accounts for unobserved time-invariant components (Jones and Rice, 2011; Angrist and Pischke, 2009). We use coarsened exact matching (Blackwell, Iacus and King, 2009) and propensity score matching (Rosenbaum and Rubin, 1983) to show that results are robust against the choice of the matching method.

To our knowledge, we are the first to evaluate the impact of a waiting time target in the mental health care context. Evidence has shown that providers do respond to waiting time targets in line with its intended objective (Propper et al., 2010; Besley, Bevan and Burchardi, 2009; Propper et al., 2008). Studies are however limited to state-level analyses in the area of physical health care. We contribute to the existing literature in a number of ways. First, our study moves beyond the state-level by analysing patient individual waiting times. This allows us to control for potential changes in case mix over time and further assures that both groups have been exposed to the same institutional setting. We analyse the probability of waiting below target at patient level and aggregate waiting times at provider level to analyse changes at different percentiles of the waiting time distribution. Data at provider level further allow us to test for some unintended provider responses to the target policies which have been investigated in the past (Propper et al., 2010; Kelman and Friedman, 2009). Second, we choose a control group with no access to EIP services in a certain travel distance. For this, we create a novel dataset on the regional distribution of EIP and standard care services across England and calculate travel distances for patients. Third, we combine our difference-in-difference approach with non-parametric matching. Pre-processing the data through matching leads to less model dependence and reduced statistical bias in the regression analysis (Ho et al., 2007). Finally, the EIP target operates in a different institutional setting which may lead to different responses to performance targets. In contrast to single-event surgical procedures provided in hospitals, we focus on services which are provided by stand-alone multidisciplinary teams within the community that deliver treatment in regular sessions over a period of up to three years (NICE, 2015). Also, the need for treatment in the case of psychosis is urgent rather than elective. Unlike target policies in the past, the EIP target is not accompanied with aggressive penalties but relies on the response of providers to the publication of performance data. Hence, we provide evidence on provider's responses

to performance targets without direct financial penalties (Smith et al., 2009; Propper and Wilson, 2003).

Our work will be of relevance to policymakers as it informs the future development of the English target policy and its potential international adaptation. We do not only provide novel information about EIP service availability and travel distances within the English NHS but also reveal and compare waiting times for both EIP and standard care patients for a large national cohort of first-episode psychosis patients. Hence, this study contributes to a still ongoing discussion whether specialised EIP services are superior to standard care in providing early access to care (Marwaha et al., 2016).

5.2 Methods

5.2.1 Difference-in-difference model

We use a difference-in-difference approach at the patient level to extract the effect of the EIP target on the probability of waiting below target (Y). For patient i in provider p at time t, we estimate the model detailed in Equation 5.1.

Equation 5.1:

$$Y_{ipt} = \alpha + \beta POST_t + \theta TREAT_{ip} + \mu\left(TREAT_{ip} \times POST_t\right) + \gamma X_{ipt} + \sigma_p + \varepsilon_{ipt}$$

$TREAT_{ip}$ is a dummy variable indicating whether the patient received EIP care, and $POST_t$ is a dummy variable for whether the patient was referred in the post-policy period. X_{ipt} is a set of patient-level characteristics to account for time-varying differences in patient severity across the treatment and control groups and mitigate the effects of compositional changes over time. It contains the variables age, male, single, non-white, unemployed, no fixed accommodation, neighbourhood deprivation quintile, overall disease severity, severity of psychotic symptoms, schizophrenia diagnosis, first-episode psychosis cluster, referral priority and referral source. Fixed effects σ_p for 58 mental health providers control for any time-invariant differences. ε_{ipt} represents the idiosyncratic error.

The coefficient $\hat{\mu}$ yields the difference-in-difference estimate of the policy effect. It can be interpreted as the population average treatment effect which represents the expected gain from the target policy for an individual randomly selected from the

treated population (Jones and Rice, 2011). We expect the probability of EIP patients to wait below the target to increase in the post-policy period ($\hat{\mu} > 0$). We estimate Equation 5.1 as a linear probability model using ordinary least squares regression. For the linear probability model, the regression is inherently heteroscedastic which is why we use robust standard errors that are clustered at the provider level.

In a second step, we aggregate our data at the provider level with one observation per provider, per quarter and per treatment. We weighed each observation of provider p in quarter t and treatment by a weight equal to each provider's number of EIP and standard care patients in a given quarter as a share of all EIP and standard care patients in that quarter. We analyse the policy effect at different percentiles of the waiting time distribution using ordinary least squares regression. Further, we look at some potentially unintended effort substitution of providers due to the increased target pressure. Providers could, for example, decrease the length of treatment of existing patients or accept fewer patients onto the caseload in order to free up resources and use the additional resources to improve target performance. Therefore, we analyse changes in length of treatment and in the number of newly accepted patients onto the caseload.

5.2.2 Pre-processing the data through matching

The credibility of the difference-in-difference approach in identifying the policy effect depends on the comparability of the treatment and control group in terms of observed as well as unobserved characteristics. In our case the assignment to EIP and standard care is not random. Patients access services through various routes (Singh and Grange, 2006). Most commonly they will be referred by a health professional, or patients may self-refer. Whereas EIP services are exclusive to first-episode psychosis patients between the ages of 16 and 35, standard care is not limited to psychotic conditions and patients may enter services at all ages. Hence, we expect patients in the treatment group to be younger and having a more severe or further developed psychotic condition than standard care patients.

We use matching as a non-parametric method to balance the treatment and control group in terms of potentially confounding pre-treatment control variables before applying our regression model. We perform two different well-established matching

methods: coarsened exact matching (CEM) and propensity score matching (PSM). The CEM algorithm performs exact matching on coarsened data to determine matches. Coarsening means that substantively indistinguishable values are grouped together and get assigned the same numerical value. CEM then sorts all observations that have identical values for all the coarsened pre-treatment covariates into strata and discards all observations within any stratum that do not have at least one observation for each unique value of the treatment variable (Blackwell, Iacus and King, 2009). However, the more covariates there are to be matched, the less likely it is to find a suitable control unit. As a consequence, unmatched treatment units have to be excluded from the analysis and the estimated treatment effect is redefined to the area of common support (Jones and Rice, 2011). In contrast, PSM is an approximate matching method that identifies control units which are close to the treated unit in terms of the propensity score, i.e. the probability of being treated conditional on the covariates (Rosenbaum and Rubin, 1983). This less restrictive method allows for more treatment units to remain in the final estimation sample. We conduct a nearest-neighbour propensity score matching with replacement and enforcing common support (without caliper). Controls with identical (tied) propensity scores were also matched to the nearest neighbour.

In both approaches, we match on patient demographic factors (age, male, single, non-white, neighbourhood deprivation quintile) as well as on variables related to the patient's psychotic condition (severity of psychotic symptoms, schizophrenia diagnosis, first-episode psychosis cluster). For the two continuous variables (age and HoNOS score as severity measure), we use the automated coarsening to perform CEM. That is that the bin size was chosen automatically since we do not have a theory about meaningful breaks within the data. Matched units were assigned a weight which was entered as an inverse probability weight to the regression based on Equation 5.1. Any residual difference in the groups after matching was accounted for by the patient characteristics vector in the model. We assessed balance by t-tests of mean differences for individual covariates, and the reduction in standardized percentage bias (Rosenbaum and Rubin, 1985).

5.2.3 Validation of the difference-in-difference approach

The difference-in-difference method assumes common time trends for both the treated and the control group (Jones and Rice, 2011). This means that in the absence of treatment, the average change in the outcomes would be the same for treated as for untreated individuals. If the assumption is violated, the estimated treatment effect would be confounded with a natural time trend. We examine the assumption by testing whether linear pre-policy trends are statistically different between the treatment and the control group. If both groups have common trends prior to the policy, then there is a reasonable expectation that outcomes would also change post-policy at similar rates in the absence of the intervention (Ryan, Burgess and Dimick, 2015; Dimick and Ryan, 2014). Hence, we re-run the regression based in Equation 5.1 including a full set of quarter dummies and an interaction of the dummies with the treatment indicator to model differential trends for treatment and control groups.

The assumption would further be violated if waiting times already changed prior to the policy implementation, in anticipation of the policy change. In chapter 4, we showed that anticipatory effects were likely to have happened. In October 2014, EIP services were officially announced to be affected by a target. We therefore omit the two quarters from October 2014 to the start of implementation in April 2015 from the analysis.

Another requirement for our difference-in-difference approach to be valid is that the comparison group is not affected by the intervention. That is, the target policy does not spill-over from EIP services to standard care services (Ryan, Burgess and Dimick, 2015). Since mental health providers may offer both, EIP and standard care, there is a possibility of spill-over effects in two directions. First, providers may re-allocate resources to enhance EIP target performance at the expense of poorer standard care performance. Second, the increased effort to improve access for EIP patients will lead to improvements in access for standard care patients as well. To investigate the possibility of any spill-over effects we make use of the fact that some providers in our sample offer standard care only. Whereas providers offering both service models and thus experiencing target pressure for their EIP patients may spill-over resources, providers offering standard care only are less likely to be affected by the EIP target policy. Hence, we repeat our main analysis with a control group that is limited to

patients being with providers that only offer standard care, to see whether we observe the same policy effect as for the full sample.

Additionally, we compare standard care outcomes pre- and post-policy for providers that offer both service types (treatment) with those that offer standard care only (control). The model is identical to Equation 5.1 with the only difference being the treatment indicator. We use the same matching procedure, outcome variables and estimation methods as introduced above.

5.3 Data and measures

5.3.1 Sample

We use patient-level data from the MHLDDS introduced in section 1.4.1. We define the pre-policy period from April 2011 to September 2014 (14 quarters), and post-policy from April 2015 to November 2015 (3 quarters). The period of anticipation from October 2014 to March 2015 was omitted.

In accordance with the policy guideline, our treatment group includes patients aged 16 to 35 years and being referred to an EIP service (NHS England, 2015). Standard care patients are identified by having had a community mental health care episode within the study period. To select EIP-eligible patients from this group, we combined a number of criteria which have been used in previous literature (Kirkbride et al., 2017; Tsiachristas et al., 2016). Standard care patients must have had either a diagnosis of schizophrenia, been classified into the first-episode psychosis cluster, or reported problems associated with hallucinations and delusions. Further, we limit our control group to EIP-eligible patients that had no access to EIP services within 15 kilometres travel distance. We assume that a patient who is actually eligible for EIP care but would have to face a long travel distance to receive it would rather be treated by a comparable standard care service nearby. This patient would, however, not necessarily be different from an EIP patient in terms of severity of condition and need of treatment.

We use item 6 of the HoNOS which was introduced in section 1.4.1 and discussed more detailed in chapter 3 as our main measure of psychotic symptom severity. Since the measure was important to ensure comparability between groups in terms of

symptom severity, we excluded patients with missing HoNOS records from the analysis.

5.3.2 Outcome measures

The policy guideline monitors the time from referral to treatment (NHS England, 2015). Treatment is defined as the patient's acceptance onto the caseload and the assignment of a care coordinator. Thus, we measure referral-to-treatment waiting time as the days from referral to care coordinator assignment (see section 1.4.2 for more details). Based on this, we created a dummy that equals 1 if the waiting time was 14 days or less, and 0 otherwise. Length of treatment is measured as the number of days from start to end of the first EIP or standard care episode (recurrent episodes not included). We use the logarithm of waiting time and length of treatment to account for the right-sided skewness.

5.3.3 Service availability and travel distances

The MHLDDS provides information on the mental health provider the patient was receiving care from and the type of care (EIP or standard care). However, no information is available on how many EIP and standard care teams a provider has and which of the teams the patient received care from. In order to identify providers that offer both or only one of the service models as well as to calculate travel distances for patients, we generated a novel dataset on the number and location of EIP and standard care teams per provider across England. We manually researched all provider websites to collect address information of all relevant service teams and double-checked whether the identified teams were registered as a site with an NHS (or care) provider based on information published online by NHS Digital. Based on this list, we calculated travel distances from the patient's place of residence to the nearest EIP team (which is not necessarily the one a patient was receiving care from). We measured distance in a straight line from the geographical centroids of the 2001 LSOA to the grid reference of the service's postcode using Stata 14 MATA.

5.4 Results

5.4.1 Descriptive statistics

In total, we identified 17,472 EIP and 23,554 EIP-eligible standard care patients. We included 5,625 (32%) EIP patients with valid HoNOS records. From the 12,404 (53%) standard care patients with a valid HoNOS record, we selected 3,702 (30%) that had no access to EIP care. In Appendices A19 and A20 we compare characteristics of the included and excluded patients. Patients excluded with missing HoNOS had a longer waiting time but also showed fewer other indicators of a psychosis such as a schizophrenia diagnosis or a first-episode psychosis cluster episode which may indicate that these patients are not clearly psychotic patients and are better excluded. Standard care patients with access to EIP (excluded) were more likely to live in the most deprived neighbourhoods.

Table 5.1 compares sample characteristics of both groups before and after matching. Before matching, t-tests indicate the groups to be highly imbalanced on all observed characteristics. The EIP group was on average three years younger and more likely to be male, single, non-white, and from more deprived neighbourhoods. EIP patients also had more severe problems with hallucinations and delusions (HoNOS 6 score) and were more likely to be diagnosed with schizophrenia or allocated to the first-episode psychosis care cluster. Although some differences in group means remain after matching, the observed mean bias between the two groups reduced substantially from 39.1 to 17.1 after CEM and 4.9 after PSM, respectively. PSM seems to have performed better particularly in balancing the psychosis related characteristics.

Table 5.1: Sample characteristics before and after matching

	Unmatched		Matched controls	
Patient characteristic	Treated	Controls	CEM	PSM
Age (mean)	22.7	26.0***	22.4*	22.5*
Male (%)	0.66	0.48***	0.66	0.64
Single (%)	0.95	0.89***	0.98***	0.96
Non-White ethnicity (%)	0.32	0.20***	0.19***	0.33
Least deprived quintile (%)	0.11	0.17***	0.13**	0.14***
Second least deprived quintile (%)	0.14	0.19***	0.14	0.14
Third least deprived quintile (%)	0.18	0.23***	0.17	0.20**
Fourth least deprived quintile (%)	0.23	0.22	0.23	0.20***
Most deprived quintile (%)	0.34	0.19***	0.32	0.32
HoNOS 6 score (range 0-4, mean)	1.99	1.51***	1.66***	1.78***
Schizophrenia diagnosis (%)	0.20	0.06***	0.03***	0.18
First-episode psychosis cluster (%)	0.72	0.11***	0.47***	0.72

Note: CEM = Coarsened exact matching; PSM = Propensity score matching.

Table 5.2 summarises the proportion below target and mean waiting times by treatment status. Independent of the matching approach, EIP patients had a significantly higher chance of waiting below target during the whole study period. Also, mean waiting times are considerably shorter for EIP patients compared to EIP-eligible standard care patients.

Table 5.2: Proportion below target and mean waiting times by treatment status

	Proportion below target		Waiting time in days	
	Treated	Control	Treated	Control
Unmatched	0.289	0.209***	48.6	81.7***
Coarsened exact matching	0.289	0.202***	48.6	106.8***
Propensity score matching	0.289	0.205***	48.1	105.0***

Note: * p<0.05, ** p<0.01, *** p<0.001 for p-values of t-tests of mean differences between groups.

There are 58 providers in the sample with an average of 3 EIP teams and 13 standard care teams. 13 providers offered standard care only. Figure 5.1 maps the distribution of EIP and standard care (CMH for community mental health) services across England. The average travel distance of EIP patients to their nearest EIP service was 11 kilometres with a minimum of 0.9 and a maximum of 87 kilometres. 50% lived no more than 7 kilometres, 75% no more than 15 kilometres, and 90% no more than 25 kilometres away from the nearest EIP service. Travel distance to the nearest EIP service is shorter for patients in most deprived neighbourhoods (8 kilometres) compared to 12 to 13 kilometres for EIP patients from the least deprived neighbourhoods.

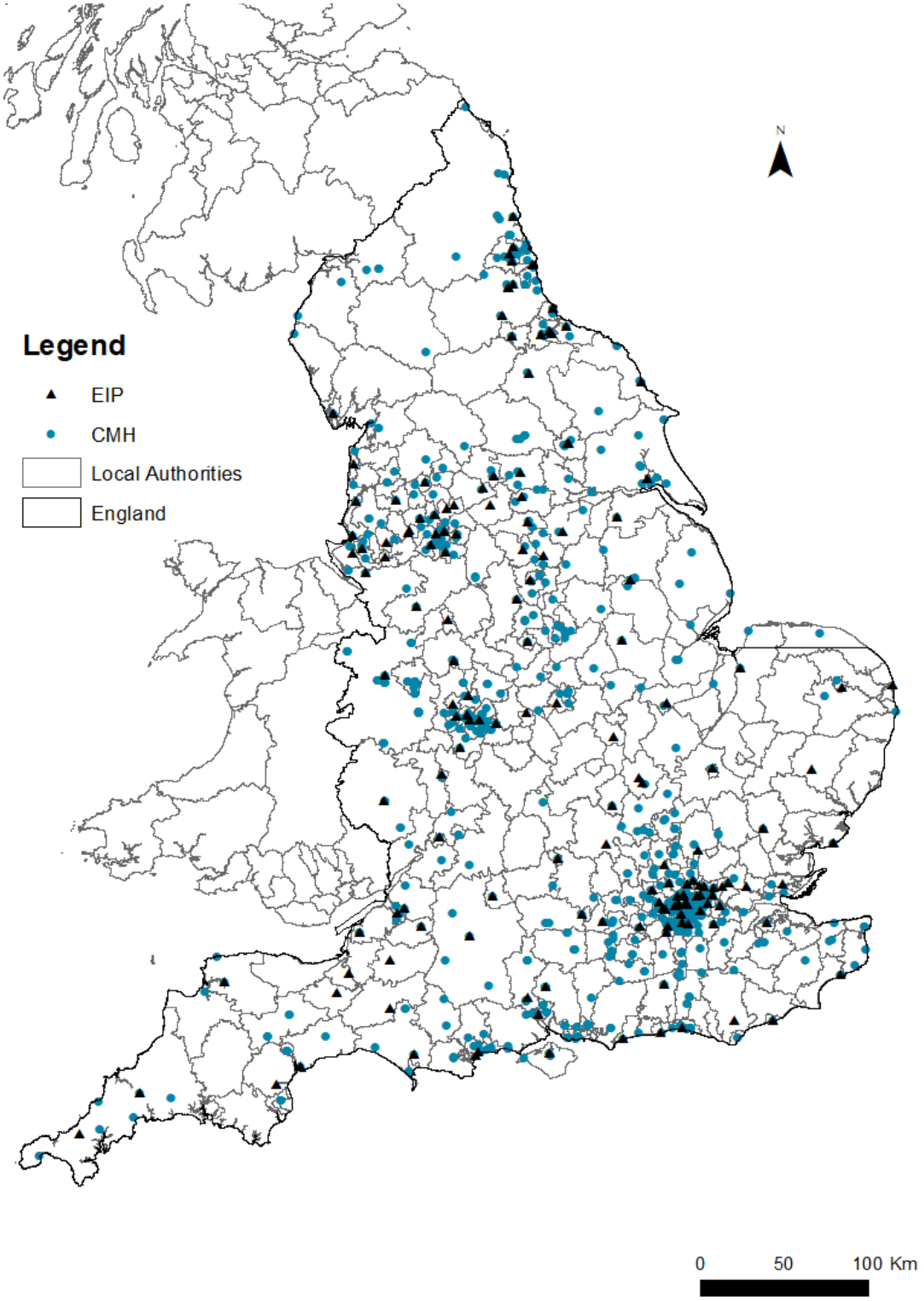
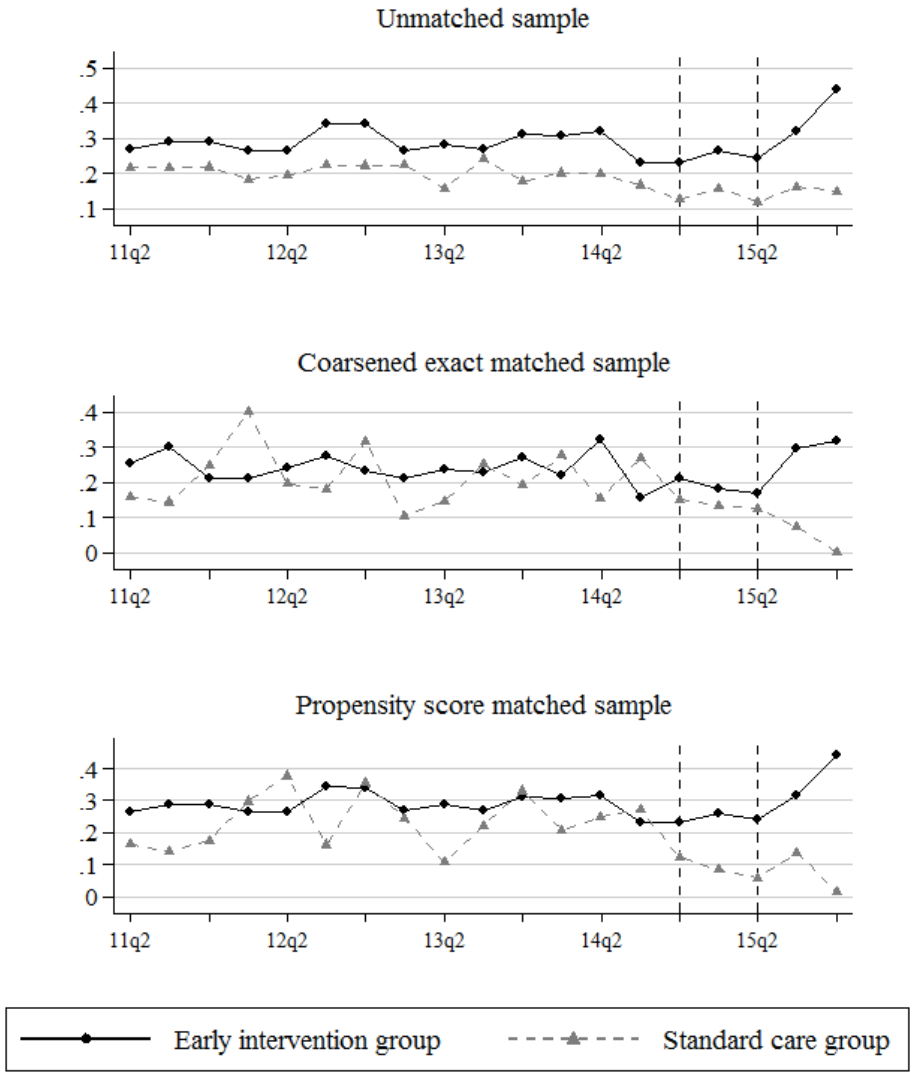


Figure 5.1: Regional distribution of EIP and standard care (CMH for community mental health) service availability in England

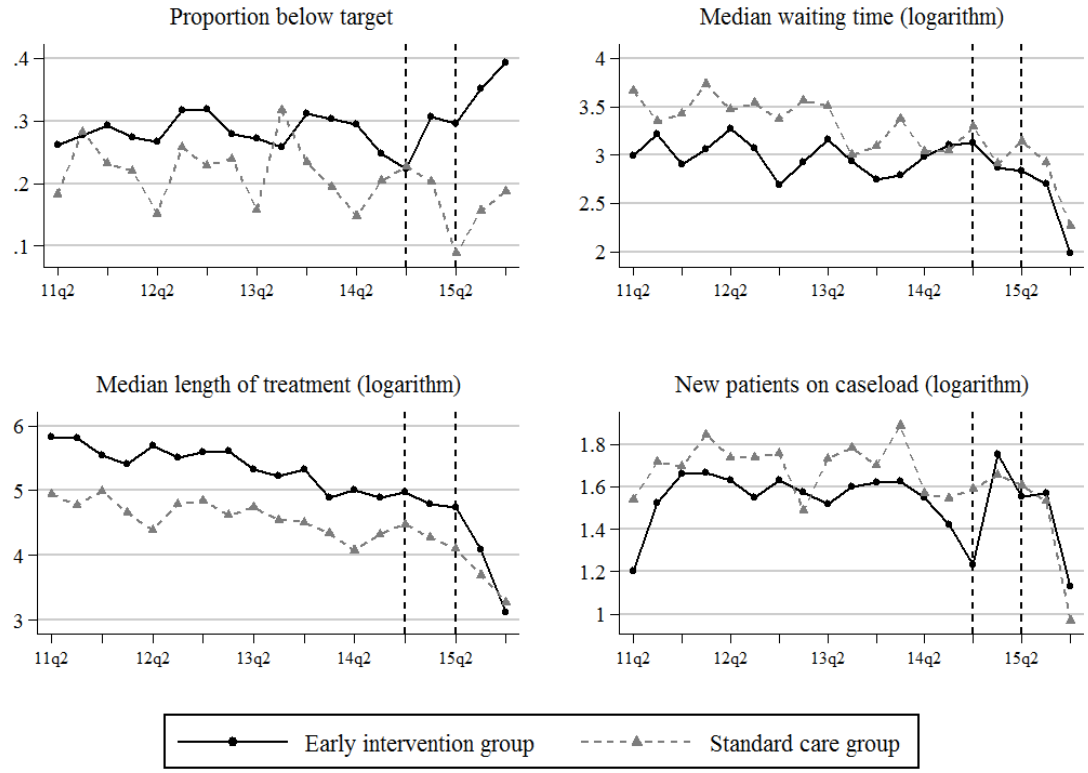
Figure 5.2 visualises pre- and post-policy trends of the probability of waiting below target for EIP and standard care patients before and after matching. Trends are quite stable and parallel between the groups between 2011 and 2013. We observe a slight downward trend in outcomes for both groups starting around the second quarter of 2014. Whereas this downward trend continued for the control group post-policy, the probability of waiting below target increased for EIP patients after the policy implementation.



Note: Vertical dashed lines indicate start of anticipation period (Oct14) and policy implementation (Apr15).

Figure 5.2: Pre- and post-policy trends by treatment group before and after matching

In Figure 5.3, we present pre- and post-policy trends of outcomes aggregated at the provider level (based on the propensity score matched sample). We observe a similar downward trend in the proportion of patients waiting below target shortly before the start of the anticipation period and a strong increase post-policy for both groups as in the patient-level case. Again, the EIP group exceeded its pre-policy levels whereas the standard care group recovered to their pre-policy levels before the downward trend. For median waiting time (logarithm) and median length of treatment (logarithm), we see a constant downward pre-policy trend for both groups which continued during the period of anticipation and increased post-policy. There is no clearly identifiable trend in pre-policy numbers of new patients accepted onto the caseload for both groups. It appears that numbers dropped slightly after the anticipation of the policy change.



Notes: Vertical dashed lines indicate start of anticipation period (Oct14) and policy implementation (Apr15). Based on propensity score matched sample.

Figure 5.3: Provider-level pre- and post-policy trends in outcomes by treatment group

5.4.2 Estimation results

Table 5.3 reports the patient-level estimation results from Equation 5.1. We find a significant positive post-policy effect for EIP patients on the probability of waiting below target independent of the matching method. EIP patients had a 12.2 to 19.5 percentage point higher chance of waiting below target post-policy compared to standard care patients.

Table 5.3: Patient-level difference-in-difference results of the EIP target policy effect on the probability to wait below target

	(1) Unmatched sample		(2) Coarsened exact matching		(3) Propensity score matching	
Post-policy	-0.50	(0.036)	-0.115	(0.051)	-0.121	(0.069)
EIP patient	0.02	(0.040)	0.032	(0.042)	0.016	(0.051)
Post-policy for EIP	0.122*	(0.049)	0.172**	(0.059)	0.195**	(0.073)
Observations	8,393		3,712		6,873	

Note: * p<0.05, ** p<0.01, *** p<0.001. Regression based on Equation 5.1. Pre-policy: Apr11 to Mar15; post-policy: Apr15-Nov15. Oct14-Mar15 omitted. Cluster robust standard errors in parentheses.

The linear predictions of the probability of waiting below target based on the difference-in-difference estimation are presented in Table 5.4. In the unmatched sample, the probability of waiting below target for EIP patients increased from 27% prior to the EIP target implementation to 34% after the policy implementation. In contrast, the probability of standard care patients decreased from 25% pre-policy to 20% post-policy. Probabilities are similar after coarsened exact matching and propensity score matching.

Table 5.4: Linear predictions of the probability of waiting below target

Unmatched sample	Pre-policy		Post-policy	
	Mean	95% Conf. Int.	Mean	95% Conf. Int.
EIP care	0.27	[0.23-0.30]	0.34	[0.29-0.38]
Standard care	0.25	[0.20-0.29]	0.20	[0.12-0.27]
CEM sample	Pre-policy		Post-policy	
	Mean	95% Conf. Int.	Mean	95% Conf. Int.
EIP care	0.24	[0.20-0.29]	0.30	[0.24-0.36]
Standard care	0.21	[0.17-0.25]	0.10	[-0.01-0.21]
PSM sample	Pre-policy		Post-policy	
	Mean	95% Conf. Int.	Mean	95% Conf. Int.
EIP care	0.26	[0.21-0.32]	0.34	[0.28-0.40]
Standard care	0.25	[0.20-0.30]	0.13	[-0.01-0.26]

A main limitation of the linear probability model is that the fitted values of our difference-in-difference estimate will not necessarily be in the [0,1] interval (Cameron and Trivedi, 2010). Appendix A21 shows that the predicted probabilities are below zero at the lower end of the distribution.

We observe a similarly consistent effect on the proportion of waiting below target at the provider-level, independent of the matching method (see Table 5.5, panel 1). The proportion of EIP patients waiting below target increased by 12.0 to 15.4 percentage points per provider post-policy. However, there was no policy effect on the median waiting time (panel 2). The CEM-matched estimates indicate that median waiting times were significantly lower for EIP patients compared to standard care patients.

Table 5.5: Provider-level difference-in-difference results of the EIP target policy effect on various outcomes

(1) Proportion below target	(1) Unmatched sample		(2) Coarsened exact matching		(3) Propensity score matching	
Post-policy	-0.036	(0.024)	-0.051	(0.026)	-0.051	(0.028)
EIP patient	0.019	(0.122)	0.064*	(0.032)	0.071	(0.054)
Post-policy for EIP	0.120*	(0.048)	0.154**	(0.048)	0.131*	(0.051)
Observations	1527		1400		1468	
(2) Median waiting time (logarithm)	(1) Unmatched sample		(2) Coarsened exact matching		(3) Propensity score matching	
Post-policy	-0.522*	(0.203)	-0.610**	(0.215)	-0.391*	(0.188)
EIP patient	-0.324	(0.630)	-0.518*	(0.195)	-0.431	(0.240)
Post-policy for EIP	-0.094	(0.226)	0.118	(0.248)	-0.047	(0.229)
Observations	1392		1214		1303	
(3) Median length of treatment (logarithm)	(1) Unmatched sample		(2) Coarsened exact matching		(3) Propensity score matching	
Post-policy	-0.964***	(0.112)	-0.888***	(0.131)	-0.927***	(0.131)
EIP patient	0.354	(0.756)	0.411	(0.212)	0.658*	(0.270)
Post-policy for EIP	-0.278	(0.154)	-0.121	(0.164)	-0.329	(0.173)
Observations	1527		1400		1468	
(4) New patients on caseload (logarithm)	(1) Unmatched sample		(2) Coarsened exact matching		(3) Propensity score matching	
Post-policy	-0.541**	(0.156)	-0.484***	(0.139)	-0.490***	(0.116)
EIP patient	-0.357	(0.708)	-0.540**	(0.202)	-0.121	(0.205)
Post-policy for EIP	0.151	(0.181)	0.137	(0.159)	0.143	(0.153)
Observations	1527		1400		1468	

Note: * p<0.05, ** p<0.01, *** p<0.001. Regression based on Equation 5.1. Pre-policy: Apr11 to Mar15; post-policy: Apr15-Nov15. Oct14 to Mar15 omitted. Cluster robust standard errors in parentheses.

But for all samples, median waiting times significantly decreased for standard care patients post-policy. We also could not find any policy effect for other parts of the waiting time distribution such as the 25th and 75th percentile or the mean (results not reported). We find evidence that median length of treatment decreased post-policy for standard care patients but not for EIP patients (panel 3). We observe the same effect for the 75th but not for the 25th percentile of the distribution (results not reported). Similarly, we find that standard care providers accepted fewer patients onto their caseloads after the EIP target introduction compared to before without any effect on the number of new EIP patients accepted onto the caseload (panel 4).

5.4.3 Validation checks

In the following we discuss the results from the validation checks detailed in section 5.2.3. The analysis of pre-policy trends showed no significant difference between the two comparison groups. Appendix A22 presents the treatment specific referral quarter estimates for both the CEM and the PSM matched samples. From the non-significant pre-policy trends, we conclude that the common trends assumption is likely to hold. We do, however, observe significantly different trends during the time of anticipation which confirms our approach to exclude the quarters of anticipation from the main analysis.

Results from the test of potential spill-overs from EIP to standard care services are presented in Appendix A23. We observe the same positive policy effect on the probability of waiting below target when limiting the control group to patients that were with providers offering standard care only for both matching approaches (panel 1). It needs to be noted, that the number of controls is very small due to the additional exclusion criterion (683/776 controls after CEM/PSM). Comparing standard care outcomes of providers offering both service models to those offering standard care only, did not show any significant differences in post-policy trends (see Appendix A23, panel 2). Overall, we conclude that the impact of any spill-over effects if any was small.

5.5 Discussion

Access to specialist services at the early stages of psychosis is critical to successful treatment and recovery. EIP services are internationally recognised as supporting

timely provision of evidence-based care to psychotic patients. However, in times of financial constraints EIP services may not always be able to meet the desired standards of providing rapid access for patients in need. To tackle increasing waiting times, the English government pioneered the introduction of a waiting time target for EIP services. This chapter examined the effectiveness of this target policy in improving access for first-episode psychosis patients. We make use of a difference-in-difference design which is a well-established method to evaluate the impact of health policy interventions in the absence of randomized controlled trial data. We find the EIP target to be effective in increasing the number of patients waiting below target in the first six months of its implementation. However, waiting times across the whole distribution have not changed differently compared to standard care patients. We find no evidence that providers freed up resources by shortening treatment for existing EIP patients or accepted fewer EIP patients onto the caseload due to the target pressure.

Our research moves beyond previous work on the effectiveness of waiting time targets which is limited to country level comparisons as we are able to compare two patient groups being treated within the same institutional setting. This allows us to measure and compare waiting times at the patient level and thus adequately control for changes in case mix over time and between groups. The challenge lies in ensuring comparability between the groups in terms of variables that may also be associated with waiting time. We select control patients with no access to EIP services within a certain travel distance to improve comparability of groups. Furthermore, the combination of matching and the difference-in-difference design allows us to balance the groups on observed as well as unobserved confounders. Whereas the regression model accounted for any remaining imbalances after matching through adding additional covariates, the non-parametric matching helped to reduce model dependence and statistical bias. We found our matching approach to reduce bias in observed characteristics between the two groups substantially. Validation checks further indicated that the common trends assumption was likely to hold and potential spill-overs between EIP and standard care patients was negligible.

There are some limitations to our research. First, our post-policy period is relatively short due to the fact that the collection for MHLDDS temporarily stopped in November 2018 in order to introduce a revised dataset version from April 2016 onwards which is

not yet available for research. Hence, we are only able to look at the first six months of the implementation process. Over time, effects may either become larger once more providers respond to the target policy at later stages, or effects may disappear over time as providers only temporarily focus on the newly introduced target. In future, as more service types are expected to be affected by similar waiting time targets, the interplay between the response to the different targets by a provider will be important to investigate. Second, the t-statistic for our difference-in-difference estimate and in consequence its significance level may be overestimated due to the 13 pre-policy quarters being serially correlated as discussed by Bertrand, Duflo and Mullainathan (2004). To solve the potential problem of serial correlation, future work could aggregate the time series into two periods – pre-policy and post-policy. The difference-in-difference analysis is then performed on the averaged outcome in a panel of two time periods. Alternatively, one could use the variation across providers to estimate the variance-covariance matrix and compute standard errors from this estimated matrix. This method will produce consistent estimates of the standard error as the number of providers goes to infinity. Third, despite various measures to improve comparability between the groups, our results may still be driven by group differences that we were not able to account for but had an impact on patient waiting times. Nevertheless, there is no other potential control group to our knowledge which would be better suited to analyse our research question. Since the policy was introduced nationwide, we cannot exploit regional variation in policy implementation. At the same time, England is to date the only country collecting data which allow the measurement of EIP waiting times. Hence, a comparison at country level, for example using Scotland as control group, as performed in previous literature (Propper et al., 2010, 2008) would not be possible in our case. Having said this, our measured treatment effect is defined for patients under common support – so for patients that were comparable in terms of observed characteristics. We cannot conclude from our results to what extent the estimated effect is generalisable to the whole population of EIP patients. However, using the national administrative database allowed us to draw our estimation sample from a nationally representative patient cohort including a large number of mental health providers across England. This is an advantage compared to

existing literature in the area of psychosis which usually relies on much smaller, regionally limited patient cohorts from only one or two providers.

Our research will be of great relevance to policymakers not only in England but internationally. Waiting times are a persistent policy concern in many countries and its importance is growing, particularly in the area of mental health. We provide novel evidence on the responsiveness of mental health providers towards waiting time targets. As such, our work translates well established knowledge and experience from the physical health area to the mental health context which entails a separate group of providers acting in different incentive systems and providing different treatment concepts. We show that targets can be an effective means to improve access to mental health care. Our research can help inform the future development of the EIP target and its expansion to other areas of mental health in England, as well as informing policymakers in other countries considering the introduction of a similar policy.

6. Conclusion

6.1 Research overview

This thesis analysed various concepts of waiting times for first-episode psychosis patients in the context of the implementation of a waiting time target for EIP services in the English NHS. Waiting times are a persistent concern of policy makers in publicly funded health care systems such as the English NHS. On the one hand, waiting lists can serve to stock available demand and optimise utilisation of the scarce supply of resources such as skilled staff and medical equipment. But on the other hand, a patient's health status may deteriorate during the time of waiting and will make the awaited treatment less likely to be successful. For patients, the time of waiting matters, as they may experience pain, anxiety, disability, and restrictions in their job and family commitments. In the case of psychosis, patients usually experience a significant amount of suffering as symptoms lead to a high degree of disability, and social isolation. Timely access to specialised care is considered a key priority in successful treatment of psychosis. Despite this, little evidence exists about waiting times for people with psychosis at various stages of their care pathway.

We analysed a number of different well-established waiting time measures and discussed each measure's strengths and limitations in the context of the analysis. We made use of econometric methods that allowed us to adequately account for the count data properties of waiting time (generalised linear models and quantile regression) as well as the problem of right censoring of samples in waiting time studies (duration analysis). We applied quasi-experimental methods to solve the evaluation problem when analysing policy effects using observational data (matching and difference-in-difference).

In chapter 2, we investigated inequalities in DUP associated with socioeconomic deprivation in a national cohort of first-episode psychosis patients in England based on the MHLDDS. We used a generalised linear model to account for non-linearity in DUP and looked at inequalities across the whole distribution of DUP using quantile regression. Although there is a large body of literature on DUP in England, we were the first to use administrative data to build a nationally representative patient and provider sample. This is particularly important when looking at socioeconomic deprivation as it allows one to include patients and providers from all levels of

deprivation. Further, we advanced the existing literature methodologically. We moved beyond the common analysis of correlations in this area as our regression model accounted for a large number of potential confounders such as demographic factors, disease history, and service use variables. We were also the first to use count data model techniques and quantile regression in this context. We found evidence of significant inequalities in DUP regarding the level of socioeconomic deprivation with considerable variations between and within mental health care providers. Patients living in the second, third, and fourth deprived neighbourhood quintiles faced a longer DUP than patients from the least deprived neighbourhoods. Inequalities were more prevalent in higher quantiles of the DUP distribution. Also, unemployment – as a patient-level measure of socioeconomic status – showed to be a significant contributing factor to a longer DUP. At the same time, having been in contact with mental health care services prior to the psychosis start significantly reduced DUP.

Chapter 3 contributed to the still ongoing discussion of whether waiting time is associated with worse treatment outcomes. We explored the impact of waiting times on patient outcomes in the context of EIP services in England using a national cohort of EIP patients from the MHLDDS. Whereas a large body of literature investigated the relationship between waiting time and outcomes in the physical health care context, little is known about delays in accessing specialist care and treatment outcomes in the case of psychosis. 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. In contrast to previous studies that looked at outcomes immediately after a single treatment event, we looked at patient outcomes after twelve months adjusting for treatment intensity during this time period. As an outcome measure, we used the routinely collected HoNOS. The methodological challenge in this context is the potential endogeneity of waiting time as the patient's outcome twelve months after treatment will not only depend on her waiting time but also her level of severity before the waiting started. We controlled for baseline HoNOS at treatment start, previous service use, and treatment intensity to account for this potential endogeneity. Using HoNOS further allowed us to advance the analysis by estimating the impact of waiting time on a clinically reliable and significant change in HoNOS. Our results showed that longer waiting time was significantly associated with a deterioration in patient

outcomes twelve months after acceptance for treatment for patients that are still in EIP care. Effects were strongest for waiting times longer than three months. Although effect sizes were small to moderate, we found the change in HoNOS to be reliable and clinically significant. Patients with shorter EIP treatment periods were not affected.

Chapters 4 and 5 focused on the recently introduced EIP waiting time target and providers' potential responses to it. Performance measures such as waiting time targets incentivise providers to improve the targeted outcome dimension. Even without direct financial penalties, providers may expect reputational damage from poor performance which makes them work towards an improved target performance. At the same time, being under pressure to serve a number of objectives, providers may also take unintended actions to make target performance look better than it actually is. For acute health care trusts, it has been shown that providers indeed respond to enforced waiting time targets in both intended and unintended ways. Mental health care trusts, however, comprise a different group of providers acting in different market structures and incentive systems (e.g. payment systems). Therefore, it is important to investigate whether the same changes in behaviour can be observed for mental health trusts compared to acute physical health trusts. We looked at both, intended as well as unintended, effects of the EIP target policy. Since this is the first waiting time target that has ever been introduced for mental health services, we provide novel evidence in this context.

In chapter 4, we analysed potential anticipatory effects of the EIP target policy in a sense that providers may have changed behaviour before the actual implementation of the policy in order to avoid poor performance from the start. EIP treatment is characterised by long enduring treatment processes with relatively inflexible pathways which is why the re-allocation of resources may require a longer period of adaptation. Hence, anticipatory behaviour of mental health providers is likely to happen and important to explore in order to be considered in the actual evaluation of the policy effect itself. The analysis was based on a national cohort of EIP patients from the MHLDDS. We investigated whether changes in waiting time were associated with the announcement of the EIP target policy (intended effect). Additionally, we explored two potentially unintended changes in behaviour providers may have adapted to improve target performance. First, we analysed whether changes in waiting time varied

by the referral priority or disease severity (re-prioritisation). To explore gaming behaviour, we tested whether providers initially focused on improving time from referral to EIP caseload acceptance but at the same time took longer to assign patients to a care coordinator. Although the care coordinator is crucial to the effective delivery of treatment it was announced only later in time as being required in order to stop the waiting time clock. We used duration analysis to overcome the challenge of right-censoring in waiting time studies and account for a patient's time in study. Our results showed that waiting times improved already in the years leading towards the target implementation. So, providers may indeed have changed their behaviour already in anticipation of the target policy. Providers accepted patients earlier for EIP treatment following the first announcement of the EIP target in October 2014. Time until the assignment of a care coordinator also improved after the second announcement of the care coordinator requirement in February 2015. We found no evidence that providers gamed the target and also only little indication of re-prioritisation.

In chapter 5, we evaluated the effects of the EIP waiting time target for a national cohort of first-episode psychosis patients from the MHLDDS. We compared patients being treated by EIP services (treatment) with EIP-eligible patients receiving care from standard community mental health services (control) pre- and post-policy. Combining non-parametric matching with a difference-in-difference approach allowed us to account for observed as well as unobserved group differences. We analysed the probability of waiting below target at patient level and looked at changes in waiting time at different percentiles of the distribution at provider level. We restricted our control group to patients with no access to EIP services in the certain travel distance. For this, we created a novel dataset which provides information on the number and location of EIP and standard care services of each mental health trust across England and calculated patients' travel distances to the nearest EIP service. Our results showed that the EIP waiting time target was effective in increasing the likelihood of waiting below target for EIP patients compared to standard care patients. However, the waiting time distribution did not significantly change post-policy. We found no evidence that length of treatment decreased post-policy for EIP patients due to the increased target pressure. Also, providers did not accept fewer EIP patients onto the caseload.

6.2 Implications for research, policy and practice

6.2.1 The relevance of waiting times for first-episode psychosis

The research presented in this thesis has implications for a number of stakeholders such as policymakers, commissioners, providers, mental health service users and their caring families as well as agencies outside the health system. The relevance of waiting times for first-episode psychosis patients to specialist services is multi-dimensional. First, providing timely and equitable access is not only one of the main corner stones in modern health systems in general (Willcox et al., 2007), it also has been particularly articulated as a major policy aim in the recent mental health policy initiative in the English NHS (Department of Health, 2014a). Especially when waiting times are unequally distributed between certain patient groups, as we could show in chapter 2 for socioeconomically disadvantaged patients, the desired policy aims are threatened. Second, previous research has undoubtedly shown that early access to specialist services in the case of psychosis is crucial in order to promote treatment engagement and recovery as well as the risk of relapse in the long run (Penttilä et al., 2014). In chapter 3, we confirmed that the negative impact on treatment outcomes can also be shown for patients waiting for a care coordinator even after being accepted onto the caseload. Shorter waiting times may therefore contribute to improved health outcomes, reduced health service utilisation and associated costs in the future. Hence, EIP waiting times are also of interest for commissioners who have to manage limited budgets. Third, psychosis has a significant social and economic dimension which makes it of relevance for policymakers and agencies outside the health care system. People suffering from acute psychotic phases have difficulties to fulfil their family and work commitments. Although the prevalence is relatively low, the costs of lost employment are tremendous in the case of psychosis (Kirkbride et al., 2012b). The longer patients wait for treatment the harder it will be to not only recover in physical terms but also in social terms. In chapter 3, we showed that waiting time had the second largest, negative impact on the social outcomes sub domain (symptomatic outcome deterioration being the first). Re-integrating patients into society demands resources from the social care system. If re-integration fails, patients may become homeless or criminal which adds interactions with the justice system as well. Last but not least, patients and their caring families are interested in short waiting times to adequate services as psychosis is

associated with a high degree of disability, anxiety and discomfort particularly when untreated.

6.2.2 EIP waiting times and services availability – opening the black box

The EIP target policy has been introduced without any evidence base of the current situation of EIP patients. Information on waiting times was based on a measure of DUP drawn from small, regionally limited cohorts within one or two providers. The target policy is however based on the referral-to-treatment waiting time concept for which no evidence exists in the mental health context. Also, there exists no published information on the number of EIP services and their distribution across England. Our research opens this black box and makes waiting times and service availability for a nationally representative cohort of EIP patients transparent. We show and discuss different measures of waiting times which policymakers can compare and draw conclusions for future policy initiatives from. Spanning the years immediately before the policy introduction until half a year after, our research builds a foundation for policymakers to actually evaluate the impact of the EIP target policy by being able to observe changes from pre- to post-policy.

We showed that regardless of the measure used there has been a long tail of very long waiting patients. We also found that socioeconomic inequalities are more likely to occur for the longest waiting patients and negative impacts on outcome prospects are largest for waiting times above three months. Alongside this, our research showed that the EIP target helped to increase the proportion of patients waiting below target but without improving the waiting time distribution as a whole. This implies, that the very long waiting patients did not benefit from this kind of policy. Future policies should aim to target the very long waiting patients in order to not further foster already existing inequalities along the waiting time distribution. Overall, we found that mean waiting times improved over the years towards the policy introduction. Policymakers may conclude that the growing awareness created by the ongoing public discussion of the importance of early intervention in psychosis in combination with the additional resources being made available may have been successful and should be continued in the future.

In chapter 5, we not only measured referral-to-treatment waiting times for EIP patients but also compared them to waiting times for EIP-eligible standard care patients. These results contribute to the still ongoing discussion about the superiority of EIP services compared to standard care. We showed that despite the national roll-out of EIP services since the early 2000's, still a significant proportion of EIP-eligible patients are receiving care from standard care rather than EIP services. However, waiting times are on average worse for patients being treated at standard care services even before the EIP target policy was introduced. At the same time, we showed that patients face travel distances of up to 87 kilometres to the nearest EIP service (11 kilometres on average). Therefore, EIP services may not always be available in a manageable travel distance for all patients. Our results provide an evidence base for policymakers to decide how to shape the future relationship between EIP and standard care. Either it is accepted that standard care can compensate EIP treatment in areas where EIP services are not available for patients. In this case, they should however underlie the same waiting time target to not disadvantage patients using these services. Or if the policy aim is to provide EIP treatment for all patients in need of it, service availability clearly has to be enriched. At the same time, our findings are of relevance to patients and their caring families as they can take better informed decisions about where to seek help.

6.2.3 Different concepts of waiting times and their implications

Waiting time is one of the most important indicators within the NHS in England to measure and quantify access of patients to health care services. However, our research showed that there exist plenty of different concepts of waiting time and each has to be interpreted in the corresponding context. In the area of psychosis, the concept of DUP dominates the existing literature. As it encompasses the patient's waiting experience from the first symptoms onwards, it includes the period of help-seeking which is crucial in the case of psychosis. It is this help-seeking period that is most likely influenced by factors outside the health care system such as level of education, degree of disease insight, knowledge about where to seek help, or the availability of family and other support networks. Therefore, implications drawn from DUP studies not only address stakeholders within the health system but also other policy areas such as the education sector, the justice system, or the social care system as well as patients and

their caring families. In chapter 2, for the first time, we showed that it is possible to measure DUP using the MHLDDS as an administrative dataset which allowed us to draw conclusions from a nationally representative sample. We found that socioeconomic deprivation is a contributing factor to a prolonged DUP, independent of a number of other patient characteristics. Understanding the factors that contribute to a longer DUP can help to develop initiatives for promoting access that target patient groups which are at particular risk of being left behind. Education campaigns could be launched to improve awareness of professionals which are typically the first point of contact for psychotic patients such as general practitioners, social workers, or teachers. Improved knowledge about the signs of early psychosis may help identify potential patients early and encourage those in contact to refer patients promptly to specialist services (Lloyd-Evans et al., 2011). Reducing stigma associated with psychosis through information campaigns for young people and their families in schools or in mainstream media may further promote early help-seeking (Connor et al., 2016). Campaigns should be tailored to target people from socioeconomically deprived areas.

If the focus of interest lies, however, on the health care system and how health policy interventions such as the EIP target should be designed and evaluated, other waiting time concepts will be better suited. We developed a way to measure a form of "inpatient" waiting time for psychotic patients which has been used pre-dominantly in the analysis of the impact of waiting time on outcomes in the past. In chapter 3, we showed that the time from caseload acceptance to care coordinator assignment can lead to a deterioration in treatment outcomes after adjusting for treatment intensity. On the one hand, this is of interest for providers. If they aim to improve patient experience and treatment prospects, it is not only sufficient to accept patients quickly onto the caseload, but it is also vital to assign a care coordinator who ensures that a tailored care plan can be implemented. On the other hand, results are of importance for policymakers. The definition of the care coordinator assignment as the endpoint of the targeted waiting time appears to be a relevant choice if the policy's aim is to prevent a waiting time induced deterioration of outcomes. However, this concept excludes the time from referral to EIP caseload acceptance which is also of relevance particularly from a patient perspective. In chapter 4, we provide a comparison between time from referral to EIP caseload acceptance and from EIP caseload acceptance to care

coordinator assignment. Since the referral may have happened in a different care system, i.e. the primary care sector or the justice system, it can be of interest for providers as well as for policymakers to see by which proportion each of the components contributed to the total waiting time. Based on this information, initiatives targeted at reducing waiting times can be better tailored. If waiting times are mainly caused by delays from referral to caseload acceptance, providers would rather have to focus on improving referral pathways and communication with their referrers rather than putting resources into more available care coordinators and vice versa. A promising approach could be the increased implementation of single access points which serve as a first point of contact for patients seeking help and triage patients to the right service according to their needs.

To conclude, each waiting time concept has its own justification and can provide important implications for policy and practice. It depends on the question of interest when deciding which measure to prefer. At the same time, we showed that a one-sided focus on one concept brings the danger of providers focusing on this part of the patient's waiting experience only to the exclusion of other parts of the waiting experience. Transparency through the routine publication of waiting time statistics will certainly help to limit such unintended gaming behaviour and give providers the opportunity to compare their own performance to others. Also, patients and caring families are able to identify good performing providers and choose where to seek help accordingly. Fostering transparency requires, however, that data structures improve to allow the routine collection of the various waiting time concepts. A first step has already been taken as a consequence of the EIP waiting time target. The most recent version of the MHLDDS which was not available to be used for this research includes a standardised measure of referral-to-treatment waiting time for EIP services.

6.2.4 Intended and unintended effects of waiting time targets

The research presented provides important evidence for the operation and future development of performance targets in the mental health context. The EIP target is the first target of its kind that has been introduced for mental health providers. We found that mental health providers respond to waiting time targets in a similar way as has been shown for the physical health context. This means that targets can also be

effective in a setting where services are provided by stand-alone multidisciplinary teams within the community; where treatment is delivered in many sessions over a period of several years; and where the need for treatment is acute rather than elective. Also, target policies in the past have been famous for the tough penalties managers had to face when they failed to meet the targets. In contrast, the EIP target was not accompanied with aggressive penalties. Hence, our results show that a change in provider behaviour can also be incentivised by non-monetary penalties such as reputational damage through the publication of performance data. Alongside this, we found that the public release of different levels of information about the detailed design of the target was associated with a corresponding change in provider behaviour even before the actual policy implementation. That is, providers may adapt behaviour already in anticipation of a policy change which has to be considered when evaluating the actual policy impact.

One key result of our analysis is that providers seem to respond to targets in the exact way as it is defined. That is, the EIP target was effective in increasing the proportion of patients waiting below the 14-day target. However, waiting times did not improve overall at any part of the distribution. Patients that are waiting above the target may be in danger of waiting even longer as resources are used for patients that could still be treated within target. Furthermore, we observed differences in waiting times by priority and severity without any clear pattern. Mostly, low priority patients seemed to have benefitted from improvements in waiting time. One explanation is that providers started improving target performance by treating patients which are less severely affected and hence may demand fewer resources first. However, we found only little evidence that this happened at the expense of more severely affected patients who required more resources to be kept engaged in treatment and attend scheduled appointments. We found no evidence that providers shifted resources from patients already in treatment towards new patients that were about to breach the target by reducing length of treatment for EIP patients or accepting fewer EIP patients onto the caseload.

Overall, our results provide important insights for the future development of the waiting time target strategy. The target is planned to be increased for EIP patients and extended to other mental health services in the next few years. From our research, we

conclude that this can be a fruitful strategy to improve access to mental health services. However, the exact definition of the target is crucial to its expected effects. Also, other non-targeted outcomes should be monitored carefully to assess the extent of potential negative effects. Once providers have to respond to different targets for different services, the interplay between the performance in these different services will be worth investigating in the future as well.

6.3 Limitations and areas for future research

There are a number of limitations to the work presented in this thesis which will be discussed in the following, alongside the outline of opportunities for future research.

6.3.1 Data availability and missingness

The length of our study periods had to be limited due to data availability. Throughout all the chapters, we used MHLDDS data from April 2011 and to November 2015. Earlier versions of the dataset were less reliable in terms of data quality which is why we decided not to use them for our analysis. Further, data collection stopped temporarily in November 2015. A new version of the dataset which also includes variables that support the measurement of referral-to-treatment waiting times for EIP patients was launched from April 2016 onwards but was not available for this research. This was particularly of relevance in chapter 5 where we could only measure the effect of the EIP target after the first six months of its implementation which is a short post-policy period. Once, future releases of the new dataset become available the long-term effects of the EIP target policy should be investigated. Similarly, the follow-up period in chapter 3 could be increased with more data being available in order to investigate the long-term effects of waiting times on patient outcomes.

Given data availability, one main limitation in working with the MHLDDS is the relatively high degree of missing data for some of the variables that were of interest for our research. The MHLDDS is designed for the primary purpose of administering provider payments. Hence, providers who are responsible for delivering the data will focus efforts particularly on data items that are relevant to their payment which are not

necessarily the items that are of interest for research. Consequently, coding quality of demographic information, diagnoses or the HoNOS is poor.

If information on the main explanatory variable of interest was missing, we followed a complete-case approach by dropping patients with missing values from the analysis. The complete-case analysis is generally preferable in dealing with missing data in multiple regression analysis as it produces unbiased estimates of the coefficient of interest and the residual variance (Jones, 1996). However, exclusions due to data missingness become an issue for the interpretation of results if missing information is systematically related to the unobserved value itself (missing not at random) (Leurent et al., 2018). This may have affected our waiting time measures. Particularly in chapter 2, the measurement of DUP relied on the psychosis related variables such as the emergent psychosis date and the start date of antipsychotic medication which was only available for a small proportion of the potential first-episode psychosis cohort. Comparisons with similar studies made us believe that we could still generate a representative sample that exceeded numbers from any previous studies. If the coding quality was, however, systematically different for patients from different socioeconomic groups, this may have affected our results. For the other chapters, the care coordinator assignment served as our main endpoint in order to measure waiting time. If a patient was not assigned to a care coordinator by the end of our data availability, we were not able to distinguish whether the patient was still waiting or whether the information on the care coordinator assignment of this patient was missing. Hence, we had to exclude these patients from the analysis in chapters 3 and 5 which may have led to an underestimation of the very long waiting times especially towards the end of data availability. The extent to which this has happened could be followed up once future data releases are available. This would also allow a separate analysis of the very long waiters – for example in what characteristics they differ from short waiting patients. Policymakers may be interested in this since our research showed that the negative impact of waiting time on outcomes is particularly affecting the long waiters and also socioeconomic inequalities increase at the higher end of the waiting time distribution.

In chapter 3, our outcome variable of interest was the HoNOS. The poor coding quality of HoNOS has been discussed in the literature already (Jacobs, 2009). Hence, we had

to exclude patients with missing HoNOS observations which reduced sample size considerably. Our strategy to reduce the number of exclusions was to widen the time window in which we measure the HoNOS for a patient. That means although we aimed to measure follow-up HoNOS scores twelve months after treatment start, we allowed for a measurement window of thirty days before and after the end of follow-up to increase the number of follow-up HoNOS scores.

If information was missing on categorical explanatory variables which were not of main interest for the analysis, we followed the missing-group approach suggested by Cohen and Cohen (1975) to deal with missing information. That is, we created a separate group for patients with missing values. This approach has the advantage that it uses all the available information and avoids the risk of selection bias by dropping subjects for which missing information is not random. Also, a lower power due to the considerable reduction in sample size can be prevented by this procedure. The missing-group method produces unbiased regression parameter estimates as in the case of the complete-case analysis. However, it overestimates the residual variance and results in considerably larger standard errors (Jones, 1996).

Missing information also challenged our sample selection procedures. Diagnostic information which would usually be used to identify patients relevant to the study is poorly coded in the MHLDDS for several reasons which we outlined in section 1.4.3. In chapter 2, we identified first-episode patients by the psychosis related date variables (emergent and manifest date). Although this was necessary to measure DUP of each patient, we noted small patient numbers over the three-year study period which we can only explain by poor data quality of the psychosis variables. In the following chapters, we therefore used a different strategy to identify relevant patients which resulted in larger sample sizes overall. Next to diagnostic information, we used a number of additional variables in order to identify patients such as whether they received treatment by an EIP team; reported problems with hallucinations and delusions (HoNOS item 6 > 0) or were allocated to the first-episode psychosis mental health care cluster. This procedure may have, however, led to some inaccuracies in our samples. On the one hand, we may have selected patients which were actually not first-episode psychosis patients. On the other hand, we may have missed patients which actually would have been relevant to our study. We could show that our study samples had a

similar composition in terms of demographic characteristics as compared to first-episode psychosis cohorts in other studies. Also, total patient numbers seemed plausible regarding the prevalence of first-episode psychosis in England. Additionally, we investigated the recently developed opportunity to link MHLDDS data to Hospital Episodes Statistics (HES) which provide data on inpatient admissions within the NHS in order to improve diagnostic information. However, this was not fruitful for three reasons. First, additional information was limited to patients that had an inpatient admission at some point which may be selective for the more severely affected patients or those with severe co-morbidities. Second, we found that diagnostic information in HES for the relevant patients was not necessarily better coded than in MHLDDS. Third, if a patient could be traced in both datasets information on the same variables did not often match between the two datasets and decisions would have to be taken on how to deal with conflicting data. Nevertheless, the linkage between the two datasets provides a range of opportunities to be exploited in the future. For example, HES also provides a linkage to general practitioners data as well as mortality statistics which offers opportunities to investigate mortality outcomes for EIP patients as well as to enrich analyses of referral pathways from primary care.

Another point to mention is that data missingness may vary between providers which may in turn be related to provider performance such as waiting times. By excluding patients with missing information from the analysis we may also exclude providers with poorer overall performance. The exclusion of patients in our research did however never result in a large number of providers being excluded.

6.3.2 Estimating waiting time

Another main limitation of our research is that the procedures we developed to estimate waiting times are based on certain assumptions. In chapter 2, we estimated DUP using the recorded emergent date and treatment start date. It is a well discussed limitation of the DUP concept in general that the definition of the two points may vary across studies which leads to considerably different lengths in DUP (Register-Brown and Hong, 2014; Large et al., 2008; Singh, 2007). We chose a conservative approach by using the emergent date as the start point compared to the prodrome or manifest date. This may have led to an underestimation of DUP and may therefore also explain that our estimates were shorter compared to other studies. Results were, however, robust against the change of

the start points. It is also recognised in the literature that these dates have to be reconstructed once the patient is in contact with services. Unfortunately, the MHLDDS provides no more detail on the methods being applied to reconstruct these dates and what training the clinical teams received with regard to this. If methods varied between providers, we were able to account for this variation through the application of provider fixed effects. However, we were not able to capture any variation if clinicians within the same provider were recording dates differently. In our case, this would have influenced results if clinicians within a provider would record dates for patients from socioeconomically more deprived areas differently to those from less deprived areas. We cannot rule out that this could have happened not at least because the reliability of the reconstructed information relies on the self-report of the patient. It is possible that patients from more deprived neighbourhoods may systematically report their symptom history differently from others due to lower educational levels or a lack of insight into their condition. This may explain why we found the most deprived neighbourhoods to have a shorter DUP than the other deprivation quintiles which contradicts findings of a clear socioeconomic gradient within the physical health literature. Future research could aim to compare waiting time measures that are independent of the patient's self-report such as the referral-to-treatment waiting time with DUP by socioeconomic status of the patient to see whether there are any systematic differences.

In chapter 3, we used a variant of inpatient waiting time which we defined as the time from acceptance onto the EIP caseload until the assignment of a care coordinator. We note some inaccuracies that may have occurred. First, we aimed to look at the patient's first EIP episode as we believe that waiting times for subsequent episodes may be different. However, the MHLDDS does not indicate a patient's first EIP episode which may have happened before our data availability. We observed patients that had treatment starts before the actual start of the data collection which helped us to exclude these. However, we have no means to check the reliability and completeness of this information for all patients. We tried to get around this by using one year of data for information on the patient's disease history. This allowed us to ensure that patients did not receive EIP treatment within the past twelve months. Second, we assumed that the assignment of the care coordinator marks the start of effective treatment. Our approach is confirmed by the EIP target guideline which defines the same endpoint to stop the

waiting time clock. It is however likely that the patient did receive treatment (at least antipsychotic medication) already in advance of the care coordinator assignment. Hence, waiting time in our case is not defined as the time until any first treatment but rather the implementation of an effective care plan. If any previous treatment influenced the patient's outcomes prospects we would have not been able to account for this. At the same time, the simple assignment of the care coordinator does not mean that the tailored care plan was indeed implemented. So, the actual waiting time may as well have been underestimated. If longer waiting time does indeed have negative effects on outcomes, we would have estimated a lower bound of the effect.

The waiting time concept used in chapter 3 excludes any waiting time that occurred between the first service contact (referral) and the specialist's decision to treat. For the analyses in chapters 4 and 5 we developed a procedure to estimate referral-to-treatment waiting time as it is also targeted according to the EIP target policy. The challenge lies in identifying the referral that was relevant to start the waiting time for the patient for acceptance onto the EIP caseload. Patients could have a number of referrals followed by a number of different team episodes and other service contacts which may have happened simultaneously. Again, we chose a conservative approach in estimating waiting times by assuming that the referral closest to the start of the care spell was the relevant one. This may have underestimated waiting time if earlier referrals were relevant to the EIP episode. Further, we excluded patients from the analysis that were in contact with a different mental health service within the same spell before they were referred to an EIP service. Since we are the first, to use the MHLDDS to calculate referral-to-treatment waiting times we were not able to validate our results in comparison to previous studies. Kirkbride et al. (2017) were the first to estimate referral-to-treatment waiting times for a regionally limited cohort of EIP patients. We found that estimates of both studies were comparable at the 25th percentile. However, median waiting time was consistently longer in our sample due to a higher proportion of very long waiting patients. The future MHLDDS release will include a measure of referral-to-treatment waiting times which allows a comparison of our estimates with future waiting times. It will be worthwhile to repeat the analyses presented in this thesis once the data are available.

6.3.3 The evaluation problem

Working with observational data to explain causal relationships between any kind of treatment on specified outcomes brings the methodological challenge of the so called "evaluation problem" with it. The problem is that at any particular point in time only one of the potential outcomes (with or without treatment) for an individual can be observed (Rubin, 1974). Since the assignment to treatment is usually not random in observational studies, treatment and control group may differ in observed or unobserved factors which are also correlated with the outcome. We apply a number of methods to overcome this challenge but also have to acknowledge each approaches' limitations.

In chapters 2 to 4, we rely on methods for selection on observables using regression analysis. Our results are based on the key assumption that, conditional on the set of covariates, selection into treatment is independent of the outcomes of interest (Heckman and Robb, 1985). The MHLDDS provides a rich set of covariates to control for. In all our models, we include demographic characteristics (age, gender, marital status, ethnicity), socioeconomic variables (accommodation and employment status, socioeconomic deprivation at neighbourhood level), and disease related variables (diagnosis, mental health care cluster, severity of disease measured by HoNOS, number of physical and mental co-morbidities, and previous service use). Nevertheless, we cannot rule out that any unobserved heterogeneity remained and may have confounded our results. In chapter 2, factors such as drug abuse, family history of psychosis, or a poor social network are likely to be differently distributed across quintiles of socioeconomic deprivation but may also be correlated with help-seeking behaviour and hence influence DUP. If these characteristics would be more likely in more deprived neighbourhoods and at the same time influence DUP negatively, we would have estimated a lower bound of the socioeconomic gradient. In chapter 3, waiting time as our covariate of interest may be related to some factors of unobserved severity which at the same time may be related to treatment outcomes twelve months after treatment start. Although we accounted for pre-treatment HoNOS, previous service use and treatment intensity, these measures may have been imperfect. Particularly, our measure of treatment intensity was simply a count of several types of service contacts such as inpatient stays and outpatient episodes. Our data provide no information on the actual

intensity of the service contacts. Treatment intensity will however have played a role in the recovery process of the patient. For example, longer waiting patients may have developed the need for more intense treatment which helped them to recover better than shorter waiting patients. But at the same time, the long waiting could have also impacted their ability to benefit from treatment as patients disengaged from treatment during the time of waiting. Future research could consider an instrumental variable approach to account for the omitted variable bias once a suitable instrument that is correlated with waiting time but not with treatment outcomes becomes available. Moscelli, Siciliani and Tonei (2016) recently used hospital-level aggregated waiting times to instrument patient waiting times when analysing the impact on mortality and readmissions. This requires hospital-level waiting times to not be correlated with in-hospital mortality or readmission after controlling for patient severity. However, provider-level waiting times may be correlated with other quality measures such as the ability to engage patients in treatment which in turn may also be related to treatment outcomes. Whereas treatment engagement may not play a significant role in coronary bypass surgery as investigated by Moscelli and colleagues, it does in the case of psychosis as will be discussed in section 6.3.5. Also, future work could consider including a measure of priority (as has been used in chapter 4) and interacting it with waiting time to explore differential effects of waiting time on outcomes for patients of different priority.

In chapter 4, we found the announcement of the EIP target policy to be associated with changes in waiting time and prioritisation patterns. According to the policy guidelines, providers were explicitly encouraged to improve waiting times long before the actual implementation of the policy. Visiting an EIP service in York also showed us that providers are preparing for the policy introduction long in advance. Nevertheless, we cannot conclude from our results that the observed changes are indeed caused by providers anticipating the policy change. Other confounders may have contributed to a change in waiting time such as the composition of the patient cohort. Although this would have been factors which we did not account for in our covariate vector.

Chapter 5 tried to overcome the weaknesses in chapter 4 by allowing for selection on unobservables in a difference-in-difference approach. However, our control group may not have been perfect to extract the causal policy effect. The groups were different in observed characteristics particularly with regard to the psychosis related variables.

Pre-processing the data through non-parametric matching improved but did not achieve perfect balance on observables. Also, there was a chance of spill-overs happening between the two service models within one provider. However, our tests did not find much evidence that potential spill-overs happened. But these tests can only provide an approximation. Future research should aim to investigate the potential interactions across services within a provider due to enforced performance targets in more detail. Also, other control groups may be identified in the future to enhance the analysis. To the best of our knowledge, this is the best possible control group for the nationally introduced policy. Should other comparable countries such as Scotland start collecting EIP waiting times in the future, a cross-country comparison could be a fruitful extension of this research. With a suitable control group, the analysis of unintended effects of the target policy according to the research ideas presented in chapter 4 could be repeated as well to see to what extent our observed associations are indeed caused by the policy anticipation. Unless a better suitable control group becomes available, some other alternatives to the difference-in-difference approach as applied in this research can be considered in future work to relax the common trends assumption. For example, a synthetic control group can be constructed as a weighted average of the available control units (O'Neill et al., 2016). The weights are chosen based on levels of covariates and outcomes to those of the treated unit prior to the intervention. Alternatively, a lagged-dependent-variable approach (LDV) allows one to adjust for pre-treatment outcomes and covariates with a parametric regression model. This may however lead to bias if the common trends assumption does hold (Angrist and Pischke, 2009).

Despite the advantages of the matching method, it reduces our study sample to those patients that could be matched. Hence, our estimated treatment effect is redefined for patients under the common support. We cannot conclude from our results to what extent the estimated effect is generalizable to the whole population of EIP patients. However, using the national administrative database allowed us to draw our estimation sample from a nationally representative patient cohort including a large number of mental health trusts across England. Alongside this, we estimated an average treatment effect which represents the expected gain from the target policy for an individual randomly selected from the treated population. In future research, it may be

worthwhile to allow for heterogeneity in the treatment effect to see which patient groups potentially benefitted more than others form the policy.

It needs to be noted that for the difference-in-difference approach linearity is central (Jones and Rice, 2011) which is challenged by the non-linearity of our outcome measure waiting time. We solved this in two ways. First, we looked at the probability of waiting below target which allowed us to estimate a linear probability model. Second, we aggregated data at provider level to estimate the policy effect on waiting times at different percentiles of the distribution including the median which is less influenced by a tail of long waiters. However, the aggregation of data may have been accompanied by a loss of information. Future research could aim to implement the more general changes-in-changes approach proposed by Athey and Imbens (2006) which relaxes the additivity assumption.

6.3.4 HoNOS as a routinely collected measure of severity

Most of our analyses rely on the HONOS as the main measure of disease severity. However, the measure comes with its limitations. First, HoNOS is a clinician-rated outcome measure and therefore demonstrates the clinician's judgement of the patient's condition which may not necessarily match the patient's perception. Also, judgements for the same patients may vary between clinicians which we cannot control for since we do not know the responsible clinician for a HoNOS record. In the physical health context, the analysis of patient-reported outcomes gains increasing importance. However, in the context of psychosis this is more difficult as patients may not always be able to judge their condition due to the impact of their symptoms or medication and the lack of insight into their illness. Also, judgements may differ depending on whether the patient is currently experiencing an acute episode or not. To our knowledge, patient-reported outcomes for psychotic patients are not available for analysis at the moment.

Second, HoNOS is not specific to psychosis. Although it was developed for the outcomes measurement of severe mental illnesses some psychosis specific measures such as the Scale for the Assessment of Positive Symptoms (SAPS) or the Scale for the Assessment of Negative Symptoms (SANS) (Andreasen, 1984, 1983) may be more sensitive in detecting variations in positive and negative psychotic symptoms. These

measures are however not routinely collected in administrative data and only used in small-scale field studies so far. With increasing opportunities of data linkage, it may be possible in future research to link one of these first-episode psychosis cohorts from field studies to the MHLDDS in order to enhance outcome measures.

Third, there are some limitations to the concept of reliable and clinically significant change concept as applied to our data in chapter 3. Since the distribution of HoNOS scores is unknown in the general population, our reference groups are psychotic patients with higher HoNOS scores on average. This leads to cut-off points to define reliable and clinically significant changes being quite high and changes will only be evident for observations at the extreme ends of the severity distribution. Moreover, there is the danger that our measured effect reflects a phenomenon known as "regression to the mean" (Barnett, van der Pols and Dobson, 2005). This means that patients who enter mental health services in a crisis or acute psychotic phase will record high HoNOS scores. But those patients are also more likely to record a lower subsequent rating regardless of the time waited.

Finally, as mentioned earlier, HoNOS is one of the more poorly coded variables in the MHLDDS. This first of all leads to sample size reduction if HoNOS is a main variable of analysis as in chapter 3. We found that patients with coded HoNOS scores have higher proportions of mental health service contacts in the past. So, the more past service contacts, the higher the probability of being in the study sample. This could have limited the external validity of our results if past service experience would affect the relationship between waiting for treatment and treatment outcomes. If patients learned coping strategies during previous service contacts which helped them to deteriorate less during the time of waiting, we would have underestimated the negative effect of waiting on outcomes. But if patients with more service contacts in the past are also the ones for which the condition worsens even more during waiting, we would have overestimated the impact of waiting on outcomes. Future work should investigate the role of previous service use in the interplay of waiting time and outcomes. Also, future research could use multiple imputation to address the problem of missing HoNOS data when making inferences about waiting times on treatment outcomes (Leurent et al., 2018). As additional HoNOS measurements become available for the same patients in future data releases, these could be analysed using a repeated measures model. This would allow the analysis of changes in outcomes over time whilst taking the

correlation between subsequent measures into account (Steele, 2011). This would also provide an alternative to the fairly arbitrarily chosen follow-up period of 12-months since multiple time points within the defined follow-up period can be investigated. Another area for future research could be the investigation of alternative outcome measures such as length of treatment, time to first hospitalisation, mortality or costs of health care utilisation. The interpretation of these outcomes is however more difficult in the case of psychosis. A shorter length of treatment or a longer time to inpatient admission may be desirable from a short-term cost perspective. But in very severe cases of psychosis, an early hospital treatment can actually help reduce the likelihood of the condition worsening or prevent the patient from harming herself or others. A cost saving perspective should therefore overlook the long-term effects including potential externalities which requires more years of data. Drawing any policy implications from such kind of analyses has to consider the complexities that come along with it.

6.3.5 The patient as co-producer in the production of mental health

The final limitation that shall be discussed here is related to the fact that patients are co-producers of their health. Psychotic patients experiencing hallucinations and delusions often lack insight into their illness, or experience fear of being stigmatised and are therefore not able to attend appointments or effectively engage in treatment (Gronholm et al., 2017). In chapter 2, this will have influenced help-seeking behaviour of patients and hence caused a longer DUP. If the effects of the patient's engagement are related to the socioeconomic background this would have also influenced our results. We hope to have accounted for this by controlling for the degree of problems with hallucinations and delusions (HoNOS item 6) in combination with the other covariates. In a similar way, this may have affected the waiting time measured in chapter 3. Although we exclude the help-seeking period in this measure, it is possible that the time until the assignment of a care coordinator is also influenced by the degree of the patient's engagement. Before a care coordinator can be assigned, a number of assessments have to be done which require the patient to attend and actively take part. For less engaged patients this may take longer. If the level of patient engagement in turn is related to the changes in outcomes twelve months after treatment start our

results would have been affected by this. We accounted for some of the effects through our treatment intensity measure which counts the number of service contacts and therefore mirrors how many appointments the patient did attend. However, we can however not distinguish the quality of the appointments and how engaged the patient was. In chapters 4 and 5, we assume that providers are able to manage patient waiting times independently from the patient's engagement in order to respond to targets. For example, we observed that low priority patients benefitted more from reductions in waiting times then high priority ones. But from our results, we cannot conclude whether this was an intentional re-prioritisation by the provider in order to meet targets or whether this was an unintentional consequence. It may as well be that providers aimed to improve referral pathways for all patients, but certain patient groups took better advantage of these changes. For example, providers may have offered rapid appointments to all patients, but the more severely affected patients may cancel appointments more often and thus cause an increase in waiting time. With the data we used it is not possible to retrieve any information on the patient's level of engagement. The best we could do is to control for the level of hallucinations and delusions as well as for other patient characteristics which may be correlated with patient engagement. It would be fruitful if future work could identify a way to measure patient engagement explicitly and analyse its impact on waiting times. The EIP target policy guideline states that it lies in the responsibility of the provider to engage the patient in attending treatments. The waiting time clock does not stop due to the patient's non-attendance. That means, provider's performance against the target will also depend on the provider's ability to effectively encourage and proactively manage patients' attendance.

Appendices

A1 Derivation of the study sample from full sample

	n	%
Patients with recorded psychosis and treatment start in study period	1,441	
with valid psychosis and treatment start* (referred to as full sample)	1,368	94.93
Excluded due to missing LSOA	6	0.44
Excluded due to missing HoNOS score	97	7.09
Excluded due to date of HoNOS more than 30 days after treatment start	356	26.02
Excluded due to fewer than 3 patients per provider	22	1.61
Final study sample	887	64.84
Within study sample:		
Number of emergent dates used as psychosis start	696	78.47
Number of manifest dates used as psychosis start	191	21.53
Number of prescription dates used as treatment start	784	88.39
Number of treatment dates used as treatment start	103	11.61
Number of emergent dates that are equal to manifest dates	192	21.65
Number of prescription dates that are equal to treatment date	466	52.54
Mean difference between manifest date and emergent date in days	14.0	
Mean difference between prescription date and treatment date in days	0.3	

Note: * Observations were dropped if treatment start happened before the psychosis start

A2 Comparison of covariates between full sample and study sample

	Study sample n = 887	Full sample n = 1,368
Duration of untreated psychosis (DUP)		
DUP ≤ 2 weeks (%)	44.3	45.8
DUP > 2 and ≤ 6 weeks (%)	18.8	19.4
DUP > 6 and ≤ 12 weeks (%)	13.4	12.9
DUP > 12 weeks (%)	23.5	21.9
Socioeconomic deprivation		
Least deprived quintile (%)	16.4	16.3
2nd least deprived quintile (%)	16.0	16.0
3rd least deprived quintile (%)	20.3	20.6
4th least deprived quintile (%)	21.5	21.4
Most deprived quintile (%)	25.8	25.7
Hallucinations and delusions (HoNOS 6)		
No problems (%)	17.1	22.4
Minor problems (%)	7.0	8.4
Mild problems (%)	20.3	20.9
Moderately problems (%)	36.8	32.8
Severe problems (%)	18.8	15.7
Previous mental health related service use		
Zero service contacts	70.4	74.0
1-10 Service contacts	22.3	19.7
Zero outpatient episodes	94.7	94.6
1-3 outpatient episodes	4.5	4.5
Zero ward stays	93.8	95.3
1-3 Ward stays	5.4	4.2
Number of physical comorbidities, mean (SD)	0.014 (0.12)	0.010 (0.11)
Number of mental comorbidities, mean (SD)	0.016 (0.16)	0.012 (0.13)
Financial year		
2012/13 (%)	33.2	40.2
2013/14 (%)	40.0	36.6
2014/15 (%)	26.8	23.3

Note: Full sample includes all patients with a valid psychosis start date and a valid prescription start date in the financial year 2012/13 - 2014/15. The study sample is based on the full sample and excludes observations with missing LSOA, missing HoNOS score (or HoNOS more than 30 days after treatment start), and providers where fewer than 3 patients were treated.

A3 Distribution of study sample across regions

Region	Full sample		Study sample	
	n	%	n	%
East Midlands	29	2.1%	17	1.9%
East of England	82	6.0%	52	5.9%
London	139	10.2%	99	11.2%
North East	3	0.2%	3	0.3%
North West	248	18.1%	163	18.4%
South East	357	26.1%	228	25.7%
South West	213	15.6%	119	13.4%
West Midlands	218	15.9%	158	17.8%
Yorkshire and The Humber	7	0.5%	4	0.5%
No information on region	72	5.3%	44	5.0%
Total	1,368	100.0%	887	100.0%

Note: Regions as defined by the Office for National Statistics.

A4 Sensitivity analysis: GLM results for different DUP start and end point definitions

GLM log-gamma regression dependent variable: DUP in days	(1) Study sample	(2) Completed observations	(3) Valid treatment date only
Socioeconomic status (ref.cat.: least deprived quintile)			
2nd least deprived quintile	0.335***	0.214***	0.412***
3rd least deprived quintile	0.414***	0.351**	0.476***
4th least deprived quintile	0.609***	0.517***	0.698***
Most deprived quintile	0.186*	-0.039	0.245
Severity of hallucinations and delusions (ref.cat.: no problems)			
Minor problems	-0.386***	-0.189	-0.217***
Mild problems	-0.075	0.132	-0.021
Moderately problems	-0.155	-0.009	-0.132
Severe problems	-0.346***	-0.248	-0.295
Previous mental health service use (ref.cat.: zero service contacts)			
1-10 Service contacts	-0.515**	-0.566**	-0.593**
>10 Service contacts	-0.918***	-0.990*	-0.688
Employment status (ref.cat.: employed)			
Unemployed	0.637**	0.615*	0.628***
Student	0.487**	0.435***	0.598***
Provider fixed effects	no	no	no
Year fixed effects	yes	yes	yes
Covariates	yes	yes	yes
Number of patients	887	658	758
Proportion of total sample	100.00%	74.18%	85.46%

Note: * p<0.05, ** p<0.01, *** p<0.001. Model (1) includes the full study sample but without controlling for provider fixed effects. Model (2) includes only observations for which the emergent date and the prescription date have been used to calculate DUP. Model (3) includes only observations from the study sample which have a treatment date recorded. Only significant covariates are shown. All models include all covariates used in the study sample model and year fixed effects. No provider fixed effects were used due to small sample sizes. Robust standard errors clustered at financial year level.

A5 Sensitivity analysis: GLM results using treatment date as end point of DUP

GLM log-gamma regression dependent variable: DUP in days	End point = prescription date (1) No provider FE	(2) With provider FE	End point = treatment date (3) No provider FE	(4) With provider FE
Socioeconomic status (ref.cat.: least deprived quintile)				
2nd least deprived quintile	0.335***	0.459***	0.413**	0.352***
3rd least deprived quintile	0.414***	0.331***	0.502***	0.308**
4th least deprived quintile	0.609***	0.410*	0.630***	0.390**
Most deprived quintile	0.186*	-0.012	0.227	-0.025
Severity of hallucinations and delusions (ref.cat.: no problems)				
Minor problems	-0.386***	-0.147	-0.341***	-0.079
Mild problems	-0.075	0.011	-0.200*	-0.152
Moderately problems	-0.155	-0.169	-0.206***	-0.230
Severe problems	-0.346***	-0.271*	-0.332**	-0.214
Previous mental health service use (ref.cat.: zero service contacts)				
1-10 Service contacts	-0.515**	-0.526**	-0.496***	-0.471*
>10 Service contacts	-0.918***	-0.929**	-0.773	-0.921**
1-3 Outpatient episodes	-0.881**	-0.713	-0.458**	-0.419
> 3 Outpatient episodes	-0.776*	-0.211	1.124**	1.424
1-3 Ward stays	-0.674	-0.726	-0.777***	-0.680*
> 3 Ward stays	-0.755	-1.417**	-5.065***	-4.725**
Employment status (ref.cat.: employed)				
Unemployed	0.637**	0.572***	0.525***	0.540***
Student	0.487**	0.456*	0.458*	0.458*
Number of patients	887	887	784	784

Note: * p<0.05, ** p<0.01, *** p<0.001. FE = fixed effects. Displayed are only significant covariates. All models include all covariates of the full model and year fixed effects. Robust standard errors at financial year level in Models (1) and (3) and at provider level in Models (2) and (4).

A6 Sensitivity analysis: GLM results after restricting age and restricting DUP

GLM log-gamma regression dependent variable: DUP in days	(1) 14-35-year-old	(2) DUP > 0	(3) DUP < 2 year	(4) DUP < 1.5 year
Socioeconomic status (ref.cat.: least deprived quintile)				
2nd least deprived quintile	0.543***	0.358**	0.480***	0.271
3rd least deprived quintile	0.442***	0.472***	0.298**	0.274**
4th least deprived quintile	0.565***	0.523***	0.412*	0.422*
Most deprived quintile	0.167	0.084	0.011	0.042
Severity of hallucinations and delusions (ref.cat.: no problems)				
Minor problems	-0.267	-0.104	-0.116	0.018
Mild problems	-0.028	0.137	0.021	0.092
Moderately problems	-0.313	-0.069	-0.150	-0.068
Severe problems	-0.514***	-0.093	-0.278*	-0.213
Previous mental health service use (ref.cat.: zero service contacts)				
1-10 Service contacts	-0.545**	-0.223	-0.541***	-0.529***
>10 Service contacts	-0.8480**	-0.190	-0.910**	-1.127***
Employment status (ref.cat.: employed)				
Unemployed	0.546***	0.439**	0.551***	0.514***
Student	0.469*	0.3600*	0.451*	0.506*
Number of patients	805	775	883	874
Proportion of total sample	90.76%	87.37%	99.55%	98.53%

Note: * p<0.05, ** p<0.01, *** p<0.001. Model (1) includes only 14 to 35-year-old patients. Models (2), (3) and (4) include observations with a DUP greater than zero, shorter than 2 years, and shorter than 1.5 years respectively. Displayed are only significant covariates. All models include all covariates of the full model, year and provider effects. Robust standard errors clustered at provider level.

A7 Sensitivity analysis: GLM results for different measures of socioeconomic status (SES)

GLM log-gamma regression dependent variable: DUP in days	(1) Small-area and patient-level SES			(2) Small-area SES only			(3) Patient-level SES only		
	Coeff.	Std. Err.	dy/dx	Coeff.	Std. Err.	dy/dx	Coeff.	Std. Err.	dy/dx
Socioeconomic status (ref.cat.: least deprived quintile)									
2nd least deprived quintile	0.459***	(0.131)	35.50	0.478***	(0.131)	34.69			
3rd least deprived quintile	0.331***	(0.099)	23.86	0.395**	(0.122)	27.44			
4th least deprived quintile	0.410*	(0.183)	30.89	0.471**	(0.167)	34.07			
Most deprived quintile	-0.012	(0.155)	-0.73	0.108	(0.141)	6.44			
Marital status (ref.cat.: single)									
Married/civil partner	-0.060	(0.136)	-4.32				-0.046	(0.123)	-3.32
Divorced/separated	-0.012	(0.314)	-0.92				0.055	(0.379)	4.16
Accommodation (ref.cat.: mainstream housing)									
Homeless	0.011	(0.217)	0.82				-0.032	(0.210)	-2.46
Institutionalised	0.077	(0.247)	6.14				0.049	(0.226)	3.89
Other accommodation	-1.081***	(0.223)	-50.89				-0.956***	(0.240)	-47.96
Not known	-0.058	(0.203)	-4.27				0.026	(0.212)	2.07
Employment (ref.cat.: employed)									
Unemployed	0.572***	(0.147)	39.98				0.513**	(0.173)	36.03
Student	0.456*	(0.203)	29.98				0.402*	(0.202)	26.63
Long-term disabled	0.270	(0.244)	16.07				0.182	(0.247)	10.70
Other employment	0.193	(0.415)	11.01				0.134	(0.469)	7.68
Not known	0.721**	(0.244)	52.14				0.631***	(0.179)	47.28

Note: * p<0.05, ** p<0.01, *** p<0.001. All models include year and provider fixed effects. Marginal effects (dy/dx) are average marginal effects in days. For factor levels, they present the discrete change from reference category. Robust standard errors (Std. Err.) clustered at provider level.

A8 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, Curtis and Beevor (1999); Wing et al. (1998).

A9 Reliable and clinically significant change

For more details, we refer the interested reader to Jacobson and Truax (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:

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

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.

A10 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			

A11 Seemingly Unrelated Regression results – dependent variable: HoNOS sub score at follow-up

	(1) Behaviour		(2) Impairment		(3) Symptoms		(4) Social	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
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).

A12 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.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
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.

A13 Comparison of results using different methods for ties

	(1) Breslow method		(2) Efron's method		(3) Exact marginal LL		(4) Exact partial LL	
a) Time-to-EIP	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Low priority from Apr11-Sept14	ref. cat.		ref. cat.		ref. cat.		ref. cat.	
Low priority from Oct14-Jan15	0.149**	(0.050)	0.178**	(0.065)	0.414***	(0.113)	0.413**	(0.144)
High priority from Apr11-Sept14	-0.048	(0.052)	-0.141*	(0.063)	-0.219**	(0.070)	-0.079*	(0.091)
High priority from Oct14-Jan15	-0.084	(0.072)	-0.073	(0.090)	-0.252	(0.145)	-0.277	(0.181)
	(1) Breslow method		(2) Efron's method		(3) Exact marginal LL		(4) Exact partial LL	
b) Time-to-CCASS	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Low priority from Apr11-Jan15	ref. cat.		ref. cat.		ref. cat.		ref. cat.	
Low priority from Feb15-Nov15	0.316***	(0.086)	0.324***	(0.088)	0.331***	(0.088)	0.354***	(0.091)
High priority from Apr11-Jan15	0.012	(0.074)	-0.001	(0.079)	-0.006	(0.077)	0.016	(0.082)
High priority from Feb15-Nov15	-0.145	(0.110)	-0.154	(0.114)	-0.161	(0.115)	-0.176	(0.120)

Note: * p<0.05, ** p<0.01, *** p<0.001. Robust standard errors (Std. Err.). Stratified by provider.

A14 Global score tests and p-values for proportional hazards

a) Time-to-EIP

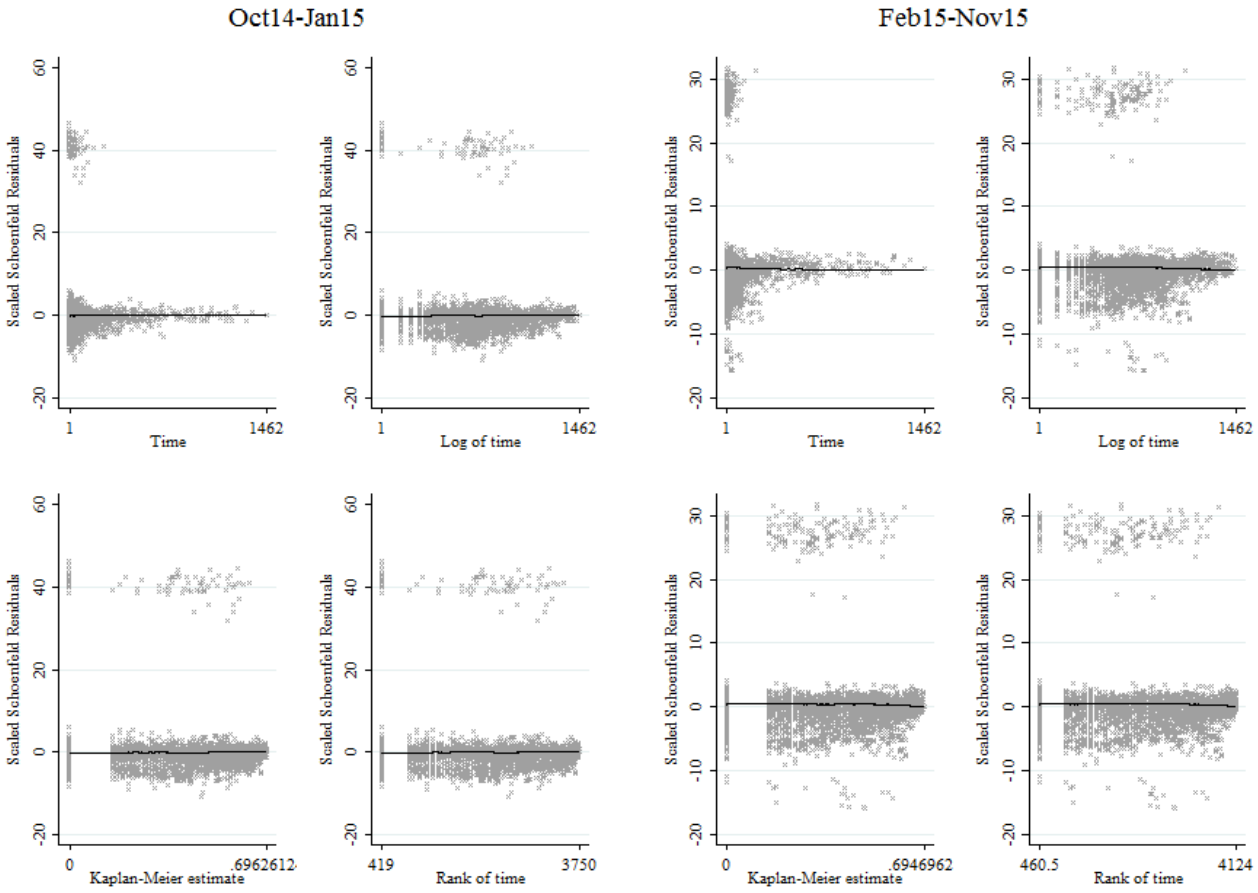
		g(t) = t		g(t) = ln(t)		g(t) = S(t)		g(t) = rank(t)	
Model specification	df	chi2	p	chi2	p	chi2	p	chi2	p
Oct14-Jan15 by priority	36	30.9	0.708	96.1	0.000	102.2	0.000	102.8	0.000
Oct14-Jan15 by severity	36	30.8	0.710	94.4	0.000	99.6	0.000	100.5	0.000
Feb15-Nov15 by priority	36	32.2	0.650	97.7	0.000	106.9	0.000	106.6	0.000
Feb15-Nov15 by severity	36	31.9	0.662	96.1	0.000	104.8	0.000	104.6	0.000

b) Time-to-CCASS

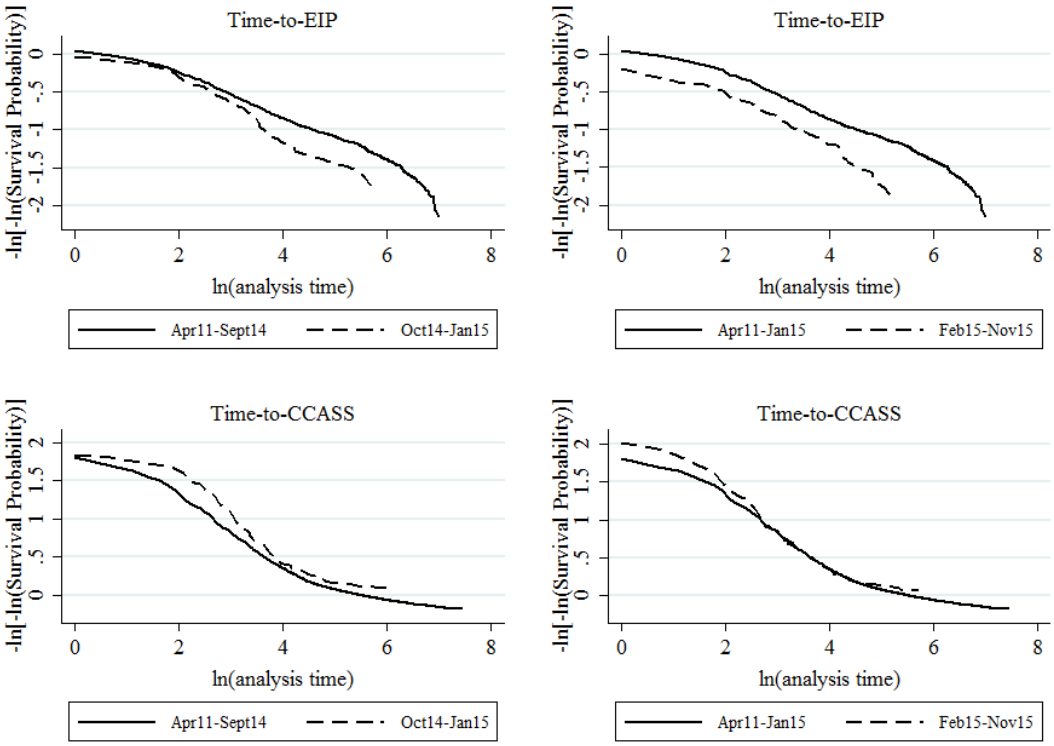
		g(t) = t		g(t) = ln(t)		g(t) = S(t)		g(t) = rank(t)	
Model specification	df	chi2	p	chi2	p	chi2	p	chi2	p
Oct14-Jan15 by priority	36	113.1	0.000	136.1	0.000	132.9	0.000	134.5	0.000
Oct14-Jan15 by severity	36	113.9	0.000	136.9	0.000	133.6	0.000	135.1	0.000
Feb15-Nov15 by priority	36	120.0	0.000	151.4	0.000	149.4	0.000	150.7	0.000
Feb15-Nov15 by severity	36	119.9	0.000	158.0	0.000	156.7	0.000	158	0.000

Note: Reported are degrees of freedom (df), chi2-statistic (chi2) and p-values (p) for score tests of the global model including all covariates for different time functions (g(t)).

A15 Scatterplot of scaled Schoenfeld residuals for the two policy announcement indicators and their lowess smooth versus four different functions of time



A16 Log-Log plots of the two policy announcement indicators



A17 Extended Cox regression results – main regressors interacted with time

	(1) Time-to-EIP		(2) Time-to-CCASS	
	Main effect	Time interaction	Main effect	Time interaction
a) Priority				
Low priority from Apr11-Sept14	ref.cat.			
Low priority from Oct14-Jan15	0.151*	0.003	-0.946	-0.001
High priority from Apr11-Sept14	-0.238**	0.005***	0.027	-0.001*
High priority from Oct14-Jan15	-0.065	-0.001	0.280	-0.001
b) Severity				
Low severity from Apr11-Jan15	ref.cat.			
Low severity from Feb15-Nov15	0.114	0.003	0.013	-0.002
High severity from Apr11-Jan15	0.033	0.005***	0.159***	-0.001*
High severity from Feb15-Nov15	-0.002	-0.002	0.128	0.001

Note: * p<0.05, ** p<0.01, *** p<0.001. Regression coefficients reported. Robust standard errors. Stratified by provider. Referral year is interacted with g(t) = t.

A18 Extended Cox regression results – main regressors interacted with log of time

	(1) Time-to-EIP		(2) Time-to-CCASS	
	Main effect	Time interaction	Main effect	Time interaction
a) Priority				
Low priority from Apr11-Sept14	ref.cat.			
Low priority from Oct14-Jan15	0.058	0.147*	-0.165	0.009
High priority from Apr11-Sept14	-0.379***	0.198***	0.029	-0.029*
High priority from Oct14-Jan15	0.060	-0.173*	0.277	-0.015
b) Severity				
Low severity from Apr11-Jan15	ref.cat.			
Low severity from Feb15-Nov15	0.136*	0.007	-0.054	-0.006
High severity from Apr11-Jan15	-0.009	0.037*	-0.143*	0.114***
High severity from Feb15-Nov15	-0.096	0.124	0.209	-0.027

Note: * p<0.05, ** p<0.01, *** p<0.001. Regression coefficients reported. Robust standard errors. Stratified by provider. Referral year is interacted with g(t) = ln(t).

A19 Comparison of patients with HoNOS score (included) and without HoNOS score (excluded)

	Treated		Controls	
Outcome variable	HoNOS	No HONOS	HoNOS	No HONOS
Proportion below target	0.26	0.18***	0.18	0.17
Waiting time in days	50.5	78.9***	82.9	110.3***
Patient characteristic				
HoNOS 6 score (mean)	-	-	-	-
Schizophrenia diagnosis (%)	0.17	0.12***	0.09	0.08
First-episode psychosis cluster (%)	0.66	0.45***	0.13	0.12*
Age (mean)	22.2	21.9***	26.3	26.2
Male (%)	0.62	0.64**	0.47	0.51***
Single (%)	0.95	0.97***	0.88	0.91***
Non-white ethnicity (%)	0.36	0.33*	0.31	0.30*
Least deprived quintile (%)	0.12	0.11*	0.13	0.11*
Second least deprived quintile (%)	0.14	0.14	0.14	0.13
Third least deprived quintile (%)	0.17	0.18	0.19	0.18
Fourth least deprived quintile (%)	0.23	0.23	0.26	0.25
Most deprived quintile (%)	0.34	0.35	0.28	0.32***

Note: * p<0.05, ** p<0.01, *** p<0.001 for p-values of t-tests of mean differences between groups.

A20 Comparison of standard care patients with no access to EIP (included) and with access to EIP (excluded)

Outcome variable	No access	Access
Proportion below target	0.21	0.19**
Waiting time in days	81.7	75.4
Patient characteristic		
HoNOS 6 score (mean)	1.51	1.52
Schizophrenia diagnosis (%)	0.06	0.08
First-episode psychosis cluster (%)	0.11	0.11**
Age (mean)	26.0	26.6***
Male (%)	0.48	0.50*
Single (%)	0.89	0.90
Non-white ethnicity (%)	0.20	0.33***
Least deprived quintile (%)	0.17	0.10***
Second least deprived quintile (%)	0.19	0.12***
Third least deprived quintile (%)	0.23	0.17***
Fourth least deprived quintile (%)	0.22	0.27***
Most deprived quintile (%)	0.19	0.34***

Note: * p<0.05, ** p<0.01, *** p<0.001 for p-values of t-tests of mean differences between groups.

A21 Observed and predicted probabilities of waiting below target

	n	Mean	Std. dev.	Min.	Max.
Unmatched sample					
observed probability	8,393	0.263	0.440	0	1
fitted probability	8,393	0.263	0.202	-0.145	0.972
Coarsened exact matched					
observed probability	3,712	0.227	0.419	0	1
fitted probability	3,712	0.218	0.205	-0.214	1.024
Propensity score matched					
observed probability	6,873	0.269	0.444	0	1
fitted probability	6,873	0.261	0.211	-0.203	0.951

A22 Referral quarter estimates from the test of common trends for proportion below target

	(1) Coarsened exact matching		(2) Propensity score matching	
Pre-policy				
11q3	0.028	(0.070)	0.089	(0.109)
11q4	0.078	(0.064)	0.024	(0.055)
12q1	0.042	(0.043)	0.124	(0.065)
12q2	-0.003	(0.045)	0.095	(0.052)
12q3	0.018	(0.042)	0.053	(0.063)
12q4	0.042	(0.046)	0.096	(0.065)
13q1	0.070	(0.057)	0.087	(0.063)
13q2	-0.047	(0.044)	0.071	(0.073)
13q3	0.057	(0.067)	0.134	(0.083)
13q4	0.057	(0.062)	0.110	(0.068)
14q1	0.093	(0.071)	0.101	(0.068)
14q2	-0.048	(0.058)	0.064	(0.070)
14q3	-0.066	(0.054)	-0.026	(0.047)
Anticipation				
14q4	-0.108	(0.055)	-0.014	(0.051)
15q1	-0.042	(0.042)	0.022	(0.054)
Post-policy				
15q2	-0.076	(0.041)	-0.087	(0.048)
15q3	-0.03	(0.056)	0.047	(0.054)
15q4	0.018	(0.083)	0.115	(0.088)
Pre-policy for EIP				
11q2 for EIP	0.052	(0.041)	0.122	(0.060)
11q3 for EIP	0.057	(0.073)	0.048	(0.089)
11q4 for EIP	-0.014	(0.066)	0.089	(0.070)
12q1 for EIP	0.005	(0.044)	0.004	(0.055)
12q2 for EIP	0.067	(0.051)	0.020	(0.057)
12q3 for EIP	0.072	(0.052)	0.110	(0.061)
12q4 for EIP	0.055	(0.060)	0.075	(0.071)
13q1 for EIP	-0.011	(0.068)	0.035	(0.068)
13q2 for EIP	0.097	(0.063)	0.055	(0.071)
13q3 for EIP	0.008	(0.066)	-0.007	(0.067)
13q4 for EIP	0.059	(0.060)	0.045	(0.056)
14q1 for EIP	-0.039	(0.073)	0.042	(0.065)
14q2 for EIP	0.157*	(0.061)	0.108	(0.067)
14q3 for EIP	0.106	(0.063)	0.120*	(0.052)
Anticipation for EIP				
14q4 for EIP	0.153*	(0.063)	0.121*	(0.050)
15q1 for EIP	0.118*	(0.045)	0.121**	(0.042)
Post-policy for EIP				
15q2 for EIP	0.140**	(0.051)	0.221***	(0.050)
15q3 for EIP	0.193***	(0.067)	0.157**	(0.055)
15q4 for EIP	0.209*	(0.093)	0.204*	(0.082)

Note: * p<0.05, ** p<0.01, *** p<0.001. "11q3" equals quarter 3 in year 2011. Reference category is 11q2. Models includes all covariates based on Equation 5.1. Cluster robust standard errors in parentheses.

A23 Test of spill-overs from EIP to standard care services

(1) Treatment = EIP patient	(a) Coarsened exact matching		(b) Propensity score matching	
Post-policy	-0.156	(0.079)	-0.279	(0.158)
EIP patient	0.086	(0.139)	0.247	(0.150)
Post-policy for EIP	0.217*	(0.086)	0.338*	(0.162)
Observations	1,879		5,797	
(2) Treatment = Provider that offers EIP and standard care	(a) Coarsened exact matching		(b) Propensity score matching	
Post-policy	-0.015	(0.040)	-0.083	(0.054)
Treatment	0.115***	(0.015)	0.066**	(0.019)
Post-policy for treatment	-0.055	(0.053)	0.043	(0.063)
Observations	4,533		10,480	

Note: * p<0.05, ** p<0.01, *** p<0.001. Pre-policy: Apr11 to Mar15; post-policy: Apr15-Nov15. Oct14-Mar15 omitted. Based on Equation 5.1 but: in model (1) control group limited to providers that offer standard care only; in model (2) treatment =1 if patients were with providers that offer EIP services additionally to standard care. Cluster robust standard errors in parentheses.

Abbreviations

CCASS	Care coordinator assignment
CEM	Coarsened exact matching
Coeff.	Coefficient
Conf. Int.	Confidence interval
df	Degrees of freedom
DUP	Duration of untreated psychosis
EIP	Early intervention in psychosis
GLM	Generalised linear model
HES	Hospital Episodes Statistics
HoNOS	Health of the Nation Outcome Scales
HR	Hazard ratio
ICD-10	International Statistical Classification of Diseases and Related Health Problems, 10th version
IMD	Index of Multiple Deprivation
LSOA	Lower super output area
MHLDDS	Mental Health and Learning Disabilities Dataset
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
PSM	Propensity score matching
RESET test	Regression Equation Specification Error Test
SES	Socioeconomic status
Std. Err.	Standard error

References

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics,* 6(4), 701-726.

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies,* 72(1), 1-19.

Abasolo, I., Negrin-Hernandez, M.A., and Pinilla, J. (2014). Equity in specialist waiting times by socioeconomic groups: evidence from Spain. *The European Journal of Health Economics,* 15(3), 323-334.

Addington, D.E., et al. (2013). Essential evidence-based components of first-episode psychosis services. *Psychiatric Services,* 64(5), 452-457.

Akaike, H. (1970). Statistical predictor identification. *Annals of the Institute of Statistical Mathematics,* 22(1), 203-217.

Allison, P.D. (2010). *Survival analysis using SAS: A practical guide*. 2 ed. North Carolina: SAS Institute.

Allison, P.D. (2014). *Event history and survival analysis*. 2 ed. Thousand Oaks, California: SAGE Publications, Inc.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders*. 5 ed. Washington D.C., London: American Psychiatric Association.

Amin, S., et al. (1999). Evaluating the Health of the Nation Outcome Scales. *The British Journal of Psychiatry,* 174.

Anderson, K.K., et al. (2018). Disparities in access to early psychosis intervention services: comparison of service users and nonusers in health administrative data. *Canadian Journal of Psychiatry,* 63(6), 395-403.

Andreasen, N. (1983). *Scale for the Assessment of Positive Symptoms (SAPS)*. Iowa City: Department of Psychiatry, University of Iowa College of Medicine.

Andreasen, N. (1984). *Scale for the Assessment of Negative Symptoms (SANS)*. Iowa City: Department of Psychiatry, University of Iowa College of Medicine.

Angrist, J.D., and Pischke, J.-S. (2009). *Mostly harmless econometrics*. New Jersey, Woodstock: Princeton University Press.

Apeldoorn, S.Y., et al. (2014). Factors contributing to the duration of untreated psychosis. *Schizophrenia Research,* 158(1-3), 76-81.

Appleby, J., et al. (2005). Do English NHS waiting time targets distort treatment priorities in orthopaedic surgery? *Journal of Health Services Research and Policy,* 10(3), 167-172.

Arnesen, K.E., Erikssen, J., and Stavem, K. (2002). Gender and socioeconomic status as determinants of waiting time for inpatient surgery in a system with implicit queue management. *Health Policy,* 62(3), 329-341.

Athey, S., and Imbens, G.W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica,* 74(2), 431-497.

Barnett, A.G., van der Pols, J.C., and Dobson, A.J. (2005). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology,* 34(1), 215-220.

Behan, C., et al. (2015). Estimating the cost and effect of early intervention on in-patient admission in first episode psychosis. *Journal of Mental Health Policy and Economics,* 18(2), 57-61.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics,* 119(1), 249-275.

Besley, T.J., and Ghatak, M. (2003). Incentives, choice, and accountability in the provision of public services. *Oxford Review of Economic Policy,* 19(2), 235-249.

Besley, T.J., Bevan, G., and Burchardi, K.B. (2009). Naming and shaming: the impacts of different regimes on hospital waiting times in England and Wales. *Discussion paper, 7306. Centre for Economic Policy Research, London, UK.*

Bevan, G., and Hood, C. (2006). Have targets improved performance in the English NHS? *The British Medical Journal,* 332(7538), 419-422.

Bevan, G., and Hamblin, R. (2009). Hitting and missing targets by ambulance services for emergency calls: Effects of different systems of performance measurement within the UK. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 172(1), 161-190.

Birchwood, M., Todd, P., and Jackson, C. (1998). Early intervention in psychosis. The critical period hypothesis. *The British Journal of Psychiatry Supplement,* 172(33), 53-59.

Birchwood, M., et al. (2013). Reducing duration of untreated psychosis: care pathways to early intervention in psychosis services. *The British Journal of Psychiatry,* 203(1), 58-64.

Blackwell, M., Iacus, S., and King, G. (2009). cem: Coarsened exact matching in Stata. *The Stata Journal,* 9(4), 524-546.

Boswell, J.F., et al. (2015). Implementing routine outcome monitoring in clinical practice: benefits, challenges, and solutions. *Psychotherapy Research,* 25(1), 6-19.

Braybrooke, J., et al. (2007). The impact of surgical wait time on patient-based outcomes in posterior lumbar spinal surgery. *European Spine Journal,* 16(11), 1832-1839.

Breitborde, N.J.K., Srihari, V.H., and Woods, S.W. (2009). Review of the operational definition for first-episode psychosis. *Early Intervention in Psychiatry,* 3(4), 259-265.

Breitborde, N.J.K., and Moe, A.M. (2017). Early intervention in psychosis in the United States: From science to policy reform. *Policy Insights from the Behavioral and Brain Sciences,* 4(1), 79-87.

Broussard, B., et al. (2013). Demographic, socio-environmental, and substance-related predictors of duration of untreated psychosis (DUP). *Schizophrenia Research,* 148(1–3), 93-98.

Cameron, A.C., and Trivedi, P.K. (2010). *Microeconometrics Using Stata*. Stata Press.

Cascio, M.T., et al. (2012). Gender and duration of untreated psychosis: a systematic review and meta-analysis. *Early Intervention in Psychiatry,* 6(2), 115-127.

Catts, S.V., et al. (2010). Is a national framework for implementing early psychosis services necessary? Results of a survey of Australian mental health service directors. *Early Intervention in Psychiatry,* 4(1), 25-30.

Cechnicki, A., et al. (2014). Duration of untreated psychosis (DUP) and the course of schizophrenia in a 20-year follow-up study. *Psychiatry Research,* 219(3), 420-425.

Cheng, C., Dewa, C.S., and Goering, P. (2011). Matryoshka Project: lessons learned about early intervention in psychosis programme development. *Early Intervention in Psychiatry,* 5(1), 64-69.

Cheng, C., et al. (2014). Rural and remote early psychosis intervention services: the Gordian knot of early intervention. *Early Intervention in Psychiatry,* 8(4), 396-405.

Cheng, S.C., and Schepp, K.G. (2016). Early intervention in schizophrenia: A literature review. *Archives of psychiatric nursing,* 30(6), 774-781.

Chong, H.Y., et al. (2016). Global economic burden of schizophrenia: a systematic review. *Neuropsychiatric Disease and Treatment,* 12, 357-373.

Cocchi, A., et al. (2018). Implementation and development of early intervention in psychosis services in Italy: a national survey promoted by the Associazione Italiana Interventi Precoci nelle Psicosi. *Early Intervention in Psychiatry,* 12(1), 37-44.

Cohen, J., and Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Oxford, England: Lawrence Erlbaum.

Collins, C.L., et al. (2015). Cardiac rehabilitation wait times and relation to patient outcomes. *European Journal of Physical and Rehabilitation Medicine,* 51(3), 301-309.

Compton, M.T., et al. (2011). The "doses" of initial, untreated hallucinations and delusions: a proof-of-concept study of enhanced predictors of first-episode symptomatology and functioning relative to duration of untreated psychosis. *The Journal of Clinical Psychiatry,* 72(11), 1487-1493.

Connor, C., et al. (2016). Don't turn your back on the symptoms of psychosis: the results of a proof-of-principle, quasi-experimental intervention to reduce duration of untreated psychosis. *BMC Psychiatry,* 16, 127.

Cookson, R., et al. (2016). Socio-economic inequalities in health care in England. *Fiscal Studies,* 37(3-4), 371-403.

Cox, D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological),* 34(2), 187-220.

Cullis, J.G., Jones, P.R., and Propper, C. (2000). Waiting lists and medical treatment: Analysis and policies. In Culyer, A.J., and Newhouse, J.P., (Eds.) *Handbook of Health Economics*. Vol. 1, Part B. Amsterdam: Elsevier, pp. 1201-1249.

Culyer, A.J., and Cullis, J.G. (1976). Some Economics of hospital waiting lists in the NHS. *Journal of Social Policy,* 5(3), 239-264.

Dawson, S., Jordan, Z., and Attard, M. (2013). Carers' experiences of seeking help for relatives with first-episode psychosis: a systematic review of qualitative evidence. *The JBI Database of Systematic Reviews and Implementation Reports*, 183 - 361.

Deb, P., Norton, E.C., and Manning , W.G. (2017). *Health econometrics using Stata*. Texas: StataCorp LLC.

Deb, P., and Norton, E.C. (2018). Modeling health care expenditures and use. *Annual Review of Public Health,* 39(1), 489-505.

Department of Health (2011). *No health without mental health: A cross-government mental health outcomes strategy for people of all ages*. Published online: https://www.gov.uk/government/publications/no-health-without-mental-health-a-cross-government-mental-health-outcomes-strategy-for-people-of-all-ages-a-call-to-action [Accessed: 16.03.2018].

Department of Health (2012). *No health without mental health: implementation framework*. Published online: https://www.gov.uk/government/publications/mental-health-implementation-framework [Accessed: 16.03.2018].

Department of Health (2013). *Mental health payment by results guidance for 2013-14*. Published online: https://www.gov.uk/government/publications/mental-health-payment-by-results-arrangements-for-2013-14 [Accessed: 16.03.2018].

Department of Health (2014a). *Closing the Gap: Priorities for essential change in mental health*. Published online: https://www.gov.uk/government/...data/.../Closing_the_gap_V2_-_17_Feb_2014.pdf [Accessed: 16.03.2018].

Department of Health (2014b). *Achieving better access to mental health services by 2020*. Published online: https://www.gov.uk/government/uploads/system/uploads/.../mental-health-access.pdf [Accessed: 16.03.2018].

Dimakou, S., et al. (2009). Identifying the impact of government targets on waiting times in the NHS. *Health Care Management Science,* 12(1), 1-10.

Dimick, J.B., and Ryan, A.M. (2014). Methods for evaluating changes in health care policy: the difference-in-differences approach. *Journal of the American Medical Association,* 312(22), 2401-2402.

Donoghue, K., et al. (2014). Cannabis use, gender and age of onset of schizophrenia: Data from the ÆSOP study. *Psychiatry Research,* 215(3), 528-532.

Doyle, R., et al. (2014). First-episode psychosis and disengagement from treatment: a systematic review. *Psychiatric Services,* 65(5), 603-611.

Drake, R.J., et al. (2000). Causes and consequences of duration of untreated psychosis in schizophrenia. *The British Journal of Psychiatry,* 177(6), 511-515.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association,* 72(359), 557-565.

Fitzpatrick, R., et al. (1998). Evaluating patient-based outcome measures for use in clinical trials: a review. *Health Technology Assessment,* 2(14).

Fortney, J.C., et al. (2005). Are primary care services a substitute or complement for specialty and inpatient services? *Health Services Research,* 40(5 Pt 1), 1422-1442.

Fridgen, G.J., et al. (2013). Help-seeking and pathways to care in the early stages of psychosis. *Social Psychiatry and Psychiatric Epidemiology,* 48(7), 1033-1043.

Fusar-Poli, P., McGorry, P.D., and Kane, J.M. (2017). Improving outcomes of first-episode psychosis: an overview. *World Psychiatry,* 16(3), 251-265.

Ghali, S., et al. (2013). Ethnic variations in pathways into early intervention services for psychosis. *The British Journal of Psychiatry,* 202(4), 277-283.

Gilbody, S.M., House, A.O., and Sheldon, T.A. (2003). Outcome measures and needs assessment tools for schizophrenia and related disorders. *Cochrane Database of Systematic Reviews,* 1, CD003081.

Goddard, M., Mannion, R., and Smith, P.C. (2000). Enhancing performance in health care: a theoretical perspective on agency and the role of information. *Health Economics,* 9(2), 95-107.

Godden, S., and Pollock, A.M. (2009). Waiting list and waiting time statistics in Britain: A critical review. *Public Health,* 123(1), 47-51.

Gravelle, H., and Siciliani, L. (2008). Is waiting-time prioritisation welfare improving? *Health Economics,* 17(2), 167-184.

Gravelle, H., and Siciliani, L. (2009). Third degree waiting time discrimination: Optimal allocation of a public sector healthcare treatment under rationing by waiting. *Health Economics,* 18(8), 977-986.

Gronholm, P.C., et al. (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.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy,* 80, 223-255.

Guloksuz, S., et al. (2016). Analyzing the duration of untreated psychosis: Quantile regression. *JAMA Psychiatry,* 73(10), 1094-1095.

Gupta, S., et al. (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.

Gutacker, N., Siciliani, L., and Cookson, R. (2016). Waiting time prioritisation: Evidence from England. *Social Science & Medicine,* 159, 140-151.

Hamilton, B.H., Hamilton, V.H., and Mayo, N.E. (1996). What are the costs of queuing for hip fracture surgery in Canada? *Journal of Health Economics,* 15(2), 161-185.

Hamilton, B.H., and Bramley-Harker, R.E. (1999). The impact of the NHS reforms on queues and surgical outcomes in England: Evidence from hip fracture patients. *The Economic Journal,* 109(457), 437-462.

Hardy, K.V., et al. (2018). Impact of age of onset of psychosis and engagement in higher education on duration of untreated psychosis. *Journal of Mental Health,* 27(3), 257-262.

Hastrup, L.H., et al. (2013). Cost-effectiveness of early intervention in first-episode psychosis: economic evaluation of a randomised controlled trial (the OPUS study). *The British Journal of Psychiatry,* 202(1), 35-41.

Heckman, J.J., and Robb, R. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics,* 30(1), 239-267.

Heckman, J.J., et al. (1998). Characterizing selection bias using experimental data. *Econometrica,* 66(5), 1017-1098.

Henderson, D.C., et al. (2000). Clozapine, diabetes mellitus, weight gain, and lipid abnormalities: A five-year naturalistic study. *American Journal of Psychiatry,* 157(6), 975-981.

Hibbard, J.H., Stockard, J., and Tusler, M. (2005). Hospital performance reports: impact on quality, market share, and reputation. *Health Affairs,* 24(4), 1150-1160.

Hirvonen, J., et al. (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., et al. (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, D.E., et al. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis,* 15(3), 199-236.

Ho, V., Hamilton, B.H., and Roos, L.L. (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, D.W., and Lemeshow, S. (2005). Assessing the fit of the model. In Hosmer, D.W., Lemeshow, S., and Sturdivant, R.X., (Eds.) *Applied Logistic Regression*. New Jersey: John Wiley & Sons, Inc., pp. 143-202.

Hosmer, D.W. (2008). *Applied survival analysis: regression modeling of time-to-event data*. 2 ed. Hoboken, New Jersey: Wiley-Interscience.

Hui, C.L.M., et al. (2013). Predictors of help-seeking duration in adult-onset psychosis in Hong Kong. *Social Psychiatry and Psychiatric Epidemiology,* 48(11), 1819-1828.

Huxley, P., and Thornicroft, G. (2003). Social inclusion, social quality and mental illness. *The British Journal of Psychiatry,* 182, 289-290.

Iyer, S., et al. (2015). Early intervention for psychosis: a Canadian perspective. *Journal of Nervous and Mental Disease,* 203(5), 356-364.

Jacobs, R. (2009). *Investigating patient outcome measures in mental health*. University of York, Centre for Health Economics: CHE Research Paper 48.

Jacobs, R., et al. (2018). Funding approaches for mental health services: is there still a role for clustering? *BJPsych Advances,* 24(6), 412-421.

Jacobson, N.S., and 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.

Januleviciute, J., et al. (2013). The impact of different prioritisation policies on waiting times: Case studies of Norway and Scotland. *Social Science & Medicine,* 97, 1-6.

Johar, M., et al. (2013). Discrimination in a universal health system: Explaining socioeconomic waiting time gaps. *Journal of Health Economics,* 32(1), 181-194.

Jones, A.M. (2007). *Applied econometrics for health economists : a practical guide*. 2nd ed. ed. Oxford: Radcliffe.

Jones, A.M., and Rice, N. (2011). Econometric evaluation of health policies. In Glied, S., and Smith, P.C., (Eds.) *The Oxford Handbook of Health Economics*. Oxford University Press: Oxford.

Jones, A.M., et al. (2013). *Applied health economics*. 2nd ed. New York: Routledge.

Jones, A.M., et al. (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.

Jones, M.P. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association,* 91(433), 222-230.

Joseph, R., and Birchwood, M. (2005). The national policy reforms for mental health services and the story of early intervention services in the United Kingdom. *Journal of Psychiatry and Neuroscience,* 30(5), 362-365.

Kaarboe, O., and Carlsen, F. (2014). Waiting times and socioeconomic status. Evidence from Norway. *Health Economics,* 23(1), 93-107.

Kaplan, E.L., and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association,* 53(282), 457-481.

Kelman, S., and Friedman, J.N. (2009). Performance improvement and performance dysfunction: An empirical examination of impacts of the emergency room wait-time target in the English National Health Service. *Journal of Public Administration Research and Theory,* 19(4), 917-946.

Kirkbride, J.B., et al. (2006). Heterogeneity in incidence rates of schizophrenia and other psychotic syndromes: Findings from the 3-center æsop study. *Archives of General Psychiatry,* 63(3), 250-258.

Kirkbride, J.B., et al. (2012a). Incidence of schizophrenia and other psychoses in England, 1950–2009: a systematic review and meta-analyses. *PLoS ONE,* 7(3), e31660.

Kirkbride, J.B., et al. (2012b). *Systematic review of the incidence and prevalence of schizophrenia and other psychoses in England*. London:: Department of Health.

Kirkbride, J.B., Stubbins, C., and Jones, P.B. (2012). Psychosis incidence through the prism of early intervention services. *The British Journal of Psychiatry,* 200(2), 156-157.

Kirkbride, J.B., et al. (2017). Waiting time variation in Early Intervention Psychosis services: longitudinal evidence from the SEPEA naturalistic cohort study. *Social Psychiatry and Psychiatric Epidemiology,* 52(5), 563-574.

Koopmanschap, M.A., et al. (2005). Influence of waiting time on cost-effectiveness. *Social Science & Medicine,* 60(11), 2501-2504.

Kramer, T.L., et al. (2003). Relationship between self-report and clinician-rated impairment in depressed outpatients. *Community Mental Health Journal,* 39(4), 299-307.

Large, M., et al. (2008). Measurement and reporting of the duration of untreated psychosis. *Early Intervention in Psychiatry,* 2(4), 201-211.

Laudicella, M., Siciliani, L., and Cookson, R. (2012). Waiting times and socioeconomic status: evidence from England. *Social Science & Medicine,* 74(9), 1331-1341.

Leurent, B., et al. (2018). Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: A tutorial. *Pharmacoeconomics,* 36(8), 889-901.

Lindsay, C.M., and Feigenbaum, B. (1984). Rationing by waiting lists. *The American Economic Review,* 74(3), 404-417.

Lloyd-Evans, B., et al. (2011). Initiatives to shorten duration of untreated psychosis: systematic review. *The British Journal of Psychiatry,* 198(4), 256-263.

Malla, A., et al. (2017). Comparing three-year extension of early intervention service to regular care following two years of early intervention service in first-episode psychosis: a randomized single blind clinical trial. *World Psychiatry,* 16(3), 278-286.

Manji, R.A., et al. (2013). Longer in-hospital wait times do not result in worse outcomes for patients requiring urgent coronary artery bypass graft surgery. *Hospital Practice* 41(3), 15-22.

Marques, E., et al. (2014). Disclosing total waiting times for joint replacement: Evidence from the English NHS using linked HES data. *Health Economics,* 23(7), 806-820.

Marshall, M., et al. (2000). *Dying to know: Public release of information about quality of health care*. Nuffield Trust Series No. 12. London, Santa Monica: Nuffield Trust and RAND.

Marshall, M., et al. (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.

Marshall, M., and Rathbone, J. (2006). Early intervention for psychosis. *Cochrane Database of Systematic Reviews,* 4, CD004718.

Marshall, M., et al. (2014). Impact of early intervention services on duration of untreated psychosis: data from the National EDEN prospective cohort study. *Schizophrenia Research,* 159(1), 1-6.

Martin, S., and Smith, P.C. (1999). Rationing by waiting lists: an empirical investigation. *Journal of Public Economics,* 71(1), 141-164.

Marwaha, S., and Johnson, S. (2004). Schizophrenia and employment - a review. *Social Psychiatry and Psychiatric Epidemiology,* 39(5), 337-349.

Marwaha, S., et al. (2016). Fifteen years on – early intervention for a new generation. *The British Journal of Psychiatry,* 209(3), 186-188.

McCabe, R., Saidi, M., and Priebe, S. (2007). Patient-reported outcomes in schizophrenia. *The British Journal of Psychiatry,* 191(50), s21-s28.

McClelland, R., et al. (2000). Validation of an outcome scale for use in adult psychiatric practice. *Quality in Health Care,* 9(2), 98-105.

McGrath, J., et al. (2008). Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiologic Reviews,* 30, 67-76.

McLennan, D., et al. (2011). *The English indices of deprivation 2010*. [Online] Department for Communities and Local Government. Available at: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010 [Accessed 21 March].

Meier-Kriesche, H.U., et al. (2000). Effect of waiting time on renal transplant outcome. *Kidney International,* 58(3), 1311-1317.

Monitor (2015). Guidance on mental health currencies and payment. [Online], [Accessed 14.10.2015].

Monstad, K., Engesæter, L.B., and Espehaug, B. (2014). Waiting time and socioeconomic status - an individual-level analysis. *Health Economics,* 23(4), 446-461.

Morgan, C., et al. (2006a). Clinical and social determinants of duration of untreated psychosis in the ÆSOP first-episode psychosis study. *The British Journal of Psychiatry,* 189(5), 446-452.

Morgan, C., et al. (2006b). Duration of untreated psychosis and ethnicity in the ÆSOP first-onset psychosis study. *Psychological Medicine,* 36(2), 239-247.

Morgan, C., et al. (2014). Reappraising the long-term course and outcome of psychotic disorders: the AESOP-10 study. *Psychological Medicine,* 44(13), 2713-2726.

Morris, S., Sutton, M., and Gravelle, H. (2005). Inequity and inequality in the use of health care in England: an empirical investigation. *Social Science & Medicine,* 60(6), 1251-1266.

Moscelli, G., Siciliani, L., and Tonei, V. (2016). Do waiting times affect health outcomes? Evidence from coronary bypass. *Social Science & Medicine,* 161, 151-159.

Naylor, C., et al. (2012). Long term conditions and mental health. [Online]. Available at: www.kingsfund.org.uk.

Nelder, J.A., and Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 135(3), 370-384.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics,* 14(4), 945-966.

NHS Digital (2017). Mental Health Services Data Set. [Online]. Available at: http://content.digital.nhs.uk/mhsds

NHS England (2015). *Guidance to support the introduction of access and waiting time standards for mental health services in 2015/16*. London: NHS England.

NHS England, and 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. [Online], National Institute for Health and Care Excellence (NICE), Version from 12 February 2014. Available at: nice.org.uk/guidance/cg178.

NICE (2015). Psychosis and schizophrenia in adults. Quality standard. [Online], National Institute for Health and Care Excellence (NICE), Version from 12 February 2015. Available at: nice.org.uk/guidance/qs80.

Nikolova, S., Sinko, A., and Sutton, M. (2015). Do maximum waiting times guarantees change clinical priorities for elective treatment? Evidence from Scotland. *Journal of Health Economics,* 41, 72-88.

Nikolova, S., Harrison, M., and 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., et al. (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.

Nolin, M., et al. (2016). Early Intervention for Psychosis in Canada: What Is the State of Affairs? *Canadian Journal of Psychiatry,* 61(3), 186-194.

Nordentoft, M., and Albert, N. (2017). Early intervention services are effective and must be defended. *World Psychiatry,* 16(3), 272-274.

Norman, R.M., and Malla, A.K. (2001). Duration of untreated psychosis: a critical examination of the concept and its importance. *Psychological Medicine,* 31(3), 381-400.

O'Donoghue, B., et al. (2016). Neighbourhood characteristics and the incidence of first-episode psychosis and duration of untreated psychosis. *Psychological Medicine,* 46(07), 1367-1378.

O'Neill, S., et al. (2016). Estimating causal effects: considering three alternatives to difference-in-differences estimation. *Health Services & Outcomes Research Methodology,* 16, 1-21.

Oudhoff, J.P., et al. (2007). The acceptability of waiting times for elective general surgery and the appropriateness of prioritising patients. *BMC Health Services Research,* 7 (no pagination)(32).

Parabiaghi, A., et al. (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.

Park, R.E. (1966). Estimation with heteroscedastic error terms. *Econometrica,* 34(4), 888-888.

Pedersen, T.J., Bogh, L.N.B., and Lauritsen, J.M. (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).

Penttilä, M., et al. (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.

Perkins, D.O., et al. (2005). Relationship between duration of untreated psychosis and outcome in first-episode schizophrenia: a critical review and meta-analysis. *The American Journal of Psychiatry,* 162(10), 1785-1804.

Pirkis, J.E., et al. (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).

Poyraz, C.A., et al. (2015). Factors associated with the duration of untreated illness among patients with obsessive compulsive disorder. *Comprehensive Psychiatry,* 58, 88-93.

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.

Propper, C. (1990). Contingent valuation of time spent on NHS waiting lists. *The Economic Journal,* 100(400), 193-199.

Propper, C. (1995). The disutility of time spent on the United Kingdom's National Health Service waiting lists. *The Journal of Human Resources,* 30(4), 677-700.

Propper, C., and Wilson, D. (2003). The use and usefulness of performance measures in the public sector. *Oxford Review of Economic Policy,* 19(2), 250-267.

Propper, C., et al. (2008). Did 'Targets and Terror' reduce waiting times in England for hospital care? *B.E. Journal of Economic Analysis & Policy,* 8(2), Article 5.

Propper, C., et al. (2010). Incentives and targets in hospital care: Evidence from a natural experiment. *Journal of Public Economics,* 94(3), 318-335.

Quintana, J.M., et al. (2011). Waiting time for cataract extraction: Predictive factors and influence on outcomes. *Journal of Cataract and Refractive Surgery,* 37(1), 19-26.

Rabe-Hesketh, S., and Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata*. 3rd ed. College Station, Tex.: Stata Press.

Ramsey, J.B. (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., et al. (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., and Hong, L.E. (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., and Priebe, S. (2012). Measuring patient-reported outcomes in psychosis: conceptual and methodological review. *The British Journal of Psychiatry,* 201(4), 262-267.

Revier, C.J., et al. (2015). Ten-year outcomes of first-episode psychoses in the MRC ÆSOP-10 study. *Journal of Nervous and Mental Disease,* 203(5), 379-386.

Robinson, D., et al. (2003). Effect of the UK government's 2-week target on waiting times in women with breast cancer in southeast England. *The British Journal of Cancer,* 89(3), 492-496.

Rosenbaum, P.R., and Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika,* 70(1), 41-55.

Rosenbaum, P.R., and Rubin, D.B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician,* 39(1), 33-38.

Rubin, D.B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology,* 66(5), 688-701.

Ryan, A.M., Burgess, J.F., and Dimick, J.B. (2015). Why we should not be indifferent to specification choices for difference-in-differences. *Health Services Research,* 50(4), 1211-1235.

Sari, I., et al. (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.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics,* 6(2), 461-464.

Seidlitz, A., et al. (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, S.S., and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika,* 52, 591-611.

Shapiro, S.S., and Francia, R.S. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association,* 67(337), 215-216.

Sharma, A.P., Siciliani, L., and Harris, A. (2013). Waiting times and socioeconomic status: Does sample selection matter? *Economic Modelling,* 33, 659-667.

Siciliani, L., Moran, V., and Borowitz, M. (2014). Measuring and comparing health care waiting times in OECD countries. *Health Policy,* 118(3), 292-303.

Siciliani, L. (2016). Waiting times: Evidence of social inequalities in access for care. In Sobolev, B., Levy, A., and Goring, S., (Eds.) *Data and measures in health services research*. Boston, MA: Springer US, pp. 1-17.

Singh, S.P., and Grange, T. (2006). Measuring pathways to care in first-episode psychosis: a systematic review. *Schizophrenia Research,* 81(1), 75-82.

Singh, S.P. (2007). Outcome measures in early psychosis: Relevance of duration of untreated psychosis. *The British Journal of Psychiatry,* 191(50), s58-s63.

Sinko, A., et al. (2016). *Models for waiting times in healthcare: comparative study using Scottish administrative data.* Working paper. University of Manchester: Institute of Health Sciences.

Smith, P.C. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration,* 18(2-3), 277-310.

Smith, P.C., et al. (2009). *Performance measurement for health system improvement: Experiences, challenges and prospects*. Cambridge University Press.

Steele, F. (2011). Multilevel discrete-time event history models with applications for the analysis of recurrent employment transitions. *Australian and New Zealand Journal of Statistics,* 53(1), 1-20.

Su, S., et al. (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, J.Y., et al. (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., et al. (2017). Exploring the use of routine outcome monitoring in the treatment of patients with a psychotic disorder. *European Psychiatry,* 42, 89-94.

Thornicroft, G. (2011). Physical health disparities and mental illness: the scandal of premature mortality. *The British Journal of Psychiatry,* 199(6), 441-442.

Tsiachristas, A., et al. (2016). Economic impact of early intervention in psychosis services: results from a longitudinal retrospective controlled study in England. *BMJ Open,* 6(10), e012611.

Tuominen, U., et al. (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., et al. (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.

Valmaggia, L.R., et al. (2015). Duration of untreated psychosis and need for admission in patients who engage with mental health services in the prodromal phase. *The British Journal of Psychiatry,* 207(2), 130-134.

Vos, T., et al. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet,* 390(10100), 1211-1259.

Willcox, S., et al. (2007). Measuring and reducing waiting times: a cross-national comparison of strategies. *Health Affairs,* 26(4), 1078-1087.

Williams, R., and Doessel, D.P. (2001). *The economics of mental health care*. Hampshire, England: Ashgate.

Wing, J.K., et al. (1998). Health of the Nation Outcome Scales (HoNOS). Research and development. *The British Journal of Psychiatry,* 172(1), 11-18.

Wing, J.K., Curtis, R.H., and Beevor, A. (1999). Health of the Nation Outcome Scales (HoNOS). Glossary for HoNOS score sheet. *The British Journal of Psychiatry,* 174(5), 432-434.

Woodcock, T., Alan, J.P., and Bell, D. (2013). The impact of changing the 4 h emergency access standard on patient waiting times in emergency departments in England. *Emergency Medicine Journal,* 30(3), e22.

Wykes, T., and Callard, F. (2010). Diagnosis, diagnosis, diagnosis: towards DSM-5. *Journal of Mental Health,* 19(4), 301-304.

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.