The performance of mental healthcare providers in England

Valerie Moran

PhD

University of York

Economics

November 2015

Abstract

This thesis investigates the performance of mental health providers in England on resource use (length of inpatient stay and costs) and quality (readmission rates and patient outcomes). Under a new payment system, it is intended that a national tariff (price) based on national average costs will be introduced and a part of future payments will be contingent on outcomes. Therefore, providers will have incentives to control costs and improve patient outcomes. We investigate the potential to achieve these aims using two nationally representative patient-level data sets: Hospital Episode Statistics (HES) and the Mental Health Minimum Data Set (MHMDS).

We utilise multilevel models, which allows us to isolate the residual variation in our response variable attributable to providers. Residual variation is quantified using Empirical Bayes (EB) methods and comparative standard errors are used to rank providers to make inferences about performance. We model length of stay (LOS) using a Poisson model; costs using a log-linear model and a generalized linear model (GLM) with a gamma distribution and log link; outcomes using ordered probit and linear models; and costs and outcomes simultaneously using a bivariate model. We employ a comprehensive range of patient and provider covariates.

Demographic, diagnostic, severity and treatment variables are key drivers of LOS and costs. Worse outcomes are associated with severity and better outcomes with older age and social support. Provider-level emergency readmission rates are associated with lower LOS and formal admissions with higher LOS. Provider-level variables have negligible effects on outcomes but a notable effect on costs. Ranking providers by residual variation suggests some providers can improve performance. Providers performing below average face financial instability under a national tariff and when a part of payment is linked to outcomes. The correlation in provider-level residual costs and outcomes is miniscule suggesting that cost-containment and outcome improving efforts by providers should not conflict.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my PhD supervisor, Professor Rowena Jacobs and the members of my Thesis Advisory Panel: Professor Luigi Siciliani and Professor Simon Gilbody for their advice and guidance throughout this project. I would like to acknowledge the inputs of Mrs Anne Mason as a co-author on Chapter 2. This work received financial support in the form of a Centre for Health Economics (CHE) studentship. My thanks also to Dr. Pauline Allen for her support and encouragement during the final stages of my PhD.

My deepest gratitude to my parents, husband, brother and sister-in-law for their love, support and encouragement throughout the duration of my study.

Author's declaration

I declare that this thesis is my own original work, and hereby explicitly state my contribution to parts of the work that are co-authored. I am the sole author of Chapters 1 and 6.

Chapter 2 is co-authored with Professor Rowena Jacobs and Mrs. Anne Mason, Centre for Health Economics (CHE). I contributed to the chapter by conceiving and designing the research, cleaning and analysing the data, interpreting the results and drafting early and final versions. The main data set analysed in Chapter 2 is the Hospital Episode Statistics (HES) data. HES are Copyright ©1998/99 - 2012/13, re-used with the permission of The Health & Social Care Information Centre (HSCIC). All rights reserved. Earlier versions of the research contained in Chapter 2 were presented at a Health, Econometrics and Data Group (HEDG) seminar, 19th June 2013 at the University of York and at the Health Economics Study Group (HESG) meeting, Sheffield, 9th January 2014. Chapter 2 has been accepted for publication, conditional on minor revisions, as a peer-reviewed paper entitled "Variations in performance of mental health providers in the English NHS: An analysis of the relationship between readmission rates and length of stay" in the journal *Administration and Policy in Mental Health and Mental Health Services Research*. Chapters 3, 4 and 5 are co-authored with Professor Rowena Jacobs, CHE. I contributed to these chapters by conceiving and designing the research, cleaning and analysing the data, interpreting the results and drafting early and final versions. The main data set used in Chapters 3, 4 and 5 is the Mental Health Minimum Data Set (MHMDS). Under a Data Sharing Agreement with the HSCIC the MHMDS data set is released on condition that it is not shared with any third party. Copyright © 2011/12 – 2012/13, re-used with the permission of HSCIC. All rights reserved.

Earlier versions of the research contained in Chapter 3 were presented at a HEDG seminar, 19th November 2014, a seminar hosted by the International Centre for Mental Health Social Research, University of York, 24th November 2014, and at the Twelfth Workshop on Costs and Assessment in Psychiatry, Venice, 28th March 2015. My participation in the Workshop was financially supported by a scholarship awarded by the organiser; the International Centre of Mental Health Policy and Economics (ICMPE). Chapter 3 has been submitted to the journal *Health Economics*.

Chapter 4 has been published as a peer-reviewed paper entitled "Comparing the performance of English mental health providers in achieving patient outcomes" in the journal *Social Science and Medicine, (140), 127-135.* Earlier versions of the research contained in Chapter 4 were presented at a HEDG seminar, 4th June 2014; the first European Health Economics Association (EUHEA) student-supervisor and early career researcher conference, Manchester, 3rd September 2014; and the Partnership of Junior Health Analysts workshop, Leeds, 17th April 2015.

Chapter 5 has been submitted to the journal *European Journal of Health Economics*.

Chapter 1. Introduction

1.1. *Introduction*

1.1.1. Overview of mental healthcare in England

Mental health constitutes an important area of research due to its large and often neglected impact on individuals, society and the economy. Around one in four people suffer from a mental health problem each year (NHS England 2014b) and mental health problems contribute to 23% of the total burden of illness in the UK (The Centre for Economic Performance's Mental Health Policy Group 2012). Consequently, the economic cost of mental illness is estimated to be in the region of £100 billion annually – equivalent to the cost of the entire English National Health Service (NHS) (NHS England 2014b). Despite the huge disease and economic burden, only 13 percent of the NHS budget is spent on mental health (NHS England 2014b).

In England, mental healthcare is provided in both primary and secondary care settings. The focus of this thesis is secondary mental healthcare, which is provided by NHS, private and voluntary providers. There are currently 59 NHS providers - known as Mental Health Trusts (NHS Choices 2015b). Until April 2013, mental health services were commissioned (organised and purchased) by Primary Care Trusts (PCTs) who also provided some mental health services. Since April 2013, the commissioning of mental health services is primarily the responsibility of Clinical Commissioning Groups (CCGs) with specialised mental health services commissioned by NHS England – an executive non-departmental public body of the Department of Health. The majority of mental healthcare is provided in outpatient or community-based settings (Health and Social Care Information Centre 2014), with inpatient care primarily reserved for patients in the most acute phase of mental illness with circumstances or care needs that cannot be appropriately met in a less restrictive setting (Smith et al. 2015).

Mental health services have historically been funded through block contracts, with funding determined by historic allocations or dependent on available budgets (Monitor and NHS England 2013a). This payment system does not necessarily incentivise providers to deliver efficient levels of care by controlling cost or increasing levels of activity (Jacobs 2014) nor has payment been aligned to patient needs or outcomes (Monitor and NHS England 2013a). Provider payment is undergoing reform with a move from block contracts towards the National Tariff Payment System (NTPS), formerly known as Payment by Results (PbR) - the activity-based provider payment system used to reimburse providers of secondary physical healthcare. The NTPS for mental health is described more fully in Section 1.2.2.

Mental health has been a key focus of recent government policy and also featured prominently in the Five Year Forward View, which sets out a vision for the NHS (NHS England 2014b). A key objective highlighted in recent policy documents has been parity of esteem between mental and physical health (Department of Health 2011; Department of Health 2014; Department of Health and NHS England 2014; NHS England 2014b). A number of policies have been introduced in order to achieve this goal including patient choice of mental health provider and mental health team (NHS England 2014a) and waiting times standards with up to £120 million in additional funding earmarked to implement new access and waiting time targets (Department of Health and NHS England 2014). The introduction of the NTPS has the potential to establish greater parity of esteem between mental and physical health and facilitate the policies of patient choice and waiting time standards. The disparity in payment systems between mental and physical healthcare risk a diversion in resources from mental health towards physical health where better activity data has made the return on investment of limited budgets more transparent to commissioners (Department of Health and NHS England 2014; Jacobs 2014). Therefore, while the introduction of the NTPS to mental health is meant to be budget neutral (Mason and Goddard 2009), it may encourage increased spending on mental health as commissioners will have a greater transparency over the marginal investment in mental health compared to physical health services. NHS England has placed an onus on CCGs to increase spending on mental health in real terms during 2015/16, and ensure growth in spending will at least equal each CCGs allocation increase (NHS England 2014c).

The introduction of the NTPS will place increased emphasis on provider's performance in relation to resource use and patient outcomes. Provider's (lagged) costs will inform the prices they will be paid to treat patients with a possible prospect of basing prices on the average costs across all providers while it is also intended that a part of payment will be linked to patient outcomes (Department of Health Payment by Results team 2013b). This motivates an investigation of the performance of mental healthcare providers in England to understand the extent to which there are variations in performance and if these can be explained by observable patient and provider factors. The remaining or residual variation in performance can then be interpreted as potentially amenable to actions on behalf of providers of care and commissioners to the extent that commissioners or policymakers can incentivise changes in provider behaviour.

1.1.2. Overview of thesis

This thesis investigates the performance of mental healthcare providers in England in the context of the introduction of the NTPS to mental healthcare. The relative performance and resource use of mental health providers in England is comparatively under-researched and this thesis makes an important contribution to the limited evidence base. Provider performance is assessed in relation to resource use in the form of length of inpatient stay and costs of care as well as patient outcomes. Our approach to measuring provider performance draws on Shleifer's (1985) theory of yardstick competition whereby a given provider is rewarded based on its performance compared to the average in a group of similar providers.

There is a compelling case for measuring the performance of mental health providers; resources are limited and the sector is under pressure to increase productivity (Monitor 2013). Mental health providers are currently operating in a tight financial climate. Since 2011/12, spending on mental health has decreased in real terms, despite increases in demand for mental health services (Smith et al. 2015). Across all mental health trusts, there was a cut in funding of 2.3% in real terms between 2011/12 and 2013/14, with the budgets of some providers falling by more than 10% (Lintern 2014a). For the financial year 2014/15, the economic regulator (Monitor) recommended nominal price adjustments for use in local negotiations as -1.5% for acute services and -1.8% for non-acute services including mental health with the differential justified by the need for acute services to meet the costs of implementing the recommendations of the Francis and Keogh reports (Monitor and NHS England

2013a). Nevertheless, improvements in care standards as recommended in these reports are as relevant for non-acute and mental healthcare as for acute care. The relatively larger cut in prices for non-acute services was viewed by many as indicative of the “institutional bias” towards acute trusts endemic in the funding system (Lintern 2014b; Lintern 2014c). Moreover, many mental health providers are not convinced that commissioners will fulfil aspirations to increase mental health funding for 2015/16 (Appleby, Thompson and Jabbal 2015; Lintern 2015a) due to the uncertain financial environment (Lintern 2015b).

The budget cuts in recent years have led to reductions in mental health staff and beds. Data from the majority (52) of Mental Health Trusts reveal reductions in nursing staff of 6% and doctors of 2% from 2011/12 to 2013/14 (Lintern 2014a). There was a 4% reduction in beds between 2011/12 and 2013/14 (Lintern 2014a) with concern that these were not balanced by increased capacity in community-based services (Edwards 2014) where demand has increased (Ahmed et al. 2015). There is evidence of increasing demand pressures on available beds with levels of bed occupancy running at over 100%, which raises concerns about care quality and patient safety (Royal College of Psychiatrists 2011; The Commission to review the provision of acute inpatient psychiatric care for adults 2015; Williams et al. 2014). Admission thresholds have increased (Csipke et al. 2014; Sabes-Figuera et al. 2012; The Commission to review the provision of acute inpatient psychiatric care for adults 2015) to the extent that patients are being admitted under the Mental Health Act (MHA) in order to access inpatient treatment (House of Commons Health Committee 2013). High pressure on beds is also evidenced by increased out-of-area placements: between 2011/12 and 2013/14 the number of patients travelling beyond their local NHS trust area to access emergency mental health treatment rose by 132% (The Commission to review the provision of acute inpatient psychiatric care for adults 2015). Much of the spare capacity for these placements is provided by the private sector, which is adding to the financial pressures placed on NHS providers (Ahmed et al. 2015). Placements far from a patient’s home may negatively affect their care and outcomes by removing them from their immediate support network. Despite these constraints, Monitor has identified scope to achieve productivity gains of between £0.5 billion to £1.3 billion in mental health with reductions in (LOS) length of stay cited as a key lever to realize some of these financial gains (Monitor 2013). Therefore it is clear that the mental health sector is under pressure to make the most of available resources by reducing inefficiencies while safeguarding patient outcomes.

The thesis is comprised of four main chapters. In Chapters 2 and 3 we examine variations in resource use across mental health providers in terms of inpatient LOS and costs. In the absence of cost data, LOS can serve as a good proxy for cost as it is a key driver of hospital costs, especially in mental health where care is staff-intensive (Mason et al. 2011). In Chapter 2, we examine variation in inpatient LOS among mental health providers in England, with a particular focus on the relationship between LOS and quality as reflected by provider emergency readmission rates. In Chapter 3, we investigate the performance of mental health providers in England in relation to cost efficiency by explaining variations in costs due to observable patient risk factors and comparing the unexplained variation in provider-level costs across providers. The objective of Chapter 4 is to explore what factors contribute to variations in patient outcomes across providers and whether providers differ systematically in terms of performance on unexplained residual variation in outcomes. Chapter 5 considers the relationship between costs and outcomes in order to determine if there is an evident trade-off between provider performance objectives of cost control and outcome improvement.

Our dependent variable of interest in Chapter 2 is inpatient LOS. LOS is a key driver of resource use and there are wide variations in LOS among mental healthcare providers in England. LOS has been identified as a key instrument with which to realise productivity gains (Monitor 2013). Under the NTPS providers will have an incentive to reduce LOS in order to control cost but if patients are discharged too early with inadequate follow-up in the community, this can have a detrimental effect on quality of care and patient outcomes. We explain variations in LOS using a comprehensive set of admission-level, patient-level and provider-level characteristics with a particular focus on the relationship between LOS and emergency readmission rates. Unexplained provider-level variation in LOS is captured by a random effect and quantified using Empirical Bayes (EB) techniques to make inferences about provider performance. The largest drivers of increased LOS at admission level are in-hospital death, a primary diagnosis of psychosis, formal detention, discharge to social care and the oldest age group (65 years and over). At a patient-level, Black ethnicity is associated with the largest increase in LOS. At a provider-level, the proportion of formal admissions under the MHA has a large positive association with LOS. The provider emergency readmission rate has a strong negative association with LOS implying that providers with high emergency readmission rates are associated with a significantly shorter LOS. Variations in residual LOS are evident across providers indicating that a number of providers have potential to improve performance on LOS, which will gain importance under a national tariff in the NTPS.

Chapter 3 investigates variations in resource use across mental health providers in terms of the costs associated with a period of care in a hospital or community setting to make assertions about provider performance. Under the NTPS, it is intended that future tariffs (national prices) will be based on the national weighted average costs of admitted and non-admitted care and initial assessments (to ascertain if a patient will enter secondary care and be reimbursed under the NTPS). This will provide an incentive for providers to control costs and increase efficiency. We cost mental healthcare activity across both hospital and community-based settings that will be reimbursed under the NTPS for public providers of specialist mental healthcare in England. We compare variations in costs across providers and explain these variations using a comprehensive set of risk adjustment variables to ascertain what factors are associated with higher or lower costs. The risk adjustment variables include the 21 care clusters that are the units of activity for which payment will be made under the NTPS as well as demographic, treatment and social variables. Residual variation in costs is quantified using EB methods and compared across providers to provide insights into which providers have the potential to make financial surpluses or losses under the new payment system. Results show that the care clusters do not explain all variation in costs. Clusters reflecting greater severity and need are associated with higher costs. Admission under the MHA and having care co-ordinated under the Care Programme Approach (CPA) (a method of assessing, planning and reviewing the needs of a person with severe mental illness) are indicative of higher costs and may be picking up aspects of severity not adequately captured by the care clusters. The key demographic cost drivers are Black ethnicity, older age, and male gender. Variables measuring provider type, size, capacity and formal admissions are also found to be associated with costs. We find evidence of differentials in provider performance with a number of providers demonstrating higher or lower residual costs independent of observable patient risk-factors controlled for. This residual cost should be susceptible to cost-controlling actions on the part of providers, which will gain heightened significance under national prices in the new payment system.

Under the NTPS for mental health, it is intended that a part of provider payment will be linked to provider's performance on patient outcomes. Chapter 4 explores provider performance in relation to patient outcomes and the potential for the NTPS to incentivise improved performance. Outcomes are measured using a Clinician Rated Outcome Measure (CROM) – the Health of the Nation Outcome Scales (HoNOS). We apply the concept of Reliable and Clinically Significant Change (RCSC) to a pair of HoNOS scores recorded at the beginning and end of a period of care for which payment is made. The majority of observations are classified with a stable outcome while relatively small proportions experience outcomes classed as clinically significant deteriorations or improvements. We model the ordered outcome variable using a hierarchical ordered probit model. Risk adjustment covariates reflect demographic, need, severity and social indicators. A hierarchical linear model is also estimated with the follow-up total HoNOS score as the dependent variable and the baseline total HoNOS score included as a risk-adjuster. Provider performance is captured by a random effect that is quantified using EB methods. We find that worse outcomes are associated with higher severity and better outcomes with older age and social support. High baseline HoNOS scores (worse outcomes) are predictive of high follow-up HoNOS scores (worse outcomes). In terms of provider variables, mental health beds have a positive association with mental health outcomes while bed occupancy is associated with worse outcomes. After adjusting outcomes for various risk factors, variations in performance are still evident across providers. This suggests that when an element of provider payment becomes contingent on patient outcomes, some providers may gain financially whilst others may lose.

The introduction of the NTPS to mental healthcare brings opportunities for providers to control costs and improve patient outcomes. However, there may potentially be a trade-off between these two objectives as improving outcomes may require additional resources to be expended and controlling costs may negatively impact patient outcomes. Chapter 5 brings together the research of chapters 3 and 4 by examining both costs and outcomes together to ascertain if incentives to control costs provided by the new payment system can be achieved without compromising patient outcomes. We estimate a bivariate multilevel model with both cost and follow-up HoNOS scores as the dependent variables. A log-linear model is the estimation model of choice for costs and a linear model is used to model outcomes. The use of a bivariate multilevel model allows us to estimate costs and outcomes simultaneously using a set of risk adjustment variables specific to each outcome. Risk adjustment variables reflect those used in Chapters 3 and 4 covering demographic, treatment, severity and social indicators. We calculate the correlation between the residual variation in costs and outcomes at the provider-level and plot the pairwise relationship between the residual responses in order to categorise providers in terms of how they perform on both residual costs and outcomes. We find little evidence of a correlation between residual costs and outcomes at the provider-level, which suggests that concerns regarding a trade-off between costs and outcomes may not be warranted, based on current evidence. Providers fall rather evenly into four groups: high costs/better outcomes, high costs/worse outcomes, low costs/worse outcomes, and low costs/better outcomes. This suggests differential responses on the part of providers to the preliminary incentives contained within the NTPS.

This chapter first outlines the economic theory underpinning the empirical analyses; explains the policy context motivating the thesis; describes the main data set used for three of the thesis chapters; outlines the methodology common to the chapters; and provides the main contributions to research of the thesis.

1.2. *Economic Theory*

Agency theory provides a useful framework to underpin our empirical analyses. This posits that a principal delegates specific activities to an agent who receives an award upon satisfactory execution of these tasks. The delegation of responsibilities will be unproblematic if the principal has full information or if the objectives of the principal correspond to those of the agent (Smith et al. 2012). In healthcare, problems in the principal – agent relationship can arise due to asymmetry of information. The principal will have less information than the agent on the nature of the production function and external circumstances influencing costs or outcomes. This means that the agent can incur costs above the optimal level or deliver outcomes below the optimal level and the principal will be unable to distinguish whether suboptimal costs or outcomes are due to the amount of effort employed by the agent or exogenous factors outside of the agent's control. In general, the agent will exert less effort than the principal would choose as the utility of the agent decreases as the level of effort (e.g. to increase activity or efficiency) increases (Burgess et al. 2011; Smith et al. 2012).

In order to address these opposing objectives, the principal can use the agent's reward or payment as an instrument to increase effort. Simply rewarding the agent (in our case a mental health provider) based on historical or anticipated costs will not incentivise efficient performance. Instead, Shleifer (1985) proposes that the price a provider receives is related to the costs of similar providers in the same sector and this will simulate a competitive environment whereby the provider will have an implicit incentive to increase effort. The motivation underlying this theory of "yardstick competition" is that providers wish to maximise profits, but tend to exert minimum effort unless profits are at stake. The principal or regulator can exploit this profit motive by basing the price any one provider receives on the average of the costs incurred by all providers. Therefore, if a provider reduces costs when a competitor does not, a profit is incurred; similarly, if a provider does not reduce costs relative to its competitors, it faces a loss. This approach overcomes the information asymmetry on the part of the regulator as it does not need to know each provider's specific production function, merely the costs incurred by all providers in the sector. Moreover, if the regulator observes specific exogenous characteristics that explain variations in costs across providers, then these can be accounted for in a cost regression to give a more accurate price signal (Shleifer 1985). This approach can also apply to the comparisons of outcomes achieved by a group of providers.

1.3. *Policy context*

1.3.1. Provider payment for mental health in an international context

Internationally, prospective activity-based payment is increasingly used to pay providers. In acute physical healthcare, the use of prospective payment systems based on casemix classification systems has been associated with reductions in unit costs (Street et al. 2011) and improvements in some aspects of quality of care (Or and

Hakkinen 2011). Prospective activity-based payment remains limited in mental healthcare (Mason and Goddard 2009) as diagnosis is often one of the main variables underpinning the unit of activity for which payment is made but diagnosis is not a strong predictor of resource use in mental health (McCrone 1994; McCrone and Phelan 1994; McCrone 1995).

Provider payment systems for mental healthcare that incentivise both cost efficiency and quality improvement have been a policy focus in a number of countries. Some countries such as the Netherlands have included psychiatric care in the prospective activity-based payment system for inpatient and outpatient care (Kobel et al. 2011). This activity-based payment system takes account of the type of care and treatment provided as well as diagnosis (Forti et al. 2014). Cost control is incentivised by nationally agreed unit prices and the system also incentivises quality improvements that lead to lower resource consumption (Swan-Tan et al. 2011). Other countries have implemented prospective payment but chosen an alternative payment unit to diagnosis casemix groups, such as the United States which reimburses psychiatric inpatient care under Medicare using a per diem system as LOS is an important determinant of inpatient cost (Mason and Goddard 2009). This system links payment to average cost in order to encourage efficiency, while aspects of the payment system are also designed to prevent adverse effects on quality of care (Mason and Goddard 2009).

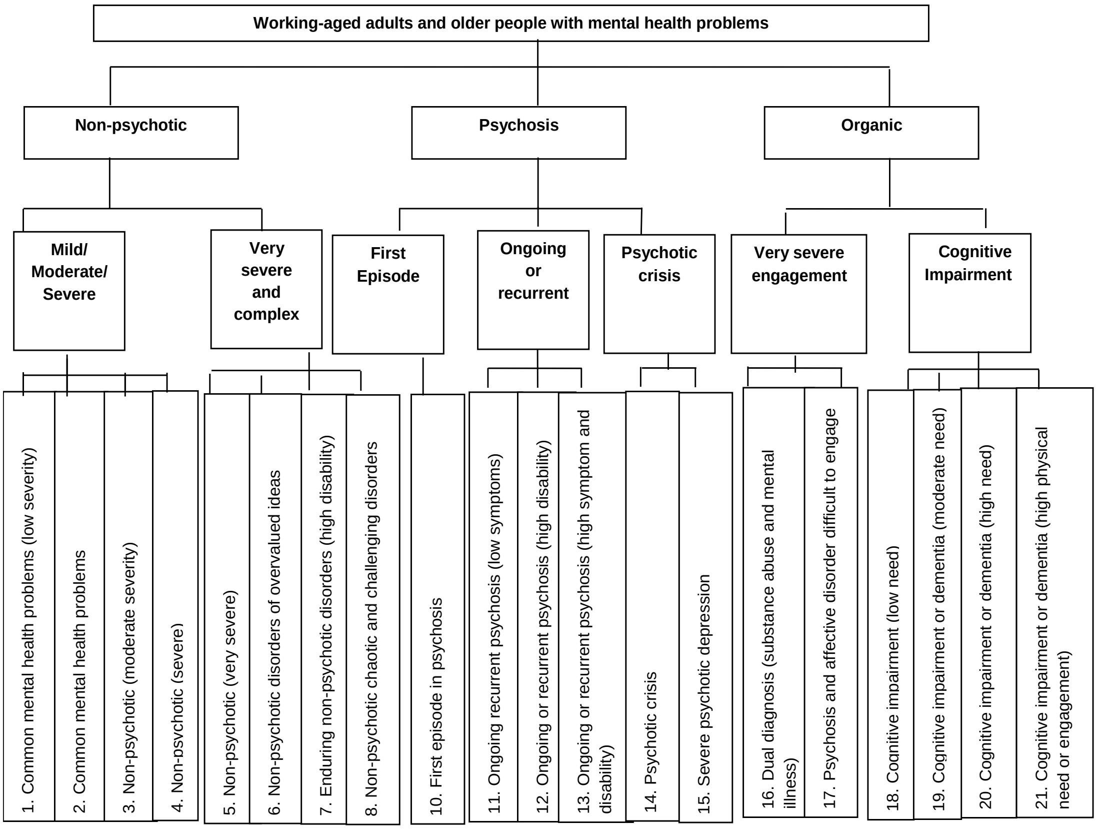
A number of countries including Australia, Canada (Ontario) and New Zealand have developed casemix classification systems specific to mental health that have incorporated information on patient severity, functioning, and legal status as well as diagnosis. The Australian and New Zealand systems included outcomes in the form of the HoNOS. In both countries provider factors were shown to drive cost variations rendering the classification systems unsuitable for provider payment although was this was an explicit objective in Australia only (Buckingham et al. 1998; Gaines et al. 2003). The classification system developed in Ontario has also not been used to fund mental health services although this is a future intention (Mason and Goddard 2009).

1.3.2. The National Tariff Payment System (NTPS) for Mental Health

In an international context, England has made considerable advances towards introducing prospective activity-based payment to mental healthcare. The system used in the acute physical healthcare sector – the NTPS – has been extended to mental health. Under the NTPS in the acute physical healthcare sector, activity is grouped into Healthcare Resource Groups (HRGs) that represent groups of patients that are clinically and economically homogenous and these are the currencies (units of activity) for which payment is made. A national fixed price or tariff per HRG is set by the economic regulator (Monitor) and this corresponds to the national average (lagged) cost of treating patients in a particular HRG. The national tariff gives providers an incentive to increase the efficiency of care provision in order to avoid making a loss and increase surpluses. The tariff also negates the need for price negotiations between commissioners and providers and contracts can focus more on the quantity and quality of care provided (Fairbairn 2007; Jacobs 2014). Therefore, proponents of the introduction of the NTPS to mental health have alluded to potential benefits arising from increased efficiency and quality of mental healthcare (Evans-Lacko et al. 2008).

A new classification system has been developed for mental health with a primary focus on patient severity, an important predictor of mental health resource use (Trauer 2010b). The currencies for which payment will be made under the NTPS for mental health are 21 care clusters which are grouped into three superclasses corresponding to non-psychotic, psychosis and organic mental illness (Figure 1.1).

Figure 1.1 Relationship of care clusters to each other



Users of mental healthcare services are allocated to a cluster by clinicians using the Mental Health Clustering Tool (MHCT). The MHCT incorporates items from HoNOS (Wing, Curtis and Beevor 1996) and the Summary of Assessments of Risk and Need (SARN) (Self R. 2008) in order to provide all relevant information to allocate individuals to clusters (Monitor and NHS England 2013b). Part 1 of the MHCT encompasses the HoNOS items, which provide information on current problems in terms of the severity of symptoms experienced by the user during the two weeks preceding the MHCT assessment. Part 2 of the clustering tool encompasses the SARN items, which assess historical problems that occur less frequently or sporadically. All MHCT items are rated from 0 (no problem) to 4 (severe to very severe problem) (Monitor and NHS England 2013d). Table 1.1. presents an overview of the MHCT items, parts and rating.

Table 1.1 Overview of Mental Health Clustering Tool (MHCT)

Rating	Part	Item
0 = no problem 1 = minor problem requiring no action 2 = mild problem but definitely present 3 = moderately severe problem 4 = severe to very severe problem Rate 9 if Not Known	**1: Current Ratings.** For scales 1-13, rate the most severe occurrence in the previous two weeks.	1. Overactive, aggressive, disruptive or agitated behaviour
		2. Non-accidental self-injury
		3. Problem-drinking or drug-taking
		4. Cognitive problems
		5. Physical illness or disability problems
		6. Problems associated with hallucinations and delusions
		7. Problems with depressed mood
		8. Other mental and behavioural problems
		9. Problems with relationships
		10. Problems with activities of daily living
		11. Problems with living conditions
		12. Problems with occupation and activities
		13. Strong unreasonable beliefs that are not psychotic in origin
	2. Historical Ratings Scales A-E rate problems that occur	A. Agitated behaviour/expansive mood

	in an episodic or unpredictable way. Include any event that remains relevant to the current plan of care.	B. Repeat self-harm
		C. Safeguarding other children & vulnerable adults
		D. Engagement
		E. Vulnerability

Source: Monitor and NHS England (2013d). National Tariff 2014/15 Payment System Annex 7C Mental health clustering tool booklet. London, Monitor

Upon completion of a MHCT assessment, clinicians identify a care cluster that corresponds to the needs of the service user. An electronic algorithm has been developed to aid clinicians in this task, which provides a probability of a service user being assigned to a particular cluster. A clinician is however able to override the algorithm allocation and the ultimate allocation decision is based on clinical judgement (Monitor and NHS England 2013a). If a clinician is unable to identify an appropriate cluster, the service user is allocated to a variance cluster (cluster 0) and the reasons for this course of action are recorded by the clinician (Monitor and NHS England 2013b). It is expected that the use of cluster 0 will decline with time as the clustering process becomes embedded in clinical practice (Monitor and NHS England 2013b) and the data suggests that this is indeed the case.

Unlike in the acute physical healthcare sector where the currencies are primarily based on procedures, the currencies in mental healthcare are based on patient characteristics and need. Diagnosis is not explicitly taken into account in the allocation of service users to clusters except for the three super-classes and the same diagnosis can be associated with several clusters, depending on level of need. This may result in considerable variation in casemix within and between the clusters (Jacobs 2014). To date, there has not been independent validation of the care clusters in terms of their clinical and resource homogeneity (Jacobs 2014).

The clusters are mutually exclusive and it is possible to allocate a service user to only one cluster at any time. If a patient changes to a new cluster, the previous cluster can no longer be used for payment purposes (Monitor and NHS England 2013b). The packages of care provided following allocation to a cluster is decided at the local level between the clinician and service user (Department of Health Payment by Results team 2013b; Monitor and NHS England 2013b). Clusters do not correspond to a particular care setting, so that care takes place in the most clinically appropriate, cost-effective and least restrictive care setting possible (Department of Health Payment by Results team 2013b; Jacobs 2014). The aspiration is that each cluster will have a fixed national price or tariff (Department of Health Payment by Results team 2013b) calculated from the national weighted average cost of admitted and non-admitted care for a cluster along with the national average cost for an initial assessment (Jacobs 2014). This will provide a strong incentive for providers to control costs, for example by treating patients in non-admitted (outpatient, community) care settings and reducing more costly admitted care, as providers with costs above the tariff will incur financial losses while those with costs below the tariff will make a surplus. It is likely that additional top-up payments or alternative funding arrangements in addition to the core cluster payment will be established to cover the cost of more specialised services (Monitor and NHS England 2013b). Nevertheless, the current policy emphasis is on the local delivery of the NTPS in mental health with recommendations to adapt national guidance on the System to suit local implementation and to develop and negotiate local cluster prices (Department of Health Payment by Results team 2013a; Monitor and NHS England 2014).

While the use of a national tariff introduces a strong incentive for providers to control costs, this may come at the expense of patient outcomes and quality of care. In order to prevent any detrimental effect on quality of care and patient outcomes in the drive to increase efficiency, a suite of quality and outcome measures have been developed that will be mandated in contracts (Department of Health Payment by Results team 2013b). These include a CROM based on the MHCT/HoNOS as well as quality indicators drawn from data collected in the Mental Health Minimum Data Set (MHMDS) (see Section 1.3.1.) covering treatment, accommodation and data quality (Department of Health Payment by Results team 2013b). In addition, a Patient Related Outcome Measure (PROM) and Patient Related Experience Measures (PREMS) are under development for future use (Department of Health Payment by Results team 2013b). It is important that the CROM is complemented by PROMs and PREMs as a clinician-rated measure may be more susceptible to manipulation or gaming with providers tempted to over-report improvements in order to improve performance and ultimately payment whereas PROMs and PREMs are likely to be more transparent (Jacobs 2010; Trauer 2010b; Yeomans 2014).

It is envisioned that as the payment system evolves, commissioners and providers will agree on a component of payment that will be conditional on outcomes achieved (Department of Health Payment by Results team 2013b). This means that clinicians will have a direct impact on the funding that their organisation receives through their work to deliver high quality care and to achieve better outcomes (Department of Health Payment by Results team 2013b). The move towards linking payment to outcomes in mental health mirrors the introduction of Best Practice Tariffs in acute physical healthcare, with outcomes measured using PROMs (Monitor and NHS England 2013a). In order to receive the payment, providers must attain best practice criteria in terms of achieving an average health gain that is not significantly below the national average and meeting data submission standards. Non-fulfilment of these criteria means that the provider receives a price 10% below the best practice price (Monitor and NHS England 2013e).

The impetus for moving towards the NTPS in mental health is based on a need to increase transparency, link funding to local mental health needs (and away from historical block contracts), reduce variation in mental health services, enhance personalisation and choice, and achieve value for money (Department of Health 2010). Extending the NTPS to mental health also helps to place mental healthcare on an equal footing with physical healthcare in order to achieve parity of esteem. The production of more tangible results under the NTPS should help to protect mental health funding and prevent disinvestment in favour of acute care where it is easier for commissioners to assess what they are spending their budgets on (Jacobs 2014).

There are some important differences between the design and implementation of the NTPS in physical and mental healthcare: 1) implementation of the NTPS in mental health has not been as well resourced as in physical health; 2) a primary goal for mental health is to relate payment to outcomes and quality, rather than just activity as has been the case for the majority of physical health; 3) the care clusters were uniquely developed for mental health and presented a completely different and novel approach to classifying activity, whereas the classification system used in physical health was adapted from already established payment systems in other countries. Thus, mental health has faced greater challenges, which has likely adversely affected data quality in the initial years and the timelines for implementation.

1.4. *Data*

The primary data sources used in the thesis are administrative patient-level data sets with national coverage for England. Our main data set for Chapter 2 is Hospital Episode Statistics (HES) - a patient-level administrative data set of all admissions, outpatient appointments and Accident and Emergency (A&E) attendances at NHS hospitals in England. As HES is only used in Chapter 2, it is fully described therein along with a range of provider-level variables used in Chapter 2. Our main data set for the remaining three chapters is the MHMDS. In order to avoid unnecessary repetition in the description of this data set, it is described fully in Section 1.3.1. We also give a brief overview of a small number of provider-level variables included in sensitivity analyses in Chapters 3 and 4 in Section 1.3.2.

1.4.1. The Mental Health Minimum Data Set (MHMDS)

The MHMDS is a patient-level data set with national coverage for England. The data set was introduced in April 2000 to facilitate the collection of clinical data in mental healthcare at a national level to support clinical audit, service planning and management. Three years later, it became mandatory for providers of specialist, including elderly, mental health services funded by the NHS to deliver MHMDS data on a quarterly and annual basis. The MHMDS contains data on all the care and treatment received by a service user from the first referral to specialist mental healthcare to the final discharge. This treatment can take place in hospital and community-based settings including inpatient, outpatient, and day care and also encompasses contacts with different mental health teams and health professionals in the community. Data is also collected on primary and secondary diagnoses, whether the patient is under the CPA or treatment is provided under a section of the MHA. There is also information on clinical and social outcomes (HoNOS, employment, accommodation). The variables included in the MHMDS have evolved over the years and we use Version 4.0 which covers 2011/12 and 2012/13 and differs to previous versions with the inclusion of information pertaining to the new payment system including MHCT ratings and the care cluster assigned to the patient.

Following referral to specialist mental healthcare by a GP or self-referral, or following a request from police or social services, service users will undergo an ‘initial assessment’ (Monitor and NHS England 2013b). The patient is assessed in order to determine if they need treatment in specialist care and require allocation to a care cluster or their needs can be adequately met in an alternative care setting (such as primary care or other services) (Monitor and NHS England 2013b). This initial assessment is not reimbursed under the care clusters but as a separate currency (unit of activity for which payment is made) (Monitor and NHS England 2013b). If a service user enters specialist care they undergo a MHCT assessment and are allocated to a cluster. This allocation is reviewed at regular intervals to ensure that the cluster continues to adequately meet the needs of the service user. It is recommended that MHCT assessments take place at CPA or other formal care reviews, and on occasions when a change in a services user’s needs necessitates a significant modification of planned care, for example if a service user is admitted to inpatient care (Monitor and NHS England 2013d). Following the MHCT assessment, the service user may remain in the same cluster or move to a different cluster. The interval between MHCT assessments is referred to as a ‘Cluster Review Period’ (CRP) and this is the unit of observation for the analyses in Chapters 3-5. The CRP forms the basis of contracts and prices agreed between commissioners and providers (Department of Health Payment by Results team 2013b; Monitor and NHS England 2013b) and so it is appropriate to analyse costs associated with CRPs. Moreover, the MHCT assessment process facilitates the appraisal of whether there has been an improvement in a service user’s wellbeing during a period of care (Department of Health Payment by Results team 2013b) implying that the CRP is a suitable entity for measuring and analysing outcomes. Maximum review periods have been recommended for each of the clusters (Table 1.2) (Department of Health Payment by Results team 2013b).

Table 1.2 Maximum review periods

Cluster Number	Cluster label	Cluster review period (CRP) (Maximum)
0	Variance	6 months
1	Common mental health problems (low severity)	12 weeks
2	Common mental health problems	15 weeks
3	Non-psychotic (moderate severity)	6 months
4	Non-psychotic (severe)	6 months
5	Non-psychotic (very severe)	6 months
6	Non-psychotic disorders of overvalued ideas	6 months
7	Enduring non-psychotic disorders (high disability)	Annual
8	Non-psychotic chaotic and challenging disorders	Annual
9	Blank cluster	Not Applicable
10	First episode in psychosis	Annual
11	Ongoing recurrent psychosis (low symptoms)	Annual
12	Ongoing or recurrent psychosis (high disability)	Annual
13	Ongoing or recurrent psychosis (high symptom and disability)	Annual
14	Psychotic crisis	4 weeks
15	Severe psychotic depression	4 weeks
16	Dual diagnosis (substance abuse and mental illness)	6 months

17	Psychosis and affective disorder difficult to engage	6 months
18	Cognitive impairment (low need)	Annual
19	Cognitive impairment or dementia (moderate need)	6 months
20	Cognitive impairment or dementia (high need)	6 months
21	Cognitive impairment or dementia (high physical need or engagement)	6 months

Source: Department of Health Payment by Results team (2013b). Mental Health Payment by Results Guidance for 2013-14. Leeds, Department of Health.

The MHMDS includes a number of variables that can be used to risk-adjust patient costs and outcomes. Risk adjustment controls for differences in patient casemix when comparing the costs and outcomes of treatment (Smith and Street 2013). We cannot simply assume that providers working in the same specialty, such as mental healthcare, treat patients with a homogenous casemix. A patient's risk factors are likely to influence the costs incurred and outcomes achieved by providers, meaning that a comparison of providers treating patients with different risk profiles will result in a misleading picture of relative provider performance in costs and outcomes. Variables considered as risk-adjusters should not be related to the treatment and should be measured prior to or at the onset of treatment (Dow, Boaz and Thornton 2001). Potential risk-adjusters should explain variation in the dependent variable of interest and there should be some degree of uniformity in the relationship between the dependent and risk adjustment variables across providers. It is also necessary to guard against using too many risk adjustment variables as this could mask genuine variation in costs and outcomes attributable to providers (Dow, Boaz and Thornton 2001) that we want to measure as performance.

It is important to highlight issues of data completeness and data quality in the MHMDS that may be partly explained by our usage of data for 2011/12 and 2012/13 which cover the initial years of the development and implementation of the NTPS in mental health. The allocation of patients to care clusters commenced in 2011 and the mandatory use of the clusters as the basis for contracting mental health services for working-age and older adults was introduced only in 2012. Particular issues that pertain to our analyses concern the limited coding of variables measuring primary and secondary diagnoses that inhibits the use of these variables in Chapters 3-5 as well as a considerable reduction in estimation sample size in Chapter 4 arising from a lack of follow-up HoNOS ratings. Moreover, it is opportune to underline that there are shortcomings to the Reference Cost data used in Chapters 3 and 5, which is fully described in Section 3.3.1 of Chapter 3. These shortcomings relate to high variability in the costs reported both within and between providers as well as missing data. Again, these issues may relate to the use of data collected at a nascent stage of the development of the payment system.

1.4.2. Provider variables

We do not include provider-level variables in the main analyses in Chapters 3 and 4, as our objective is to control only for observable patient factors that may lead to variations in costs and outcomes across providers and allow provider factors to be captured in the provider-level residual and be indicative of performance. Nevertheless, it is of interest whether variables reflecting provider governance and capacity constraints have an effect on patient costs and outcomes. Therefore, in sensitivity analyses in Chapters 3 and 4 we include a number of provider-level variables in the estimation models including Foundation Trust (FT) status, mental health beds, mental health bed occupancy and proportion of formal admissions (in Chapter 3 only). We source these data from the websites of the Health and Social Care Information Centre (HSCIC) – the national provider of information, data and IT systems for health and social care in England – and NHS England.

1.5. *Methods*

1.5.1. Multilevel generalized linear model (GLM)

In chapters 2-4, we use a random effects multilevel GLM to model responses that are not normally distributed. A link function relates the conditional mean to the covariates and multivariate normal random effects while a distribution is used to specify the relationship between the variance and the mean (Jones 2010; Rabe-Hesketh, Skrondal and Pickle 2004; Skrondal and Rabe-Hesketh 2009). As we use a different dependent variable in each chapter, different distributions and links are utilised which are described in more detail in each chapter. Advantages of a multilevel GLM are that predictions are made on the original scale of the dependent variable so it is not necessary to alter the dependent variable to facilitate model estimation or interpretation of results while heteroskedasticity can be accommodated through the choice of distributional family (Jones 2010).

1.5.2. Comparisons of provider performance: Empirical Bayes (EB) prediction of the Random Effects

Having obtained estimates of the model parameters and treating them as the true parameter values, we can predict values of the provider random effects using EB techniques. This allows us to quantify the residual variation (i.e. the unexplained variation, which remains after taking account of all the variables in our model) and compare this residual variation across providers in terms of the response variable. EB predictions combine the prior (normal) distribution with the likelihood to obtain the posterior distribution given the observed responses. The EB estimates of the provider-level random effects are "shrunken" towards the mean of the posterior distribution with the degree of shrinkage determined by the relative information available on the group. A high level of shrinkage reflects relatively little information about the group (the number of patients is small for a particular provider or the patient-level variance is large relative to the provider-level variance). Therefore this shrinkage is desirable as it means less weight is placed on units with less data (Steele 2008). In order to compare the residual variation across providers we use comparative standard errors. We assume a normal posterior distribution and known model parameters in order to form Bayesian credible intervals using the posterior mean and posterior standard deviation. The posterior standard deviation is commonly used as a standard error of prediction for multilevel GLMs (Skrondal and Rabe-Hesketh 2009). The EB estimates of the provider-level random effects are ranked and graphically displayed. The residuals represent provider departures from the overall mean, so a provider whose confidence interval does not overlap the line at zero is said to differ significantly from the average at the 5% level. However, it is not possible to conclude that two providers whose confidence intervals fail to overlap are statistically significantly different from each other at the 5% level as the confidence intervals are too wide (Goldstein and Healy 1995). The width of the confidence interval associated with a particular provider depends on the standard error of that provider's residual estimate, which is inversely related to the size of the sample (Steele 2008) so wider confidence intervals signify relatively smaller numbers of observations in that provider.

1.6. *Contribution to research*

This thesis makes a contribution to the current evidence base in a number of ways. The research offers a comprehensive and rigorous analysis of the performance of mental healthcare providers in England in terms of resource use (LOS, costs) and patient outcomes (including readmission rates). To date, such analysis is lacking due to the paucity of national data on mental healthcare in community settings. This meant that the majority of studies have primarily relied on data with limited geographical, provider or patient samples, restricting the generalisability of results. This research overcomes this problem by using large, administrative, nationally representative data sets – HES and MHMDS, the latter of which has only recently become available for research purposes. This means that the MHMDS has not been commonly used for research and this thesis makes an important contribution in that regard. The use of HES data allows us to conduct a comprehensive study of resource use and quality of care in inpatient mental health settings. This is complemented by the MHMDS, which contains information on specialist mental healthcare and allows us to examine the entire care pathway encompassing outpatient and community mental healthcare as well as inpatient mental healthcare. Both HES and the MHMDS contain patient-level data and provider identifiers which allow the use of multilevel models in order to make inferences about the influence of different levels on the dependent variable of interest – another innovative feature of this research. Both HES and the MHMDS contain a comprehensive set of variables spanning demographic, treatment, need, and social indicators which are complemented by provider-level variables from additional sources in order to explain variations in resource use and outcomes. This moves this research beyond current literature in this field which considers a more limited range of covariates. Residual variation in the dependent variable of interest is quantified using EB techniques and comparative standard errors are used to compare residual variation across providers. To the best of my knowledge, this thesis presents the first attempts to measure and compare provider performance in mental healthcare using EB methods. The thesis also offers other innovations with regard to methodological applications to mental health data including the use of a cross-classified model, which reflects patient movement between providers, and the use of a bivariate multilevel model to analyse patient costs and outcomes simultaneously. Additional original contributions include the use of provider-level emergency readmission rates calculated using HES data, the calculation of RCSC in HoNOS scores for nationally representative data, and the costing of specialist mental healthcare activity across the entire care pathway, again at a national level.

The research presented in this thesis will be of potential benefit to policymakers by informing the design of the NTPS for mental healthcare. The research is the first attempt to investigate the ability of the care clusters to explain variations in costs for a period of care. This work also provides insights into whether observable patient and provider factors influence cost and outcome variations. The research explores how provider payment can be linked to outcomes and if providers may potentially face a trade-off between controlling costs and improving patient outcomes. Therefore, the results of these analyses can aid the development and implementation of the NTPS for mental health to ensure that providers are adequately reimbursed for providing good quality care tailored to patient's needs that is cost-efficient and improves patient's outcomes.

Chapter 2. Variations in performance of mental health providers in the English NHS: An analysis of the relationship between readmission rates and length of stay

2.1. *Introduction*

Using LOS as a proxy for cost and resource use (Martin and Smith 1996) we explore variation in LOS across providers and the extent to which patient and provider characteristics explain this variation. LOS is a key driver of hospital costs, especially when care is staff-intensive as is the case in mental health (Mason et al. 2011). Differences in LOS can reflect differences in patient needs, but can also be indicative of differences in treatment philosophies and practice patterns (Horgan and Jencks 1987). One of the risks of the NTPS is that the incentives to generate efficiencies through reducing unit costs and LOS may have unintended consequences such as skimping on quality (Jacobs 2014). A major challenge in monitoring the quality of mental healthcare lies in utilising hospital-based data to make inferences about both hospital and community care (Lakhani et al. 2005). Hospital emergency readmission rates are increasingly used as a performance measure and as a basis for hospital reimbursement (Laudicella, Li Donni and Smith 2013) and can act as a good proxy measure for inferences about both hospital and community mental healthcare.

The aim of this chapter is to examine variation in LOS among mental health providers in England, in particular the relationship between LOS and quality as reflected by provider emergency readmission rates. The study makes a unique contribution to research in two ways. First, the analysis uses three levels at which factors influence LOS, by considering admission-, patient- and provider-level variables. Second, the chapter uses a cross-classified model to explore variation in LOS and tests the sensitivity of this modelling approach by estimating a three-level hierarchical model to see if results diverge when there is a small degree of cross-classification in the data. We quantify the residual variation in LOS at the provider-level using EB estimates with comparative standard errors to compare performance across providers.

2.2. *Findings from previous literature*

We conducted a literature search for studies investigating LOS in mental healthcare using the following databases: EconLit, Embase, Medline, OvidMedline, and PsychInfo. Search terms included: “mental health”, “psychiatry”, “length of stay”, “readmission”, “inpatient”, and “performance”.

In the studies reviewed, sample size ranged from 56 (Rothbard and Schinnar 1996) to 327,797 (Harman, Cuffel and Kelleher 2004) patients. While some of the studies (Abas, Vanderpyl and Robinson 2008; Chung et al. 2010; Compton, Craw and Rudisch 2006; Dausey, Rosenheck and Lehman 2002; Hodgson, Lewis and Boardman 2000; Huntley et al. 1998; Imai et al. 2005; Lerner and Zilber 2010; Oiesvold et al. 1999; Padgett et al. 1994; Pertile et al. 2011; Rothbard and Schinnar 1996; Stevens, Hammer and Buchkremer 2001; Williams et al. 2014; Wolff et al. 2015b; Zhang, Harvey and Andrew 2011) included all psychiatric admissions irrespective of diagnosis, other studies (Chung et al. 2013; Douzenis et al. 2012; Fong Chan and Lieh Yan 2010; Harman, Cuffel and Kelleher 2004; Jacobs et al. 2015; Lay, Lauber and Rossler 2006; Peiro et al. 2004) explicitly focused their analyses on inpatients with particular diagnoses of serious mental illness, most commonly schizophrenia, bipolar and major depressive disorders.

The majority of studies (Abas, Vanderpyl and Robinson 2008; Chung et al. 2010; Compton, Craw and Rudisch 2006; Dausey, Rosenheck and Lehman 2002; Douzenis et al. 2012; Fong Chan and Lieh Yan 2010; Harman, Cuffel and Kelleher 2004; Huntley et al. 1998; Imai et al. 2005; Lay, Lauber and Rossler 2006; Padgett et al. 1994; Peiro et al. 2004; Pertile et al. 2011; Rothbard and Schinnar 1996; Williams et al. 2014) used multiple linear regression. Some studies modelled the log of LOS due to the skewed nature of LOS (Abas, Vanderpyl and Robinson 2008; Chung et al. 2010; Compton, Craw and Rudisch 2006; Lay, Lauber and Rossler 2006; Pertile et al. 2011; Rothbard and Schinnar 1996). Another common methodology employed by studies was a Cox regression (survival analysis) (Hodgson, Lewis and Boardman 2000; Lerner and Zilber 2010; Oiesvold et al. 1999; Stevens, Hammer and Buchkremer 2001).

A number of authors (Chung et al. 2013; Lay, Lauber and Rossler 2006; Wolff et al. 2015b; Zhang, Harvey and Andrew 2011) were interested in examining the predictors of long LOS with the definition of long LOS varying from greater than 12 days (Zhang, Harvey and Andrew 2011) to greater than 300 days (Lay, Lauber and Rossler 2006). Logistic regression was used to model a dependent variable equal to one if the LOS was considered long according to the definition used. Poisson (Jacobs et al. 2015) and zero-truncated negative binomial regression (Wolff et al. 2015b) were alternative methodologies employed. Several studies employed multilevel models with either fixed (Jacobs et al. 2015) or random (Chung et al. 2013; Chung et al. 2010; Harman, Cuffel and Kelleher 2004; Pertile et al. 2011; Williams et al. 2014) effects. Only a minority of studies (Huntley et al. 1998; Wolff et al. 2015b) utilised a split sample validation design whereby a part of the sample was used to estimate the model and the model was validated using the remaining sample observations.

For the purposes of our study, determinants of LOS for psychiatric inpatient care can be classified in terms of admission, patient or provider characteristics. Table 2.1 gives an overview of the main characteristics considered in previous studies and their relationship with LOS according to the three levels considered in our analysis.

Table 2.1 Literature on characteristics associated with length of stay (LOS)

Variable	**Direction of association (Reference)**
Admission-level characteristics	
Physical co-morbidities	**Positive:** (Douzenis et al. 2012) **Negative:** (Chung et al. 2013)
Diagnosis of psychosis	**Positive:** (Chung et al. 2010; Hodgson, Lewis and Boardman 2000; Huntley et al. 1998; Jacobs et al. 2015; Lay, Lauber and Rossler 2006; Lerner and Zilber 2010; Oiesvold et al. 1999; Peiro et al. 2004; Pertile et al. 2011; Tulloch, Fearon and David 2011)
Co-morbid diagnosis of Substance Misuse	**Negative:** (Compton, Craw and Rudisch 2006; Harman, Cuffel and Kelleher 2004; Huntley et al. 1998; Jacobs et al. 2015; Stevens, Hammer and Buchkremer 2001)
Legal status (compulsory admission)	**Positive:** (Jacobs et al. 2015; Kallert, Glockner and Schutzwohl 2008; Lerner

	and Zilber 2010; Pertile et al. 2011; Williams et al. 2014) **Negative:** (Compton, Craw and Rudisch 2006; Tulloch, Fearon and David 2011)
Social support	**Positive:** (Oiesvold et al. 1999) **Negative:** (Fong Chan and Lieh Yan 2010; Tulloch, Fearon and David 2011)
Prior service use	**Positive:** (Huntley et al. 1998; Lerner and Zilber 2010; Stevens, Hammer and Buchkremer 2001; Williams et al. 2014) **Negative:** (Dausey, Rosenheck and Lehman 2002; Jacobs et al. 2015; Rothbard and Schinnar 1996)
Deprivation	**Positive:** (Abas, Vanderpyl and Robinson 2008; Jacobs et al. 2015) **Negative:** (Dekker et al. 1997)
Age	**Positive:** (Chung et al. 2013; Fong Chan and Lieh Yan 2010; Hodgson, Lewis and Boardman 2000; Huntley et al. 1998; Jacobs et al. 2015; Oiesvold et al. 1999; Pertile et al. 2011) **Negative:** (Chung et al. 2010; Peiro et al. 2004; Stevens, Hammer and Buchkremer 2001) **Non-linear:** (Harman, Cuffel and Kelleher 2004; Horgan and Jencks 1987; McCrone and Lorusso 1999)
Patient-level characteristics	
Gender	**Positive for males:** (Chung et al. 2013; Chung et al. 2010; Rothbard and Schinnar 1996) **Positive for females:** (Hodgson, Lewis and Boardman 2000; Oiesvold et al. 1999; Pertile et al. 2011; Tulloch, Fearon and David 2011; Wolff et al. 2015b)
Ethnicity	**Positive for Black ethnicity:** (Jacobs et al. 2015; Padgett et al. 1994) **Positive for Jewish ethnicity:** (Lerner and Zilber 2010)
Provider-level characteristics	
Hospital type	**Positive for psychiatric hospital:** (Chung et al. 2013)

Hospital capacity	**Positive:** (Chung et al. 2013; Chung et al. 2010; Imai et al. 2005; Tulloch, Fearon and David 2011)
Human resources of healthcare professionals	**Negative:** (Chung et al. 2013; Chung et al. 2010; Imai et al. 2005)
Readmission rates	**Positive:** (Korkeila et al. 1998; Wolff et al. 2015b) **Negative:** (Appleby et al. 1993; Boden et al. 2011; Figueroa, Harman and Engberg 2004; Lin et al. 2006; Sytema and Burgess 1999)

At an admission-level, diagnostic (primary and secondary diagnoses), treatment (prior service use, involuntary admission) and socioeconomic (social support, deprivation) variables are reported as being significantly associated with LOS. In terms of diagnostic variables, the presence of physical co-morbidities as well as a diagnosis of psychosis has been found to be positively associated with longer LOS while a co-morbid diagnosis of substance misuse disorder is generally reported as reducing LOS (Table 2.1). There is less consensus in the literature regarding the effect of treatment and socioeconomic characteristics. A longer LOS may arise from an involuntary admission if such admissions are indicative of greater severity of illness. Tulloch (2011) report that being married is associated with shorter LOS. Rothbard and Schinnar (1996) posit that prior service use may be associated with a shorter LOS as patients who are familiar with the mental health system can be treated and discharged quicker. However, prior service use may also be indicative of greater severity of illness thus resulting in a longer LOS (Huntley et al. 1998). Abas et al. (2008) found that greater levels of socioeconomic deprivation in the inpatient's neighbourhood of residence were associated with extended hospitalization after adjustment for demographic factors and primary diagnosis but not after adjustment for comorbid diagnosis, chronicity, function, and severity. Jacobs et al. (2015) report a longer LOS for patients from more deprived neighbourhoods, with a larger effect for patients with bipolar disorder. Dekker et al. (1997) reported negative correlations between LOS and deprivation characteristics which they surmised resulted from deprived areas having a larger number of patients who are frequently readmitted for a short time. McCrone and Lorusso (1999) showed a non-linear relationship between LOS and age. A longer LOS for older people may be somewhat related to the availability of social supports as well as the availability of and access to continuing health and social care (Bryan 2010; Pertile et al. 2011). Moreover, presentations by elderly patients may be medically complex due to a higher risk of medical co-morbidities and adverse reactions to medications (Pertile et al. 2011).

At a patient-level, black ethnicity is associated with a longer LOS compared to white or Asian ethnicity (Jacobs et al. 2015; Padgett et al. 1994). There is no clear direction regarding the relationship between LOS and gender.

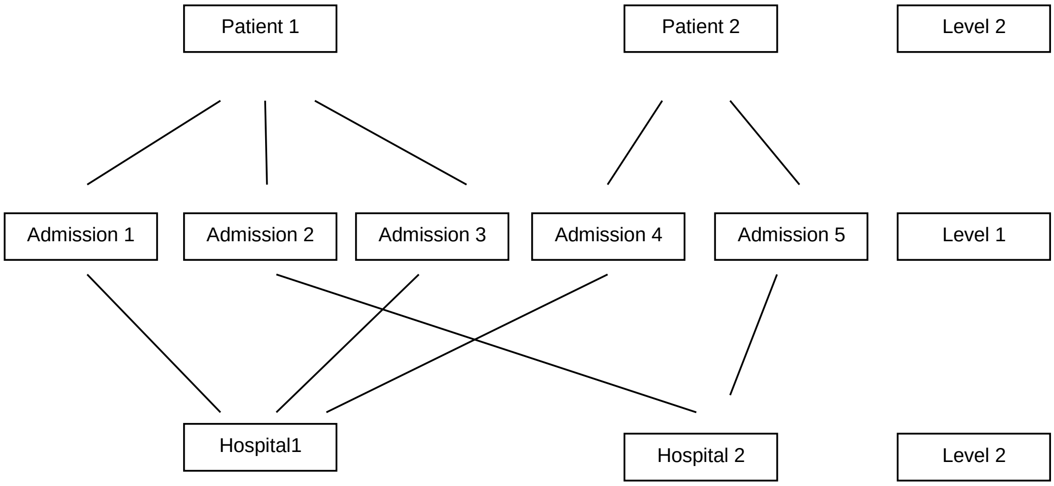
At a provider-level, hospital capacity is positively associated with LOS (Chung et al. 2010) while levels of human resources in terms of healthcare professionals have shown a negative relationship with LOS (Imai et al. 2005). The positive relationship between LOS and hospital capacity is likely due to a desire or need to keep bed occupancy levels high and may also be related to the provider payment method (e.g. per diem) and hospital efforts to increase revenues (Chung et al. 2010). Lower numbers of human resources in terms of healthcare professionals may be indicative of cost-cutting efforts on the part of hospitals that consequently reduce quality of care and increase LOS (Imai et al. 2005).

2.3. *Methods*

2.3.1. Study sample

Our data exhibits a multilevel structure with admissions nested in patients, who are nested in hospitals. Some patients have multiple admissions (spells) the majority of which are to the same hospital, but approximately 2% of patients (accounting for 4% of admissions) have admissions to different hospitals. We use a cross-classified model to reflect this non-hierarchical data structure. Cross-classified data occurs when lower-level units relate to more than one distinct higher-level unit. Lower level units will then be connected to a pair or group of higher level units resulting in two or more higher level units or classifications being crossed (Leckie 2013) as shown in Figure 2.1.

Figure 2.1 Cross-classified data structure



2.3.2. Estimation model

Our choice of estimation model is a three-level GLM that can be written as:

$$g^{-1}\{E[\,y_{ijk} \mid x_{ijk,}\, v_{jk}\, u_k]\} = \boldsymbol{\beta}\mathbf{X}_{ijk} + u_k + v_{jk} \equiv \eta_{ijk} \qquad (1)$$

where $\mathbf{X}_{ijk}$ is a column vector of admission, patient and hospital characteristics, u_k are level 3 random intercepts or hospital specific effects, v_{jk} are level 2 random intercepts or patient specific effects, $g^{-1}(.)$ is the link function and η_{ijk} is the linear predictor. The conditional expectation of the response, given the covariates and the random effects is:

$$\mu_{ijk} \equiv \{E[\,y_{ijk} \mid x_{ijk,}\, v_{jk,}\, u_k]\} = g(\boldsymbol{\beta}\mathbf{X}_{ijk} + u_k + v_{jk}) = g(\eta_{ijk}) \qquad (2)$$

The random effects are considered multivariate normal with strictly exogenous covariates (Skrondal and Rabe-Hesketh 2009). Conditional independence of the responses is assumed with conditional distributions drawn from the exponential family (Skrondal and Rabe-Hesketh 2009). The conditional variance is given by:

$$\mathrm{Var}\,(y_{ijk} \mid u_k, v_{jk}) = \varphi_{ijk}\mathrm{V}(\mu_{ijk}) \qquad (3)$$

where φ_{ijk} is a dispersion parameter and $\mathrm{V}(\mu_{ijk})$ is a variance function specifying the relationship between the conditional variance and conditional expectation. As our response variable (LOS) can be evaluated as count data, a Poisson distribution with a log link is specified. The variance function $V(\mu_{ijk}) = \mu_{ijk}$ and the dispersion parameter $\varphi_{ijk} = 1$. The Poisson distribution assumes that equi-dispersion is present implying that the conditional mean is equal to the variance. For some spells of care the conditional variance may exceed the mean so the assumption of equi-dispersion is too restrictive. Therefore we allow for an extra binomial variation parameter to allow for over- or under-dispersion. Statistical significance is tested at the 5%, 1% and 0.1% levels.

2.3.3. Empirical Bayes (EB) prediction of the Random Effects

Having obtained estimates of the model parameters and treating them as the true parameter values, we can predict values of the level 3 or hospital random effects u_k using EB techniques as outlined in Section 1.4.2 of Chapter 1. This allows us to quantify the residual variation (i.e. the unexplained variation which remains after taking account of all the variables in our model) and compare this residual variation across hospitals in terms of LOS. As we use a log link, we can interpret provider performance in days of LOS by calculating the exponentiation of the EB estimates.

2.3.4. Estimation method

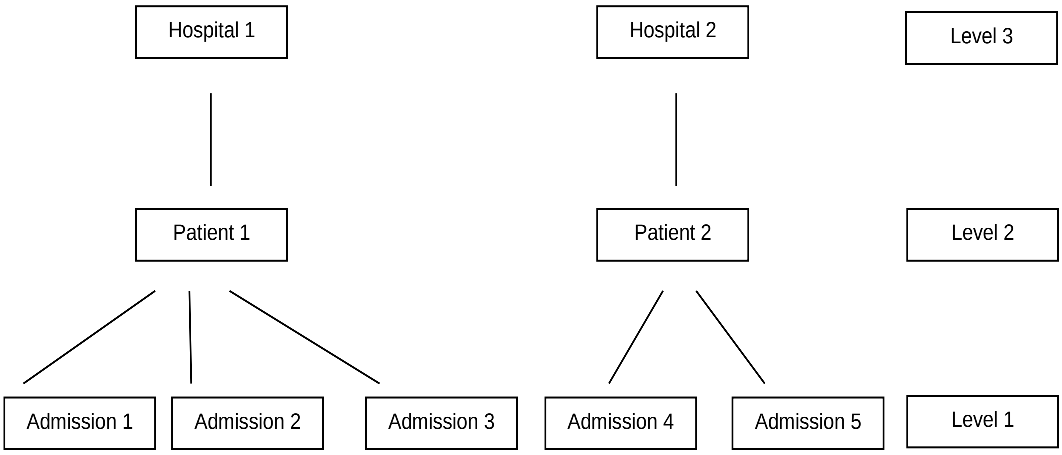
The cross-classified model is estimated using the Monte Carlo Markov Chain (MCMC) method. MCMC utilises simulation methods to produce parameter estimates (Browne 2012). For the cross-classified model presented here, the chain is first run for 5,500 iterations until the Markov chain converges and is then run for an additional 350,000 iterations. Parameter estimates and standard errors are based on the means and standard deviations of the estimates produced during each of the 350,000 iterations (Leckie and Charlton 2012).

The coefficients on the predictor variables are expressed using Incidence Rate Ratios (IRRs). The IRR represents exponentiated coefficients that can be given a multiplicative interpretation (Cameron and Trivedi 2011). Therefore, a coefficient greater than one signals that the variable exerts an upward pressure on LOS and a coefficient less than one a downward pressure. The quantitative effect of a variable is calculated as (IRR-1)*100.

2.3.5. Sensitivity Analyses

As only a small proportion of our data sample is affected by a cross-classified structure, we consider how the treatment of our data sample as a three-level hierarchy affects the estimation results. Figure 2.2 displays the three-level hierarchical data structure of admissions nested in patients who in turn are nested in hospitals in graphical form.

Figure 2.2 Three-level data structure



Our use of both three-level and cross-classified models allows us to investigate if the models produce consistent results when only a small portion of the sample does not exhibit a strict hierarchy. The three-level model is estimated using restricted iterative generalized least squares (RIGLS) which corresponds to restricted maximum likelihood (Goldstein 1989).

We also re-estimate the cross-classified model using a data set that excludes the observations with zero LOS (patients who are admitted and discharged on the same day or day cases) to test if admissions with a positive LOS better reflect resource use.

The models were estimated in MLwiN 2.29 (Rabash et al. 2009) using the *runmlwin* command (Leckie and Charlton 2012) in Stata 13.0 (StataCorp 2013).

2.4. *Data*

2.4.1. Data sources and coverage

We drew on the literature on the determinants of LOS for psychiatric inpatient admissions to inform our choice of independent variables in our models. The independent variables comprise a range of admission-, patient- and provider-level variables that are likely to influence LOS. Admission- and patient-level variables were sourced from HES, a patient-level administrative data set of all admissions, outpatient appointments and A&E attendances at NHS hospitals in England. HES information is stored as a large collection of separate records - one for each period of care - in a secure data warehouse and it is managed by the HSCIC. Our study used HES data for 2009/10 and 2010/11. Sources for provider-level variables include HSCIC (variables sourced from HES), Hospital Activity Statistics, the Care Quality Commission, and the Department of Health Staffing Survey. These data are all publicly available on the websites of the respective organisations. The Hospital Activity Statistics are published by NHS England - an executive non-departmental public body of the Department of Health. The Care Quality Commission is the independent regulator of all health and social care services in England.

Our data set consists of 63 public mental healthcare providers comprising Mental Health Trusts, Care Trusts and Primary Care Trusts (PCTs). Mental Health Trusts provide health and social services for people with mental health problems, in particular specialist services for people with severe mental health problems (NHS Choices 2015b). Care Trusts provide a range of services including social care and mental health services. PCTs provide the equivalent full set of mental health services as Mental Health Trusts, but are unable to become FTs. FTs differ from other NHS Trusts in that they are independent legal bodies and have different governance arrangements. They are not subject to the same levels of performance management and have significant financial freedoms (NHS Choices 2015a).

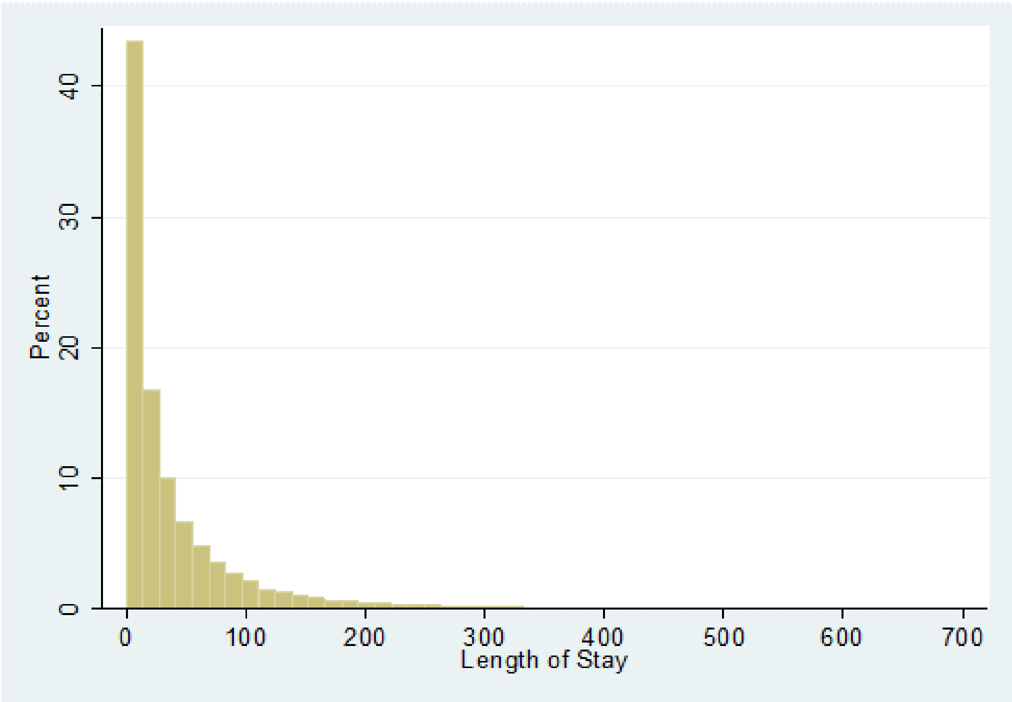
We selected our study sample by trimming episodes of care to cover only patients with mental disorders treated by mental healthcare providers. More specifically, we dropped observations for admissions to PCTs without a record of an ICD-10 F chapter (Mental and Behavioural Disorders) code or a HRG Version 3.5 T code (Mental

Health); observations for patients admitted prior to 1st April 2009 so that the data set consists only of patients with finished episodes that were admitted during 2009/10 and discharged during 2009/10 or 2010/11; and admissions with incorrectly coded age.

2.4.2. Dependent variable

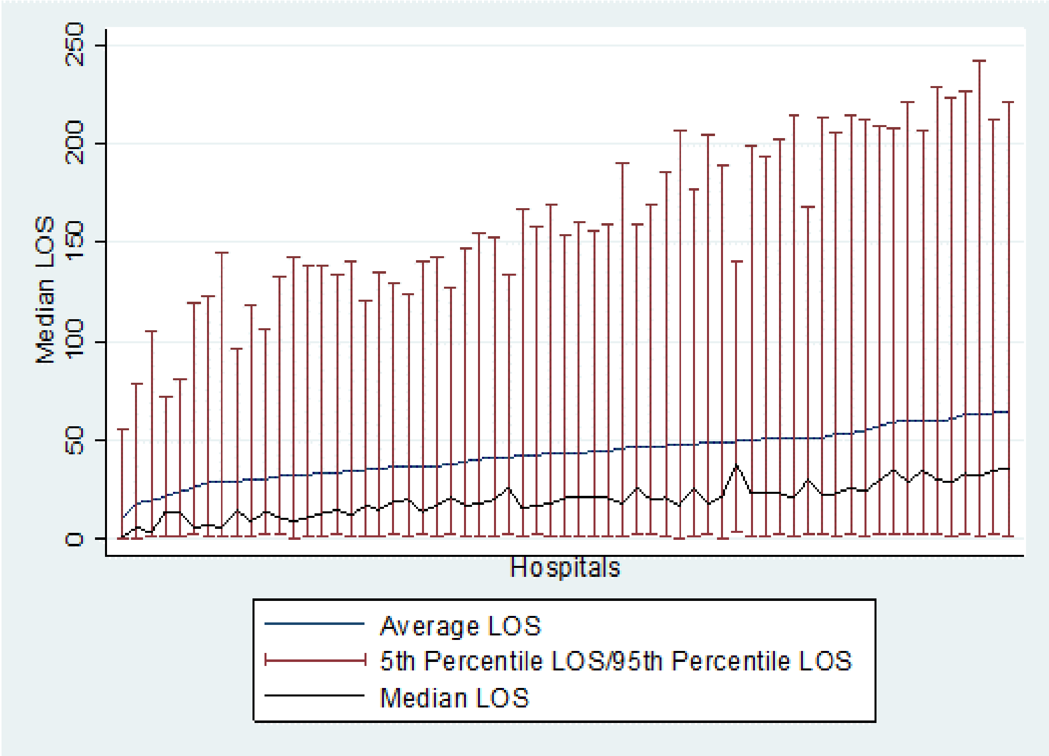
The dependent variable LOS is measured by the time elapsed between admission and discharge dates. LOS per admission ranges from a minimum of 0 days to a maximum of 708 days (Figure 2.3).

Figure 2.3 Length of Stay (LOS) by patient admission



There is substantial variation in LOS between providers (Figure 2.4).

Figure 2.4 Variation in length of stay (LOS) between providers



2.4.3. Independent variables

Level 1 relates to an admission i.e. a period of care in one provider. Level 1 independent variables reflect the diagnostic, treatment and socioeconomic characteristics of patients included in the study and can potentially change from admission to admission. Patients can be transferred to or from another hospital provider. This indicator may be a proxy for patient casemix because providers may specialize in the treatment of certain diagnoses and patients may be transferred if the provider they were originally admitted to cannot meet the needs of the patient. Patient death in hospital captures if the patient died during a particular admission. It reflects the proportion of all admissions with a reason for discharge coded as death. Patient death in hospital is a relatively rare event but it can act as an indicator of the quality of care provided (Department of Health and Human Services Agency for Healthcare Research and Quality 2002). Co-morbidities are measured using the total number of secondary diagnoses recorded for an admission. We include a number of variables that describe the ICD-10 mental health chapter codes that represent the most common primary diagnoses recorded. Severity is reflected in psychiatric history represented by one or more previous psychiatric admissions. Another severity variable indicates if patients have been formally detained under the MHA. Marital status and a record of carer support signal the extent of social support available to patients. Information on deprivation is captured by the Index of Multiple Deprivation (IMD). The IMD is measured at Lower Layer Super Output Area (LSOA) level and subsequently assigned to patients on the basis of residency. LSOAs are a geographic hierarchy with a minimum population of 1000 and a mean of 1500 (Health and Social Care Information Centre 2013). The IMD has seven domains, of which the IMD Income Domain is included in our analysis. The purpose of this Domain is to capture the proportions of the population experiencing income deprivation in an area (Noble 2008). A higher score for the IMD Income Domain indicates a greater proportion of the population in the area in which the patient lives experiences income deprivation. The scores for the Income Domain are rates which we multiplied by 100 to ease interpretation. So, for example, if an LSOA scores 0.72 in the Income Deprivation Domain, this means that 72% of the LSOA's population is income deprived. We indicate if an admission has been discharged to social care and this will include any delayed discharge which may increase LOS. Age ranges from 3 to 104 years and is divided into 5 categories to capture any non-linearity in the relationship between LOS and age. Age category 2 (18-39) is the reference category.

Level 2 relates to variables measured at the patient-level that do not change from admission to admission. Patient-level independent variables cover demographic characteristics of patients. Gender is measured as a dummy variable with females as the reference category. Patient ethnicity is categorized into White (the reference category), Asian, Black and Other ethnicity (e.g. mixed race, or unknown ethnicity).

Level 3 variables are measured at the provider-level and vary only for admissions and patients in different hospitals. Provider-level independent variables describe provider type and capacity, proportion of formal admissions, emergency readmission rate, co-morbidities recorded by a provider, quality of care and human resources. Two dummy variables are included in the models to indicate if a provider is 1) a Mental Health Trust and/or 2) has FT status.

The variable “total available beds” provides a measure of hospital size. Total bed occupancy provides an indication of utilisation of available bed capacity and reflects average bed occupancy over a quarterly time period for 2010/11 and an annual time period for 2009/10. Human resources variables are measured as the percentage of medical staff from total Full Time Equivalent (FTE) staff and the percentage of nursing staff from total FTE staff. Nurses make up a higher percentage of total FTE staff. The proportion of formal admissions under the MHA provides information on patient severity. We include a variable on formal admissions at provider-level as well as admission-level as we expect providers to have different thresholds for detention. Similarly, we include a provider-level variable measuring the average number of co-morbidities recorded by a provider. This complements the admission-level co-morbidity variable and controls for systematic under- or over-recording of co-morbidities by providers.

The study utilises emergency readmission rates for mental health providers which have not been calculated nationally before. We calculated rates for mental health providers using HES data, following a methodology used for acute providers (Health and Social Care Information Centre 2011). However, we adapted this to include readmissions treated by mental health specialities. The HSCIC excludes mental health speciality in its standard calculation of emergency readmission rates for acute providers. In the calculation of readmission rates for mental health providers, the numerator is based on a pair of admissions – the discharge (index) admission and the next readmission to reflect emergency admissions within 28 days of discharge from hospital. The readmission includes cases where the patient dies but excludes those with a main speciality of obstetrics or learning disability upon readmission, and those with a diagnosis of cancer (other than benign or in situ) or chemotherapy for cancer coded anywhere in the admission. The denominator excludes day cases, admissions with a discharge coded as death, admissions with obstetric and learning disability specialities and those with a diagnosis of cancer or chemotherapy treatment for any form of cancer in the 365 days prior to admission.

Quality of care is also represented by a number of variables upon which providers are performance managed by the regulator, the Care Quality Commission. Crisis Resolution and Home Treatment (CRHT) teams provide intensive home-based support for people in mental health crises in their own home and stay involved until the problem is resolved (Care Quality Commission 2009). An aim of CRHT teams is to prevent hospital admissions; therefore access to CRHT teams can provide an indication of the level of gate-keeping available. This indicator is measured using the number of admissions to the Trust's acute wards (excluding admissions to psychiatric intensive care units) that were "gate-kept" by the CRHT teams as a percentage of the total number of admissions to the Trust's acute wards (excluding admissions to psychiatric intensive care units). The indicator CPA 7 day follow-up measures the extent to which people under adult mental illness specialities on CPA receive follow-up (by phone or face-to-face contact) within seven days of discharge from psychiatric inpatient care. Providers are judged to have "achieved" this indicator if at least 95% of patients receive timely follow-up post-discharge. The patient experience score is based on five domains: access and waiting; safe, high quality, coordinated care; better information, more choice; building relationships; and clean, comfortable, friendly place to be and a higher score indicates a more positive experience (Care Quality Commission 2010). We hypothesise that efforts by providers to drive down LOS may be associated with commensurate declines in quality.

2.5. *Results*

2.5.1. Descriptive statistics

Table 2.2 presents the descriptive statistics for our data sample, presented according to the levels (admissions; patients; providers) with reference categories in brackets.

Table 2.2 Descriptive statistics

Variable	Source	Mean	Standard Deviation	Min	Max
LOS (days)	Derived from HES	43	66	0	708
Admission-level variables (n=133,156)					
Patient transfer-in	HES	0.210	0.407	0	1
Patient transfer-out	HES	0.063	0.243	0	1
Patient death in hospital	HES	0.009	0.096	0	1
Total number of comorbidities	Derived from HES	1	1	0	17

Primary diagnosis of psychosis	Derived from HES	0.156	0.363	0	1
Primary diagnosis organic disorder	Derived from HES	0.057	0.231	0	1
Primary diagnosis mood disorder	Derived from HES	0.203	0.402	0	1
Primary diagnosis substance misuse disorder	Derived from HES	0.094	0.291	0	1
Primary diagnosis neurotic disorder	Derived from HES	0.054	0.226	0	1
Primary diagnosis personality disorder	Derived from HES	0.054	0.227	0	1
Formally detained under the MHA	HES	0.128	0.334	0	1
Carer support recorded	HES	0.072	0.259	0	1
Married/civil partner	HES	0.186	0.389	0	1
One or more previous psychiatric admission	HES	0.407	0.491	0	1
Income Deprivation	HES	20	13	0	83
Discharge to social care	HES	0.059	0.236	0	1
Age Category 1 (under 18)	HES	0.048	0.213	0	1
Age Category 2 (18-39)	HES	0.380	0.485	0	1
Age Category 3 (40-49)	HES	0.190	0.392	0	1
Age Category 4 (50-64)	HES	0.156	0.363	0	1
Age Category 5 (65+)	HES	0.226	0.418	0	1
Patient-level variables (n=90,980)					
Patient gender: male	HES	0.506	0.500	0	1
Patient ethnicity: White	HES	0.848	0.359	0	1
Patient ethnicity: Asian	HES	0.045	0.206	0	1
Patient ethnicity: Black	HES	0.052	0.221	0	1
Patient ethnicity: Other	HES	0.056	0.230	0	1
Provider-level variables (n=63)					
Foundation Trust (FT)	HES	0.625	0.484	0	1
Mental Health Trust	HES	0.928	0.258	0	1
Total available beds	HAS	512	252	14	1237
Total bed occupancy (%)	HAS	85.1	5.7	63.5	97.9
Proportion of formal admissions under the MHA	HSCIC	0.184	0.075	0.044	0.650
Emergency readmission rate by provider	Derived from HES	0.115	0.034	0.053	0.226
Average comorbidities recorded by provider	Derived from HES	1	1	0	4
CPA 7 day follow-up (%)	CQC	97.1	2.8	82.7	100
Patient experience total score	CQC	298.7	10.7	273.5	325.8

Access to Crisis Resolution Home Treatment (CRHT) team (gatekeeping) (%)	CQC	95.6	5.5	71.9	100
Percentage of medical staff from total Full Time Equivalent (FTE) staff	DoH SS	5.7	2	1.3	10.9
Percentage of nurses from total FTE staff	DoH SS	32.3	4.1	19.4	40

HES: Hospital Episode Statistics; HAS: Hospital Activity Statistics; HSCIC: Health and Social Care Information Centre; CQC: Care Quality Commission; DoH SS: Department of Health Staffing Survey.

Our estimation sample consists of 133,156 admissions in 90,980 patients that are treated in 63 hospitals. In-hospital death is a relatively rare event - it affects only 1% of admissions in this data set. The admissions in our sample had, on average, one co-morbidity recorded, but there was sizeable variation with some admissions recording no co-morbidities and up to seventeen co-morbidities recorded for others. Mood disorder was recorded for 20% of admissions, making it the most common primary diagnosis followed by psychosis (16%) and substance misuse disorder (9%) while the primary diagnoses of organic, neurotic and personality disorders accounted for 6% or less of admissions. Given high admission thresholds it is somewhat surprising that only 16% of observations are coded with a diagnosis of psychosis. This may be partly explained by the relatively sizeable coding of observations (around 20%) with "Unknown and unspecified causes of morbidity" (ICD-10 R69X). Given the challenges of diagnosing mental illness (Timimi 2014) and the reluctance on the part of some clinicians to attach diagnostic labels to patients (Ben-Zeev, Young and Corrigan 2010; Sartorius 2002; Timimi 2014), the numbers of admissions with a diagnosis of psychosis may be underreported in our data. Approximately 13% of admissions were involuntary (i.e. the individual was detained under the MHA). Again, we might expect a larger proportion of observations to be formally detained under the MHA given the high severity thresholds for admission. However, a recent study (Jacobs et al. 2015) using HES data to investigate LoS for people with serious mental illness reports 19.3% of the sample as formally detained. Given that we include a broader range of diagnoses in our sample, the number of admissions under formal detention in our study does not appear unreasonable.

In terms of social support, almost one-fifth (19%) of the sample was married or had a civil partner while less than one-tenth (7%) had a record of carer support. However, the latter may underestimate the true extent of carer support as it only reflects patients for whom there is a formal record. Almost half (41%) of the sample have a history of psychiatric treatment. The income deprivation variable has a minimum of 0 and a maximum of 83%, with a mean of 20%, which implies that the average admission was from a neighbourhood where 20% of residents experienced income deprivation. On average, 6% of admissions were discharged to a social care setting. Just over half of patients in our sample were male (51%). White ethnicity accounts for the majority of the sample (85%), followed by Other (6%), Asian (5%) and Black (5%).

The majority (93%) of providers in the sample are Mental Health Trusts and almost two-thirds (63%) of providers have FT status. Compared with those treated in Mental Health Trusts without FT status, it is interesting to note that individuals in our data set who are treated by FTs are less likely to be transferred in from another hospital, have less comorbidity, and are less likely to have psychosis or a substance misuse disorder. They are less likely to have been formally detained, to have previous psychiatric admissions or be male, and more likely to be aged 65 years or over, or of White ethnicity.

The mean proportion of formal admissions under the MHA is 0.18 but this varies widely across providers from 0.04 to 0.65 suggesting that providers have different formal admission thresholds or different types of local populations. The mean emergency readmission rate is 0.12 but for one provider approximately one in twenty patients is readmitted while for another almost one in four patients is readmitted within 28 days. The providers in our sample have on average been successful in achieving the indicator measuring CPA 7-day follow-up with a mean score of 97% and a maximum score of 100%. However a minimum score of 83% indicates that some providers failed to achieve adequate patient follow-up. On average, 6% of FTE staff are medical staff while 32% of FTE staff are nurses.

2.5.2. Estimation results

Table 2.3 presents the estimates of the cross-classified (baseline) and three-level (sensitivity analysis) models. In the following paragraphs we discuss the results of the cross-classified (baseline) model.

Table 2.3 Estimates of cross-classified and three-level models

	Observations per group			
	Number of Observations	Minimum	Average	Maximum
Level 3: Hospital	63	234	2113.6	5377
Level 2: Patient	90,980	1	1.5	86
Level 1: Admission	133,156			
	Cross-classified model		Three-level model	
	IRR	Standard Error	IRR	Standard Error
Constant	4.908	0.319***	6.087	7.287
Patient transfer-in	1.150	0.003***	1.065	0.009***
Patient transfer-out	0.660	0.002***	0.836	0.012***
Patient death in hospital	1.862	0.021***	1.454	0.047***
Total number of comorbidities	1.142	0.001***	1.098	0.004***
Primary diagnosis of psychosis	1.406	0.005***	1.493	0.017***
Primary diagnosis organic disorder	1.196	0.008***	1.049	0.018**
Primary diagnosis mood disorder	1.231	0.005***	1.022	0.012
Primary diagnosis substance misuse disorder	0.795	0.006***	0.450	0.008***
Primary diagnosis neurotic disorder	0.903	0.006***	0.667	0.013***
Primary diagnosis personality disorder	1.116	0.007***	0.826	0.017***
Formally detained under the MHA	1.477	0.005***	1.603	0.017***
Carer support recorded	0.732	0.005***	0.850	0.020***
Married/civil partner	1.097	0.007***	0.907	0.010***
One or more previous psychiatric admission	0.815	0.002***	0.890	0.008***
Income Deprivation	0.992	0.000***	0.996	0.000***
Discharge to social care	2.010	0.008***	2.008	0.027***
Age Category 1 (under 18)	1.131	0.023***	1.345	0.035***
Age Category 3 (40-49)	1.129	0.011***	1.010	0.012
Age Category 4 (50-64)	1.381	0.015***	1.220	0.015***
Age Category 5 (65+)	2.049	0.021***	1.493	0.019***
Patient gender: male	0.993	0.008	1.023	0.009**
Patient ethnicity: Asian	1.171	0.019***	1.065	0.021**
Patient ethnicity: Black	1.314	0.019***	1.165	0.021***
Patient ethnicity: Other	0.731	0.007***	0.858	0.016***
Foundation Trust (FT)	1.028	0.080	0.975	0.052

Mental Health Trust	1.041	0.087	1.364	0.116***
Total available beds	1.000	0.000	1.000	0.000
Total bed occupancy (%)	1.006	0.001***	1.006	0.004
Proportion formal admissions under the MHA	3.608	1.533***	2.692	0.731***
Emergency readmission rate by provider	0.052	0.039***	0.372	0.284
Average comorbidities recorded by provider	0.897	0.046***	0.967	0.032
CPA 7 day follow-up	1.007	0.001***	1.006	0.010
Patient experience total score	1.003	0.001***	1.004	0.002
Access to Crisis Resolution Home Treatment (CRHT) team (gatekeeping)	0.996	0.001***	0.996	0.005
Percentage of medical staff from total Full Time Equivalent (FTE) staff	0.997	0.007	0.998	0.015
Percentage of nurses from total FTE staff	0.989	0.001***	0.990	0.006
Random Effects Parameters	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**
Level 3: Hospital	0.088	0.017	0.026	0.005
Level 2: Patient	1.355	0.007	0.733	0.008
Overdispersion parameter	41.960	0.000	41.960	0.268

*****p<0.001, **p<0.01, *p<0.05**

Admission-level variables with a significant positive association with LOS include transfer-in, inpatient death, number of co-morbidities, a primary diagnosis of psychosis, organic, mood or personality disorders, formal detention, married/civil partner, discharge to social care and age less than 18 years or over 39 years. Of these, the variables measuring death in hospital, a diagnosis of psychosis, formal detention, discharge to social care and age of 65 years or over have the largest significant effects on LOS. More specifically, patient death is associated with an 86% increase in LOS. A primary diagnosis of psychosis is associated with an increase of LOS of 41% while detention under the MHA is associated with an increased LOS of almost 50%. Patients discharged to social care are associated with a LOS twice the length of those who are discharged elsewhere. Similarly, age 65 years or over is associated with a doubling of LOS. The association of a longer LOS with inpatient death, psychosis and formal detention is likely to reflect greater disease severity among these admissions. The variables measuring a primary diagnosis of organic disorder and mood disorder are also associated with relatively large increases in LOS of around 20%. Admissions transferred from another provider and total number of comorbidities are associated with a higher LOS of 14-15%. A primary diagnosis of personality disorder and marriage/civil partner are associated an increased LOS of 10-11%.

Admission-level variables with a statistically significant negative association with LOS include transfer-out, a primary diagnosis of substance misuse or neurotic disorders, a record of carer support, psychiatric treatment history and income deprivation. A primary diagnosis of substance misuse disorder is associated with a 20% reduction in LOS. This finding may be because the presence of substance misuse disorder without a mental disorder diagnosis precludes detention under the MHA which we find to be associated with a longer LOS. A primary diagnosis of neurotic disorder is associated with a reduced LOS of 10%. A record of carer support is associated with a 27% reduction in LOS. Patients may be discharged earlier if a carer is available to provide care at home. Transfer-out reduces LOS in the order of 34% while previous psychiatric treatment is associated with an 18% decrease in LOS, possibly because services are familiar with the care of these patients.

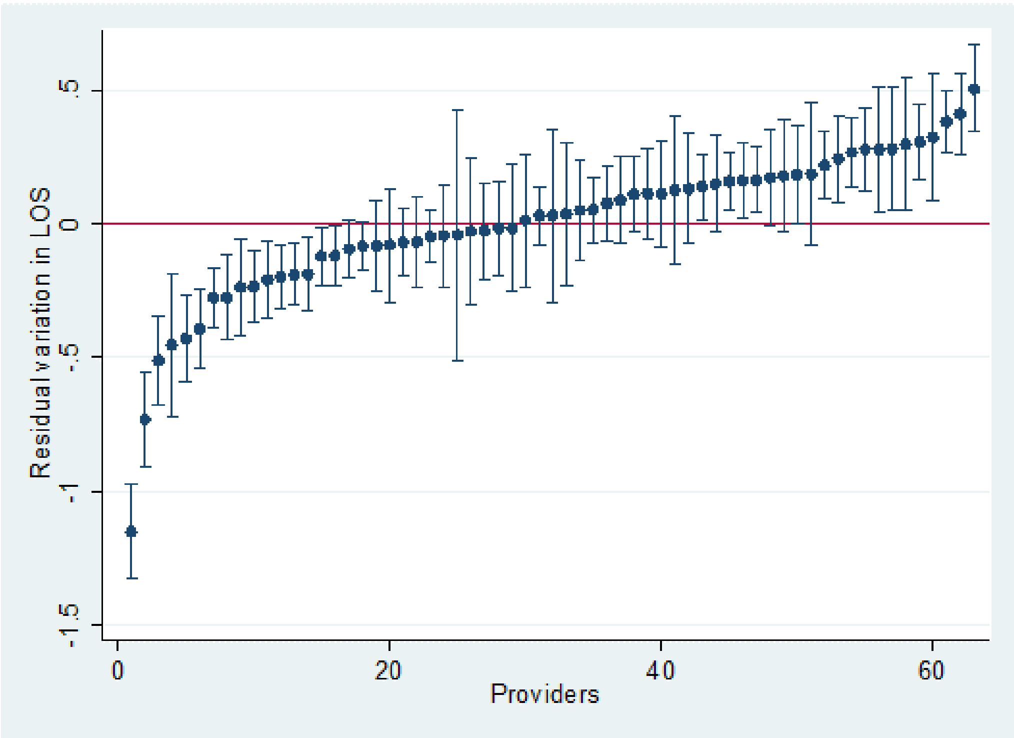
All of the patient-level variables have a statistically significant association with LOS with the exception of male gender. Black and Asian ethnicities are associated with a longer LOS compared to White ethnicity with Black ethnicity associated with a 31% increase and Asian ethnicity a 17% increase in LOS. The only patient-level variable with a statistically significant negative association with LOS is Other ethnicity which is associated with a 27% reduction in LOS compared to White ethnicity.

In terms of provider-level variables, the emergency readmission rate is associated with a large reduction in LOS of around 95%. The variable measuring the proportion of admissions under the MHA exerts a strong upward pressure on LOS of almost four times in the cross-classified model. The average number of comorbidities recorded by a provider is associated with a reduction in LOS of 10%. The variables measuring total bed occupancy, CPA 7 day follow-up, patient experience total score, access to a CRHT team, and the percentage of nurses from total Full Time Equivalent (FTE) staff have small effects on LOS of around 1% or less.

2.5.3. Provider-level residual variation

After controlling for the admission-, patient- and provider-level variables included in the model, there remains some residual variation in LOS as captured by the provider random effects. Figure 2.5 presents the EB estimates of the provider-level residual variation from the cross-classified model. While the majority of providers do not differ significantly from zero, there are a number of providers with a statistically significant higher or lower LOS compared to the average. Hospitals above (below) the line at zero have higher (lower) residual LOS compared to the average, i.e. the most (least) unexplained variation in LOS after controlling for observable characteristics. This implicitly assumes that the model has controlled for all known factors driving LOS and the remaining variation is due to a range of unobserved factors, one of which may be inefficiency.

Figure 2.5 Empirical Bayes (EB) estimates of residual variation in length of stay (LOS) in the cross-classified model



The EB estimates also suggest that relative to the average performing hospital with respect to unexplained variation in LOS (i.e. the residual), the worst performing hospital has a higher LOS of almost 1 day, while the best performing hospital has a lower LOS of almost 1 day in the cross-classified model (Figure 2.5) due to factors not considered in the model such as variations in efficiency, suggesting that there is scope for some providers to improve their relative performance.

An examination of the provider with the lowest residual LOS compared to the average, reveals that before conditioning on the admission-, patient- and provider-level variables in the model, the mean LOS of this provider is the second-lowest of the group. A perusal of the descriptive statistics of this provider does indicate that it treats a somewhat different case-mix to other providers. In particular, this provider has relatively low proportions of admissions with a diagnosis of psychosis as well as low levels of formal admissions at both admission- and provider levels. As these variables are identified as significant drivers of LOS, this may help to explain why the provider has the lowest residual LOS and emerges as an outlier in Figure 2.5.

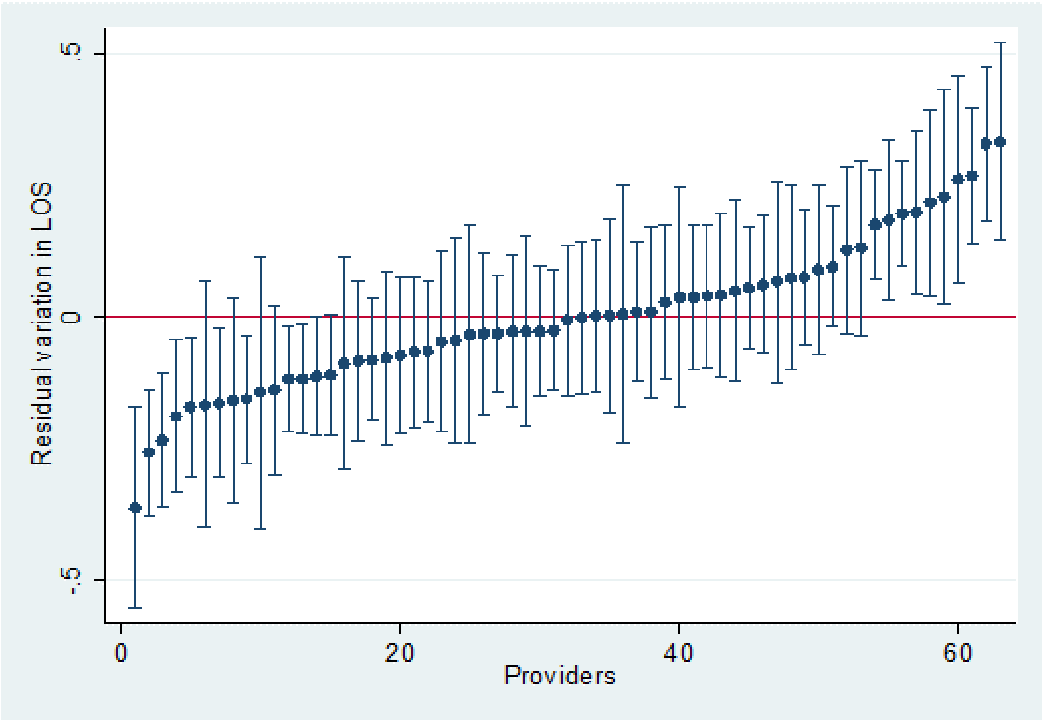
2.5.4. Sensitivity analysis

The results of the three-level model largely agree with the cross-classified model for the admission- and patient-level variables, although there is a minor tendency for the magnitudes of effects to be smaller in the three-level model. The most pronounced differences between the two models lies at the provider-level with more variables reaching statistical significance in the cross-classified model as this specification results in more accurate standard errors, especially for variables measured at higher levels (Leckie 2013).

Figure 2.6 shows the EB estimates from the three-level model in which the worst performing hospital has a higher LOS of almost half of a day and the best performing hospital has a lower LOS of around one-third of a day.

Figure 2.6 demonstrates the effect of not accounting for the cross-classified nature of the data. When the data are modelled as three-level, the cross-classified nature of the data is not recognised. Therefore, if a patient attends two different providers, that patient is counted twice. This means that the total number of admissions within a patient is reduced and the within-patient variance or variability is increased. This subsequently causes a greater degree of shrinkage as is evident in Figure 2.6 compared to Figure 2.5. Almost all (62/63) providers are affected by cross-classification so almost all are affected by the higher shrinkage from specifying a three-level model. Therefore, all are brought closer to the overall mean and relative differences are reduced. When the cross-classified nature of the data is correctly modelled, the shrinkage is less and the provider estimates are closer to those predicted by the data. Therefore, the outlier provider is more distinct in Figure 2.5 as its estimate is closer to the posterior distribution (i.e. that predicted by the data) than to the prior distribution.

Figure 2.6 Empirical Bayes (EB) estimates of residual variation in length of stay (LOS) in the three-level model



When admissions with a zero LOS were excluded from the analysis, the results of the cross-classified model remained robust, with the exception of the variable measuring if a provider is a Mental Health trust, which became statistically significant at the 0.1% level and is associated with an increase in LOS of 23%.

2.6. *Discussion*

2.6.1. Contribution to the current evidence base

This chapter has sought to investigate the main drivers of variations in LOS for mental health providers in England. The largest drivers of increased LOS at admission level are in-hospital death, a primary diagnosis of psychosis, formal detention, discharge to social care and the oldest age group (65 years and over). The first three of these factors are likely to reflect greater disease severity (or need) among these admissions. In line with previous literature (McCrone and Lorusso 1999) we find evidence of a non-linear relationship between LOS and age with younger and older age groups having positive coefficients. At a patient-level, Black ethnicity is associated with the largest increase in LOS and this finding is supported by previous literature in this field (Jacobs et al. 2015; Padgett et al. 1994). At a provider-level, the proportion of formal admissions under the MHA has a large positive association with LOS while the provider-level emergency readmission rate is associated with a large reduction in LOS.

We contribute to the current evidence base in a number of ways. The use of three-level and cross-classified models has allowed us to exploit the multilevel nature of a patient-level data set with national coverage – HES. Our results reveal that when a small proportion of the sample exhibits a cross-classified structure, three-level and cross-classified models provide somewhat similar results with differences most pronounced at the highest level of the data. Therefore, it is important to correctly model the cross-classified data structure in order to avoid misleading inferences. HES data provides rich information on a wide range of variables related to admission-, patient- and provider-level attributes which enables us to move beyond current literature in this field which considers a more limited range of variables. Moreover, we include provider-level emergency readmission rates calculated using HES data – a valuable addition as the HSCIC does not routinely calculate emergency readmission rates for mental healthcare providers. This allows us to investigate the relationship between variations in LOS and provider quality of care as measured by the emergency readmission rate – another novel contribution to the current evidence base.

2.6.2. Policy implications

We find that the provider emergency readmission rate has a strong negative association with LOS implying that providers with high emergency readmission rates are associated with a significantly shorter LOS. A plausible explanation is that providers may be compromising quality of care resulting in readmission and this is reflected in resource use in terms of LOS. Therefore, our findings lend some credence to the argument that, in the absence of clear guidelines on optimal LOS, decisions regarding duration of hospitalisation could be driven by economic rather than clinical considerations (Capdevielle and Ritchie 2008). Internationally, psychiatric LOS has experienced a downward trend corresponding to a decrease in psychiatric beds (OECD 2015) with shorter LOS associated with community-based mental health systems (Sytema, Burgess and Tansella 2002). Yet, there lacks a clear consensus on what constitutes an optimal LOS or indeed on best practice in this area (Capdevielle and Ritchie 2008). A high emergency readmission rate may indicate an inadequate provision of mental health support in the community. It may also represent poor quality inpatient care during the index admission, in particular in relation to inadequate discharge preparedness (Durbin et al. 2007). Many previous studies have investigated the relationship between LOS and readmissions at the individual patient-level for mental health. Shorter initial hospital stays have been shown to be related to higher readmissions (Appleby et al. 1993; Boden et al. 2011; Canadian Institute for Health Information 2008; Figueroa, Harman and Engberg 2004; Lin et al. 2006; Tulloch, David and Thornicroft 2015). Nevertheless, a recent study (Wolff et al. 2015b) has found readmission to be associated with an increase in LOS while an association between a long LOS and an increased risk of multiple readmissions has been reported by Korkeila et al. (1998). Efforts to reduce costs may drive shorter LOS (Capdevielle and Ritchie 2008; Lin et al. 2006) but risk compromising the quality of care leading to readmission which can in fact increase overall costs (Lin et al. 2006). Readmissions that take place within a relatively short period after discharge may be negatively associated with LOS due to the need for a longer inpatient stay to stabilise symptoms and provide adequate treatment. On the other hand, readmissions taking place within a longer period following discharge may be more likely to reflect the influence of factors beyond inpatient hospitalization, such as effective transitional care, the availability of community and family supports, access to primary care, housing and continued access and adherence to prescribed medications. This implies an important role for adequate discharge planning in protecting against early readmission (Durbin et al. 2007).

Our finding that shorter LOS comes at the expense of higher emergency readmission rates raises concerns of a 'revolving door' phenomenon of recurring hospitalisations with little effect (Williams et al. 2014) that can undermine a policy of strong community care and has long-term cost and quality implications. Internationally, readmission rates have garnered policy focus as a result of an increased awareness of the need to achieve value for purchasers in terms of quality and cost (Burgess Jr. and Hockenberry 2014). This has led to the introduction of high-powered incentives in the form of financial penalties imposed on hospitals for levels of readmission that are deemed inappropriate (Burgess Jr. and Hockenberry 2014). In England, providers in the acute sector are not reimbursed for readmissions within 30 days of discharge under the NTPS, and based on these results, such a policy may also be pertinent in the mental healthcare sector if it were to discourage reductions in LOS to such an extent as to have a detrimental impact on quality.

2.6.3. Limitations and future research

There are several limitations to this research. In order to gain a more comprehensive picture of the performance of mental healthcare providers it is necessary to model the entire care pathway across different settings. The majority of mental healthcare takes place in community-based settings and inpatient care is usually reserved for crisis stabilisation. Thus, by focusing on a relatively narrow segment of the care process we may misrepresent the true performance of mental healthcare providers. Moreover, consideration of the entire care pathway is likely to provide important insights into the interplay of other factors such as the range of outpatient and community-based services received, accommodation status, and crisis planning among others, which could not be considered in this model, but which may influence inpatient LOS. Future analysis using HES linked to the MHMDS would allow us to investigate provider performance across the entire care pathway. Our results find an association between reduced LOS and higher emergency readmission rates but we cannot infer a causal relationship. Moreover, while we have highlighted differences in residual variation across providers that we interpret as differences in provider performance, we cannot provide definitive reasons why some providers perform better than others once we account for observable admission-, patient- and provider-level variables. Nevertheless, the identification of providers with above- and below-average performance is in itself a useful exercise as this type of benchmarking highlights potential problems or potential efficiency savings allowing hospitals and regulators to undertake in-depth investigations to address such issues . This will become more pertinent if a national tariff is introduced as part of the NTPS as LOS can be viewed as a proxy for cost and providers will likely face pressures to reduce relatively more expensive inpatient care in order to control costs.

Chapter 3. Investigating variations in costs and performance of English mental healthcare providers

3.1. *Introduction*

Under the NTPS mental health providers will be paid a prospective fixed price for care provided in a given care cluster. This will incentivise providers to control costs and may encourage care provision in cost-effective community-based settings rather than more costly inpatient settings. The aim of this research is to investigate the performance of mental health providers in England in relation to cost efficiency. We explain variations in costs due to observable patient risk factors. We assume that the unexplained variation in costs is amenable to efficiency enhancing behaviour on the part of providers. We compare residual variation in costs across providers to assess performance in achieving cost efficiency and to provide insights into which providers may potentially gain or lose under the new payment system.

We add to the existing literature on mental health costs in several ways. Firstly, we go beyond the remit of using risk adjustment to explain variations in mental health costs by explicitly comparing the performance of mental healthcare providers in terms of residual cost variation. This complements recent literature in the physical acute sector by extending similar methodologies to mental healthcare. Additionally, we improve upon existing studies in mental healthcare by using a large, nationally representative patient-level data set and exploit the richness of this data set by applying multilevel models which allow us to report the variation in costs explained by various levels of analysis. We utilise a multilevel log-linear model and a multilevel GLM to estimate costs and adopt EB methods to quantify provider random effects in order to interpret provider performance. We use a comprehensive set of explanatory variables spanning demographic, treatment, and social variables and provide a first insight into how well the new currency developed to implement the NTPS in mental health performs in predicting resource use.

3.2. *Literature review*

3.2.1. Review of evidence on risk adjustment of mental health costs

We reviewed previous literature investigating risk adjustment of costs in mental healthcare by searching a number of databases including EconLit, Embase, Medline, OvidMedline, and PsychInfo using the following search terms: “mental health”, “psychiatry”, “costs”, “expenditures”, “risk adjustment”, and “performance”. The majority of the studies we retrieved are from the US and England but we also review notable studies from Australia and New Zealand which describe the development of psychiatric casemix classification systems. While some of the studies we review include children and adolescents in the data sample and analysis, we focus on the results for adults only as this corresponds to the coverage of our data.

Many of the US-based studies investigated the adequacy of diagnosis based classification systems developed for use in acute physical healthcare in predicting mental health costs (Ettner et al. 2001; Ettner et al. 1998; Ettner and Notman 1997; McGuire et al. 1987; Mitchell et al. 1987). Other studies had a main objective of developing casemix classification systems for inpatient psychiatric care only (Cromwell et al. 2005; Drozd et al. 2006) or inpatient and outpatient psychiatric care (Sloan et al. 2006). This was also the prime purpose of the research conducted in the Australian and New Zealand studies (Buckingham et al. 1998; Gaines et al. 2003). Moreover, in Australia, the intention was to use the system as a basis for reimbursing providers. In England, research was undertaken to inform resource allocation, specifically for the mental health component of a funding formula for general practitioners (Sutton et al. 2012). Other studies from England conducted research to identify the predictors of costs associated with the use of mental health services by people with psychosis (McCrone et al. 1998; McCrone, Johnson and Thornicroft 2001). Williams et al. (2014) investigated the factors associated with the costs of an inpatient stay as part of an evaluation of an intervention to reduce hospital LOS.

The studies from the US use a range of administrative data sources including from the federal Medicare (Cromwell et al. 2005; Drozd et al. 2006; Mitchell et al. 1987) and Medicaid (Ettner et al. 2001; Ettner and Notman 1997; Robst 2009) health insurance programmes, the Veterans Health Administration (VHA) (Montez-Rath et al. 2006; Sloan et al. 2006), and private insurance claims data (Ettner et al. 1998; McGuire et al. 1987). These data sources vary considerably in terms of coverage with Medicare primarily covering over-65s and people on disability benefits; Medicaid, low-income groups; the VHA is primarily comprised of males and certain mental disorders such as schizophrenia and substance abuse are more prevalent; while private health insurance is provided by employers to employees and their families whom may have less severe forms of mental illness.

Sample size varied from 147 patients (McCrone et al. 1998) to 914,225 patients (Sloan et al. 2006). Costs covered care delivered in hospital inpatient settings only (Cromwell et al. 2005; Drozd et al. 2006; McGuire et al. 1987; Mitchell et al. 1987; Williams et al. 2014) or hospital inpatient and outpatient settings (Ettner et al. 2001; Ettner et al. 1998; Ettner and Notman 1997; Kapur, Young and Murata 2000; Sloan et al. 2006) as well as wider community mental healthcare settings (Buckingham et al. 1998; Gaines et al. 2003; McCrone et al. 1998; McCrone, Johnson and Thornicroft 2001; Montez-Rath et al. 2006; Robst 2009; Sutton et al. 2012). The measurement unit for cost was average (McGuire et al. 1987; Mitchell et al. 1987); total (Ettner et al. 2001; Ettner et al. 1998; Ettner and Notman 1997; Sloan et al. 2006; Williams et al. 2014); per diem (Buckingham et al. 1998; Cromwell et al. 2005; Drozd et al. 2006; Gaines et al. 2003); or for a defined six (McCrone et al. 1998; McCrone et al. 2001) or twelve (Kapur, Young and Murata 2000; Robst 2009; Sutton et al. 2012) month period.

Risk-adjusters included diagnosis or diagnosis-related classification groups (Buckingham et al. 1998; Cromwell et al. 2005; Drozd et al. 2006; Ettner et al. 2001; Ettner et al. 1998; Ettner and Notman 1997; Kapur, Young and Murata 2000; McCrone, Johnson and Thornicroft 2001; McGuire et al. 1987; Mitchell et al. 1987; Robst 2009; Sloan et al. 2006; Sutton et al. 2012; Williams et al. 2014) as well as demographic variables such as age (Buckingham et al. 1998; Cromwell et al. 2005; Drozd et al. 2006; Ettner et al. 2001; Ettner et al. 1998; Ettner and Notman 1997; Gaines et al. 2003; Kapur, Young and Murata 2000; McCrone et al. 1998; McCrone, Johnson and Thornicroft 2001; Robst 2009; Sutton et al. 2012; Williams et al. 2014), sex (Ettner et al. 2001; Ettner et al. 1998; Ettner and Notman 1997; Gaines et al. 2003; Kapur, Young and Murata 2000; McCrone, Johnson and Thornicroft 2001; Robst

2009; Sutton et al. 2012), ethnicity (Gaines et al. 2003; Kapur, Young and Murata 2000; McCrone, Johnson and Thornicroft 2001; Robst 2009; Sutton et al. 2012) and marital status (Kapur, Young and Murata 2000; McCrone, Johnson and Thornicroft 2001; Sutton et al. 2012). Additional risk adjustment variables have reflected functioning (Buckingham et al. 1998; Cromwell et al. 2005; Drozd et al. 2006; Kapur, Young and Murata 2000; McCrone et al. 1998; McCrone, Johnson and Thornicroft 2001), need (Buckingham et al. 1998; Gaines et al. 2003; McCrone, Johnson and Thornicroft 2001), treatment (Buckingham et al. 1998; Cromwell et al. 2005; Drozd et al. 2006; Gaines et al. 2003; McCrone et al. 1998; Williams et al. 2014) and social (Kapur, Young and Murata 2000; McCrone et al. 1998; McCrone, Johnson and Thornicroft 2001; Sloan et al. 2006; Sutton et al. 2012; Williams et al. 2014) factors. Provider-level variables in the form of provider type and ownership (Cromwell et al. 2005; Drozd et al. 2006; McGuire et al. 1987), site (as a proxy for provider) (Buckingham et al. 1998), teaching status, size, occupancy rate, urban location, and area hospital wage rates (Drozd et al. 2006) were also considered. Indicators for day of stay were included in two studies (Cromwell et al. 2005; Drozd et al. 2006) to investigate the potential use of declining block pricing in a payment system.

A number of study samples included non-utilisers of mental health services resulting in cost data containing a considerable proportion of zero observations. This necessitated the use of two-part models with either logit (Ettner et al. 2001; Ettner et al. 1998; Ettner and Notman 1997) or probit (Kapur, Young and Murata 2000) models used to model the probability of incurring mental health expenditures in the first stage. For the second stage, the level of expenditures was modelled using either untransformed (Ettner et al. 2001) data or transforming data to its log (Montez-Rath et al. 2006; Sutton et al. 2012) or square-root (Ettner et al. 1998; Ettner and Notman 1997; Montez-Rath et al. 2006) levels or using a non-linear exponential least squares regression (Kapur, Young and Murata 2000). If data was transformed, a smearing estimator equivalent to the mean of the squared residuals was used to retransform the predicted estimates to the mean of the original expenditure distribution (Ettner et al. 1998; Ettner and Notman 1997). Studies with samples covering only users of mental health services used weighted least squares regression models (Sloan et al. 2006) or a linear regression model with expenditures modelled on the raw scale (Buckingham et al. 1998; Gaines et al. 2003; McCrone, Johnson and Thornicroft 2001; Robst 2009;

Williams et al. 2014) or with log or square root transformations (Cromwell et al. 2005; Drozd et al. 2006; McCrone et al. 1998; Montez-Rath et al. 2006; Robst 2009). In order to avoid transforming the dependent variable, GLMs were used by a number of authors. Montez-Rath et al. (2006) estimated three GLMs on untransformed cost: normal with identity link (equivalent to Ordinary Least Squares (OLS)); gamma with log link; and gamma with square-root link. Sutton et al. (2012) used a GLM with a log link in the second part of a two-part model. Williams et al. (2014) estimated a multilevel model of admissions nested within patients with random intercepts included for patients. Model performance in terms of predictive ability was compared using R-squared (Buckingham et al. 1998; Cromwell et al. 2005; Ettner et al. 1998; Ettner and Notman 1997; Gaines et al. 2003; Kapur, Young and Murata 2000; Mitchell et al. 1987; Sloan et al. 2006); Mean Absolute Prediction Error (MAPE) (Ettner et al. 1998; Ettner and Notman 1997; Montez-Rath et al. 2006; Sloan et al. 2006); Root Mean Square Error (RMSE) (Montez-Rath et al. 2006); Predictive Ratios (PRs) (Kapur, Young and Murata 2000; Montez-Rath et al. 2006; Robst 2009); Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) (Robst 2009); and Coefficients of Variation (CV) (Buckingham et al. 1998; Gaines et al. 2003).

With regard to study findings, Diagnosis Related Groups (DRGs) or other classification systems designed for use in physical healthcare did not prove successful in explaining a substantial part of variation in mental health costs (Ettner et al. 2001; Ettner et al. 1998; Ettner and Notman 1997; McGuire et al. 1987; Mitchell et al. 1987). The use of additional control variables such as demographic, co-morbidities (Ettner et al. 2001; Ettner et al. 1998; Ettner and Notman 1997; Kapur, Young and Murata 2000) as well as social (homelessness) (Kapur, Young and Murata 2000), previous costs (Kapur, Young and Murata 2000), functioning (Cromwell et al. 2005; Drozd et al. 2006; Kapur, Young and Murata 2000), severity (Cromwell et al. 2005; Drozd et al. 2006) and treatment (Cromwell et al. 2005; Drozd et al. 2006) improved the predictive ability of models. A systematic review (Wolff et al. 2015a) of drivers of mental health inpatient resource use found that the most relevant patient characteristics were age, major diagnostic group, experiencing psychotic or affective symptoms, risk, legal problems, and ability to perform activities of daily living. Non-patient characteristics associated with inpatient resource use included day of stay

(Cromwell et al. 2005; Drozd et al. 2006) and treatment site (Buckingham et al. 1998).

Table 3.1 provides an overview of the direction of association between cost and various patient characteristics. For several variables, the relationship was consistent across numerous studies.

Table 3.1 Literature on characteristics associated with mental health costs

Variable	**Direction of association (Reference)**
Age	**Positive:** (Buckingham et al. 1998; Cromwell et al. 2005; Drozd et al. 2006; McCrone, Johnson and Thornicroft 2001) **Negative:** (Kapur, Young and Murata 2000; Robst 2009)
Functioning	**Negative:** (Kapur, Young and Murata 2000; McCrone et al. 1998)
Previous costs	**Positive for inpatient costs and negative for outpatient costs:** (Kapur, Young and Murata 2000)
Diagnosis	**Positive:** (Buckingham et al. 1998; Drozd et al. 2006; Kapur, Young and Murata 2000; Robst 2009)
Deficits in Activities of Daily Living (ADL)	**Positive:** (Buckingham et al. 1998; Cromwell et al. 2005; Drozd et al. 2006)
Detox	**Positive:** (Drozd et al. 2006)
Electroconvulsive Therapy (ECT)	**Positive:** (Cromwell et al. 2005; Drozd et al. 2006)
Legal status (compulsory admission)	**Positive:** (Buckingham et al. 1998; Gaines et al. 2003; Williams et al. 2014)
Severity (measured using HoNOS)	**Positive:** (Buckingham et al. 1998; Gaines et al. 2003)
Aggressive/disruptive/suicidal behaviour	**Positive:** (Buckingham et al. 1998; Cromwell et al. 2005)
Single, divorced, widowed, or living alone	**Positive:** (McCrone et al. 1998; Sutton et al. 2012)

Unemployment	**Positive:** (Sutton et al. 2012)
More years of education	**Negative:** (McCrone, Johnson and Thornicroft 2001)
No fixed residence or unsettled accommodation	**Positive:** (Sutton et al. 2012; Williams et al. 2014)
Prior service use	**Positive:** (Sutton et al. 2012; Williams et al. 2014)
Gender	**Positive for males:** (McCrone, Johnson and Thornicroft 2001; Robst 2009)
Ethnicity	**Positive for ethnic minorities compared to whites:** (Gaines et al. 2003; Robst 2009)

A number of studies also examined the influence of provider factors. An early study (McGuire et al. 1987) found that after adjusting for differences in average cost between providers using DRGs, some variation remained with private psychiatric hospitals and hospital-based substance abuse facilities having the highest adjusted cost per case and general hospitals without specialist psychiatric units having much lower average adjusted costs. Conversely, Cromwell et al. (2005) found no statistical differences in costs between acute general hospitals with distinct psychiatric units and either public or private psychiatric hospitals. This study also found that higher costs were associated with higher area wage rates as well as teaching status while lower costs were associated with a higher average daily psychiatric census, and a higher facility share of Medicaid plus Medicare SSI eligible (low-income) patients. The initial part (days 2-5) of the inpatient stay was associated with higher costs while an inpatient stay of less than 24 hours was associated with lower costs (Cromwell et al. 2005). Drozd et al. (2006) report that a classification model comprising only facility characteristics (a weekend indicator, ownership, teaching status, size, occupancy rate, urban location, area hospital wage rates and Medicare disproportional share ratio) and day of stay explained only 23% of daily cost variation compared to an alternative model with 16 groups that used five major DSM-IV categories and stratified by age, illness severity, deficits in daily living activities, dangerousness, and use of electroconvulsive therapy (ECT) that explained 40% of daily cost variation. Projects to develop casemix classification systems for mental health services in Australia and New Zealand revealed that patient-level costs were driven by provider factors such as individual clinical practice, resource availability and the types of services available (Buckingham et al. 1998; Gaines et al. 2003).

In terms of methodology, Montez-Rath et al. (2006) showed that transforming the data to its square-root was optimal with a square-root normal model having the lowest RMSE, MAPE and bootstrap confidence interval values, and a GLM with gamma distribution and square-root link having the best PRs. However, the latter model had convergence problems with small samples. The authors maintain that the advantageous performance of the square root models may be due to the ability of this transformation to address high comorbidity levels in the study sample by introducing a form of interaction while the gamma distribution helps to address the long tail typical of cost data. Similarly, the results of Robst (2009) suggested that a model using the square root of expenditure fit the data best. Nevertheless, the PRs implied that modeling untransformed expenditures may be sufficient on larger samples while the log transformation was advantageous for groups with very high or low expenditures.

3.2.2. Review of selected literature on variations in costs and provider performance in acute care

In a separate review we selected recent studies that have analysed patient costs and provider performance in acute physical healthcare. The studies reviewed tend to focus on a particular speciality or disease (Cooper et al. 2007; Hvenegaard et al. 2009; Kristensen et al. 2010; Laudicella, Olsen and Street 2010; Olsen and Street 2008). Nevertheless, there are studies that did not limit the sample to patients treated in a particular specialty. Daidone and Street (2013) analysed the costs of the entire population (12,154,599) of patients admitted to 163 hospitals in England in order to investigate the costs associated with specialist care. Gaughan et al. (2012) analysed variations in costs across English NHS providers for ten treatments[1]in order to investigate if above- or below-average cost performance is consistent across

[1] Medical: acute myocardial infarction; childbirth; stroke. Surgical: appendectomy; breast cancer (mastectomy); coronary artery bypass graft; cholecystectomy; inguinal hernia; hip replacement and knee replacement

departments/specialities for a particular hospital. Carey (2000) used data on a sample of 526,117 patients treated in 24 medical centres operated by the VHA in the US.

Studies used two-level multilevel model with fixed (Gaughan et al. 2012; Hvenegaard et al. 2009; Kristensen et al. 2010; Laudicella, Olsen and Street 2010; Olsen and Street 2008) or random (Carey 2000; Cooper et al. 2007; Daidone and Street 2013; Olsen and Street 2008) effects. Costs were commonly modelled on a log (Carey 2000; Cooper et al. 2007; Gaughan et al. 2012; Hvenegaard et al. 2009; Olsen and Street 2008) or raw (Daidone and Street 2013; Hvenegaard et al. 2009; Laudicella, Olsen and Street 2010) scale in a linear regression. Daidone et al. (2013) modelled costs using a GLM with a gamma distribution and square root link. Cooper et al. (2007) utilised two-part models with cost modelled as 1) a log-normal and 2) a gamma distribution with log link in the second part.

Dependent variables were measured as total patient cost during a fiscal year (Carey 2000; Cooper et al. 2007) or during a hospital stay (Gaughan et al. 2012; Hvenegaard et al. 2009; Kristensen et al. 2010; Laudicella, Olsen and Street 2010; Olsen and Street 2008). Due to the large number of patients and HRGs in the study of Daidone and Street (2013), the dependent cost variable was measured as the individual patient's cost compared to the average cost of all patients assigned to the same HRG. All studies included age and gender as patient-level explanatory variables while additional explanatory variables included casemix classification variables and additional diagnostic variables (Carey 2000; Gaughan et al. 2012; Hvenegaard et al. 2009; Kristensen et al. 2010; Laudicella, Olsen and Street 2010; Olsen and Street 2008); health and treatment variables (Cooper et al. 2007; Daidone and Street 2013; Gaughan et al. 2012; Hvenegaard et al. 2009; Laudicella, Olsen and Street 2010; Olsen and Street 2008); socioeconomic characteristics (Daidone and Street 2013; Hvenegaard et al. 2009; Laudicella, Olsen and Street 2010); and quality variables (Gaughan et al. 2012). A number of studies included provider variables including volume of patients treated (Gaughan et al. 2012; Kristensen et al. 2010; Laudicella, Olsen and Street 2010); beds (Carey 2000); teaching status (Carey 2000; Gaughan et al. 2012; Laudicella, Olsen and Street 2010); staff (Laudicella, Olsen and Street 2010); input price index (Kristensen et al. 2010; Laudicella, Olsen and Street 2010);

specialisation (Gaughan et al. 2012; Kristensen et al. 2010); and quality (Gaughan et al. 2012) variables.

Two studies compared various model specifications. Cooper et al. (2007) found that the two-part models provided the best fit to the data. Daidone and Street (2013) compared OLS to 1) OLS with a log or square root transformation applied to costs; 2) GLM; and 3) Finite Mixture Models (FMM). The latter failed to converge while log-linear models were found to be imprecise. A relatively consistent finding was that diagnosis-based classification systems emerged as important predictors of cost variation (Carey 2000; Hvenegaard et al. 2009; Kristensen et al. 2010) and that much of the heterogeneity in patient costs between providers was due to casemix (Olsen and Street 2008). Some, but not all, types of specialist care was found to be associated with higher costs (Daidone and Street 2013). Provider-level variables were found to have no (Carey 2000; Gaughan et al. 2012) or negligible (Kristensen et al. 2010; Laudicella, Olsen and Street 2010) effects in terms of explaining cost variations. The ranking of providers in terms of fixed or random effects revealed significant differences in cost-containment performance (Carey 2000; Daidone and Street 2013; Gaughan et al. 2012; Hvenegaard et al. 2009; Laudicella, Olsen and Street 2010; Olsen and Street 2008). Hvenegaard et al. (2009) found that provider rankings were sensitive to choice of functional form with department rankings differing according to whether costs were modelled on a raw or log-transformed scale as the use of log-transformed costs gave less weight to cost outliers which centred on certain specialised providers. In contrast, Daidone and Street (2013) found that provider rankings were not sensitive to the choice of linear OLS or GLM specifications. Gaughan et al. (2012) reported that, after controlling for patient- and provider-level variables, a small number of hospitals had higher average costs across multiple treatments compared to their peers implying that these providers may face financial instability under a national tariff or fixed price system.

This literature review informs our study in several ways. Firstly, the studies on risk adjustment of mental health costs suggest that classification systems based on diagnosis tend to perform poorly in adequately explaining variation in mental health costs and it is necessary to use additional demographic, treatment, and socioeconomic variables. We move beyond previous studies by risk-adjusting mental healthcare costs by using a classification system developed specifically for mental healthcare and supplement this with a range of demographic, treatment and socioeconomic variables. We also examine the influence of provider-level variables in a sensitivity analysis. Secondly, we utilise random effects to make assertions about relative provider performance thus extending methods applied in physical healthcare to mental healthcare.

3.3. *Data*

We use two main data sets for our analysis; the Department of Health Reference Cost data is used to construct the dependent cost variable, while the MHMDS contains information on the clustering process as well as our independent risk adjustment variables.

3.3.1. Reference Cost data

We use the Reference Cost data published by the Department of Health to construct our dependent cost variable. The Reference Cost data is submitted by NHS providers to the Department of Health and provides an indication of the costs of providing mental health services. Privately owned providers do not submit Reference Cost data despite providing care for NHS patients and being reimbursed under the NTPS. A considerable proportion of mental healthcare is provided by the private sector and the costs of this care may differ significantly from the costs of NHS provision (Jacobs 2014).

Reference cost data for the mental healthcare clusters were first collected in a pilot exercise that complemented the main Reference Cost collection in 2011 (Department of Health 2012). While the clusters are independent of setting, cluster activity is costed according to admitted- and non-admitted cluster days as well as initial assessments. Therefore, each provider reports Reference Cost data for each cluster disaggregated into the per diem cost associated with admitted and non-admitted care respectively. Of the 55 providers in our database that report Reference Cost data, one provider does not report costs for admitted care for 2011/12 while two providers do not report costs for admitted care for 2012/13. Six providers do not report costs for non-admitted care for 2011/12. These providers are included in the analysis with costs reflecting only the admitted or non-admitted part of the care pathway as appropriate.

Figure 3.1 displays the mean unit cost per day of admitted care for 2011/12 and 2012/13 for the 52 providers in our database that report these costs for both 2011/12 and 2012/13.

Figure 3.1 Average unit cost per day of admitted care, by provider, n=52

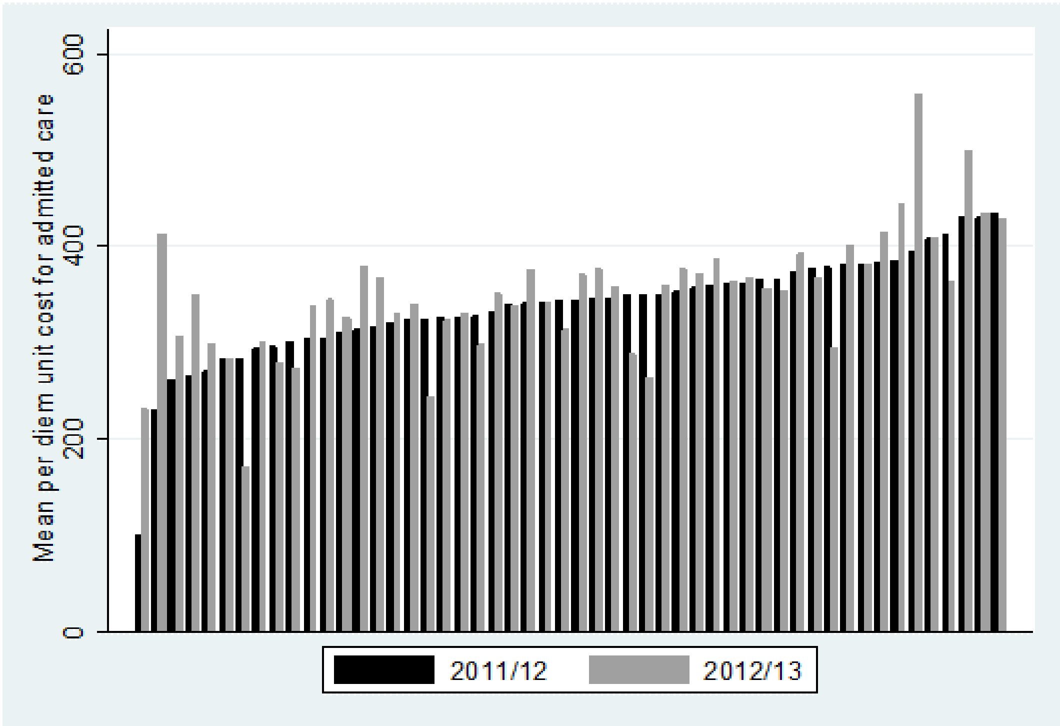
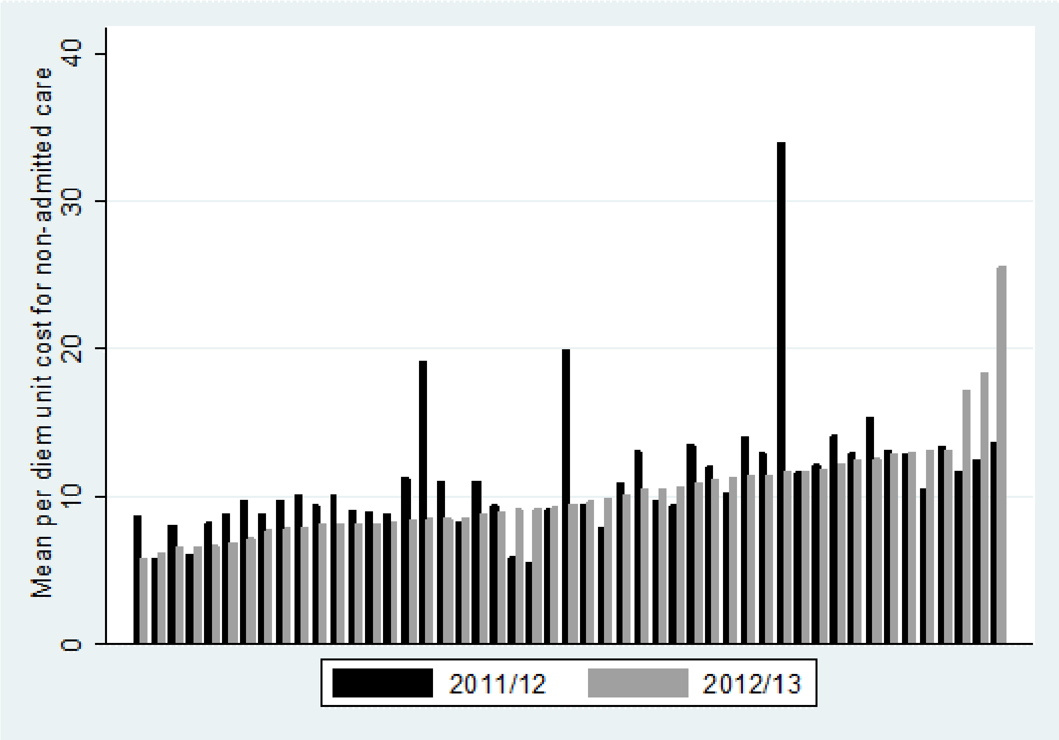


Figure 3.2 displays the mean unit cost per day of non-admitted care for 2011/12 and 2012/13 for the 49 providers in our database that report these costs for both 2011/12 and 2012/13.

Figure 3.2 Average unit cost per day of non-admitted care, by provider, n=49



Variation across providers is evident and this appears to be more pronounced for admitted care. As may be expected, costs increased in 2012/13 but there are some providers with higher costs in 2011/12 and these higher costs are particularly marked for non-admitted care which may suggest that data quality also improved in 2012/13.

Concerns regarding the quality of the Reference Cost data submitted during the first data collection have been raised (PriceWaterhouseCoopers 2012). These concerns relate to variations in unit costs within clusters and between providers; variations in the number of service users and LOS in each cluster; and missing data from Reference Cost returns. The low implementation rate of Patient Level Information Costing Systems (PLICS) in mental healthcare compared to acute care has also been noted with 52% of mental health providers implementing or planning to implement PLICS in 2011 compared to 87% of acute trusts (PriceWaterhouseCoopers 2012). A greater implementation of PLICS would increase the accuracy and reliability of Reference Cost data – a necessity for implementation of a national price or tariff per cluster.

Taking account of issues of Reference Cost data quality, we compared Reference Cost data for 2011/12 and 2012/13 by provider and omitted data for outliers defined as greater than 4 times the cost reported in the previous (for 2012/12 data) or following year (for 2011/12 data) (n=102,121). This resulted in dropping one provider with consistently high costs for all clusters across both years. We also dropped observations with inpatient days in the 99th percentile (>=48 days) for Cluster 1 (n=383) and observations with inpatient days in the 99th percentile (>=74 days) for Cluster 2 (n=450) as these clusters cover common mental health problems and due to pressure on mental health beds we would not expect patients in these clusters to have such long lengths of stay.

While the care cluster currencies cover most services for working age adults and older people, some services such as children and adolescent, drug and alcohol, and specialist mental health services are not included and will be reimbursed under separate non-cluster currencies. Table 3.2 outlines the coverage of mental health services in Reference Cost data.

Table 3.2 Coverage of mental health services in Reference Cost data

Service	**Included in cluster Reference Cost data**	**Included in non-cluster Reference Cost data**	**Excluded from Reference Cost data**
Approved social worker services*	Yes		
Assertive outreach teams	Yes		
Crisis accommodation services	Yes		
Crisis resolution and home treatment teams	Yes		
Early intervention in psychosis services from age 14	Yes		
Eating disorder services (adult, excluding tertiary eating disorders)	Yes		
Emergency clinics or walk in clinics	Yes		

Emergency duty teams (which are not emergency assessments e.g. for sectioning under the Mental Health Act)*	Yes		
Homeless mental health services	Yes		
Local psychiatric intensive care units	Yes		
Mental health counselling and therapy	Yes	Yes	
Psychology	Yes	Yes	
Psychotherapy	Yes	Yes	
A&E mental health liaison services (psychiatric liaison)		Yes	
Autism and Asperger syndrome		Yes	
Child and Adolescent Mental Health Services (CAMHS)		Yes	
Drug and alcohol services		Yes	
Eating disorder services (children and adolescents)		Yes	
Forensic and secure mental health services		Yes	
Learning disability services in high dependency or high secure units		Yes	
Mental health services provided under a GP contract		Yes	
Perinatal mental health services (mother and baby units)		Yes	
Primary diagnosis of drug or alcohol misuse		Yes	
Specialised addiction services		Yes	
Specialist psychological therapies (admitted patients and specialised outpatients)		Yes	
Specialised eating disorder services		Yes	
Improving access to psychological therapies (IAPT)		Yes**	
Acquired brain injury			Yes
Complex or treatment resistant disorders in tertiary settings			Yes
Gender dysmorphia			Yes

Learning disability services not provided in high dependency or high secure units			Yes
Specialist mental health services for deaf people			Yes
Neuropsychiatry			Yes

* these services are only included in clusters where NHS funded, otherwise they are excluded

** other specialist teams

Source: Department of Health (2012). Reference Costs 2011-12. Leeds, Department of Health.

3.3.2. Risk adjustment covariates

Demographic variables include age, gender, ethnicity and marital status. We categorise age based on deciles in order to capture any non-linearities in the relationship between age and cost and use age 18-34 years as the reference category. Ethnicity is also categorised to represent the main ethnic groups in the data: White, Black, Asian and Other with White set as the reference category. Gender is represented by a dummy variable with males equal to one. Information on severity and treatment are captured by variables reflecting if a patient has care co-ordinated under the CPA or has been admitted to hospital under the MHA. Around 40% of observations for the CPA and MHA variables were missing but we coded these as zero and make the assumption that these observations have not been subject to the MHA or under CPA.

We include dummy variables for the 21 care clusters to investigate the extent to which these explain variations in cost. We use the cluster with the lowest cost (Cluster 1) as the reference category. The MHMDS contains data for a small area level geographic marker, the Lower Layer Super Output Area (LSOA) of the individual. LSOAs have a minimum population of 1000 and a mean of 1500 (Health and Social Care Information Centre 2013). The LSOA codes can be matched to data from the Census or the Index of Multiple Deprivation (IMD) in order to enable variables reflecting various domains of deprivation to be used in the analysis. We include information on income deprivation using the IMD Income Deprivation (Noble 2008) variable. This

Domain measures the proportions of the population experiencing income deprivation in an area (Noble 2008). A higher score for the IMD Income Domain indicates a greater proportion of the population in the area in which the patient lives experiences income deprivation. The scores for the Income Domain are rates which we multiplied by 100 to allow interpretation as percentages. Observations include those with a CRP that starts in 2011/12 or in 2012/13 so a dummy variable is included to capture the year that the cluster started in order to control for factors such as inflation or differential coding of costs with 2011/12 used as the reference category.

3.3.3. Sensitivity Analysis

Provider-level variables are not included in the main analysis, in order to allow us to control only for observable patient factors that may lead to variations in costs across providers and allow provider factors to be captured in the provider-level residual and be indicative of performance. As a sensitivity analysis we include a number of provider-level variables reflecting provider governance and capacity constraints to investigate if these variables can explain variations in cost. These variables include provider size as measured by the number of available mental health beds, percentage occupancy of mental health beds, and whether the provider has FT status. We would expect that providers with a higher number of beds may have lower costs due to economies of scale effects. Providers with high occupancy rates (above the optimum of 85%) (Royal College of Psychiatrists 2011) may have higher costs if high rates result in patients being discharged early and subsequently readmitted. Providers with FT status have more autonomy and control over their finances so can be expected to be associated with higher financial performance (Verzulli, Jacobs and Goddard 2011) and hence lower costs. We also include a variable measuring the proportion of admissions under the MHA by provider. Recent research has revealed statistically significant differences in compulsory admissions between providers in England, after controlling for a large number of explanatory variables (Weich et al. 2014) and we expect compulsory admission to be positively associated with cost. We do not include any provider-level variables capturing quality as we want these to be reflected in the measure of provider performance.

3.4. *Methods*

3.4.1. Construction of dependent cost variable

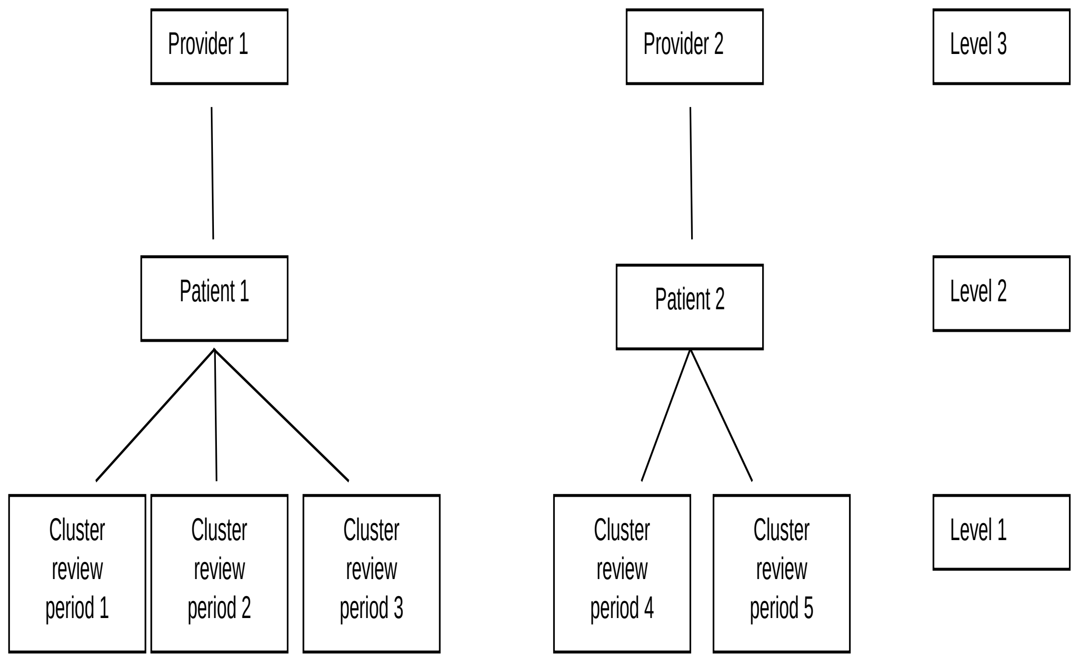
In order to construct our dependent variable, we first measured all activity during a CRP that corresponds to mental health services covered under the care cluster currencies (see Table 3.2). This activity relates to admitted care and non-admitted care. Admitted care refers to inpatient hospital stays while non-admitted care includes care delivered by mental health teams, NHS day care, consultant outpatient episodes, acute home-based care, NHS mental healthcare homes, and community contacts with healthcare professionals. For each observation (CRP), we calculated the total number of days in admitted and non-admitted care. These LOS variables for admitted and non-admitted care were then multiplied by the per diem unit costs for admitted and non-admitted care for the particular care cluster and provider in order to construct a variable reflecting the total cost associated with a CRP. We use the 2011/12 Reference Cost data for activity between 1 April 2011 and 31 March 2012 and the 2012/13 Reference Cost data for activity between 1 April 2012 and 31 March 2013. For activity between 1 April 2011 and 31 March 2013 we calculate a weighted average cost that reflects the number of days during a CRP in each year. It is important to highlight that the use of cost data reported at a provider level, albeit disaggregated by cluster and admitted and non-admitted care will conceal the true variation in cost that would be evident in data reported at the patient level.

We do not adjust costs using the Market Forces Factor (MFF). The Market Forces Factor (MFF) is an estimate of unavoidable costs faced by healthcare providers due to their geographical location. Each NHS provider is allocated an MFF value and this informs a payment index that is used to adjust provider payment in order to avoid financial (dis)advantage due to geographical location (Monitor and NHS England 2013c). Our primary interest is in variations in cost performance across providers and in comparing performance, our objective is to control only for factors that vary at CRP- or patient-levels, rather than at provider-level. Moreover, we are interested in examining the role of cluster costs in the design of the NTPS and the variation in costs that would explain whether it's possible to obtain an accurate price signal for payment. As the MFF would not be considered in the calculation of a national tariff, it is therefore excluded from this assessment.

3.4.2. Multilevel models

A patient can have more than one CRP and the maximum number of CRPs per patient is 43. This means that our data is characterised by a multilevel structure with three levels: CRPs clustered in patients clustered in providers (Figure 3.3).

Figure 3.3 Multilevel data structure



As our data is not a random sample of the general population but rather covers all users of specialist mental healthcare in England, we do not face the problem of dealing with a large number of zero cost observations and so avoid the use of a two-part model.

We adopt two estimation approaches: 1) a linear model with the log of total cost as the dependent variable, and 2) a multilevel GLM with untransformed total cost as the dependent variable. As our dependent variable is highly skewed, we transform it by taking logs in order to achieve a normally distributed variable. This is preferable for making inferences about provider performance as EB techniques make the assumption that the prior distribution of the residuals is normal. However, in order to interpret the model coefficients in terms of the arithmetic mean of the dependent variable in the original monetary units of cost, retransformation from the log scale is required. Direct transformation in the form of exponentiation of the model coefficients can result in biased estimates as $E\{\ln(\mathrm{Y})\}$ does not necessarily equal $\ln\{E(\mathrm{Y})\}$(Montez-Rath et al. 2006). The use of a multilevel GLM allows us to easily interpret model estimates in terms of the arithmetic mean in monetary terms as it does not necessitate the transformation and subsequent retransformation of the dependent variable.

We estimate the following three-level log-linear model for CRP i in patient j in provider k:

$$y_{ijk} = \alpha + \beta X_{ijk} + u_k + v_{jk} + \varepsilon_{ijk} \quad (1)$$

where y_{ijk} is the dependent cost variable, X_{ijk} represents a vector of risk adjustment covariates at the cluster-review- and patient-levels, u_k is the provider-level random intercept, v_{jk} is the patient-level random intercept and ε_{ijk} is the error term at the CRP-level. The coefficients for the log of total cost dependent variable can be interpreted in terms of a percentage change in the geometric mean of total cost which can be calculated as (exp(β) – 1)*100. For the majority of covariates measured as dummy variables, this is the percentage change in the geometric mean resulting from a change in the variable from zero to one. For the continuous IMD Income Deprivation variable, the coefficient can be interpreted as the percentage change in the geometric mean in total cost resulting from a one unit change in this variable.

We estimate a three-level GLM with a gamma distribution and a log link. More specifically we estimate the following multilevel GLM for CRP i in patient j in provider k:

$$g\,\{E\,[y_{ijk} \mid \mathrm{X}'_{ijk}, u_k\, v_{jk}]\} = \mathrm{X}'_{ijk}\beta + u_k + v_{jk} \equiv \eta_{\mathrm{ijk}},\, y_{ijk} \sim gamma \quad (2)$$

where y_{ijk} is the vector of responses from the gamma distributional family, X'_{ijk} is a vector of risk adjustment covariates for the fixed effects β. $\mathrm{X}'_{ijk}\beta + u_k + v_{jk}$ is the linear predictor, also denoted as η_{ijk}; g (.) is the link function and is assumed to be invertible so that

$$E\,(y_{ijk} \mid X'_{ijk}\, u_k\, v_{jk}) = g^{-1}\,(X'_{ijk}\beta + u_k + v_{jk}) = \exp\,(\eta_{ijk}) = \mu_{ijk} \qquad (3)$$

Model coefficients for the GLM model(s) can be interpreted as average marginal (or partial) effects. All but one of our independent variables are dummy variables so coefficients can be interpreted in terms of average effects measuring discrete change i.e. the change in the total cost of a CRP as the independent variable changes from zero to one, holding all other variables at their mean value. The coefficient on the independent IMD variable can be interpreted in terms of the change in the total cost of a CRP arising from a one unit change in the IMD score. Statistical significance is tested at the 5%, 1% and 0.1% levels.

3.4.3. Comparison of provider performance

In order to compare the residual variation across providers we predict the random effects u_k from the log-linear model using EB estimates with comparative standard errors as described in Section 1.4.2 of Chapter 1. We calculate the percentage difference in the EB estimates of provider-level residual variation for the best and worst performing providers compared the average performing provider as $(\exp(u_k - u_0) - 1)*100$ where u_0 refers to the average provider.

The models are estimated in Stata 13.0 (StataCorp 2013) using the *meglm, margins,* and *predict* commands and in MLwiN 2.29 (Rabash et al. 2009) using the *runmlwin* command (Leckie and Charlton 2012) in Stata 13.0 (StataCorp 2013).

3.5. *Results*

3.5.1. Dependent variables

Figure 3.4 shows our untransformed dependent variable – total cost for CRPs. The graph shows that there is considerable variation both within and between providers.

Figure 3.4 Dependent variable, total cost per Cluster Review Period (CRP) by provider, n=55

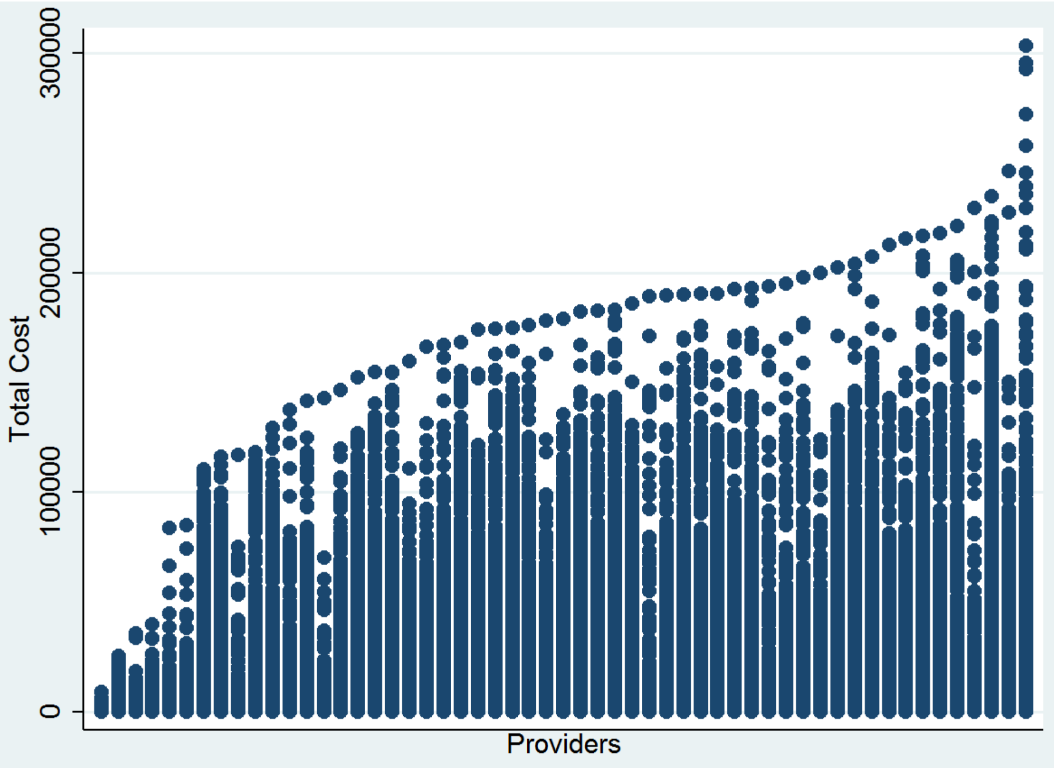
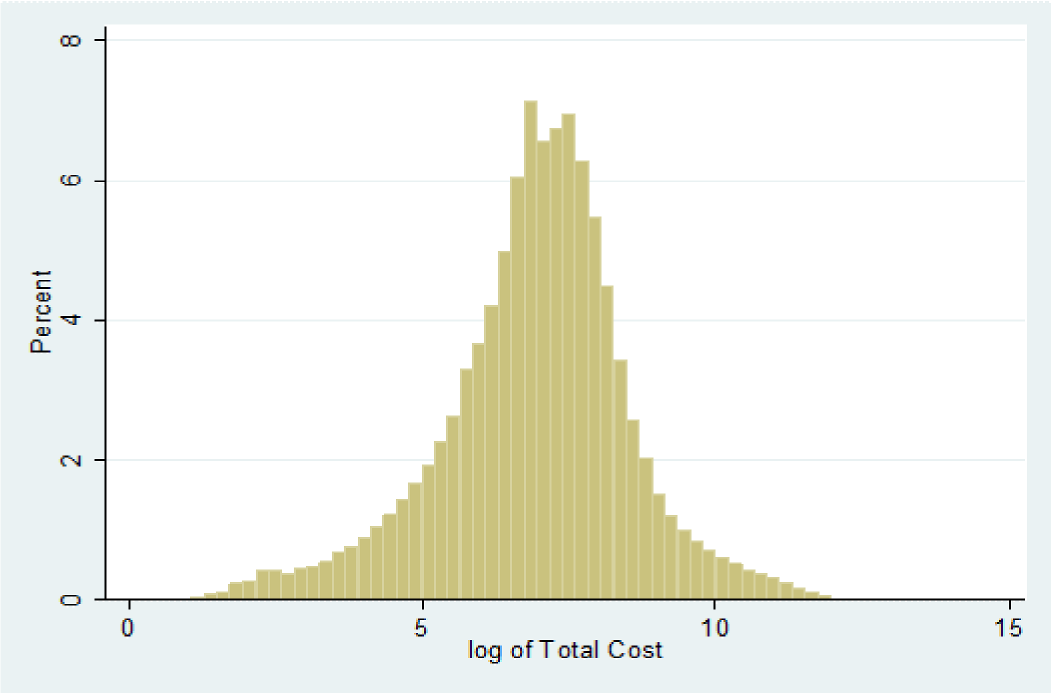


Figure 3.5 displays the log transformation of the total cost per CRP variable, which reflects a normal distribution.

Figure 3.5 Dependent variable, log of total cost per Cluster Review Period (CRP)



3.5.2. Descriptive statistics

Table 3.3 displays the descriptive statistics for our dependent and independent variables for the estimation sample of 689,404 observations with reference categories in brackets.

Table 3.3 Descriptive statistics (n=689,404)

Variable	Mean	Standard Deviation	Min	Max
Total cost of a CRP	3448	9783	0.99	303131
Log of total cost of a CRP	6.92	1.62	0.01	12.62
[White ethnicity]	0.877	0.328	0	1
Asian ethnicity	0.045	0.208	0	1
Black ethnicity	0.047	0.211	0	1
Other ethnicity	0.031	0.173	0	1
[Age category 1 (18-34)]	0.204	0.403	0	1
Age category 2 (35-46)	0.191	0.393	0	1
Age category 3 (47-62)	0.207	0.405	0	1
Age category 4 (63-79)	0.204	0.403	0	1

Age category 5 (80+)	0.195	0.396	0	1
Gender [Female]	0.436	0.496	0	1
Married/civil partner	0.331	0.471	0	1
Admitted under the MHA	0.087	0.282	0	1
Under CPA	0.411	0.492	0	1
Cluster 0: Variance	0.011	0.102	0	1
[Cluster 1: Common mental health problems, low severity]	0.040	0.195	0	1
Cluster 2: Common mental health problems	0.050	0.219	0	1
Cluster 3: Nonpsychotic, moderate severity	0.117	0.321	0	1
Cluster 4: Non-psychotic, severe	0.088	0.284	0	1
Cluster 5: Non-psychotic, very severe	0.032	0.175	0	1
Cluster 6: Non-psychotic disorders of overvalued ideas	0.017	0.128	0	1
Cluster 7: Enduring non-psychotic disorders	0.039	0.193	0	1
Cluster 8: Non-psychotic chaotic and challenging disorders	0.036	0.186	0	1
Cluster 10: First episode in psychosis	0.027	0.163	0	1
Cluster 11: Ongoing recurrent psychosis, low symptoms	0.090	0.286	0	1
Cluster 12: Ongoing or recurrent psychosis, high disability	0.064	0.245	0	1
Cluster 13: Ongoing or recurrent psychosis, high symptom/disability	0.045	0.208	0	1
Cluster 14: Psychotic crisis	0.028	0.166	0	1
Cluster 15: Severe psychotic depression	0.010	0.102	0	1
Cluster 16: Dual diagnosis, substance abuse and mental illness	0.016	0.126	0	1
Cluster 17: Psychosis and affective disorder difficult to engage	0.022	0.148	0	1
Cluster 18: Cognitive impairment, low need	0.098	0.297	0	1
Cluster 19: Cognitive impairment or dementia, moderate need	0.108	0.310	0	1
Cluster 20: Cognitive impairment or dementia, high need	0.044	0.204	0	1
Cluster 21: Cognitive impairment or dementia, high physical need	0.019	0.135	0	1
CRP started in 2012/13 [CRP started in 2011/12]	0.423	0.494	0	1
Income Deprivation	17.97	11.785	0	77
Provider-level variables, n=681,027				
Foundation Trust (FT)	0.74	0.44	0	1
Number of mental health beds	516	230	50	1010
Mental health beds occupancy (%)	88.31	5.30	63.9	99.6
Proportion of formal admissions	0.27	0.09	0.06	37.40

In terms of our risk-adjusters, the majority (88%) of observations are of White ethnicity with Black ethnicity and Asian ethnicities representing around 5% of observations and Other ethnicities 3%. Age ranges from 18 to 110. The majority of observations are of female gender with males accounting for around 44% of observations. One-third of observations are married or have a civil partner. 9% of observations had an admission under the MHA prior to or at the beginning of entry to a cluster while around 40% were under CPA. In terms of the 21 clusters, Cluster 3 is the most common with around 12% of observations followed by Clusters 19 and 18 with 11% and 10% respectively. 43% of observations started a CRP in 2012/13. On average, the observations in our study lived in an area where 18% of the population experienced income deprivation but this ranges from 0% to 77%.

The estimation sample size is reduced to 681,027 observations for the sensitivity analysis including provider variables due to missing data on these additional variables for 4 providers. Almost three-quarters (74%) of providers have FT status. On average, the providers in our sample have just over 500 beds but there is considerable variation ranging from 50 to over 1,000 mental health beds. There is also variation between providers regarding bed occupancy. The average occupancy rates is 88% - just over the recommended rate of 85% (Royal College of Psychiatrists 2011) but some providers are operating with spare capacity with the lowest occupancy rate around 64% and other providers operating at almost full capacity with an occupancy rate of almost 100%. For the average provider, just over one-quarter admissions is under the MHA, but over one-third of admissions are under the MHA for one provider.

A Hausman test confirmed our preference for the random-effects model (chi-squared (33) = 32.79, Prob>chi-squared = 0.4775).

3.5.3. Estimation results

Table 3.4 displays the estimation results for the three-level log-linear model and GLM.

Table 3.4 Estimates of three-level log-linear model and generalized linear model (GLM)

	Observations per group			
	Number of Observations	Minimum	Average	Maximum
Level 3: Provider	55	33	12535	54060
Level 2: Person	413,568	1	1.7	43
Level 1: CRP	689,404			

	Log-linear		GLM	
Log likelihood	-1222897.6		-5970609.1	
Variable	**Coefficient**	**Standard Error**	**Coefficient**	**Standard Error**
Married/civil partner	0.009	0.004*	-20.02	16.45
Asian ethnicity	0.026	0.009**	114.65	35.55**
Black ethnicity	0.083	0.010***	423.74	36.36***
Other ethnicity	0.031	0.011**	70.27	42.41
Age category 2 (35-46)	0.086	0.006***	255.90	23.91***
Age category 3 (47-62)	0.147	0.006***	480.70	23.91***
Age category 4 (63-79)	0.295	0.007***	1123.07	27.37***
Age category 5 (80+)	0.181	0.008***	585.04	30.56***
Gender	0.011	0.004**	115.0	14.94***
Admitted under the MHA	0.681	0.008***	4272.19	41.37***
Under CPA	0.231	0.005***	1007.00	17.53***
Cluster 0: Variance	0.287	0.019***	1864.02	75.52***
Cluster 2: Common mental health problems	0.378	0.012***	1386.41	43.89***
Cluster 3: Nonpsychotic, moderate severity	0.686	0.010***	2518.38	39.62***
Cluster 4: Non-psychotic, severe	1.019	0.011***	3804.10	43.70***
Cluster 5: Non-psychotic, very severe	1.324	0.013***	5230.34	56.07***
Cluster 6: Non-psychotic disorders of overvalued ideas	1.285	0.016***	4836.18	65.19***
Cluster 7: Enduring non-psychotic disorders	1.282	0.013***	4783.26	52.75***
Cluster 8: Non-psychotic chaotic and challenging disorders	1.349	0.013***	5281.83	55.97***
Cluster 10: First episode in psychosis	1.684	0.014***	6444.12	62.46***
Cluster 11: Ongoing recurrent psychosis, low symptoms	1.034	0.011***	3678.16	43.81***
Cluster 12: Ongoing or recurrent psychosis, high disability	1.468	0.012***	5449.89	50.87***

Cluster 13: Ongoing or recurrent psychosis, high symptom/disability	1.721	0.013***	6602.53	57.82***
Cluster 14: Psychotic crisis	2.012	0.014***	7898.02	66.78***
Cluster 15: Severe psychotic depression	1.627	0.020***	6651.91	80.61***
Cluster 16: Dual diagnosis, substance abuse and mental illness	1.531	0.017***	5997.82	69.70***
Cluster 17: Psychosis and affective disorder difficult to engage	1.881	0.015***	7052.80	68.18***
Cluster 18: Cognitive impairment, low need	0.186	0.011***	39.43	42.18
Cluster 19: Cognitive impairment or dementia, moderate need	0.550	0.011***	1556.22	42.85***
Cluster 20: Cognitive impairment or dementia, high need	0.810	0.013***	3275.40	51.80***
Cluster 21: Cognitive impairment or dementia, high physical need	0.682	0.016***	2954.41	63.47***
Income Deprivation	0.000	0.000*	-0.65	0.64
CRP started in 2012/13	-0.494	0.004***	-1655.33	15.02***
Constant	5.934	0.057***	1151.89	1.01***
Random Effects	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**
Level 3: Provider	0.170	0.033	0.039	0.001
Level 2: Person	0.291	0.004	0.436	0.002
Level 1: CRP	1.768	0.004		

*****p<0.001, **p<0.01, *p<0.05**

As may be expected given the relatively large sample size most variables are statistically significant. The majority of variables have a positive effect on the cost of a CRP. The results of both models correspond closely in terms of sign and magnitude of coefficients with the exception of married/civil partner, which is statistically significant in the log-linear model but not in the GLM. Other variables that are statistically significant in the log-linear model but not in the GLM include Other ethnicity and Income Deprivation. Variables with the largest effects in both models include Black ethnicity, older age, admission under the MHA and care clusters 10 and 13-17.

In the log-linear model, Black ethnicity is associated with a 9% increase in the cost of a CRP compared to White ethnicity. Observations aged 63-79 are associated with CRPs that are 34% more costly than CRPs for observations aged 18-34. Admission under the MHA is associated with increased costs of almost 100% while the care clusters 10 and 13-17 are associated with cost increases ranging from 362% (Cluster 16) to almost 648% (Cluster 14).

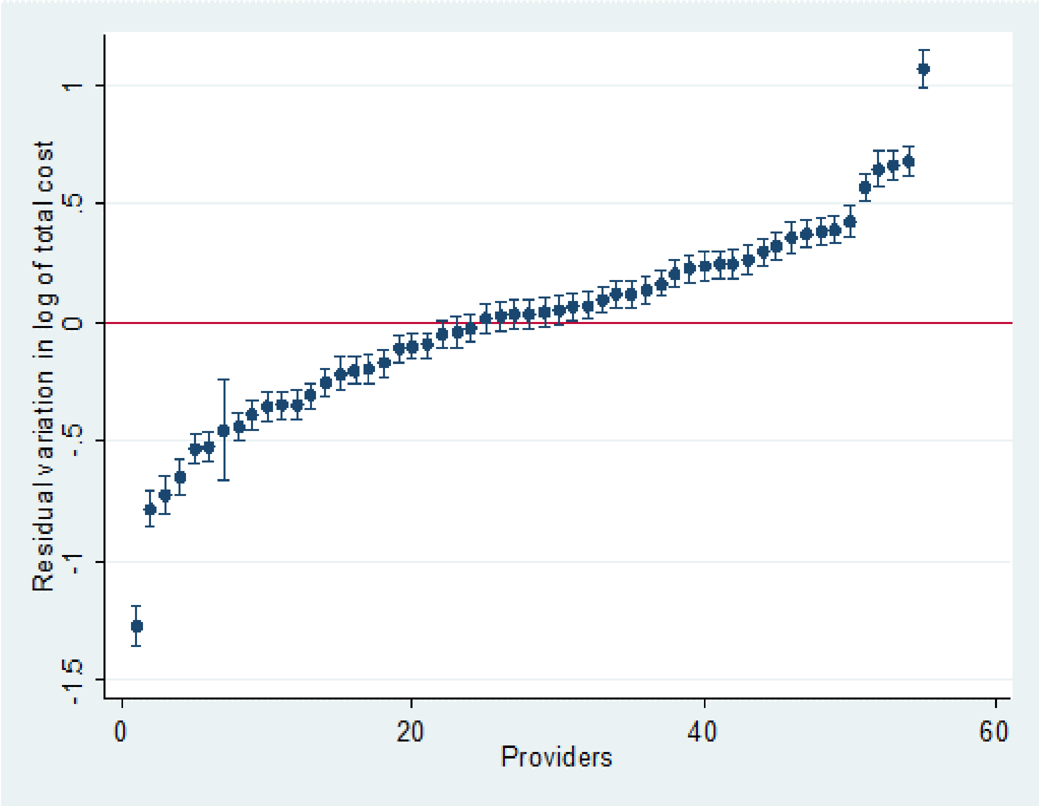
For the GLM, Black ethnicity is associated with an increased cost of a CRP of £424 compared to White ethnicity. Older age is associated with higher cost with age of 63-79 years associated with an increased cost of £1,123 and age 80 years and above associated with an increased cost of £585 compared to the age 18-34. Admission under the MHA is associated with an increase in costs of £4,272. The care clusters are broadly increasing in cost within the broad diagnostic groupings shown in Figure 1.1 in Chapter 1. In particular, Clusters 10 and 13-17 are associated with considerably higher costs compared to Cluster 1; Cluster 10 is associated with an increased cost of £6,444 and Cluster 17 is associated with a higher cost of £7,053 compared to Cluster 1. The variable capturing if the CRP started in 2012/13 is associated with a reduction in the cost of a CRP of 39% in the log-linear model and £1,655 in the GLM.

3.5.4. Provider-level residual variation

Around 8% of the residual variation in log of total cost is at the provider-level.

Figure 3.6 displays the EB predictions of the provider-level random effects for the log-linear model. The graph show that a number of providers consistently have higher or lower costs compared to the average performing provider after controlling for observable risk-factors. The provider performing best in terms of cost-containment has residual costs 72% below the average while the worse performing provider has residual costs 194% above the average performing provider.

Figure 3.6 Variation in provider-level residual variation from the log-linear model



3.5.5. Sensitivity analysis

Table 3.5 shows the results of the sensitivity analysis that included a number of provider-level variables we expect to be associated with cost.

Table 3.5 Estimates from sensitivity analysis including provider-level variables

	Observations Per Group			
	Number of observations	**Minimum**	**Average**	**Maximum**
Level 3: Provider	51	489	13353.5	54060
Level 2: Person	407385	1	1.7	43
Level 1: CRP	681,027			
	Log-linear		**GLM**	
Log-likelihood	-1207545		-5897662.9	
Variable	**Coefficient**	**Standard Error**	**Coefficient**	**Standard Error**
Married/civil partner	0.009	0.004*	-13.50	8.93
Asian ethnicity	0.026	0.010**	8.20	19.30
Black ethnicity	0.085	0.010***	185.72	19.96***

Other ethnicity	0.032	0.011**	-21.13	23.11
Age category 2 (35-46)	0.087	0.006***	143.35	12.94***
Age category 3 (47-62)	0.149	0.006***	265.49	12.93***
Age category 4 (63-79)	0.296	0.007***	613.11	14.74***
Age category 5 (80+)	0.182	0.008***	331.70	16.54***
Gender	0.010	0.004**	65.02	8.09***
Admitted under MHA	0.670	0.008***	2275.57	20.78***
Under CPA	0.230	0.005***	508.77	9.16***
Cluster 0: Variance	0.287	0.019***	989.46	39.66***
Cluster 2: Common mental health problems	0.377	0.012***	765.22	23.64***
Cluster 3: Nonpsychotic, moderate severity	0.685	0.010***	1369.59	20.92***
Cluster 4: Non-psychotic, severe	1.018	0.011***	2052.37	22.47***
Cluster 5: Non-psychotic, very severe	1.325	0.013***	2839.59	28.46***
Cluster 6: Non-psychotic disorders of overvalued	1.290	0.016***	2633.31	34.00***
Cluster 7: Enduring non-psychotic disorders	1.281	0.013***	2600.45	26.82***
Cluster 8: Non-psychotic chaotic and challenging disorders	1.348	0.013***	2879.71	28.49***
Cluster 10: First episode in psychosis	1.683	0.014***	3509.89	31.46***
Cluster 11: Ongoing recurrent psychosis, low	1.029	0.011***	2012.14	22.76***
Cluster 12: Ongoing or recurrent psychosis, high disability	1.466	0.012***	2983.98	25.46***
Cluster 13: Ongoing or recurrent psychosis, high symptom/disability	1.715	0.013***	3581.40	28.50***
Cluster 14: Psychotic	2.007	0.014***	4282.05	32.78***
Cluster 15: Severe psychotic depression	1.623	0.020***	3623.61	41.75***
Cluster 16: Dual diagnosis, substance abuse and mental illness	1.520	0.017***	3241.07	36.05***
Cluster 17: Psychosis and affective disorder difficult to engage	1.874	0.016***	3828.30	34.60***
Cluster 18: Cognitive impairment, low need	0.184	0.011***	23.39	22.85
Cluster 19: Cognitive impairment or dementia, moderate need	0.547	0.011***	838.81	22.91***
Cluster 20: Cognitive impairment or dementia, high need	0.813	0.013***	1788.40	27.28***
Cluster 21: Cognitive impairment or dementia, high physical need	0.693	0.016***	1646.44	34.01***
Income Deprivation	0.000	0.000*	-0.68	0.35
CRP started in 2012/13	-0.490	0.004***	-881.38	7.69***
Foundation Trust (FT)	-0.216	0.132	-259.93	9.97***

Number of mental health beds	0.000	0.000	-0.82	0.02***
Mental health beds occupancy (%)	-0.002	0.009	25.40	0.78***
Proportion of formal admissions	-0.222	0.641	231.32	50.10***
Constant	6.33583	0.820***	292.02	1.03***
Random Effect	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**
Level 3: Provider	0.170	0.034	0.062	0.066
Level 2: Person	0.287	0.004	0.430	0.439
Level 1: CRP	1.769	0.004		

*****p<0.001, **p<0.01, *p<0.05**

None of the provider-level variables are statistically significant in the log-linear model but they are all significant in the GLM model. For the log-linear model, the magnitudes of the cluster review- and patient-level variables are similar to the baseline model while the coefficients of these variables in the GLM model are reduced when provider variables are included compared to the baseline model. In the GLM model, the number of mental health beds and mental health bed occupancy are associated with relatively small effects on costs with the former exercising downward pressure on costs and the latter upward pressure. On the other hand, FT status and the proportion of formal admissions at the provider-level are associated with sizable effects on costs; providers with FT status are associated with reduced costs of a CRP of £260, while a one-unit increase in the proportion of formal admissions is associated with an increased cost of a CRP of £231. The residual variation in log of costs at the provider-level remains 8%.

3.6. *Discussion*

This chapter has provided the results of a preliminary exercise in costing mental health activity that will be reimbursed under the NTPS. We have compared costs across providers and attempted to explain variations in these costs due to observable patient (and provider) factors that would be expected to be beyond the control of the provider. Furthermore, we provide insight into the extent to which the care clusters explain variation in costs. Despite controlling for a wide range of variables, we find evidence of residual variation in costs at the provider-level and this suggests that a number of providers have above average costs and may face financial instability when a national tariff is introduced.

Our findings on the drivers of mental health costs echo those of previous studies. We find that Black ethnicity is associated with higher costs compared to White ethnicity. This supports the findings of previous studies (Eagar et al. 2004; Robst 2009; Sutton et al. 2012) that found that ethnic minorities have higher costs. Age has a non-linear relationship with costs with older age categories associated with higher costs. Males are associated with higher costs compared to females. Admission under the MHA is a key cost driver and it may be the case that this variable is picking up aspects of severity not adequately captured by the care clusters. While the care clusters do not explain all variation in costs, the direction of the effects of the cluster variables does appear intuitive with the clusters reflecting higher severity and need associated with higher costs. CRPs that started in 2012/13 are associated with lower costs compared to those that started in 2011/12. This may reflect improved coding of the cost data in 2012/13. Our sensitivity analysis that considered provider-level variables revealed that the number and percentage occupancy of mental health beds are associated with relatively small effects on mental health costs but FT status and the proportion of formal admissions at provider-level are associated with notable negative and positive effects respectively on mental health costs.

Previous literature on mental health costs has underlined the inadequacies of classification and payment systems based primarily on diagnosis to accurately predict a large proportion of mental health costs and the need to consider a wider range of variables, in particular those reflecting patient need, social circumstances and treatment. While we do not consider a classification system based primarily on diagnosis, our study nevertheless continues in a similar vein in that it shows that the classification system developed for the NTPS in mental health is not sufficient by itself to explain variations in mental health costs and other factors are important cost drivers. Moreover, even after controlling for all of these variables, there still remains considerable residual variation in costs and this varies across providers with a small number of providers continuing to have higher than average costs. It would be too simplistic to label these providers as being inefficient as there are a number of important factors that we haven't considered in this analysis. Firstly, those providers with higher residual costs may be providing better quality care and Chapter 5 explores the relationship between cost and outcomes to try to provide some insight into whether higher cost is associated with better outcomes. In the acute physical healthcare sector, the NTPS primarily incentivises higher levels of activity. For implementation of the NTPS in mental health, a set of quality indicators and outcome measures that commissioners and providers can use in setting contracts are under development (Department of Health Payment by Results team 2013b) so that quality of care will not be sacrificed in the drive to increase activity and contain costs. Moreover, the introduction of national fixed prices for each cluster should allow contract negotiations to focus more on quality and there will also be more of an onus on providers to demonstrate good outcomes in order to distinguish themselves in a more competitive market (Yeomans 2014). Secondly, providers with higher residual costs may be treating a certain casemix of patients that we haven't been able to fully account for. A limitation of our set of risk adjustment variables is that they exclude diagnosis due to poor data coding in the MHMDS. While classification systems based primarily on diagnosis perform poorly in predicting the majority of mental health costs, this does not mean that diagnosis should be ignored entirely and several studies have found that diagnosis can explain some variation in costs with more severe diagnoses such as psychoses being associated with higher costs (Buckingham et al. 1998; Drozd et al. 2006). The clustering method does not explicitly take diagnosis into account and it is likely that the clusters are very variable in terms of diagnosis and casemix (Jacobs 2014; Yeomans 2014). It may also be the case that some patients have treatment-resistant variants of mental illness which implies that they will be consuming large amounts of care and resources to little avail (Jencks et al. 1987). If certain providers have a higher case-load of such patients this could well explain their unexplained higher costs. Thirdly, the variations in residual costs may be a reflection of variations in practice. Practice variations may be more prevalent in psychiatric care compared to physical healthcare and the introduction of the NTPS will shed insight into variation in practice as different clinicians are likely to make different MCHT ratings (Yeomans 2014). If practice variations are leading to inefficiencies in resource use, then it is likely that a national tariff will help to reduce inappropriate resource use while adequate training and supervision of clinicians in the use of the MHCT will help to prevent the clustering process exacerbating such problems. However, some variations in practice may be warranted and the classification system could partly address this by incorporating psychiatric procedures such as rehabilitation, detoxification, and intensive inpatient care to mirror the use of medical/surgical procedures in HRGs used in the acute sector (Oyebode 2007). Fourthly, poor cost data may lead to certain providers appearing to have above-average costs. Concerns have been raised as to the reliability of cluster costing data (Capita 2013; PriceWaterhouseCoopers 2012). Accurate costing will be imperative under a national tariff as this assumes that providers face the same cost structures and have the same prospects to make cost reductions but the tariff will not act as an accurate price signal if it is based on imperfect data (Jacobs 2014). It is intended that PLICS will be introduced for use by mental health providers on a developmental basis in 2016 leading to eventual mandatory use by 2020 (Monitor 2015).

In order to assess its potential to successfully facilitate the introduction of the NTPS to mental healthcare in England the care clusters classification system may require refinements. Despite being implemented the system has not been independently evaluated (Jacobs 2014) and such an exercise would inform and potentially improve the system. From an international perspective, the fact that the system is being used to inform contracts between commissioners and providers is both innovative and progressive as several countries have developed psychiatric classification systems but have not implemented these in a provider payment system. Moreover, the care clusters classification system is only one aspect of the NTPS in mental health. Any weaknesses of the care clusters approach can potentially be addressed by other aspects of the payment system. An important consideration for the refinement of the NTPS in mental health will be the outlier policy used so that any providers attracting high-cost patients, not adequately accounted for by the classification system, will not be penalised. A case in point may be in relation to the MHA as we find that the proportion of formal admissions at the provider-level is associated with a considerable increase in costs. Caution has been advised about the use of legal status in a classification and payment system as it may inadvertently increase involuntary treatment (Buckingham et al. 1998). This may be a legitimate concern in England as while use of the MHA is tightly regulated, it has been suggested that the MHA is used to acquire access to an inpatient bed due to high demand pressures on beds (House of Commons Health Committee 2013). However, providers do have different thresholds and capacities for admission under the MHA and if this is not recognised in the payment system it could potentially leave some providers facing financial risk.

If the NTPS for mental health does not adequately address legitimate reasons for cost variations among providers then there is a danger of inducing undesirable behaviours on the part of providers. These could include “dumping’ more expensive patients and treating more of those patients expected to incur less resources in order to reduce costs. Alternatively, providers may move patients into more expensive clusters and it could be argued that this may be relatively easier in mental healthcare where clinicians themselves will be the coders as opposed to acute physical healthcare where coders are external. However, the existence of a small number of clusters may mitigate this somewhat and the use of audit should also help to deter such practices (Jacobs 2014; Yeomans 2014). The extent to which providers may be tempted to engage in “gaming” the system may also depend on how much revenue they will receive from the NTPS. As noted earlier, not all mental health services will be reimbursed under the NTPS and even if providers make a loss on the NTPS services this may be balanced by a surplus on non-NTPS services. However, continual losses from NTPS may then encourage a shift away from providing these services and increased specialisation in non-NTPS services.

Chapter 4. Measuring and comparing the performance of English mental health providers in achieving patient outcomes

4.1. *Introduction*

A routinely collected mental health outcome measure is necessary to underpin any prospective payment system linked to outcomes. There is great variation between countries in the implementation and use of mental health outcome measures. In some countries, routine outcome measurement is centrally driven while in others, several relatively autonomous systems of outcome measurement co-exist. Similarly, the choice of instrument differs across countries and appears to be determined by local factors with some countries using CROMs and others PROMs (Trauer 2010a).

While it has been mandatory for English mental health providers to collect routine outcome data in the form of HoNOS since 2003, completion rates have been suboptimal (Jacobs 2009; Slade 2010), partly because commissioning of services was not based on outcomes and there were no financial consequences to non-compliance with outcome reporting requirements (Slade 2010). The NTPS for mental health changes the incentives faced by providers as HoNOS is an integral part of the classification system used for payment, while it is envisioned that provider performance on patient outcomes (based on HoNOS data) will also influence payment. Moreover, a national tariff per cluster will help to focus contracts between commissioners and providers on quality rather than price (Fairbairn 2007; Yeomans 2014). Information on quality and outcomes will help to guide commissioners as to which providers are performing best in terms of patient outcomes (Department of Health Payment by Results team 2013b) and this should encourage providers to distinguish themselves in order to secure contracts in a more competitive market (Yeomans 2014). More transparent provider outcomes will also inform patient choice in mental health services (NHS England 2014a) which is facilitated by the move away from block budgets to prospective activity-based funding.

This chapter explores the potential for the NTPS for mental health to incentivise provider performance in relation to patient outcomes by examining what factors contribute to variations in outcomes measured using HoNOS and, after controlling for these variables if systematic differences in performance across providers remain. We move beyond previous international studies that compare provider performance on patient outcomes by using a large, nationally representative patient-level data set – the MHMDS. To date, the performance of English mental healthcare providers in relation to outcomes has not been researched in a systematic and rigourous manner. Yet, the availability of HoNOS data in the MHMDS allows a rigourous performance assessment of English mental healthcare providers to be undertaken. The richness of the MHMDS is exploited using multilevel models, which enables us to consider the influence of different levels of analysis on outcomes. Residual variation is quantified using EB methods which allows us to compare provider performance.

4.2. *Risk adjustment and comparisons of provider performance in physical and mental healthcare*

The routine collection of health outcome data offers the potential to make comparisons across health professionals or organisations in order to answer questions such as "are providers meeting minimum standards of performance?" and "how are providers performing relative to others doing the same kind of work?" (Smith and Street 2013). With adequate risk adjustment, it is possible to examine the performance of providers using routinely collected patient outcome data (Gutacker et al. 2013b). Sections 4.2.1 and 4.2.2 provide an overview of the data and methods used to compare risk-adjusted outcomes across providers in both the physical and mental healthcare sectors.

4.2.1. Physical healthcare

A recent study (Nuttall, Parkin and Devlin 2015) describes a methodology for the casemix adjustment of PROMs data collected by NHS funded providers in England in order to facilitate the comparison of outcomes between providers of elective surgery. This methodology consists of two stages of which the first stage involves regressing the post-operative health outcome of an individual patient against their pre-operative health outcome, condition-specific factors and characteristics unrelated to their condition. The model also includes a provider effect that can be considered fixed or random. In the second stage, indirect standardization is used to generate a measure of providers' performance relative to other providers. National average health outcomes are multiplied by a provider-specific variable ρ_j which measures how provider j performs relative to the national average. A value of ρ_j greater than one signifies that the provider performs better than average while a value of ρ_j less than one signals worse than average performance. Provider performance in terms of casemix adjusted and unadjusted health outcomes can then be compared.

A separate study (Gutacker et al. 2013b) using PROMs data collected by NHS funded providers in England outlines a methodology to compare providers on the separate dimensions of the EQ-5D using hierarchical ordered probit models. Health status can be considered a latent variable as it is not directly observed but is instead inferred from patient's responses to the EQ-5D questionnaire which contain three categories; 1 = no problems, 2 = some problems, 3 = extreme problems. A hierarchical ordered probit model was employed with pre- and post- treatment outcomes regressed against a set of time invariant patient-level risk adjustment variables with a dummy variable equal to one if the outcome is post-treatment. The model also included random intercepts for both patients and providers as well as a random coefficient on the treatment variable that varies by providers and reflects the provider effect on post-treatment outcomes. This treatment effect was predicted post-estimation and quantified using EB estimates in order to rank providers. Provider performance was also compared according to the probability of reporting a specific post-treatment outcome (m=1,2,3), based on the estimated quality effort of the provider.

4.2.2. Mental healthcare

In contrast to acute care and elective surgery, risk adjustment is relatively underdeveloped in mental health (Dow, Boaz and Thornton 2001; Hendryx, Beigel and Doucette 2001; Hermann, Rollins and Chan 2007; Rosen et al. 2010). Moreover, risk adjustment in mental health has focused mainly on payment systems, and little on risk adjustment of outcomes data for the purpose of comparing provider performance (Dow, Boaz and Thornton 2001). A number of studies of risk adjustment of mental health outcomes were identified through a search of databases including EconLit, Embase, OvidMedline and PsychInfo using the following search terms: "mental health", "psychiatry", "outcomes", "risk adjustment", and "performance".

The numbers of patients and providers examined in previous studies has been relatively small and usually restricted to one US state or geographical area. Two studies (Hendryx, Dyck and Srebnik 1999; Hendryx and Teague 2001) compared six publicly funded community mental health agencies in the US state of Washington with the sample sizes varying from 289 (Hendryx, Dyck and Srebnik 1999) to 336 (Hendryx and Teague 2001) adult users. A larger study used data from 24 state-funded mental health facilities in Florida covering a sample of almost 8,000 patients who were classified according to state certification procedures as adult disabled or in crisis (Dow, Boaz and Thornton 2001). Two studies (Kramer et al. 2001; Rosen et al. 2010) used data from the VHA. Kramer et al. (2001) compared the outcomes of 187 patients undergoing treatment for major depression disorder in three types of specialty mental health treatment settings including a VHA clinic, a clinic attached to a university teaching hospital, and a staff model managed care organisation. Rosen et al. (2010) analysed data on 986 veterans receiving inpatient or outpatient mental health/substance abuse care in one of two VHA medical centres in New England from mid-2004 to mid-2006.

Previous studies measured mental health outcomes in terms of various domains including PROMS (Hendryx and Teague 2001; Rosen et al. 2010), functioning (Dow, Boaz and Thornton 2001; Hendryx, Dyck and Srebnik 1999; Kramer et al. 2001), patient satisfaction (Dow, Boaz and Thornton 2001; Hendryx, Dyck and Srebnik 1999), quality of life (Hendryx, Dyck and Srebnik 1999) and diagnosis and severity (Kramer et al. 2001).

In terms of risk adjustment variables, all studies included information on age and gender. Additional sociodemographic variables included ethnicity (Hendryx, Dyck and Srebnik 1999; Hendryx and Teague 2001; Kramer et al. 2001; Rosen et al. 2010), marital status (Kramer et al. 2001; Rosen et al. 2010), education (Dow, Boaz and Thornton 2001; Kramer et al. 2001; Rosen et al. 2010), income (Dow, Boaz and Thornton 2001; Kramer et al. 2001), employment (Dow, Boaz and Thornton 2001; Rosen et al. 2010), homelessness and social support (Kramer et al. 2001; Rosen et al. 2010). All studies also included information on diagnosis as well as baseline measures of the dependent outcome variable. Other risk-adjusters included information on substance abuse (Hendryx and Teague 2001; Kramer et al. 2001), clinical history and status (Hendryx and Teague 2001; Kramer et al. 2001), social functioning (Hendryx and Teague 2001), physical health (Hendryx and Teague 2001), voluntary treatment and duration of community-based treatment (Dow, Boaz and Thornton 2001).

A number of the studies (Hendryx, Dyck and Srebnik 1999; Kramer et al. 2001) used split-sample model validation in the development of the risk adjustment model with the sample randomly split to allow model development using one part of the sample and model testing on the other. The motivation for this approach is the design of an external validation study (Steyerberg 2009). However, split-sample validation can be criticized on a number of grounds including: imbalances in the distribution of the outcome and predictor variables owing to the random split of the sample, which may be aggravated if these distributions are skewed; model results may be less stable given that a subset of the data sample is used; similarly validation of model performance may be unreliable as it is based on only a section of the sample and model performance may depend on the particular random sample used; while bias is also introduced as ideally model performance should be assessed based on the full sample, not a random selection (Steyerberg 2009).

Most studies (Dow, Boaz and Thornton 2001; Hendryx, Dyck and Srebnik 1999; Hendryx and Teague 2001; Kramer et al. 2001; Rosen et al. 2010) modeled outcomes using linear regression, while one study that used a measure with a binary outcome also used logistic regression (Kramer et al. 2001). Another study included provider random effects to control for the clustering of patients by site (Rosen et al. 2010). A common technique employed was to introduce risk adjustment variables sequentially into the model in order to examine the additional variance they explained in the dependent variable (Hendryx, Dyck and Srebnik 1999; Hendryx and Teague 2001; Kramer et al. 2001; Rosen et al. 2010). In a number of studies, the variables that performed best in terms of explaining variation in the dependent variable were used to predict outcomes in the final risk adjustment model (Dow, Boaz and Thornton 2001; Hendryx, Dyck and Srebnik 1999; Hendryx and Teague 2001; Kramer et al. 2001). Providers were compared by generating ranks based on unadjusted and risk-adjusted outcomes (Dow, Boaz and Thornton 2001; Hendryx, Dyck and Srebnik 1999; Hendryx and Teague 2001; Kramer et al. 2001).

Variables found to significantly predict mental health outcomes include age (Dow, Boaz and Thornton 2001; Hendryx, Dyck and Srebnik 1999; Hendryx and Teague 2001; Kramer et al. 2001; Rosen et al. 2010); gender (Rosen et al. 2010); ethnicity (Rosen et al. 2010); marital status (Rosen et al. 2010); diagnosis (Hendryx, Dyck and Srebnik 1999; Hendryx and Teague 2001; Rosen et al. 2010); education (Dow, Boaz and Thornton 2001; Rosen et al. 2010), income (Kramer et al. 2001); employment (Rosen et al. 2010); social support (Rosen et al. 2010), homelessness (Rosen et al. 2010); baseline measures of the outcome variable (Dow, Boaz and Thornton 2001; Hendryx, Dyck and Srebnik 1999; Kramer et al. 2001; Rosen et al. 2010); substance abuse (Hendryx, Dyck and Srebnik 1999; Hendryx and Teague 2001); physical health (Hendryx and Teague 2001; Kramer et al. 2001); duration of community-based care (Dow, Boaz and Thornton 2001); and involvement in decision to enter services (Dow, Boaz and Thornton 2001). Studies that entered variables sequentially to models reported that sociodemographic variables alone did not account for significant variation in outcomes (Hendryx, Dyck and Srebnik 1999; Rosen et al. 2010).

Previous studies (Dow, Boaz and Thornton 2001; Hendryx, Dyck and Srebnik 1999; Hendryx and Teague 2001; Kramer et al. 2001) have underlined the need for risk adjustment by showing that provider performance varied between unadjusted and adjusted outcomes. Moreover, the ranking of providers was found to differ according to the outcome variable used (Dow, Boaz and Thornton 2001). Hendryx and Teague (2001) conducted analyses based on different diagnostic groups and found that the significance of risk adjustment variables as well as provider performance differed according to diagnostic groups, implying that the pooling of diagnostic samples can obscure this information. However, the stratification of samples according to diagnosis will depend on diagnosis-specific sample sizes and the particular outcome of interest.

Our review of the literature in both physical and mental healthcare informs our analysis in several ways. While variables such as age, sex, ethnicity and co-morbidities are important risk-adjusters in acute physical healthcare, the literature on risk adjustment for mental healthcare highlights the need to supplement these with a broader set of covariates covering treatment and social circumstances. Moreover, baseline measures of outcome appear to be particularly strong predictor variables. We draw on the multilevel modeling techniques used in physical healthcare and directly model the provider effect in order to make inferences about provider performance in terms of patient outcomes.

4.3. *Measuring mental health outcomes using Reliable and Clinically Significant Change (RCSC)*

In measuring an outcome in terms of change in baseline and follow-up HoNOS scores, it is important to differentiate between *statistical* significance and *clinical* significance as differences that meet the criteria for statistical significance may not be clinically meaningful, particularly as sample size increases (Eisen et al. 2007). Reliable change can be ascertained using the Reliable Change Index (RCI) whereby the post-treatment score is subtracted from the pre-treatment score and divided by the standard error of the differences. If the absolute value of "t" is greater than 1.96, then change is considered statistically reliable. If a change is deemed statistically reliable, the clinical significance of the change can be established by verifying that the post-treatment score falls within the range of scores for a population with no mental health problems (Eisen et al. 2007).

The RCI was initially developed by Jacobson et al. (1984) for use in psychotherapy outcome research and later modified based on suggestions by Christensen and Mendoza (1986). The premise of the RCI is that the level of change in outcome for a particular individual should be statistically reliable in that it cannot merely be attributed to chance or measurement error. In order to judge how much change is necessary to be considered clinically significant, there is a need for a reference standard - the normal or functional population - in order to prevent arbitrary decisions. Clinical significance should refer to a range of possible outcomes and not merely a binary choice. Then, an intervention can be judged to have achieved a reliable and clinically significant change if an individual's level of functioning following the intervention means they are statistically more likely to fall within the functional, rather than the dysfunctional population (Jacobson, Follette and Revenstorf 1984). However, to date no HoNOS measurements have been taken from a sample of people with no mental health problems in the general population (Jacobs 2009).

Despite a lack of HoNOS scores for a population with no mental health problems, some researchers have calculated the RCI for HoNOS using an alternative criterion for a functional population. In a study using HoNOS to assess patient change in NHS psychotherapy and psychological treatment services, Audin et al. (2001) calculated a threshold for clinical change as the mean total assessment (baseline) score plus the mean total discharge (follow-up) score, halved (Audin et al. 2001). This enabled a categorisation of study participants into seven mutually exclusive groups as outlined in Table 4.1. In another study using HoNOS, Parabiaghi et al. (2005) similarly categorised patients into seven groups based on RCSC calculations (Parabiaghi et al. 2005) which correspond to those of Audin et al. (2001) (Table 4.1).

Table 4.1 Reliable and Clinically Significant Change (RCSC) categories, Audin et al. (2001) and Parabiaghi et al. (2005)

Patient category	**Audin et al. (2001)**	**Parabiaghi et al. (2005)**
1	Clinical/reliable deterioration	Recurrence
2	Clinical deterioration/no reliable change	Clinical deterioration
3	No clinical change/reliable deterioration	Deterioration
4	No clinical change/no reliable change	Stable
5	No clinical change/reliable improvement	Improvement
6	Clinical improvement/no reliable change	Clinical improvement
7	Clinical improvement/reliable improvement	Remission

In the absence of HoNOS scores for a population with no mental health service users, Parabiaghi et al. (2005) classified a sample of patients based on severity to identify a functional population in order to determine clinical significance (Parabiaghi et al. 2005).

Reliable change was calculated as follows:

$$RC_{index} = 1.96 \text{ x } SE_{diff} \text{ where } SE_{diff} = SD_1 \text{ x } \sqrt{2} \text{ x } \sqrt{(1-\alpha)}$$

where SD_1 is the standard deviation of the baseline observations and α is Cronbach's coefficient (a measure of internal consistency).

A clinically significant change was judged to occur when a patient's score moved from the "dysfunctional population" range into the "functional population" range, which required calculation of a cut-off point where there was an equal chance of belonging to either distribution. Functional and dysfunctional was determined according to clinical severity with two categories of severity defined: 1) severe patients defined as having scores of ≥ 3 in at least one item of HoNOS and 2) very severe patients with a score of ≥ 3 in at least two items. Non-severe patients were considered either: 1) mild with at least one item with a score =2, or 2) subclinical with a score of <2 in all items.

The clinically significant (CS) "cut off" point was calculated as:

$$CS_{cut\text{-}off} = \frac{(mean_{clin} \text{ x } SD_{norm}) + (mean_{norm} \text{ x } SD_{clin})}{SD_{norm} + SD_{clin}}$$

where $mean_{clin}$ and $mean_{norm}$ are the mean scores of the "dysfunctional population" and the "functional population", respectively and SD_{norm} and SD_{clin} are the standard deviations of the scores in these two groups. Two cut-off points were calculated: cut-off_1 that separated the group of "very severe" patients from the other service users and cut-off_2 that separated the group of subclinical subjects from the group of clinical subjects (mild, moderately severe and very severe). The cut-$off_{1 \text{ and}}$ cut-off_2 thresholds calculated from baseline data were a HoNOS score of 11 and 5 respectively.

In a follow-up study Parabiaghi et al. (2011) again applied the RCSC concept to HoNOS scores in order to evaluate clinical change. The Clinical Global Impression Scale (CGIS) was used to measure severity in the study sample. Patients were categorised into two groups based on having a CGIS rating of 1-4 (mild to moderate) or 5-7 (severe) in order to calculate the $CS_{cut\text{-}off.}$ A five-level classification was obtained: reliable and clinically significent improvement, reliable improvement, stability, reliable deterioration and reliable and clinically significant deterioration (Parabiaghi et al. 2011).

More recently, the RCSC concept has been applied to a sample of 4,146 working age and older adult HoNOS data sets provided by the Tees, Esk and Wear Valleys NHS FT in England in order to investigate the utility of the MHCT as a generic outcome measure (Speak and Hay 2012). Patient severity was classified according to HoNOS scores in order to calculate the RCI and clinical severity cut-offs. The baseline HoNOS ratings had a Cronbach alpha of 0.6 which gave a RCI of 9, a cut-off threshold of 12 (cut-off$_1$) to seperate the group of "very severe" patients from the other service users, and a cut-off threshold of 5 (cut-off$_2$) to seperate the group of subclinical subjects from the group of clinical subjects. The resulting categorisation of patients found that 92% remained stable, 0.3% showed a reliable improvement, 2.8% a clinical improvement and 3.6% were considered in remission. In contrast, 0.4% were shown to have a recurrence or clinically and statistically significant deterioration; 0.6% a clinical deterioration; and 0.1% deterioration. The authors interpret these results in light of the relatively low Cronbach alpha and relatively wide standard deviations, which made it more difficult to detect clinically significant changes in HoNOS total scores. The RCI and clinical cut-offs were also calculated for each care cluster individually. The RCI ranged from 0.569 to 0.740, cut-off$_1$ ranged from 9 to 15 and cut-off$_2$ ranged from 4 to 8.

4.4. *Data*

The main data set used for the analysis is the MHMDS, which is described in Section 1.3.1 of Chapter 1. The MHMDS data for 2011/12 and 2012/13 were cleaned to remove: duplicate observations, observations with a CRP that did not have corresponding HoNOS scores recorded, observations with age coded as less than 18 years or greater than 110 years, and observations treated by private providers. Our final estimation sample is 305,960 CRP observations for 163,611 patients treated by 57 providers.

4.4.1. Health of the Nation Outcome Scales (HoNOS)

The MHMDS contains data on HoNOS, which is used to measure mental health outcomes for our dependent variable(s).

HoNOS was developed by the Royal College of Psychiatrists' Research Unit in response to a request by the Department of Health in 1993 in order to measure progress towards the Health of the Nation target "to improve significantly the health and social functioning of mentally ill people" (Wing, Curtis and Beevor 1994). As HoNOS forms part of the MHMDS, providers of specialist adult mental healthcare are mandated to undertake HoNOS assessments.

HoNOS is comprised of 12 items, each of which is scored from 0 (no problem) to 4 (severe problem) giving a total score in the range of 0 (best) to 48 (worst). The 12 items can also be aggregated into 4 subscales/sections. Table 4.2 describes the 12 items, 4 subscales and scoring of HoNOS.

Table 4.2 Health of the Nation Outcome Scales (HoNOS) items, subscales and scoring

Item	Subscales/sections	Scoring
1. Overactive, aggressive, disruptive or agitated behaviour	Behaviour (1-3)	Each item rated on a 5-point scale: 0. no problem 1. minor problem requiring no action 2. mild problem but definitely present 3. moderately severe problem 4. severe to very severe problem Scoring yields individual item scores, subscale scores and a total score.
2. Non-accidental self-injury		
3. Problem-drinking or drug-taking		
4. Cognitive problems	Function / Impairment (4-5)	
5. Physical illness or disability problems		
6. Problems associated with hallucinations and delusions	Symptoms (6-8)	
7. Problems with depressed mood		
8. Other mental and behavioural problems		
9. Problems with relationships	Social (9-12)	
10. Problems with activities of daily living		
11. Problems with living conditions		
12. Problems with occupation and activities		

Source: Jacobs, R. (2009). Investigating Patient Outcome Measures in Mental Health. *CHE Research Paper 48.*

Ratings are made by an individual clinician (psychiatrist, nurse, psychologist, or social worker) or using a consensus rating. The rating is made on the basis of all information available to the clinician and is based on the most severe problem that arose during the two weeks leading up to the point of rating.

At the very least, a rating should be made at the beginning and end of each episode of care. Ratings are also expected to be taken at any regular review (for instance a CPA review), when a major change occurs in the patient's condition (such as an admission to or discharge from hospital) and on a bi-annual basis for long episodes of care. As the HoNOS underpins the MHCT, HoNOS scores will also be recorded when a patient changes cluster.

An outcome measure is obtained by calculating the change in a patient’s ratings at two points in time using individual item scores, the subscale scores and the total score but follow-up scores are often more difficult to record than baseline scores due to issues such as patient access and attendance or staff turnover (Jacobs 2009).

4.4.2. Risk adjustment covariates

The risk adjustment variables used in the analysis include demographic, need, treatment and social variables. Demographic information covers age, gender, ethnicity and marital status. Age ranges from 18 to 110 years and is grouped into five categories reflecting quintiles of the distribution in order to capture any non-linearities in the relationship with costs and outcomes with age 18-34 years as the reference category. Ethnicity is also categorised to represent the main ethnic groups in the data – White, Black, Asian and Other with White ethnicity treated as the reference category. Gender is represented by a dummy variable with males equal to one. Need is captured by the care cluster a patient is assigned to. This data is categorized according to 7 broad groups based on the relationship of care clusters to each other (Figure 1.1). We aggregate the individual clusters into these groups in order to make interpretation of the coefficients more meaningful. For example, it may be difficult to compare the outcomes associated with Cluster 1 (common mental health problems (low severity)) with those of Cluster 14 (psychotic crisis). Variables reflecting if a patient is under CPA or has been admitted to hospital under the MHA provides information on severity and treatment. Information on social circumstances is captured in the MHMDS by variables on employment and settled accomodation but poor coding discourages the use of these variables. Therefore, we include information on social circumstances by constructing a variable using Item 11 of the HoNOS that reflects problems with living conditions. This variable takes the form of a dummy variable equal to one if a HoNOS score of ≥2 is recorded for HoNOS Items 11. Information on socioeconomic deprivation is captured by including a variable for the IMD Income Deprivation Domain (Noble 2008) as described in Section 2.4.3 of Chapter 2 and Section 3.3.2 of Chapter 3. We also include dummy variables that capture the full time period of the CRP. Observations can have a CRP that: starts in 2011/12 and ends in 2012/13, starts and ends in 2011/12, or starts and ends in 2012/13. These variables are measured differently to those in Chapters 3 and 5 as in the cost analyses the time dummy variables act primarily to control for annual fluctuations in prices and inflation so it is more appropriate for the variables to capture the year the CRP commenced.

4.4.3. Provider variables

We include a number of provider-level variables in a sensitivity analysis to investigate if there are provider characteristics that are systematically associated with better or worse outcomes. These variables include FT status, number of mental health beds and bed occupancy. FTs are differentiated from other NHS Trusts as they are autonomous legal bodies and have different governance arrangements. They have considerable financial independence and do not undergo comparable levels of performance management (NHS Choices 2015a). The variable “mental health beds” provides a proxy for hospital size. Mental health bed occupancy provides an indication of utilisation of available capacity and reflects average mental health bed occupancy over a quarterly period from 2010/11 and an annual period for 2009/10. It is recommended that mental health bed occupancy rates not exceed 85% to avoid delays in admissions that can adversely affect outcomes (Royal College of Psychiatrists 2011) and therefore necessitate a greater outlay of resources.

4.5. *Methodology*

4.5.1. Multilevel modelling

As the MHMDS is characterised by a hierarchical structure with CRPs nested within patients who are in turn nested within providers, we utilise a multilevel modelling approach and compare results from two three-level models of CRPs in patients in providers: 1) an ordered probit model; and 2) a linear model.

1) Ordered probit model

We obtain a measure of RCSC in our HoNOS data by defining observations as severe if scores of ≥3 in at least one item of HoNOS are recorded in order to determine a clinical severity cut-off and subsequently categorise patients into five mutually exclusive groups. An ordered probit model is then employed in order to predict the probability of a particular outcome in a CRP, conditional on a set of risk adjustment variables. We assign a numerical value to each outcome so that: m=1 for reliable and clinically significant deterioration, m=2 for reliable deterioration, m=3 for stable, m=4 for reliable improvement, m=5 for reliable and clinically significant improvement. We assume a natural ordering of the outcomes in that 5 is considered a better outcome than 4 and so on. This model is expressed as:

$$Y_{ijk} = m \text{ if } \kappa_{m-1} < y^*_{ijk} \leq \kappa_m,\ \text{m} = 1,..,5 \qquad (1)$$

The threshold values are unknown and must be estimated from the data. It is not possible to identify both the constant term and all of the cut points so the constant term is excluded (Woolridge 2010). This threshold model relates the ordinal outcome to the underlying latent measure of mental health which is unobservable and will in principle be a continuous variable. Instead we observe mental health as measured by HoNOS scores which we subsequently interpret in terms of an ordered outcome.

Latent mental health y^*_{ijk} can then be described by the following equation:

$$y^*_{ijk} = x'_{ijk}\beta + u_k + v_{jk} + e_{ijk} \qquad (2)$$

where x_{ijk} is a vector of risk adjustment variables, v_{jk} is the patient random intercept and e_{ijk} is the random error for CRP i in patient j in provider k and has a zero mean and variance of one, $e_{ijk} \sim (0, 1)$. The provider effect is captured through u_k which is assumed to be random with a zero mean and constant variance, $u_k \sim (0, \sigma^2_u)$. A similar approach has previously been used to measure provider performance on patient outcomes in acute physical healthcare (Gutacker et al. 2013b).

2) Linear model

We also estimate the following linear model:

$$y_{ijk} = \alpha_{ijk} + x'_{ijk}\beta + u_k + v_{jk} + e_{ijk} \qquad (3)$$

with the follow-up HoNOS as the dependent variable y_{ijk} and the baseline HoNOS score included as an additional risk adjustment variable in the vector x_{ijk}.

Coefficients for the linear model can be interpreted as average partial effects; for covariates measured as dummy variables the coefficient represents the average effect or the change in the follow-up HoNOS score when the independent variable changes from zero to one. The coefficients on continuous variables such as the baseline HoNOS score and the IMD Income Deprivation variable can be interpreted in terms of marginal effects or the change in the follow-up HoNOS score arising from a one unit change in the continuous variable. The coefficients of the ordered probit model relate to the underlying latent mental health variable which is unobservable and not measured in any kind of natural units meaning the coefficients can only be interpreted as qualitative effects (Jones 2005).

4.5.2. Quantifying and comparing provider performance

Provider performance in both models is illustrated by ranking providers based on their relative impact on mental health status as measured by the random effect $\hat{u}_k$ which can be quantified using EB techniques. Additionally for the ordered probit model, provider performance is compared according to the probability of achieving a specific outcome (m=1, 2, 3, 4, 5), given an average set of risk-adjusters:

$$\text{Prob}\ (y_{ijk} = m|\ \bar{x} = \hat{u}_k = 0) = \Phi\ (\kappa_m - \bar{x}'_{ijk}\beta) - \Phi\ (\kappa_{m-1} - \bar{x}'_{ijk}\beta),\ \kappa_0 = -\infty,\ \kappa_5 = +\infty \qquad (4)$$

95% credible intervals are calculated around $\hat{u}_k$ based on their posterior distribution in order to compare departures of providers from the average.

The ordered probit model and associated EB estimates were estimated using the *gllamm* and *gllapred* commands (Rabe-Hesketh, Skrondal and Pickle 2004) and the hierarchical linear model and associated EB estimates were estimated using MLWiN 2.29 (Rabash et al. 2009) via the *runmlwin* command (Leckie and Charlton 2012). Both models were estimated in Stata 13.0. (StataCorp 2013).

4.5.3. Sensitivity Analyses

We conducted a number of sensitivity analyses to assess the robustness of the models to changes in variable construction and estimation sample. Firstly, we code the approximately 50% of observations with no information on MHA or CPA as missing instead of zero. We coded observations with missing information on MHA and CPA as zero as these activities are subject to regulation and scrutiny and we would expect that they would be recorded if they had taken place. Secondly, we estimate the three-level model excluding a provider that is an outlier on follow-up HoNOS scores. Thirdly, we include a number of variables at the provider-level to control for provider factors that may be associated with patient outcomes. These include provider size as measured by the number of available mental health beds, percentage occupancy of these beds, and whether the provider has FT status.

4.6. *Results*

4.6.1. Outcome variable

Following Parabiaghi et al. (2005) we calculated the RCI and clinical severity cut-off using the baseline HoNOS scores from the 1,663,894 observations comprising 583,138 patients treated by 57 NHS Mental Health Trusts that were clustered under the NTPS and had a HoNOS score recorded at the beginning (baseline) of a CRP. Based on having at least one HoNOS item ≥3, 63% of observations were classified as severe. Table 4.3 shows the descriptive statistics of the baseline HoNOS scores for the observations classified as severe and non-severe. This shows that observations classified as severe had an average total HoNOS score of 14 – over twice the average HoNOS score of 6 for observations classified as non-severe. Moreover, the maximum total HoNOS score of 24 for the non-severe group of observations was half of the maximum total HoNOS score of 48 for the observations classified as severe.

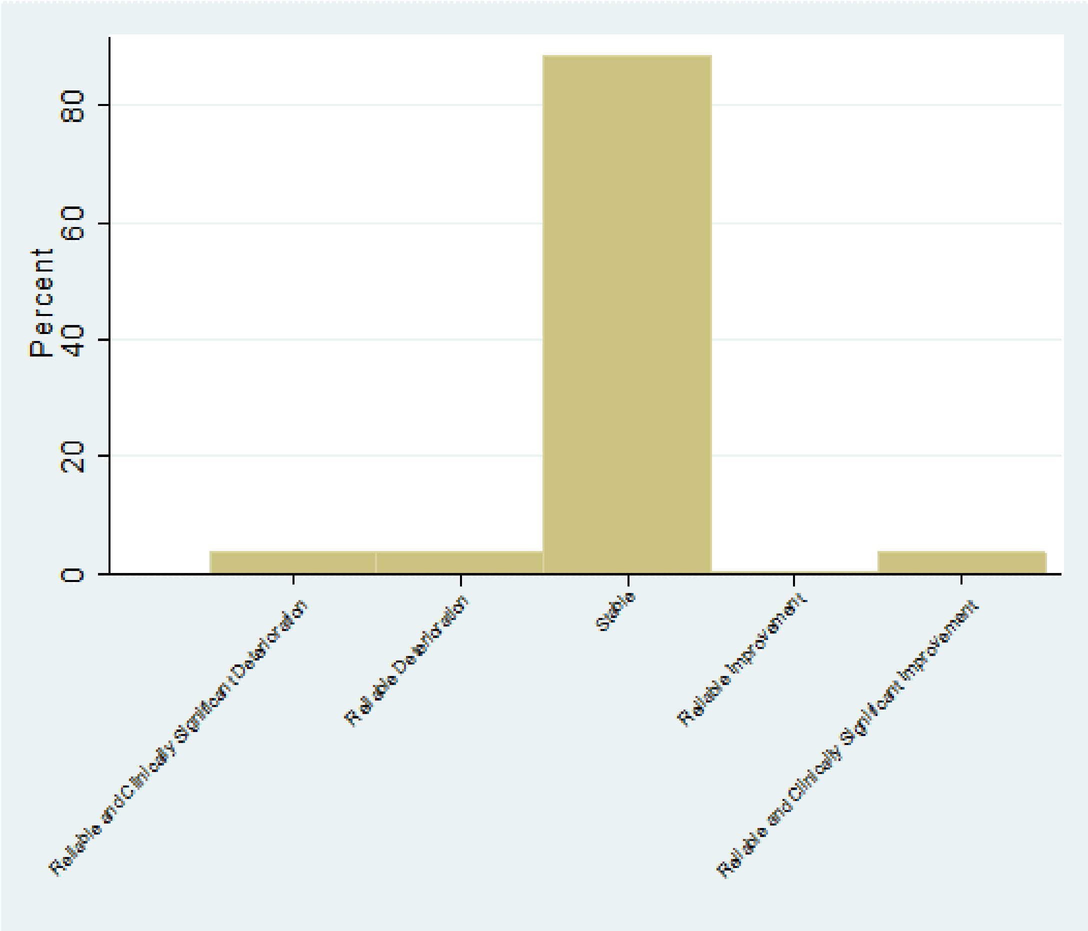
Table 4.3 Descriptive statistics of baseline Health of the Nation Outcome Scales (HoNOS) scores for severe and non-severe samples

	Obs	Mean	Standard Deviation	Min	Max
Non-severe	610,637	6.165	3.643	0	24
Severe	1,053,257	14.180	5.846	3	48

A Cronbach alpha of 0.7, which is relatively high, resulted in a RCI of 10 and a clinical cut-off of 9. This is comparable to Parabiaghi et al. (2011) who report a RCI of 8 and a clinical cut-off of 10. This indicates that a change between HoNOS baseline and follow-up scores of at least 10 is required for a change to be considered reliable (and not due to measurement error or chance) while a change in score of at least 9 is required for it to be considered clinically significant.

There was a considerable reduction in the number of observations with HoNOS scores recorded at both the beginning (baseline) and end (follow-up) of a CRP. This meant that the RCI and CScut-off were applied to only 342,288 observations encompassing 185,281 patients that had both baseline and follow-up HoNOS scores recorded. This resulted in a five-category HoNOS ordered outcome variable as shown in Figure 4.1. 88% of observations fall into "stable" category while approximately 4% record a "reliable deterioration", a "reliable and clinically significant deterioration" or a "reliable and clinically significant improvement". Less than 1% of observations have a "reliable improvement" based on this classification. The proportion of the sample categorized as "stable" is comparable to previous studies (Parabiaghi et al. 2005; Speak and Hay 2012) in which 92% of the sample was categorized as stable.

Figure 4.1 Health of the Nation Outcome Scales (HoNOS) ordered outcome variable



4.6.2. Descriptive statistics

Table 4.4 displays the descriptive statistics for the variables included in the analysis with the reference categories in brackets.

Table 4.4 Descriptive statistics

Variable [N=305,960]	Mean	Standard Deviation	Min	Max
HoNOS follow-up	11.406	6.480	0	48
HoNOS baseline	11.404	6.480	0	48
Ordered HoNOS	2.965	0.572	1	5
Married/civil partner	0.338	0.473	0	1
[White ethnicity]	0.892	0.310	0	1
Asian ethnicity	0.043	0.203	0	1
Black ethnicity	0.038	0.190	0	1

Other ethnicity	0.027	0.163	0	1
[Age category 1 (18-34)]	0.200	0.400	0	1
Age category 2 (35-47)	0.210	0.407	0	1
Age category 3 (48-61)	0.191	0.393	0	1
Age category 4 (62-78)	0.209	0.407	0	1
Age category 5 (79+)	0.190	0.392	0	1
Gender [Female]	0.434	0.496	0	1
Admitted under the MHA	0.154	0.361	0	1
Under CPA	0.465	0.499	0	1
[Cluster: Non-Psychotic Mild, Moderate, Severe (Clusters1-4)]	0.257	0.437	0	1
Cluster: Non-Psychotic Very Severe and Complex (Clusters 5-8)	0.139	0.346	0	1
Cluster: Psychosis First Episode (Cluster 10)	0.030	0.171	0	1
Cluster: Psychosis Ongoing or recurrent (Clusters 11-13)	0.231	0.421	0	1
Cluster: Psychosis Psychotic Crisis (Clusters 14-15)	0.044	0.205	0	1
Cluster: Psychosis Very Severe Engagement (Clusters 16-17)	0.019	0.136	0	1
Cluster: Organic Cognitive Impairment (Clusters 18-21)	0.254	0.435	0	1
Problems with Accommodation	0.137	0.344	0	1
Income Deprivation	18.505	11.915	0	77
[Year:2011/12]	0.189	0.392	0	1
Year: 2012/13	0.500	0.500	0	1
Year: 2011/12 and 2012/13	0.311	0.463	0	1
Provider-level variables [N=305,960]				
Foundation Trust (FT)	0.756	0.429	0	1
Mental Health Beds	564	253	40	1010
Mental Health Beds Occupancy (%)	88	5	64	100

The total HoNOS score at follow-up ranges from 0 to 48 with a mean of 11. In terms of the risk adjustment variables, the baseline total HoNOS score has a similar distribution to the follow-up total HoNOS score. Just over one-third of the sample is married or has a civil partner. The majority (90%) is of White ethnicity, while Asian, Black and Other ethnicities each account for 3-4% of the sample. Age is represented by five categories to capture any non-linearities in the relationship between mental health outcomes and age. The majority of observations are of female gender with males accounting for 44%. Around 15% of observations had an admission under the MHA before or at baseline while around 47% of observations were under CPA prior to or upon entry to the cluster.

The most common clusters in our sample relate to non-psychotic mild/moderate/severe (26%), organic cognitive impairment (25%) and psychosis ongoing or recurrent (23%). Clusters related to non-psychotic very severe and complex account for 14% while clusters for psychosis first episode, psychotic crisis and psychosis very severe engagement each account for 4% or less of the sample. 14% of the sample has problems with accommodation as recorded in baseline HoNOS scores. On average, the observations in our sample live in an area where just under 20% of the population experience income deprivation but this ranges from 0% to 77%. Observations include those with a CRP that: starts and ends in 2011/12 (19% of the sample); starts and ends in 2012/13 (50% of the sample); and starts in 2011/12 and ends in 2012/13 (31% of the sample).

In terms of provider-level variables, approximately three-quarters of the providers in our sample have FT status. The average number of mental health beds is 564 but there is considerable variation with this variable ranging from 40 to 1010. Average mental health bed occupancy is 88% - just above the recommended rate of 85% (Royal College of Psychiatrists 2011), with a minimum occupancy rate of 64% and a maximum of 100%.

4.6.3. Estimation results

Table 4.5 displays the estimation results of the ordered probit and linear models.

Table 4.5 Estimates of ordered probit and linear models

	Observations per group			
	Number of observations	Minimum	Average	Maximum
Level 3: Provider	57	2	5368	36981
Level 2: Person	163,611	1	1.9	67
Level 1: CRP	305,960			
	Ordered probit model		Linear model	
Log likelihood	-149096.96		935428.94	
Variable	**Coefficient**	**Standard Error**	**Coefficient**	**Standard Error**
HoNOS baseline			0.440	0.002***
Married/civil partner	0.031	0.006***	-0.391	0.024***
Asian ethnicity	-0.009	0.014	-0.086	0.056
Black ethnicity	-0.006	0.015	-0.249	0.061***
Other ethnicity	-0.012	0.017	-0.041	0.068
Age category 2 (35-47)	-0.006	0.009	0.236	0.035***
Age category 3 (48-61)	0.002	0.009	0.241	0.036***
Age category 4 (62-78)	0.042	0.010***	-0.348	0.039***
Age category 5 (79+)	0.049	0.012***	-0.359	0.045***
Gender	-0.030	0.006***	0.256	0.022***
Admitted under the MHA	-0.109	0.008***	0.813	0.034***
Under CPA	-0.026	0.006***	0.633	0.024***
Cluster: Non-Psychotic Very Severe and Complex (Clusters 5-8)	-0.065	0.009***	1.315	0.033***
Cluster: Psychosis First Episode (Cluster 10)	0.007	0.017	-0.279	0.062***
Cluster: Psychosis Ongoing or recurrent (Clusters 11-13)	0.033	0.008***	-0.412	0.030***
Cluster: Psychosis Psychotic Crisis (Clusters 14-15)	-0.129	0.014***	1.303	0.051***
Cluster: Psychosis Very Severe Engagement (Clusters 16-17)	-0.149	0.020***	1.734	0.075***
Cluster: Organic Cognitive Impairment (Clusters 18-21)	-0.022	0.010*	0.513	0.037***
Problems with Accommodation	0.211	0.008***	0.104	0.032**
Income Deprivation	-0.001	0.000***	0.016	0.001***
Year: 2012/13	-0.005	0.008	0.042	0.026
Year: 2011/12 and 2012/13	-0.012	0.008	-0.156	0.028***
Constant			5.622	0.116***
κ1	-1.881	0.014***		
κ2	-1.511	0.014***		
κ3	1.784	0.014***		
κ4	1.830	0.014***		
Random-effects Parameters				
Level 3	0.132	0.013***	0.570	0.119***

Level 2	0.044	0.004***	5.378	0.071***
Level 1			21.888	0.076***

*****p<0.001, **p<0.01, *p<0.05**

Variables with the same qualitative meaning have opposite signs due to the different dependent variables in the ordered probit and linear models. For the ordered HoNOS variable, the values are in ascending order which means a positive coefficient signifies a better outcome. In the linear model, the follow-up HoNOS variable is measured on a continuous scale from 1 (best) to 48 (worst) meaning that an increase signifies a worse outcome.

The results from both models show that variables associated with better mental health outcomes include married/civil partner, the two oldest age categories, allocation to the cluster for a first episode of psychosis (relative to non-psychotic mild/moderate/severe), and allocation to the cluster for ongoing or recurrent psychosis (relative to non-psychotic mild/moderate/severe). Additionally, in the linear model Black ethnicity is associated with a reduction in follow-up HoNOS scores of 0.25 points, thus positively associated with mental health outcomes. Variables associated with worse mental health outcomes in both models include male gender, admission under the MHA, having care co-ordinated under CPA, allocation to a cluster for 1) non-psychotic illness that is very severe and complex; 2) psychotic crisis; 3) psychosis very severe engagement; and 4) cognitive impairment (all relative to allocation to a cluster for non-psychotic mild/moderate/severe) and income deprivation. The linear model reveals that allocation to clusters for very severe and complex non-psychotic disorders and psychotic crisis is associated with an increased follow-up HoNOS of 1.3 points; while allocation to clusters for psychosis with very severe engagement is associated with an increase in follow-up HoNOS of almost 2 points. Moreover, in the linear model the baseline HoNOS score and age categories 2 and 3 variables are positively associated with the follow-up total HoNOS score and therefore associated with a worse mental health outcome. A one-unit increase in baseline HoNOS scores is associated with an increase of 0.44 points in follow-up HoNOS scores, while an age of 35 years or over is associated with an increase in the follow-up HoNOS score of around 0.24 points compared to an age of 18 to 34 years. The results for the variable capturing problems with accommodation were inconsistent between the two models, having a positive association with mental health outcomes in the ordered probit model but a negative association in the linear model.

The threshold values from the ordered probit model imply that a value of the latent mental health variable less than -1.881 corresponds to an outcome of “reliable and clinically significant deterioration”; a value between -1.881 and -1.511 corresponds to “reliable deterioration”; a value between -1.511 and 1.784 corresponds to “stable”; a value between 1.784 and 1.830 corresponds to “reliable improvement”; and a value above 1.830 corresponds to “reliable and clinically significant improvement”.

4.6.4. Provider performance

Around 11% of the residual variation in latent mental health unexplained by the risk adjustment variables in the ordered probit model lies at provider-level. Around 2% of the variation in the follow-up total HoNOS scores not accounted for by the risk adjustment variables in the linear model lies at provider-level. The EB estimates of this residual variation are plotted in Figure 4.2 for the ordered probit model and Figure 4.3 for the linear model with providers on the right performing better than those on the left in both figures. The figures show that a considerable number of providers (with intervals that do not encompass zero) perform better or worse than average. A larger degree of variation in the EB estimates from the linear model likely reflects the higher level of variation in the response variable.

Figure 4.2 Empirical Bayes (EB) estimates of provider-level residual variation for ordered probit model

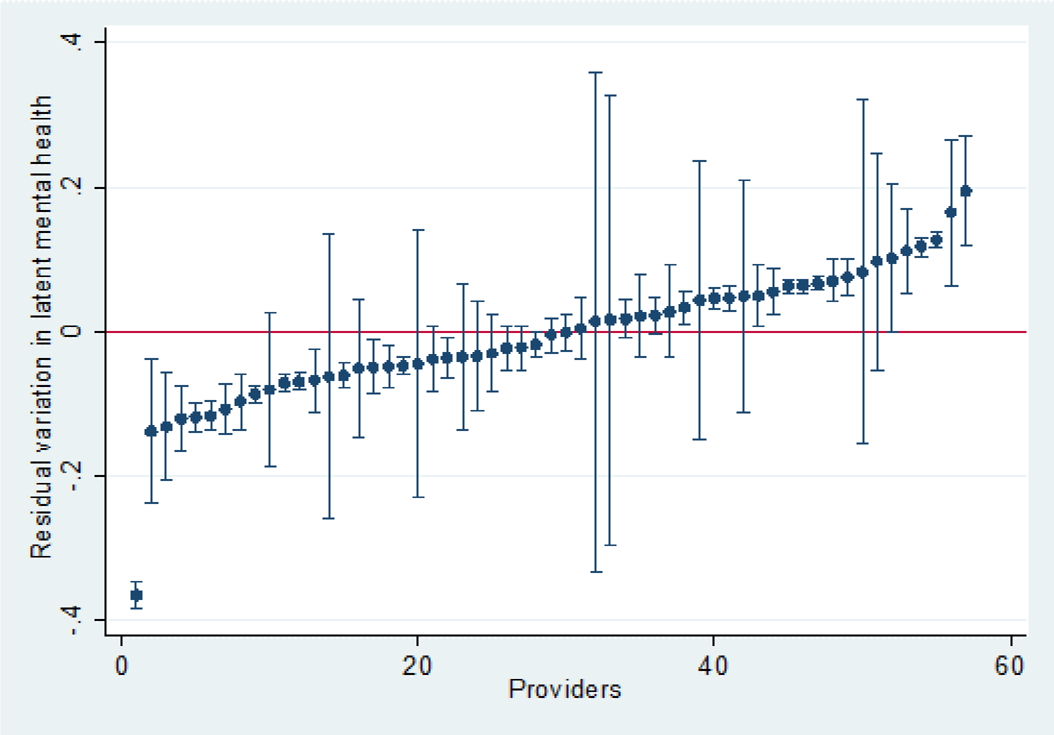
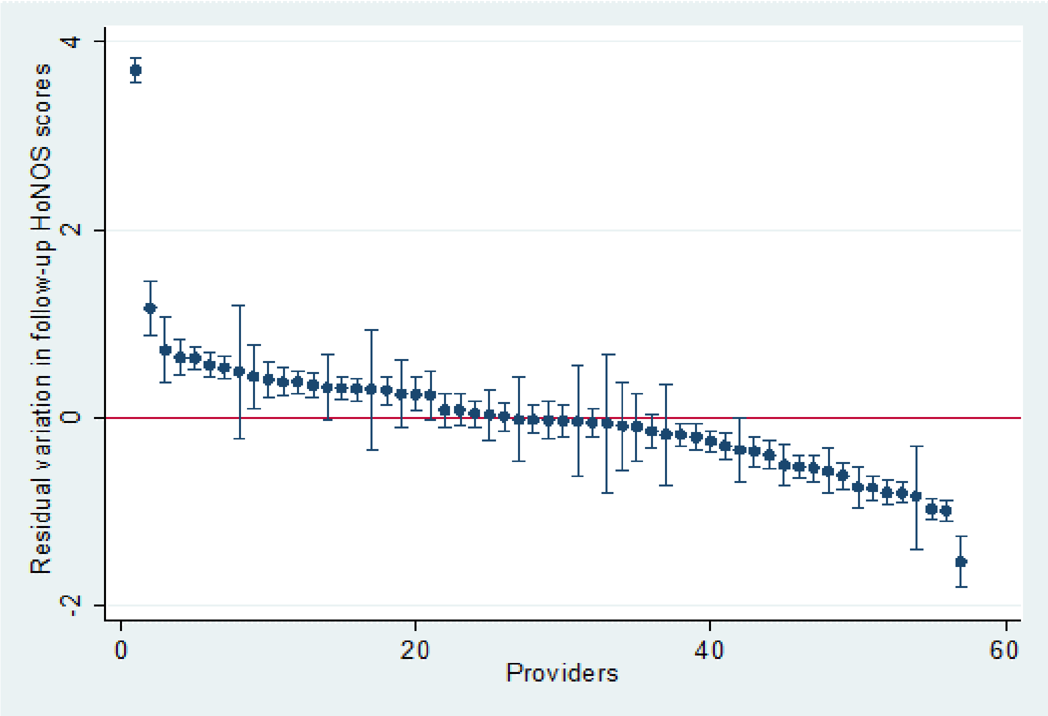
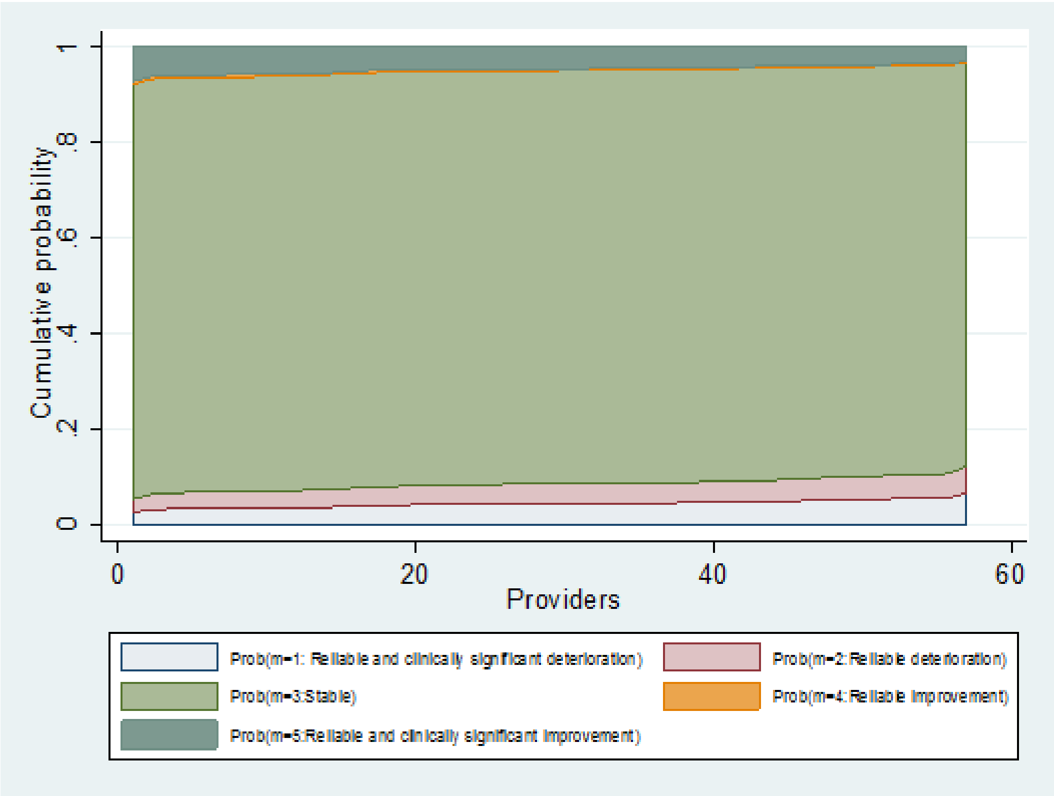


Figure 4.3 Empirical Bayes (EB) estimates of provider-level residual variation for linear model



For the model with the ordered outcome HoNOS variable, an alternative means of comparing provider performance is in terms of the cumulative probability of reporting a particular outcome for the average patient by provider. Figure 4.4 plots the cumulative probability of reporting an outcome (m=1,2,3,4,5) where providers on the left perform better than those on the right and we can see for example, that the probability of reporting the outcome of “reliable and clinically significant deterioration” (m=1) ranges from 2% to 9% while the probability of reporting the oucome of “reliable deterioration” (m=2) ranges from 5% to 16%.

Figure 4.4 Cumulative probability of reporting an outcome by provider



4.6.5. Sensitivity analyses

Observations with missing values of CPA or MHA were dropped from the estimation resulting in a reduced estimation sample of 144,063 observations with 92% of observations being subject to CPA and 33% admitted under the MHA. The results are robust to this coding change with the exception of CPA which has a positive association with mental health outcomes in both the ordered probit and linear models.

The large percentage (92%) of observations subject to CPA in this sensitivity analysis may have overestimated the effect.

The results of the linear model without the provider that emerged as an outlier in the EB estimates remain robust with the exception of year 2012/13 which became statistically significant at a 0.1% level.

Table 4.6 shows the estimation results with provider-level variables included.

Table 4.6 Estimates from sensitivity analysis including provider-level variables

	Observations per group			
	Number of observations	**Minimum**	**Average**	**Maximum**
Level 3: Provider	57	2	5367.7	36981
Level 2: Person:	163,611	1	1.9	67
Level 1: CRP:	305,960			
	Ordered Probit Model		**Linear Model**	
Log-likelihood	-149097.8		-935432.13	
Variable	**Coefficient**	**Standard Error**	**Coefficient**	**Standard Error**
HoNOS baseline			0.440	0.002***
Married/civil partner	0.031	0.006***	-0.391	0.024***
Asian ethnicity	-0.009	0.014	-0.086	0.056
Black ethnicity	-0.006	0.015	-0.249	0.061***
Other ethnicity	-0.012	0.017	-0.042	0.068***
Age category 2 (35-47)	-0.006	0.009	0.236	0.035***
Age category 3 (48-61)	0.002	0.009	0.241	0.036***
Age category 4 (62-78)	0.042	0.010***	-0.348	0.039***
Age category 5 (79+)	0.049	0.012***	-0.359	0.045***
Gender	-0.030	0.006***	0.256	0.022***
Admitted under Mental Health Act (MHA)	-0.109	0.008***	0.814	0.034***
Under Care Programme Approach (CPA)	-0.026	0.006***	0.632	0.024***
Cluster: Non-Psychotic Very Severe and Complex (Clusters 5-8)	-0.065	0.009***	1.315	0.033***
Cluster: Psychosis First Episode (Cluster 10)	0.006	0.017	-0.279	0.062***
Cluster: Psychosis Ongoing or recurrent (Clusters 11-13)	0.032	0.008***	-0.412	0.039***
Cluster: Psychosis Psychotic Crisis (Clusters 14-15)	-0.129	0.014***	1.303	0.051***

Cluster: Psychosis Very Severe Engagement (Clusters 16-17)	-0.149	0.020***	1.734	0.075***
Cluster: Organic Cognitive Impairment (Clusters 18-21)	-0.022	0.010*	0.514	0.037***
Problems with Accommodation	0.211	0.008***	0.104	0.032**
Income Deprivation	-0.001	0.000***	0.016	0.001***
Year: 2012/13	-0.006	0.008	0.042	0.026
Year: 2011/12 and 2012/13	-0.013	0.008	-0.157	0.028***
Foundation Trust (FT)	0.009	0.013	-0.129	0.231
Mental Health Beds (size)	0.000	0.000***	0.000	0.000
Bed Occupancy (%)	-0.004	0.001*	0.023	0.016
Constant			3.698	1.431*
κ_1	-2.086	0.131***		
κ_2	-1.716	0.131***		
κ_3	1.580	0.131***		
κ_4	1.626	0.131***		
Random-effects Parameters				
Level 3	0.131	0.013***	0.542	0.113***
Level 2	0.044	0.004***	5.379	0.071***
Level 1			21.887	0.076***

Results at the CRP- and patient-level are stable and similar to those presented in Table 4.5. Mental health beds have a positive association with mental health outcomes in the ordered probit model while mental health bed occupancy is associated with worse outcomes in the same model. The threshold values from the ordered probit model now imply that a value of the latent mental health variable less than -2.086 corresponds to an outcome of "reliable and clinically significant deterioration"; a value between -2.086 and -1.716 corresponds to "reliable deterioration"; a value between -1.716 and 1.580 corresponds to "stable"; a value between 1.580 and 1.626 corresponds to "reliable improvement"; and a value above 1.626 corresponds to "reliable and clinically significant improvement".

4.7. *Discussion and conclusions*

The introduction of the NTPS to mental health brings with it a future objective of linking some part of provider payment to patient outcomes. This presents a ripe opportunity to investigate the relationship between provider performance and patient outcomes in order to inform the policy context of the NTPS for mental health. The MHMDS offers a comprehensive repository of data related to the NTPS for mental health including patient cluster, outcomes, demographic and treatment variables. The availability of this data set provides an ideal resource to apply the RCSC concept in order to compare the performance of English mental health providers based on patient outcomes. The use of multilevel modelling is innovative compared to previous studies in mental health and enables us to utilise the hierarchical structure of this data set to make inferences about the influence of different levels on outcomes. Thus, it complements recent work on risk adjustment and comparisons of provider performance using PROMS for elective surgery in England (Gutacker et al. 2013b) and makes an important contribution to the relative lack of evidence in the area of mental health.

A stated goal of the NTPS for mental health is to link some element of payment to delivering particular outcomes, but this remains challenging and firm proposals on how this will be achieved have not yet been outlined (Department of Health Payment by Results team 2013b). Nevertheless, the use of a CROM based on HoNOS has been recommended for use (Department of Health Payment by Results team 2013b; Monitor and NHS England 2013a) and a calculation of the clinical significance of changes in total HoNOS scores such as that employed in this analysis could be used to evaluate changes in outcomes arising from treatment (Department of Health Payment by Results team 2013b; Speak and Hay 2012). Our findings suggest that it is feasible for providers to achieve clinically and statistically significant changes in outcomes during CRPs and there is potential to improve outcomes by linking payment to patient needs. A part of provider's payment could then be made contingent on achieving certain benchmarks measured by changes in HoNOS scores relating to clinical improvement. This would mean that providers that are in a position to achieve better outcomes as highlighted by this analysis would stand to gain financially if payment were linked to outcomes in this way. However, the potential unintended consequences of using a CROM as part of a payment system should be guarded against as clinicians may have incentives to 'game' the recording of outcomes for financial gain. A policy consideration may include ways in which data quality could be audited, although this may be difficult to do since clinical judgement underpins the outcome assessment.

While the proportion of the sample categorized as “stable” may appear high (88%), it is comparable to previous studies (Parabiaghi et al. 2005; Speak and Hay 2012) in which 92% of the sample was categorized as stable. Moreover, for some diagnoses of severe mental illness such as schizophrenia, exhibiting stability can be viewed as a favourable outcome as research has indicated that for example, it is rare to observe significant change in schizophrenia illness among patients treated on an outpatient basis (Miles et al. 2014). However, the lack of diagnostic information in the MHMDS prevents us from stratifying the different outcomes by diagnosis. Given the chronic nature of mental illness, the large proportion of patients with a stable outcome may also signal that longer time periods may be needed to realize improvements in outcomes. The Cronbach alpha of 0.7 for the baseline HoNOS scores showed a reasonable level of consistency in clinician ratings. A higher Cronbach alpha and narrower standard deviations would improve the model’s sensitivity to change and result in more observations being classified as reliably and clinically significant. Nevertheless, it has been noted that Cronbach’s alpha tends to underestimate reliability and is affected by the number of items in a given measure (Speak and Hay 2012) so a Cronbach alpha of 0.7 should not necessarily be viewed negatively.

We find that demographic variables describing age and married/civil partner have a positive association with our outcome variables. Older people may experience better outcomes as the positive symptoms of some serious mental illnesses such as schizophrenia tend to diminish with age (Hendryx, Dyck and Srebnik 1999) while social support may also improve outcome. Our finding that CPA is associated with worse outcomes may reflect that those on CPA have an enduring and severe mental disorder and thus this variable may be picking up aspects of illness severity that we are unable to control for due to the absence of diagnosis variables in our model. The positive association between HoNOS outcomes and problems with accommodation may suggest that this variable does not adequately capture issues of unsettled housing or homelessness. Patients who are admitted under the MHA are likely to be more severely ill which may reduce the probability of experiencing a better outcome although involuntary status may be more representative of high acuity at admission rather than greater severity of illness over time (Hermann, Rollins and Chan 2007). The sensitivity analysis including provider variables revealed a positive association between provider size as measured by number of mental health beds and outcomes. This supports previous work that found a positive relationship between mental health volume and performance that parallels the medical and surgical literature (Druss et al. 2004).

While it is important to control for patient need as reflected in cluster assignment, it is somewhat difficult to interpret the results pertaining to the cluster variables in this analysis as the non-psychotic mild/moderate/severe clusters are not necessarily suitable benchmarks for clusters relating to psychosis or cognitive impairment. Ideally, we would like to compare clusters within the broad categories of non-psychotic, psychotic and organic and future work needs to incorporate more suitable benchmarks for cluster comparisons.

A limitation to our set of risk adjustment variables is the inability to include information on primary diagnosis due to poor coding of this variable in the MHMDS. This may be due to the difficulty of making a definitive diagnosis in mental healthcare (Timimi 2014); a reluctance on the part of clinicians to label patients as suffering from a particular mental disorder – in part due to stigma associated with mental illness (Ben-Zeev, Young and Corrigan 2010; Sartorius 2002; Timimi 2014); and that diagnosis may not necessarily inform treatment decisions that influences outcomes (Timimi 2014). Nevertheless, "the proportion of users who have a valid ICD10 recorded" is among the quality indicators recommended for use by commissioners and providers in drawing up contracts to enable the evaluation of the quality and outcomes of services delivered by providers (Department of Health Payment by Results team 2013b). The uptake of this indicator could lead to improved coding of diagnosis in the MHMDS and it would be important to test diagnosis as a risk adjustment variable in future work should this become feasible.

Our risk adjustment methodology is also limited by the inability to include information on the clinical severity of mental illness and the goal of treatment which would subsequently affect the outcome. For example, two patients may have the same risk profile but the current goal for one of them is acute management of their symptoms while for the other the current goal is rehabilitation and functional gain.

Another limitation of this work is that it does not account for missing data which is an issue for the HoNOS records in the MHMDS. As highlighted in Section 4.4.1, follow-up HoNOS scores may be more challenging to record than baseline scores and this results in missing data and an inability to measure outcomes. Concerns regarding compliance with mental health outcome reporting requirements have been documented internationally (Trauer 2010c). The reasons for low rates of completion include: poor data quality, lack of adequate IT support, limited systematic training of staff, lack of clinician engagement due to time pressures as a result of high caseloads, and a perceived lack of benefits of outcome measurement. Commonly, outcome measurement is seen as a managerial or administrative exercise which has little direct impact on frontline services (Trauer 2010c). This has led to a recognition of the need to differentiate the use of outcome measures for research and audit versus use for direct clinical care. Promoting outcome measures as a means of shared decision making rather than as tools primarily used for audit or performance review will help to promote greater clinician engagement and uptake of outcome measures (Wolpert 2014). This will ensure that the outcome data underpinning funding and reimbursement systems is of good quality (Trauer 2010a) – an important contributer to the success of such systems.

Missing data can be considered missing completely at random (MCAR) missing at random (MAR) or missing not at random (MNAR) (Bartlett and Carpenter 2013). Data is MCAR if the probability of observations being missing is independent of the variables of interest in the model. Data can be considered MAR when the chance of observing the variable for an observation may depend on the underlying value of that variable, but given the observed data, this association no longer holds. For data to be MNAR, the probability of the value of a variable being missing for a particular observation depends on the observation's underlying value of that variable. For the MHMDS data, HoNOS may be MNAR if patients with more severe mental health problems are more likely to drop out of care and be lost to follow-up. As HoNOS is a clinician-rated measure, missing follow-up HoNOS scores may also be indicative of the quality of provider coding practices (which in turn may be associated with provider performance) and be considered MNAR. There is some evidence to suggest that routine use of outcome measures in mental healthcare is associated with aspects of provider performance in terms of reduced psychiatric inpatient admissions (Slade et al. 2006). In acute care, inferences about relative provider performance are sensitive to the assumptions made about the reasons for missing data on PROMs (Gomes et al. 2015).

While assumptions regarding the missing data mechanism can be ascertained from the data under study, these assumptions cannot be definitively verified from the observed data (Bartlett and Carpenter 2013). The issue of missing data could be addressed using multiple imputation which involves the generation of multiple imputed data sets which, conditional on the missingness assumption, correctly reflect the distribution of the missing data given the observed data. The model of interest is then fitted to each imputed data set and the results of all the iterations are averaged – for example using Rubin's rules (Rubin 1987) – for the final inference.

The issues of data completeness and data quality highlighted above may be partly explained by our usage of data for 2011/12 and 2012/13 which cover the initial years of the development and implementation of the NTPS in mental health. The allocation of patients to care clusters commenced in 2011 and the mandatory use of the clusters as the basis for contracting mental health services for working-age and older adults was introduced only in 2012.

Finally, we do not take account of provider costs associated with CRPs and it may be that providers associated with better outcomes are also associated with higher costs. As the NTPS for mental health introduces incentives for providers to control costs, it may become more challenging for providers to maintain good performance in relation to outcomes. Chapter 5 considers potential trade-offs between provider costs and outcomes.

This research will be of interest to policymakers not only in England but also further afield. We make an important contribution to the small evidence base on the performance assessment of mental healthcare providers in relation to outcomes and provide evidence to inform the continual refinement of the NTPS in mental healthcare. Previous international attempts to develop psychiatric classification systems for the purposes of provider benchmarking and reimbursement have achieved varying degrees of success (Mason and Goddard 2009) and this research adds to the existing evidence base. Our results suggest that some providers are more likely to achieve better outcomes for patients with an average set of risk-factors. This implies that if the objective of linking some element of provider payment to outcomes is realised, these providers will stand to benefit from the new financial regime, whilst others may lose.

Chapter 5. Investigating the relationship between costs and outcomes for English mental health providers: A bivariate multilevel regression analysis

5.1. *Introduction*

The aim of this chapter is to investigate the relationship between costs and outcomes for mental health providers in England to ascertain if incentives to control costs provided by the new payment system can be achieved without negatively affecting patient outcomes. As outlined in Section 1.2.2 in Chapter 1, a fixed national price can encourage providers to control costs in order to avoid financial losses and create surpluses. However, there is a risk that cost control may be achieved by reducing care quality. We estimate a multilevel bivariate model with costs and outcomes as responses and include a comprehensive set of risk adjustment covariates encompassing sociodemographic, need and treatment variables. We calculate the correlation in residual variation in costs and outcomes at the provider-level and plot the pairwise relationship between residual costs and outcomes for the providers in our sample.

Chapters 3 and 4 have investigated variations in costs and outcomes across English mental health providers separately and shown evidence of variations in performance as reflected in above-average residual costs and below-average residual outcomes. These differentials in performance may be due to not taking account of the relationship between costs and outcomes. Providers with above-average costs may be associated with above-average outcomes and vice versa. Examining the relationship between costs and outcomes is challenging due to endogeneity in the form of reverse causality: providers who spend more may produce better outcomes, but outcomes may also drive costs as patients with worse outcomes may consume more resources. A potential way to control for this endogeneity is to use an instrumental variable approach but data on a suitable instrumental variable is lacking in this context. Therefore, we do not attempt to estimate the causal relationship between costs and outcomes; rather we investigate the correlation between the residual variation in costs and outcomes for providers, after controlling for a range of risk adjustment variables.

We contribute to existing evidence in several ways. To the best of our knowledge, this chapter is the first to use a multilevel bivariate model to examine both mental health cost and outcome responses separately and simultaneously and calculate the correlation in residual variation between two responses. Our use of multilevel methods allows us to isolate the residual variation in costs and outcomes that can be attributed to providers while the estimation of a provider effect gives a quantifiable measure of provider performance. Moreover, while previous studies in mental health have used data with limited geographical or provider samples, we use a nationally-representative data set that contains data for all specialist mental health providers in England. This means that we can examine costs and quality for both admitted and non-admitted care and are not constrained to just one care setting as in several previous studies.

5.2. *Literature Review*

5.2.1. Review of literature on the relationship between costs and quality in mental healthcare

We searched a number of databases including PubMed, EconLit, Embase, Ovid MEDLINE, PsychINFO with the following search terms: “mental health”, “psychiatry”, “costs”, “expenditures”, “outcomes”, “performance” and “efficiency”. We found just four studies (Dickey and Normand 2004; Haas et al. 2013; Hendryx 2008; Schulz, Greenley and Peterson 1983) that examined the relationship between costs and quality in mental healthcare. The majority of the studies are based in the US (Dickey and Normand 2004; Hendryx 2008; Schulz, Greenley and Peterson 1983), with the exception of one study from Germany (Haas et al. 2013). These studies are summarised in Table 5.1.

Table 5.1 Literature on the relationship between costs and quality in mental healthcare

Sample	Dependent variable	Independent variables	Method	Results
(Schulz, Greenley and Peterson 1983)				
13 inpatient acute psychiatric units located in one US state: 6 in community general hospitals and 7 in public county-controlled specialty hospitals.	Cost per stay (covering only direct expenditures); staff's perception of the relative quality of the unit compared with other units.	Environmental, patient, institution, professional, and management characteristics.	Modified case approach, comparing unit characteristics across four quality and cost outcome categories: higher quality and lower cost; higher quality and higher cost; lower quality and lower cost; lower quality and higher cost.	The 13 units were almost equally distributed between the four groups. The four groups appeared to differ according to management characteristics.
(Dickey and Normand 2004)				
Patients who had visited one of 8 psychiatric emergency screening teams (EST) in one US state.	Care was labelled "better" ("poorer"): if the mean monthly medication dose was within (above) the recommended guideline dose range and patients with a record of substance abuse received (did not receive) treatment.	Mental health costs in terms of health benefits paid for by state and federal government through the Medicare and Medicaid programs; Outcomes measured by self-reported health-related quality of life, psychiatric and substance use problems and medication side effects.	Descriptive summaries of cost data. Statistical analysis of the socio-demographic and outcome data.	Patients in the "poorer care" group had higher treatment expenditures and more side effects. No statistical difference between the two groups was found for self-reported problems and mental health-related quality of life.

(Hendryx 2008)				
All US states as well as Washington, D.C.	21 performance measures covering: 1) access to information and services; 2) availability of recovery supports; 3) social circumstances; 4) readmission rates; 5) patient-reported experience and outcome measures; 6) inpatient hospitalisations; 7) morbidity; 8) mortality (suicide rates); 9) forensic mental health; 10) workforce development.	Mental health expenditures on a *per capita* and *per client* basis; mean state per capita income; severity of illness.	Correlation analysis; linear multiple regression analysis, hierarchical models using generalized estimating equations with a robust variance estimation.	After adjusting for state income and illness severity, no statistical relationship was found for 17 of the 21 performance measures.
(Haas et al. 2013)				
101 patients from a university teaching hospital in Berlin with a main diagnosis of somatoform pain disorder according to ICD-10.	Change in the Mental Health Component of the Short Form-8 (MCS-8).	Mental health costs; patient characteristics; sociodemographic variables; pain-related variables; comorbidities; and subjective illness attribution.	Minimal clinical important difference (MCID); the omitted variable version of Hausman test; OLS regression.	A trade-off between costs and outcome was found only for patients without or with only minor somatic illness attribution (77% of the sample).

The studies used data on patients treated in inpatient settings only (Haas et al. 2013; Schulz, Greenley and Peterson 1983) or on an outpatient basis with no psychiatric inpatient episodes during the study period (Dickey and Normand 2004). In contrast, Hendryx (2008) conducted a state-level analysis that included all US states as well as Washington, D.C.

All four studies used some measure of quality of care as the main variable(s) of interest. Independent variables included environmental, patient, institution, professional, and management characteristics (Haas et al. 2013; Schulz, Greenley and Peterson 1983) as well as costs (Dickey and Normand 2004; Haas et al. 2013), and expenditures on mental health (Hendryx 2008), outcomes (Dickey and Normand 2004), and mean state per capita income and severity of illness (Hendryx 2008). The studies employed various methods including primarily descriptive as well as statistical depending on the sample size. Statistical analyses included correlation analyses and multilevel modelling (Hendryx 2008) as well as linear regression (Haas et al. 2013; Hendryx 2008).

The studies found some evidence of a trade-off between costs or expenditure and outcomes but this is restricted to certain measures of quality or patient populations. Dickey and Normand (2004) found that patients associated with poorer care had higher psychiatric unadjusted treatment expenditures and higher pharmacy costs compared to their "better care" counterparts but medical care expenditures were comparable between the two groups. Hendryx (2008) found that higher spending was associated with better quality of care in terms of access and lower rates of incarceration. Haas et al. (2013) found a trade-off between costs and outcome for patients without or with only minor somatic illness attribution (77% of the sample).

The studies were characterized by several limitations including a small sample size (Dickey and Normand 2004; Haas et al. 2013; Hendryx 2008; Schulz, Greenley and Peterson 1983) which limits the analytical methods that could be employed and generalizability of study findings; and data limitations, for example the data sourced from medical records could be considered incomplete and not include all relevant variables (Dickey and Normand 2004); and performance measures collected across geographical entities (i.e. states in the US) may be inconsistently collected and reported (Hendryx 2008).

5.2.2. Review of literature on the relationship between costs and quality in acute physical healthcare

We also identified a number of studies (Carey and Burgess 1999; Gutacker et al. 2013a; Hvenegaard et al. 2011; Schreyogg and Stargardt 2010; Weech-Maldonado, Shea and Mor 2006) that examined the relationship between cost and quality in non-mental healthcare through a separate literature search. These studies are summarised in Table 5.2.

Table 5.2 Literature on the relationship between costs and quality in acute physical healthcare

Sample	Dependent variable	Independent variables	Method	Results
(Carey and Burgess 1999)				
137 non-psychiatric VA hospitals for the six fiscal years of 1988 to 1993.	Annual total variable cost, excluding the cost of physicians and dentists and the costs of long-term care.	Risk-adjusted mortality within 30 days of discharge and readmission rates within 14 days of discharge; outpatient follow-up (patients not seen) within 30 days after inpatient discharge; output measures; input prices; number of beds; and teaching status. Instrumental variables: lagged measures of quality.	OLS and Two Stage Least Squares (2SLS) equations estimated for each year of data separately.	A positive relationship between cost and the measures of quality. For mortality and readmission, this is likely a result of inadequate risk adjustment with quality measures also controlling for severity.
(Weech-Maldonado, Shea and Mor 2006)				

749 nursing homes in five US states. The nursing homes all participated in the Centers for Medicare and Medicaid Services' Multi-State Casemix and Quality Demonstration in 1996.	Log of a facility's total patient care costs.	Pressure ulcers worsening (physical outcome); mood decline (psychosocial outcome); casemix; output measures; input prices; occupancy rate; and measures of competition. Instrumental variables: cost and quality county-level variables associated with demand of nursing home care.	Weighted least squares with weights equal to the inverse of the square root of the total number of facility residents.	Pressure ulcers: increasing (decreasing) costs at the lower (higher) range of quality. Mood decline: a relatively flat curve at the lower range of quality but increasing costs after a threshold.
(Schreyogg and Stargardt 2010)				
35,279 patients from 115 VHA hospitals with an index admission during which a primary diagnosis of AMI was made during the years 2000-2006.	Individual level costs incurred during the inpatient stay.	Risk-adjusted mortality and readmission rates one year after discharge for the index hospitalization; casemix; number of beds; number of patients treated; teaching status. Instrumental variables: Medicare Wage Index and general overhead costs per day at the hospital level.	Two-stage model: 1) a generalized linear mixed model with a gamma distribution and a log-link; 2) multilevel random-effects proportional hazard models for mortality and readmission including actual costs and residuals from the first-stage.	A highly significant negative association between costs and quality (mortality and readmission conditional on not dying).

(Hvenegaard et al. 2011)				
All (3,754) patients admitted for vascular surgery in six (of eight) Danish vascular departments.	Individual level costs incurred during the inpatient stay.	Risk-adjusted mortality within 30 days of discharge; wound complications; patient demographic, casemix, and treatment variables.	Fixed effect models for costs (linear) and quality (logistic). Provider cost and quality effects measured as the difference between observed and expected cost and quality. Providers ranked according to a cost-quality ratio (provider cost effect divided by the quality effect).	Lower costs were associated with higher mortality with some evidence of a U-shaped relationship between costs and mortality but no clear association between costs and wound complications.
(Gutacker et al. 2013a)				
Individual-level data for four surgical procedures (hip replacement, knee replacement, varicose vein and groin hernia) from a minimum of 125 (for varicose veins) to a maximum of 147 (for	Individual level costs incurred during the inpatient stay.	PROMs; patient demographic; casemix and treatment variables; number of patients treated; teaching status; provider specialisation.	Multilevel linear model with costs on natural scale. Provider random effects were used to rank providers in terms of cost-containment performance with and without	A non-linear association between health outcomes and risk-adjusted costs was found for hip replacement surgery. Controlling for quality resulted in a significant improvement in relative cost performance

hernia) hospitals.			quality controls.	for some hospitals.

A common feature of these studies is that they use administrative data to study the relationship between costs and outcomes. The dependent variable in all studies is cost but studies vary in what measure of cost is used. In two studies (Carey and Burgess 1999; Weech-Maldonado, Shea and Mor 2006) cost is measured at the aggregate provider-level while the other three studies (Gutacker et al. 2013a; Hvenegaard et al. 2011; Schreyogg and Stargardt 2010) use individual level costs.

The most common measures of quality used are risk-adjusted mortality (Carey and Burgess 1999; Hvenegaard et al. 2011; Schreyogg and Stargardt 2010) and readmission rates (Carey and Burgess 1999; Schreyogg and Stargardt 2010). While two studies use mortality within 30 days of discharge (Carey and Burgess 1999; Hvenegaard et al. 2011), another (Schreyogg and Stargardt 2010) used mortality one year after discharge for the index hospitalization. Readmission rates also had different time frames: within 14 days of discharge (Carey and Burgess 1999) and within one year of discharge (Schreyogg and Stargardt 2010). Other quality measures included PROMs (Gutacker et al. 2013a); outpatient follow-up within 30 days after inpatient discharge (Carey and Burgess 1999); wound complications (Hvenegaard et al. 2011); and pressure ulcers worsening (physical outcome) and mood decline (psychosocial outcome) (Weech-Maldonado, Shea and Mor 2006).

Given the concerns about endogeneity between costs and quality measures, a number of studies (Carey and Burgess 1999; Schreyogg and Stargardt 2010; Weech-Maldonado, Shea and Mor 2006) adopted an instrumental variable approach. The choice of instrumental variables differ across studies and include lagged measures of quality (Carey and Burgess 1999), county-level variables associated with demand of nursing home care (Weech-Maldonado, Shea and Mor 2006) and the Medicare Wage Index and general overhead costs per day at the hospital level (Schreyogg and Stargardt 2010). Two studies (Gutacker et al. 2013a; Schreyogg and Stargardt 2010) utilised a multilevel modelling approach with patients at level one nested in providers at level two. Provider effects - estimated either as fixed (Hvenegaard et al. 2011) or random (Gutacker et al. 2013a) effects - were used to make inferences about performance with respect to cost and quality. Hvenegaard et al. (2011) constructed provider cost and quality effects as the difference between observed and expected cost or quality and subsequently divided the provider cost effect by the quality effect to form a cost-quality ratio which was used to rank providers. Gutacker et al. (2013a) used the provider effect to rank providers in terms of their performance in terms of cost-containment with and without quality controls.

Findings regarding the relationship between cost and quality were not consistent across the studies with Carey and Burgess (1999) reporting a positive relationship and Schreyogg and Stargardt (2010) a negative association. The remaining studies (Gutacker et al. 2013a; Hvenegaard et al. 2011; Weech-Maldonado, Shea and Mor 2006) found some evidence of a non-linear relationship between costs and quality with the pattern of the relationship dependent on the quality indicator.

A rather innovative study (Hauck and Street 2006) analysed the performance of English health authorities using a multivariate multilevel analysis. The use of a multivariate analysis enabled the investigation of 13 different objectives covering the four NHS performance domains “health outcomes”, “clinical quality”, “access” and “efficiency” simultaneously. The performance indicators included mortality, limiting long standing illness, emergency admissions, waiting times, accessibility to general practitioners (GPs), number of elective surgery episodes, day case rate, maternity costs, and psychiatry costs. A number of socioeconomic variables were included in the model to control for factors associated with population need and utilisation of healthcare. The study used multilevel models to examine data for electoral wards at level one nested within health authorities at level two and consider the correlation in performance across indicators. The results revealed that the majority of performance indicators had positive significant coefficients implying that areas with worse socioeconomic conditions were likely to have worse performance than expected given basic age-sex standardisation. Focusing on psychiatric costs in particular, significant positive correlations were found for the following variables: limiting long-standing illness for ages 0-74; waiting time for radiotherapy; accessibility to GPs; and maternity costs while a negative significant correlation was found for deaths following hospital surgery. Furthermore, 34% of the variation in performance for psychiatric costs is attributable to health authorities.

A more recent study (Gutacker and Street 2015) also used multivariate multilevel analysis to examine the performance of English NHS providers of elective hip replacement surgery. Four dimensions of performance were investigated: waiting time between referral and treatment; LOS; unplanned readmission within 30 days of discharge; and patient-reported post-operative outcomes. The study found that providers performing well on one of these dimensions tended to perform well on all four dimensions and providers who performed well were usually privately owned with a limited concentration on providing hip replacements.

5.2.3. Summary of previous literature in mental and physical healthcare

Previous studies investigating the relationship between costs and quality in both physical and mental healthcare have revealed that this is a challenging endeavour. Particular challenges relate to the availability of adequate measures of quality, small sample sizes and the endogenous relationship between costs and quality. Regarding the latter, a number of studies (Carey and Burgess 1999; Schreyogg and Stargardt 2010; Weech-Maldonado, Shea and Mor 2006) have used instrumental variables in order to consistently estimate the causal relationship. Nevertheless other studies (Gutacker et al. 2013a; Hvenegaard et al. 2011) have highlighted the inherent difficulty of addressing endogeneity including the limited availability of suitable instrumental variables. Given the challenge of finding suitable instrumental variables, we avoid the causal identification problem and motivated by a similar methodology used in previous studies (Gutacker and Street 2015; Hauck and Street 2006), we analyse costs and outcomes using two separate equations and allow for a correlation in responses. As in previous studies (Gutacker et al. 2013a; Haas et al. 2013) we measure quality in terms of an outcome measure – HoNOS. Drawing on the literature in acute physical healthcare (Gutacker et al. 2013a; Gutacker and Street 2015; Hauck and Street 2006; Schreyogg and Stargardt 2010) we use a multilevel model which allows us to examine the correlation in residual responses at provider-level to provide insight into the relationship between costs and outcomes and if a potential trade-off exists. The use of a large, nationally representative data set with individual-level data moves us beyond previous studies (Dickey and Normand 2004; Haas et al. 2013; Schulz, Greenley and Peterson 1983) in mental healthcare that were constrained by small sample sizes.

5.3. *Methods*

5.3.1. Unit of analysis

As in Chapters 3 and 4 the CRP forms the unit of observation in this analysis.

5.3.2. Multilevel modelling

As the MHMDS is characterised by a hierarchical structure with responses constituting the lowest level of the hierarchy nested within CRPs at level 2 nested within patients at level 3 who are in turn nested within providers at level 4, we utilise a multilevel modelling approach.

We estimate the following bivariate model with two response variables: costs y_{1ijk} and outcomes y_{2ijk}:

$$\begin{cases} y_{1ijk} = \alpha_1 + \beta_1 X_{1ijk} + u_{1k} + v_{1jk} + \varepsilon_{1ijk} \\ y_{2ijk} = \alpha_2 + \beta_2 X_{2ijk} + u_{2k} + v_{2jk} + \varepsilon_{2ijk} \end{cases} \quad (1)$$

$$\begin{pmatrix} u_{1k} \\ u_{2k} \end{pmatrix} \sim N(0, \Omega_u) : \begin{pmatrix} \sigma^2{}_{u1} & \\ \sigma_{u1u2} & \sigma^2{}_{u2} \end{pmatrix} \quad (2)$$

X_{1ijk} represents a vector of risk adjustment covariates for the cost equation while X_{2ijk} reflects a vector of risk adjustment covariates for the outcomes equation. The provider-level random intercepts for costs and outcomes are represented by u_{1k} and u_{2k} respectively. The individual-level random intercepts for each response are denoted by v_{1jk} and v_{2jk}; while ε_{1ijk} and ε_{2ijk} signify the error terms at the CRP-level for each response. The provider-level effects, u_{1k} and u_{2k} are both assumed to follow a bivariate normal distribution with zero mean and covariance matrix Ω_u. Our interest lies in the correlation between the residual variation in y_{1ijk} and y_{2ijk} at the provider-level which can be calculated as $r_{(x,y)} = \frac{\sigma_{u1u2}}{\sqrt{\sigma^2{}_{u1}\sigma^2{}_{u2}}}$.

Our cost response variable y_{1ijk} is modelled using a log-linear model and our outcome response variable y_{2ijk} using a linear distribution. The multilevel estimates are statistically efficient even if some observations have missing data for either response under the assumption that data is missing at random (Rabash et al. 2012).

The coefficients for the log of total cost can be interpreted in terms of a percentage change in the geometric mean of total cost calculated as $(\exp(\beta) - 1)*100$. For the majority of covariates measured as dummy variables, this is the percentage change in the geometric mean resulting from a change in the variable from zero to one. For the continuous IMD Income Deprivation variable, the coefficient can be interpreted as the change in the geometric mean in total cost resulting from a one unit change in this variable. Coefficients for the total follow-up HoNOS score can be interpreted as average partial effects; for covariates measured as dummy variables the coefficient represents the average effect or the change in the follow-up HoNOS score when the independent variable changes from zero to one. The coefficients on continuous variables such as the baseline HoNOS score and the IMD Income Deprivation variable can be interpreted in terms of marginal effects or the change in the follow-up HoNOS score arising from a one unit change in the continuous variable.

The model is estimated using restrictive iterative generalized least squares (RIGLS) which is equivalent to restricted maximum likelihood (Goldstein 1989) in MLwiN 2.29 (Rabash et al. 2009) using the *runmlwin* command (Leckie and Charlton 2012) in Stata 13.0 (StataCorp 2013).

5.3.3. Sensitivity Analysis

We performed a sensitivity analysis by excluding a provider that is an outlier on follow-up HoNOS scores.

5.4. *Data*

The data and variables used for the analysis are described fully in Chapters 3 and 4. More specifically, mental health costs are constructed using Reference Cost data as described in Section 3.3.1. in Chapter 3 and our dependent cost variable measures the log of total cost associated with a CRP. Mental health outcomes are measured using follow-up HoNOS scores; HoNOS is fully described in Section 4.4.1 of Chapter 4. In terms of risk adjustment variables, the baseline total HoNOS score is included as a risk adjustment variable for the follow-up total HoNOS score response variable following our earlier findings and those of previous studies (Dow, Boaz and Thornton 2001; Hendryx, Dyck and Srebnik 1999; Kramer et al. 2001; Rosen et al. 2010) that show that baseline outcome is a consistent predictor of follow-up outcome. Demographic information covers age, gender (with female as the reference category), ethnicity (White (reference category), Black, Asian and Other) and marital status. Age is grouped into five categories reflecting quintiles of the distribution in order to capture any non-linearities in the relationship with costs and outcomes with age 18-34 years the reference category. Variables reflecting if a patient has care co-ordinated under the CPA or has been admitted to hospital under the MHA provides information on severity and treatment. Missing values of CPA and MHA were coded as zero under the assumption that these observations were not likely subject to the MHA or under CPA as these are highly regulated activities and likely to be recorded. We include dummy variables for the 21 care clusters to investigate the extent to which these explain variations in costs and outcomes. We use the cluster with the lowest cost (Cluster 1) as the reference category. We account for income deprivation in the area of residence of observations by including a variable for the IMD Income Domain (Noble 2008) as described in Section 2.4.3 of Chapter 2 and Section 3.3.2 of Chapter 3. A dummy variable is included to capture the year (2011/12 and 2012/13) the CRP commenced (with 2011/12 as the reference category) in order to control for inflation and changes in coding practices.

5.5. *Results*

5.5.1. Response variables

Figure 5.1 shows our cost response variable and figure 5.2 our follow-up HoNOS score response variable measured at the CRP-level.

Figure 5.1 Log of total cost

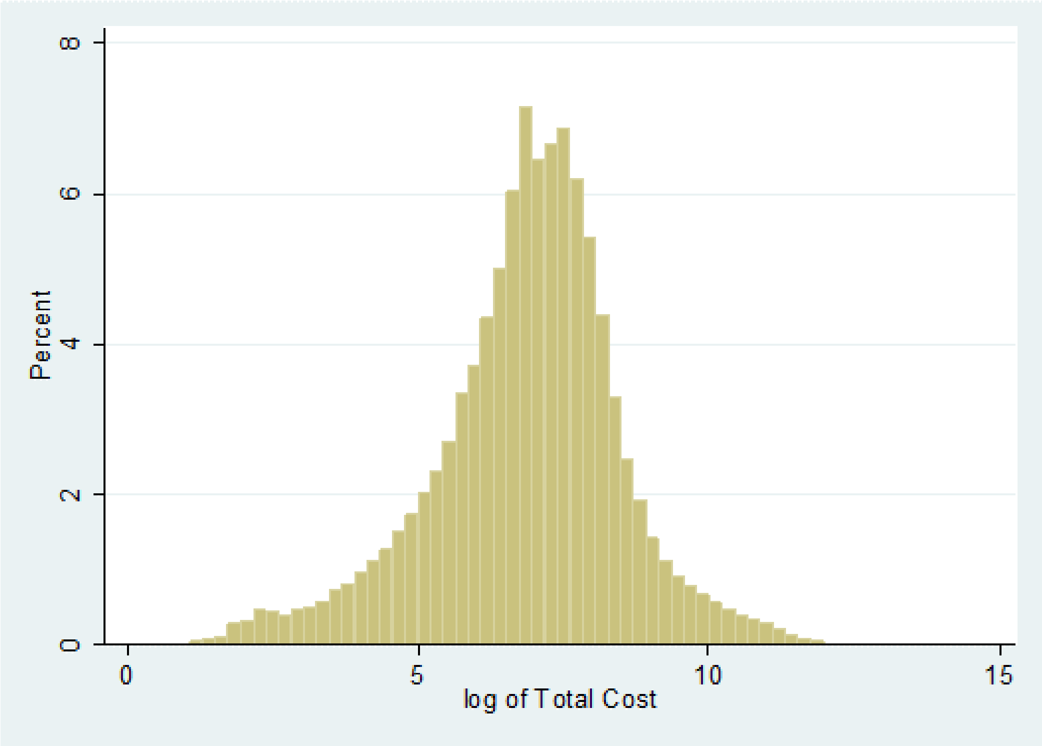
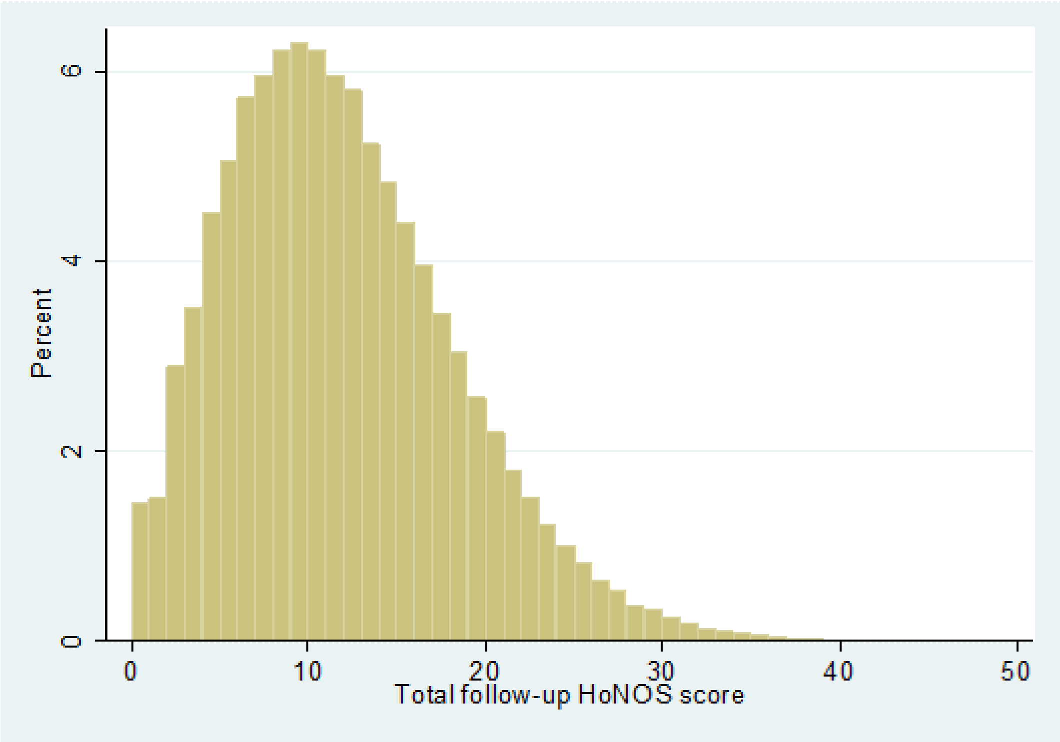


Figure 5.2 Total follow-up Health of the Nation Outcome Scales (HoNOS) score



The graphs show that both variables are approximately normally distributed although the follow-up HoNOS score variable is slightly right skewed reflecting a smaller number of observations with high scores (and more severe mental health problems).

5.5.2. Descriptive statistics

We merged data on outcomes and costs from the databases constructed for Chapters 3 and 4 using the Person and Spell identifiers in the MHMDS and the CRP start and end dates. This resulted in 854,037 observations. Our estimation sample was reduced to 697,022 observations due to missing data on the risk adjustment variables. In our estimation sample, 269,525 observations have both cost and outcome responses, 419,879 observations have the cost response only, and 7,618 have the outcome response only. As in Chapter 3, data is available for 55 providers as three providers don't report cost data and one high-cost outlier provider was excluded from the analysis. Table 5.3 displays the descriptive statistics for our estimation sample with reference categories in brackets.

Table 5.3 Descriptive statistics

Variable	Observations	Mean	Standard Deviation	Min	Max
Total cost	689404	3448.076	9783.937	0.99	303131.3
log of total cost	689404	6.919	1.615	0.01	12.62
Total HoNOS follow-up score	277143	11	6	0	48
Total HoNOS baseline score	277143	11	6	0	48
Married/civil partner	697022	0.331	0.470	0	1
[White ethnicity]	697022	0.877	0.329	0	1
Asian ethnicity	697022	0.046	0.208	0	1
Black ethnicity	697022	0.047	0.211	0	1
Other ethnicity	697022	0.031	0.173	0	1
[Age category 1 (18-34)]	697022	0.204	0.403	0	1
Age category 2 (35-46)	697022	0.191	0.393	0	1
Age category 3 (47-62)	697022	0.207	0.405	0	1
Age category 4 (63-79)	697022	0.204	0.403	0	1
Age category 5 (80+)	697022	0.194	0.395	0	1
Gender [Female]	697022	0.436	0.496	0	1
Admitted under the MHA	697022	0.088	0.283	0	1
Under CPA	697022	0.413	0.492	0	1
Cluster 0: Variance	697022	0.011	0.103	0	1
[Cluster 1: Common mental health problems, low severity]	697022	0.039	0.195	0	1

Cluster 2: Common mental health problems	697022	0.050	0.219	0	1
Cluster 3: Nonpsychotic, moderate severity	697022	0.116	0.321	0	1
Cluster 4: Non-psychotic, severe	697022	0.088	0.283	0	1
Cluster 5: Non-psychotic, very severe	697022	0.032	0.175	0	1
Cluster 6: Non-psychotic disorders of overvalued ideas	697022	0.017	0.128	0	1
Cluster 7: Enduring non-psychotic disorders	697022	0.039	0.193	0	1
Cluster 8: Non-psychotic chaotic and challenging disorders	697022	0.036	0.186	0	1
Cluster 10: First episode in psychosis	697022	0.027	0.163	0	1
Cluster 11: Ongoing recurrent psychosis, low symptoms	697022	0.090	0.286	0	1
Cluster 12: Ongoing or recurrent psychosis, high disability	697022	0.064	0.246	0	1
Cluster 13: Ongoing or recurrent psychosis, high symptom/disability	697022	0.046	0.209	0	1
Cluster 14: Psychotic crisis	697022	0.029	0.167	0	1
Cluster 15: Severe psychotic depression	697022	0.010	0.102	0	1
Cluster 16: Dual diagnosis, substance abuse and mental illness	697022	0.016	0.127	0	1
Cluster 17: Psychosis and affective disorder difficult to engage	697022	0.022	0.148	0	1
Cluster 18: Cognitive impairment, low need	697022	0.098	0.297	0	1
Cluster 19: Cognitive impairment or dementia, moderate need	697022	0.107	0.310	0	1
Cluster 20: Cognitive impairment or dementia, high need	697022	0.043	0.204	0	1
Cluster 21: Cognitive impairment or dementia, high physical need	697022	0.019	0.135	0	1
Income Deprivation	697022	17.988	11.785	0	77
CRP started in 2012/13 [CRP started in 2011/12]	697022	0.423	0.494	0	1

The total HoNOS baseline score ranges from 0-48 with a mean of 11. One-third of our sample is married or has a civil partner. The majority (88%) of observations are of White ethnicity with Asian, Black and Other ethnicities accounting for 3-5% each. Just under half (44%) of observations are of male gender. Almost one-tenth of observations were admitted under the MHA while around 41% of observations were under CPA prior to or during a CRP. On average, our sample consists of people living in an area where 18% of the population experiences income deprivation but some members of the sample live in an area where as much as 77% of the population experiences income deprivation. The majority (12%) of observations are allocated to Cluster 3 while Clusters 4, 11, 18 and 19 were also quite common accounting for 9-11% of observations. It is somewhat reassuring that Cluster 0 (the variance cluster) is one of the least populated clusters, along with Cluster 15.

5.5.3. Estimation results

Table 5.4 displays the estimation results.

Table 5.4 Estimates of bivariate model

	Observations per group			
	Number of observations	**Minimum**	**Average**	**Maximum**
Level 3: Provider	55	33	12673	54090
Level 2: Person	414092	1	1.7	43
Level 1: CRP	697022			
Log likelihood	-2065741			
	Log of total cost		**Total follow-up HoNOS**	
	Coefficient	**Standard Error**	**Coefficient**	**Standard Error**
Total HoNOS baseline score			0.388	0.002***
Married/civil partner	0.009	0.004*	-0.378	0.025***
Asian ethnicity	0.026	0.009**	-0.121	0.057*
Black ethnicity	0.083	0.010***	-0.302	0.063***
Other ethnicity	0.031	0.011**	-0.050	0.070
Age category 2 (35-46)	0.086	0.006***	0.191	0.037***
Age category 3 (47-62)	0.147	0.006***	0.169	0.036***
Age category 4 (63-79)	0.295	0.007***	-0.338	0.041***
Age category 5 (80+)	0.181	0.008***	-0.401	0.048***
Gender	0.011	0.004**	0.236	0.023***
Admitted under the MHA	0.681	0.008***	0.484	0.042***
Under CPA	0.231	0.005***	0.407	0.026***
Cluster 0: Variance	0.286	0.019***	1.094	0.128***
Cluster 2: Common mental health problems	0.378	0.012***	0.539	0.084***
Cluster 3: Nonpsychotic, moderate severity	0.686	0.010***	1.262	0.075***
Cluster 4: Non-psychotic, severe	1.019	0.011***	2.232	0.076***
Cluster 5: Non-psychotic, very severe	1.323	0.013***	3.129	0.087***
Cluster 6: Non-psychotic disorders of overvalued ideas	1.284	0.016***	3.003	0.102***

Cluster 7: Enduring non-psychotic disorders	1.280	0.013***	2.995	0.085***
Cluster 8: Non-psychotic chaotic and challenging disorders	1.347	0.013***	3.390	0.087***
Cluster 10: First episode in psychosis	1.684	0.014***	1.550	0.092***
Cluster 11: Ongoing recurrent psychosis, low symptoms	1.035	0.011***	0.125	0.076
Cluster 12: Ongoing or recurrent psychosis, high disability	1.468	0.012***	1.833	0.078***
Cluster 13: Ongoing or recurrent psychosis, high symptom/disability	1.720	0.013***	3.020	0.083***
Cluster 14: Psychotic crisis	2.011	0.014***	3.412	0.089***
Cluster 15: Severe psychotic depression	1.626	0.020***	3.562	0.118***
Cluster 16: Dual diagnosis, substance abuse and mental illness	1.528	0.017***	3.871	0.103***
Cluster 17: Psychosis and affective disorder difficult to engage	1.880	0.015***	3.373	0.095***
Cluster 18: Cognitive impairment, low need	0.186	0.011***	0.385	0.079***
Cluster 19: Cognitive impairment or dementia, moderate need	0.550	0.011***	2.310	0.079***
Cluster 20: Cognitive impairment or dementia, high need	0.808	0.013***	4.354	0.089***
Cluster 21: Cognitive impairment or dementia, high physical need	0.681	0.016***	5.462	0.113***
Income Deprivation	0.000	0.000*	0.014	0.001***
CRP started in 2012/13	-0.494	0.004***	0.195	0.020***
Constant	5.934	0.057***	4.648	0.148***
Random Effects Parameters	**Estimate**	**Standard Error**	**95% Confidence Interval**	
Level 3: Provider				
Variance: Log of total cost	0.170	0.033	0.106	0.234
Variance: Follow-up total HoNOS	0.748	0.052	-0.109	0.095
Covariance: Log of total cost, Follow-up total HoNOS	-0.007	0.159	0.436	1.060
Level 2: Person				
Variance: Log of total cost	0.291	0.004	0.284	0.298
Variance: Follow-up total HoNOS	5.407	0.013	0.038	0.089
Covariance: Log of total cost, Follow-up total HoNOS	0.063	0.073	5.263	5.551
Level 1: CRP				
Variance: Log of total cost	1.768	0.004	1.760	1.776
Variance: Follow-up total HoNOS	21.000	0.015	-0.078	-0.017
Covariance: Log of total cost, Follow-up total HoNOS	0.047	0.078	20.848	21.153

*****p<0.001, **p<0.01, *p<0.05**

The follow-up HoNOS response variable is measured on a continuous scale from 1 (best) to 48 (worst) meaning that a positive coefficient signifies a worse outcome.

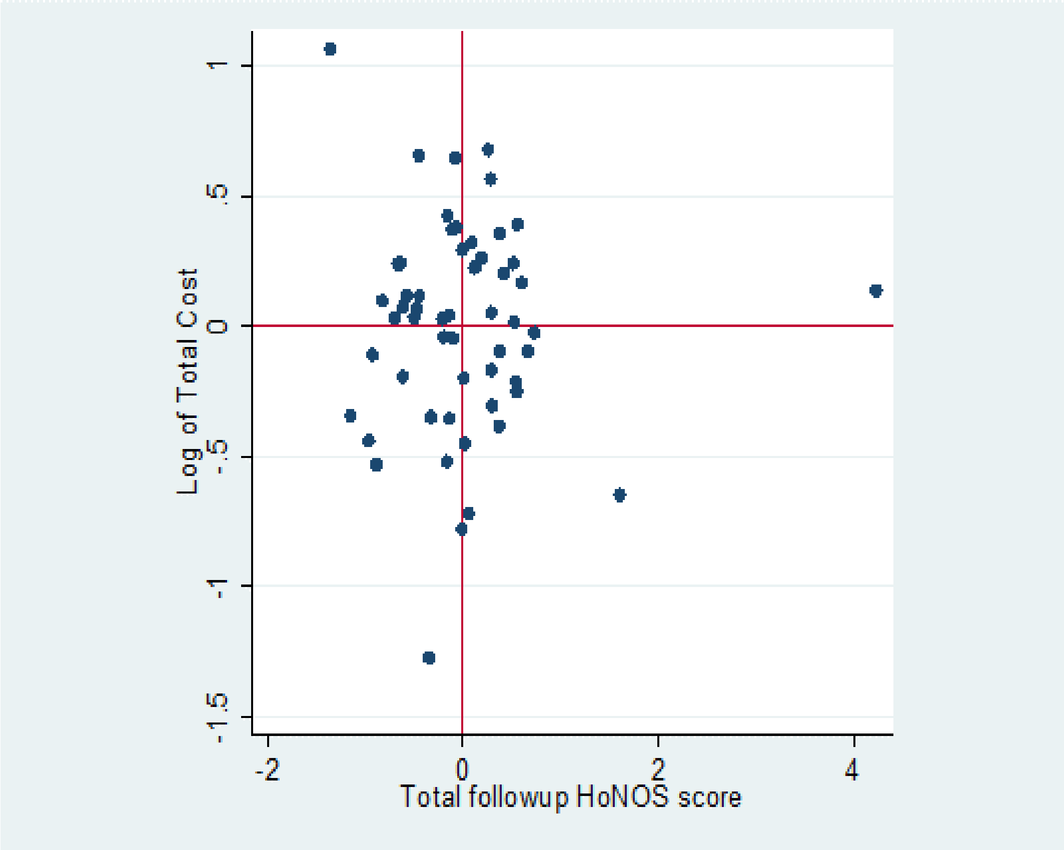
For the log of total cost response variable, many of the cluster variables are associated with the largest effects. For example, cluster 14 is associated with a 647% and cluster 17 a 555% increase in cost compared to cluster 1. Clusters 10, 13 and 15 are also associated with considerable increases of over 400% compared to cluster 1. The MHA variable is associated with a 98% increase in cost. In terms of demographic variables, Black ethnicity is associated with an increase of 9% in costs compared to White ethnicity while age of 63-79 years is associated with an increase in costs of 34% compared to age 18-34 years. CRPs that started in 2012/13 are associated with a 39% reduction in costs compared to CRPs that started in 2011/12. For the follow-up HoNOS response variables, covariates associated with an improved outcome include married/civil partner, Asian and Black ethnicities compared to White ethnicity, and older age. Marriage/civil partnership and age 80 years or over are associated with a reduced HoNOS score of around 0.4 points while Black ethnicity is associated with a reduction of 0.3 points. The positive association between married/civil partner, older age and Black ethnicity are consistent with the findings for Chapter 4. The MHA and CPA variables are associated with an increase in the follow-up HoNOS score of 0.4-0.5 points. Similar to the cost response, the clusters with higher severity are associated with greater magnitudes of effects with clusters 15, 16, 20 and 21 associated with increases of 4-5 points compared to cluster 1. A CRP that started in 2012/13 is associated with an increased HoNOS score of around 0.2 points compared to a CRP that started in 2011/12.

5.5.4. Provider-level residual variation in costs and outcomes

The correlation between residual costs and outcomes at the provider-level was calculated as -0.02 suggesting little evidence of a meaningful relationship between the two measures.

Figure 5.3 shows the pairwise plot in residual costs and outcomes for the providers in our analysis.

Figure 5.3 Pairwise plot of residual costs and outcomes for providers



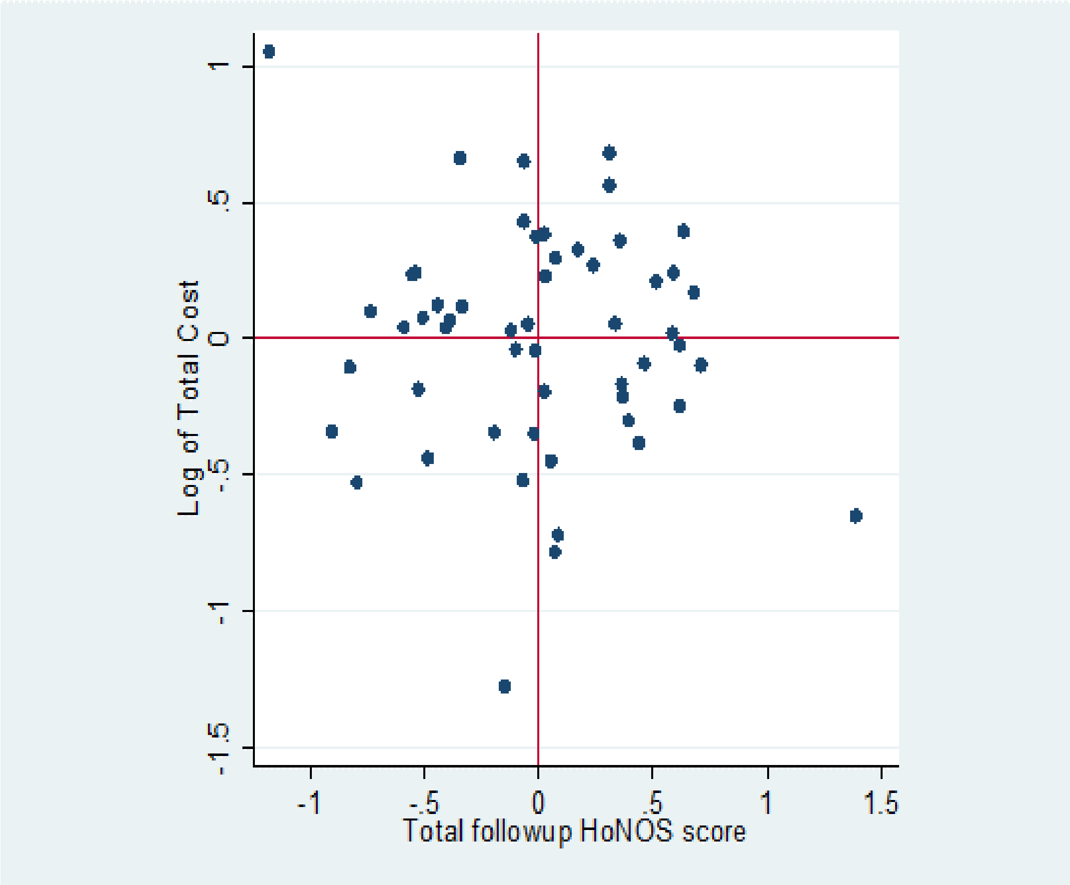
The providers fit quite evenly into four groups; those associated with 1) higher costs and lower follow-up HoNOS scores (better outcome) in the top left quadrant, 2) higher costs and higher follow-up HoNOS scores (worse outcome) in the top right quadrant, 3) lower costs and higher follow-up HoNOS scores (better outcome) in the bottom right quadrant, and 4) lower costs and lower follow-up HoNOS scores (worse outcome) in the bottom right quadrant. There is an outlier provider with a residual follow-up HoNOS score of just over 4 points above the average and slightly above-average residual costs. This outlier in respect to residual follow-up HoNOS score is consistent with the findings of Chapter 4.

The estimates of residual costs and outcomes at the provider-level follow a normal distribution with a mean of zero. The follow-up HoNOS response variable is measured on a continuous scale from 1 (best) to 48 (worst) meaning that a positive score for the residual total follow-up HoNOS score signifies a worse outcome. The residual follow-up HoNOS score varies from -1.36 to 4.23 meaning that the best performer in relation to outcomes is associated with a residual follow-up HoNOS score of 1.36 less than the average performer while the worst performer is associated with a residual follow-up HoNOS score of 4.23 greater than the average performer. For the cost response variable we can calculate the percentage difference in the EB estimates of provider-level residual variation for the top and bottom performing providers compared to the average provider as (exp(u_{1k} - u_{10}) – 1)*100 where u_{10} refers to the average provider. This shows that the provider-level variation in total cost varies from 71% below the average for the best performing provider and 194% above the average for the worst performing provider.

5.5.5. Sensitivity analysis

The exclusion of the provider with the above-average residual follow-up HoNOS score of 4.23 decreased the estimation sample to 681,305 observations. The estimation results were robust to this change with the exception of married/civil partner which loses statistical significance for the log of total cost response. The correlation between residual costs and outcomes at the provider-level became -0.09. Figure 5.4 displays the pairwise plot in residual costs and outcomes for the 54 providers in the sensitivity analysis. The residual follow-up HoNOS score reduced to 1.39 for the worst performing provider on follow-up HoNOS scores compared to the average performing provider for this response. This sensitivity analysis had little effect on the residual log of total cost which varied from 72% for the best performing provider to 191% for the worst performing provider.

Figure 5.4 Pairwise plot of residual costs and outcomes for providers in sensitivity analysis



5.6. *Discussion*

The reimbursement of mental healthcare providers in England is undergoing a considerable reform with the move towards a prospective, activity-based payment system. With future intentions to link prices to national average costs as well as linking some part of payment to patient outcomes, the new system will offer incentives for providers to deliver care more efficiently while better meeting patient needs and improving outcomes. This research has explored the relationship between costs and outcomes in order to examine the scope for providers to respond to the incentives introduced by the new payment system. The relationship between costs and outcomes has been the subject of a number of studies in both physical (Carey and Burgess 1999; Gutacker et al. 2013a; Hvenegaard et al. 2011; Schreyogg and Stargardt 2010; Weech-Maldonado, Shea and Mor 2006) and mental (Dickey and

Normand 2004; Haas et al. 2013; Hendryx 2008; Schulz, Greenley and Peterson 1983) healthcare and this research makes an important contribution to this literature.

After controlling for a range of demographic and treatment factors, we find that residual variation remains in both costs and outcomes at the provider-level. However, the correlation between residual costs and outcomes at the provider-level is miniscule, which suggests that a trade-off between cost containment and outcome improving efforts on the part of providers is not a major concern. Plotting the provider-level residual costs and outcome variables reveals that providers broadly fall into four groups with an outlier provider. This outlier provider is consistent with Chapter 4. Providers with higher than average residual costs and lower than average residual outcomes may signify poor performance but may also indicate that certain providers are treating a casemix that our model has not fully accounted for. While patient casemix is controlled for to a certain extent by the care clusters, the clustering method does not explicitly take diagnosis into account and it is likely that the clusters are very variable in terms of diagnosis and casemix (Jacobs 2014; Yeomans 2014). It may also be the case that some patients have treatment-resistant variants of mental illness which implies that they will be consuming large amounts of care and resources with little discernible changes in outcome scores (Jencks et al. 1987). If certain providers have a higher case-load of such patients this could well explain their unexplained higher costs and worse outcomes. If the higher costs are legitimate then these providers may warrant additional payments as defined by any outlier policy attached to the payment system. A number of providers are associated with better residual outcomes but also with higher residual costs. These providers in particular may face a potential trade-off between costs and outcomes and efforts to reduce costs under a national tariff may compromise outcomes if providers are induced to undertake undesirable behaviours such as skimping on patient care. A number of providers have lower than average residual costs and higher than average residual outcomes. These providers are likely to financially benefit from the new payment system if a national tariff is introduced and patient outcomes are linked to provider payment. Providers with lower than average residual costs and lower than average residual outcomes may have scope to make financial profits under a national tariff but these may be offset if payment is linked to outcomes. If providers are achieving lower costs at the expense of patient outcomes then they would warrant particular scrutiny by commissioners under quality and outcomes standards established in the contracting process.

As highlighted in Section 3.6 of Chapter 3 and Section 4.7 of Chapter 4, there are limitations regarding our data, in particular issues surrounding the quality of the Reference Cost data and missing values for HoNOS scores, which we do not take into account in this analysis. Nevertheless, this research provides a useful insight into the relationship between mental health costs and outcomes that is pertinent in the context of a new payment system for mental health providers in England.

Chapter 6. Conclusion

6.1. *Overview of research*

This thesis investigates the performance of English mental healthcare providers in terms of resource use (LOS and costs) and quality (readmission rates and patient outcomes) in the context of the introduction of a new payment system to mental healthcare. The NTPS will present providers with various incentives with regard to resource use and quality, which may subsequently affect their performance. We attempt to anticipate the incentives introduced by the NTPS and potential effects on providers by examining provider performance on LOS, costs and outcomes. It is important to investigate provider performance in the context of the introduction of the NTPS, given that mental health expenditure is not commensurate with the disease and economic burden and mental health budgets have been cut recently, despite increases in demand for services.

We employ multilevel methodology to model the various response variables including linear, log-linear, ordered probit and Poisson models. While the majority of models are hierarchical we also utilise a non-hierarchical model. We make inferences about provider performance by quantifying the provider-level random intercepts or effects using EB techniques. While this approach has been previously employed in studies assessing the performance of providers in acute physical healthcare (Gutacker et al. 2013a; Gutacker et al. 2013b), to my knowledge this is the first research to apply this method to mental health data to investigate performance of mental healthcare providers. We use data sets with national coverage that allow us to study the performance of the majority of English public mental health providers, moving us beyond previous studies in mental healthcare that have been limited in the geographical, population or patient coverage of their data.

In Chapter 2, we investigate the determinants of inpatient LOS at admission-, patient- and provider-levels using HES data. Our estimation model of choice is a cross-classified model, which recognises that patients can have admissions to different hospitals. As a sensitivity analysis, we model the data using a three-level hierarchical model that does not allow for patient movement across multiple providers in order to investigate the extent to which results differ given that only a small proportion of patients had admissions to more than one provider. Admission-level covariates include demographic, diagnostic, treatment and social variables. Patient-level covariates cover demographic information while provider-level covariates include provider type and capacity, information on staffing, and treatment and quality indicators including provider emergency readmission rates. The research is innovative in using provider-level readmission rates as previous studies (Appleby et al. 1993; Boden et al. 2011; Figueroa, Harman and Engberg 2004; Korkeila et al. 1998; Lin et al. 2006; Sytema and Burgess 1999; Wolff et al. 2015b) have investigated the relationship between LOS and readmission rates at the patient-level and this is the first study, to my knowledge, that uses provider-level readmission rates to explain variation in admission-level LOS. It is beneficial to use provider-level readmission rates as there may be systematic differences between providers in readmission rates that go beyond patient casemix. These differences may reflect not only quality of care but also wider organisational and management attributes such as discharge planning and policies, coordination between inpatient and community care and community-based follow-up. Provider-level variation in LOS unexplained by the admission-, patient- and provider-level variables is quantified using EB methods in order to compare provider performance.

Results show that at an admission-level, inpatient death, a primary diagnosis of psychosis, formal admission under the MHA, discharge to social care and the oldest age group (65 years and over) are associated with the largest increases in LOS. The variables capturing transfer-out, a primary diagnosis of substance misuse disorder, carer support recorded and psychiatric treatment history are associated with relatively large reductions in LOS at the admission-level. Of the patient-level variables, Black ethnicity is associated with the largest increase in LOS. At provider-level, the emergency readmission rate has a large negative association with LOS while the proportion of formal admissions under the MHA has a large positive association with LOS. The sensitivity analysis shows that the two models are broadly in agreement with differences most pronounced for provider-level variables. This is because the three-level model does not correctly model the dependence in the data when there is cross-classification, which leads to biased standard errors (Leckie 2013). Ranking providers by residual variation reveals significant differences with a number of providers exhibiting above- or below-average residual LOS, suggesting scope for some providers to improve performance.

In Chapter 3, we use the MHMDS to examine resource use in terms of costs associated with a CRP – the unit of activity for which payment is made under the NTPS. Cost encompasses both admitted and non-admitted care and wide variations in costs are evident across providers. We model costs using 1) a three-level log-linear model and 2) a three-level GLM with a gamma distribution and a log link. We use a comprehensive range of variables to risk-adjust costs in order to compare residual costs across providers to assess performance. Risk adjustment variables cover demographic, treatment and social characteristics of patients as well as the care cluster patients are assigned to. Provider-level variables hypothesized to be associated with cost including provider type, number of mental health beds and mental health bed occupancy, and formal admissions as a proportion of total inpatient admissions are included in the model in a sensitivity analysis. Results show consistency between the log-linear model and GLM with variables associated with higher costs including older age, Black ethnicity, admission under the MHA, and care clusters 10 and 13-17. The number and percentage occupancy of mental health beds are associated with small negative and positive effects on mental health costs but FT status and the proportion of formal admissions at provider-level are associated with a sizeable decrease and increase respectively in mental health costs. Around 8% of residual variation in costs remains at the provider-level despite controlling for a wide range of variables with an influence on cost. While a number of providers exhibit lower residual costs than average, other providers are associated with higher residual costs compared to the average. These higher costs may be associated with higher quality care or with inefficient provision of care, which may lead to financial instability under a national tariff. The large variation in costs also brings into question the ability to set a national average price (tariff), which can act as an accurate price signal for each cluster.

Chapter 4 also uses the MHMDS and examines the performance of mental health providers in relation to patient outcomes. Outcomes are measured using HoNOS scores recorded at the beginning and end of a CRP. An ordered outcome variable is constructed that measures a RCSC in HoNOS scores. This variable is subsequently modelled using an ordered probit three-level model. A linear three-level model is used to model follow-up HoNOS scores. Both estimation models include a wide range of risk adjustment covariates reflecting demographic, treatment, social and need variables. Provider-level residual variation is quantified using EB methods, and providers are ranked in order to make inferences about provider performance. We include provider-level variables measuring provider type, size and capacity in a sensitivity analysis. The models reveal a number of variables are associated with better outcomes including married/civil partner, older age and need associated with ongoing or recurrent psychosis compared to need associated with common mental health problems as captured by the care clusters. Variables associated with worse outcomes in both models include male gender, admission under the MHA, having care coordinated under CPA, income deprivation, and higher need as captured by care clusters reflecting very severe and complex non-psychotic illness, severe psychotic illness, and cognitive impairment compared to care clusters for common mental problems. In addition, in the linear model, Black ethnicity is associated with better outcomes while higher baseline HoNOS scores and problems with accommodation are associated with worse outcomes. In contrast, problems with accommodation are associated with better outcomes in the ordered probit model. These results remain stable in the sensitivity analysis, which also shows that mental health beds have a positive association and mental health bed occupancy a negative association with mental health outcomes in the ordered probit model. Residual variation across providers persists after adjusting outcomes for these risk factors, which indicates that some providers are more likely to achieve better outcomes for a patient with an average set of risk factors.

In Chapter 5, we consider costs and outcomes together by estimating a bivariate model with log of cost and follow-up HoNOS scores as dependent variables. To my knowledge, this is the first study to use a multilevel bivariate model to explore the relationship between costs and outcomes in mental healthcare. As in Chapters 3 and 4, we risk-adjust these variables using a wide-ranging set of covariates capturing demographic, need, treatment and social information to ensure impartial comparisons across providers. The correlation between the provider-level residual variation in costs and outcomes is calculated to assess if a trade-off exists between resource use and quality. The residual variation in the response variables is also quantified using EB techniques and plotted in order to categorise providers into four groups based on their performance in relation to costs and outcomes. Results regarding the direction of relationships between the dependent and independent variables mirror those of Chapters 3 and 4. Providers fall into four relatively equal groups: those associated with 1) higher costs and lower follow-up HoNOS scores (better outcome), 2) higher costs and higher follow-up HoNOS scores (worse outcome), 3) lower costs and higher follow-up HoNOS scores (worse outcome), 4) lower costs and lower follow-up HoNOS scores (better outcome). Results show a negligible correlation between residual costs and outcomes at the provider-level, which suggests that concerns about a potential trade-off between cost-containment and quality-improving efforts on the part of providers may not be warranted.

In summary, variables associated with higher costs and worse outcomes include care clusters reflecting higher severity, formal admission under the MHA, care co-ordination under CPA and male gender. Conversely, we find that older age and Black ethnicity are associated with higher resource use measured by both LOS and costs but also better outcomes. Results regarding social support and resource use are somewhat inconclusive. We find that carer support is negatively associated with LOS while married/civil partner is positively associated with resource use and is also associated with better outcomes.

With regard to provider variables, results show that providers with FT status have a negative association with costs but there is no evidence of an association with outcomes. Provider size as measured by number of mental health beds has a negative association with costs and a positive association with outcomes. This may be suggestive of economies of scale. Bed occupancy is associated with positive effects on resource use in terms of both LOS and costs and a negative effect on outcomes.

A trade-off between resource use and quality of care is reflected in our finding regarding LOS and readmission rates. We find a large negative association between admission-level LOS and provider readmission rates suggesting that providers with high readmission rates are associated with shorter LOS for admissions. This implies that efforts on the part of providers to make efficiencies in terms of LOS may have a detrimental impact on quality of care as measured by the readmission rate. In contrast, we find little evidence of a trade-off between costs and outcomes. The disparity in results may be due to the differing measures of resource use and quality and methodologies employed.

In all chapters, the greatest variation within groups is at level 3 – hospital or provider. In Chapter 2, there are on average, over 2,000 patients per hospital but the smallest hospital has 234 patients and the largest 5,377 patients. At the patient level, on average there are only 1.5 admissions per patient with some patients having just one admission, and one patient with 86 admissions. In Chapters 3-5, the minimum number of patients per provider ranges from 2 (Chapter 4) to 33 (Chapters 3 and 5). On average, there are between 5,368 and 12,673 patients per provider, with a maximum number of patients per provider of between 36,981 (Chapter 4) and 54,090 (Chapter 5). At the patient-level, the minimum number of CRPs is 1 with an average of 2 CRPs indicating little information on the group. However, some patients have up to 67 CRPs (Chapter 4).

6.2. *Implications for policy and recommendations for practice*

This research has a number of implications for policymakers, commissioners, providers and users of mental healthcare. We discuss implications in relation to the design of the classification and payment system including the role of outcomes in the NTPS, readmissions, formal admissions, and mental health funding.

6.2.1. Care clusters classification system

Our research shows that a number of providers have higher than average resource use as measured by LOS and costs for reasons unaccounted for by the variables included in our analyses including the care clusters. One possible explanation is that higher resource use is attributable to factors we have been unable to control for in our models such as these providers treating a larger proportion of patients with treatment-resistant mental illness. On the other hand, high residual levels of resource use may be interpreted as inefficient provision. If the care clusters classification system does a good job of explaining variation in cost and the payment system accounts for any legitimate cost drivers not considered in the classification system then any financial risk due to inefficiency will be borne by providers and they will have a good incentive to reduce this inefficiency in order to avoid financial deficits. Therefore, the design of the classification and payment system is crucial. We find that the care clusters do not control for all variation in costs. An implication of this for policymakers is that the care clusters may require additional refinements or the payment system will need to compensate for any weaknesses of the classification system in order to prevent providers facing unfair financial risk. Otherwise, providers may be enticed to engage in manipulating or 'gaming' cluster allocations in order to increase income. This would have implications for service users in terms of receiving either inadequate or unnecessary care. Furthermore, commissioners may end up expending extra resources for unnecessary care, which may destabilise local health economies. Additional factors may need to be considered in the classification or payment systems to reduce differentials between prices paid to and costs borne by providers. However, this must be balanced against the risk of incentivising providers to selectively treat patients based on their observable characteristics. We find that Black ethnicity and older age are associated with higher costs and LOS. Age was also found to be a key cost driver in the classification systems developed in Australia and New Zealand while the US per diem payment system also adjusts for patient age. The New Zealand classification system adjusted for ethnicity. The care cluster classification system does not take account of diagnosis. We find that a diagnosis of psychosis is associated with longer LOS which suggests it is a driver of inpatient costs, as supported by previous studies (Chung et al. 2010; Hodgson, Lewis and Boardman 2000; Huntley et al. 1998; Jacobs et al. 2015; Lay, Lauber and Rossler 2006; Lerner and Zilber 2010; Oiesvold et al. 1999; Peiro et al. 2004; Pertile et al. 2011; Tulloch, Fearon and David 2011). Previous research (Buckingham et al. 1998; Drozd et al. 2006) has also found diagnosis to predict higher inpatient costs.

It would be premature to consider the introduction of a national price or tariff per cluster before further refinement to the classification or payment systems. A national tariff is also contingent on the submission of good quality cost and activity data by providers. An implication for providers is that clinicians who are clustering patients need to be aware that this activity informs provider payment. Providers must also put in place systems to collect high-quality cluster activity and cost data. Commissioners can incentivise the collection of good-quality cluster activity data by including indicators on data quality in NTPS contracts with providers, for example the proportion of service users in each cluster. There is also an onus on the economic regulator (Monitor) to encourage the reporting of high quality cost data by providers in order to ensure that provider payment adequately reflects resource use on the part of providers. The planned introduction and eventual mandatory use of patient-level information and costing systems (PLICS) (Monitor 2015) by mental health providers will help to achieve this.

From the perspective of implementing a payment system, it is somewhat disconcerting that we found evidence of a statistical association between provider variables and the total cost of a CRP (after controlling for patient factors). This raises a concern that providers may systematically gain or lose financially based on FT status, size (number of beds), spare capacity (% occupancy) and proportion of formal admissions. The Reference Cost data that is used to calculate the cost of a CRP will underpin price in the NTPS and in theory, (baseline) payment should reflect differences in costs arising from patient characteristics alone and not be biased towards providers with particular characteristics (Fries et al. 1993).

An important consideration for policymakers is the design of an outlier policy to ensure that providers treating patients with costs that are not reflected in the classification system are not discriminated against financially. This raises the question of how outliers should be defined, for example in terms of inpatient days or per diem costs. Our finding of statistical relationships between provider-level variables reflecting inpatient care and resource use suggests that an outlier policy to reimburse legitimate cost differences between providers may be best focused on the inpatient component of a CRP. An outlier policy may therefore be designed in a similar fashion to that adopted for the NTPS in acute elective care whereby a long-stay outlier payment is activated once a LOS threshold is reached and additional days beyond this threshold are paid on a per diem basis (Mason, Ward and Street 2011). This threshold and the per diem payments could be differentiated by cluster. In clusters where relatively long inpatient stays are foreseen, these per diem payments could decline as LOS increases as is the case for inpatient psychiatric payment systems in the US (Mason and Goddard 2009).

6.2.2. Role of outcomes in the new payment system

It is intended that a part of provider payment under the NTPS will be linked to outcomes. This has several implications for policymakers. We find that some providers perform worse than average on patient outcomes and the linking of payment to performance on outcomes has the potential to drive improvements among these providers. As HoNOS is a CROM there is greater potential for manipulation of ratings in order to influence payment. Subsequently there will need to be strict regulation of the system in order to prevent the incentivisation of adverse behaviour on the part of clinicians. A particular challenge is that clinical judgement underpins the cluster allocation decision. Clinicians could be encouraged to adhere to the available care transition protocols and guidelines regarding cluster allocations (Monitor and NHS England 2013d) and deviations from the recommended care transitions could be monitored. There is also a need to collect baseline outcome measures to ensure that any targets set are appropriate. We found that there was considerable missing data on follow-up HoNOS scores and this may have implications for linking payment to outcomes if outcomes are measured using changes in HoNOS scores. While making a part of provider payment contingent on outcomes will incentivise the collection of follow-up HoNOS data, the collection of follow-up HoNOS scores could be included as a quality indicator in NTPS contracts between commissioners and providers to encourage routine collection of this data.

Linking a part of provider payment to outcomes has the potential to improve the commissioning of mental health services as the value commissioners perceive they are getting for resources spent on mental health will be more evident and transparent. Commissioners will be able to direct resources to the providers that deliver comparatively better quality and outcomes. The inclusion of quality and outcomes in contracts with the eventual linkage of payment to quality and outcome standards and targets will make it easier for commissioners to hold providers to account and penalise them if necessary for sub-standard care. This will place an onus on providers to deliver high quality care and perform well in relation to patient outcomes or risk facing financial losses. In turn, service users will stand to benefit from improved quality of care and outcomes. Moreover, the recent introduction of patient choice of provider to mental healthcare may entice service users to be more cognisant of the quality of different providers, including their performance on outcomes.

6.2.3. High readmission rates are associated with lower LOS.

We find a negative association between provider-level readmission rates and admission-level LOS. This implies that higher readmission rates are associated with shorter LOS. Under the NTPS for mental health, providers are paid a fixed prospective price for all care provided in a given cluster irrespective of setting. Therefore, the NTPS introduces an incentive for providers to reduce inpatient LOS, given that inpatient care is relatively more costly on a daily basis compared to community-based care. The current financial climate may also place pressure on providers to reduce inpatient LOS. Providers need to consider that attempts to reduce resource use in terms of LOS may have detrimental effects on quality of care and lead to higher costs in the longer run due to increased readmissions. The impacts on quality of care may be in the form of inadequate care during the index admission and insufficient preparation and planning for discharge. Deficiencies in community-based care, in particular post-discharge follow-up may also contribute to increased readmission rates. If the provider market for mental healthcare becomes more competitive under a national tariff then providers with high readmission rates may not gain contracts and may face financial losses. The introduction of a national tariff is likely to provide a strong incentive to reduce costs and LOS and some providers may have to increase efficiency in order to remain financially viable once this objective is realised.

The implication of our findings on the relationship between LOS and readmission rates for service users is the concern that multiple admissions results in a revolving door pattern of care with patients frequently admitted with little clinical benefit for their condition. In agreeing care packages following cluster allocation, it is important that providers and patients plan appropriately for periods of inpatient admission and ensure adequate community support, in particular for crisis and post-discharge services and support and providers ensure patients are informed about the availability of necessary services.

Providers also need to consider the comprehensiveness and adequacy of services provided in particular clusters – particularly those where periods of inpatient care may be anticipated – when negotiating and agreeing cluster prices with commissioners. For their part, commissioners may want to monitor readmission rates to mental health services more closely. Data on "Unplanned readmissions to mental health services within 30 days of a mental health inpatient discharge in people aged 17 and over" is included in the CCG Outcomes Indicator Set. The data for this indicator is published at CCG level and measures the number of unplanned readmissions to a mental health service within 30 days of the discharge date as a proportion of the CCG level count of discharges from a mental health inpatient service in people aged 17 and over (Health and Social Care Information Centre Clinical Indicators Team 2015). Commissioners may also want to monitor readmissions by including this indicator in contracts with providers, taking into consideration the casemix profile of providers, in particular in relation to the clusters. It can be expected that there will be different rates of readmissions in the different clusters. Performance on readmissions could be financially incentivised through its inclusion in the Commissioning for Quality and Innovation (CQUIN) payment framework. CQUIN payments incentivise the delivery of high quality and efficient care by linking a proportion of provider's reimbursement to performance on national and local goals. There are currently two national mental health CQUIN indicators related to dementia and improving the physical healthcare of people with serious mental illness in order to reduce premature mortality (NHS England 2015). It can be informative for commissioners to consider provider performance on readmissions alongside contextual indicators such as availability of mental health beds and community services (Durbin et al. 2007). Indicators on the availability of post-discharge or crisis community-based care could also be included in payment contracts or incentivised through CQUIN payments. Commissioners could withhold payment or impose financial penalties if providers exceed a certain threshold for readmissions.

Our finding on the relationship between readmission rates and LOS also has implications for policymakers and the design of the NTPS for mental health. In acute physical healthcare, providers do not receive payment for readmissions within 30 days of discharge under the NTPS. Policymakers need to consider whether a similar policy should be attached to mental health payment. In mental healthcare, it would be feasible to discriminate financial penalties according to the clusters, taking into consideration the different readmission rates in the different clusters.

6.2.4. Formal admissions

We find that formal admissions is a key driver of increased resource use. Admission under the MHA is not considered in the care clusters classification system. This may be due to a concern that it may incentivise increased formal admissions in order to increase provider payment. Moreover, depriving service users of their civil liberties would be a concern for patients if they are increasingly treated in more restricted settings, against their will. Nevertheless, providers have different thresholds for formal admissions and may face financial risk if this is not adequately taken account of in the payment system. Ideally, the clinical or patient factors driving formal admission should be adequately reflected in the classification system (Buckingham et al. 1998). Our findings suggest that these clinical factors are not adequately captured by the care clusters in and of themselves. Research has suggested that patient factors such as ethnicity and age are associated with compulsory admission (Weich et al. 2014). This suggests that further work could be done to refine the classification system. Alternatively, the payment system could compensate providers for treatment of patients that are characterised by cost drivers (e.g. Black ethnicity). Providers could be compensated through an outlier policy included in the system as described earlier. As formal admission increases LOS, providers could be paid for the additional inpatient days beyond a stipulated LOS threshold. This would ensure that patients receive adequate, good quality care and providers are sufficiently reimbursed for this.

Formal admission is also associated with inferior outcomes for patients. This implies that providers with high rates of formal admissions will also face a disadvantage when a proportion of provider payment is linked to patient outcomes. Commissioners can take into consideration previous rates of formal admissions when agreeing prices and negotiating contracts with providers. Commissioners may want to monitor rates of formal admissions of providers to see if these are stable over time and to provide insights into what factors are associated with high rates of formal admissions. Research has suggested that around 7% of variation in admission to inpatient care under the MHA is attributable to (unobservable) provider factors (Weich et al. 2014).

If rates of formal admissions are amenable to actions on the part of providers, then financial incentives could be attached to formal admissions through CQUIN or NTPS payments, such as withholding a part of payment in the event that providers exceed an agreed threshold for formal admissions or imposing financial penalties on providers in such instances. This would help to ensure that patients do not face a higher probability of formal admission conditional on the provider they attend and help to achieve better patient outcomes.

6.2.5. Funding for mental health services

The English NHS is currently experiencing a difficult financial situation. At the end of 2014/15, providers of acute physical healthcare recorded an overall net deficit of over £820 million. The situation does not look set to improve with around two-thirds of providers of acute physical healthcare forecasting deficits for 2015/16 (Appleby, Thompson and Jabbal 2015). Mental health has not been spared from the recent financial hardships with decreases in funding and subsequent cuts in staff and beds (Lintern 2014a), despite increased demand pressures (Smith et al. 2015). Moreover, mental health providers are pessimistic that increased funding commitments for 2015/16 will be met (Appleby, Thompson and Jabbal 2015).

Mental health has traditionally received less attention than physical health among policymakers and funding has not been commensurate with the prevalence of mental illness and its societal and economic impacts. Recently however, there has been an increased recognition of the need to place mental health on a par with physical health. The goal of parity of esteem has started to be realised with the extension of polices such as patient choice, and access and waiting times targets to mental health. The introduction of the NTPS to mental health is another key policy lever to achieve parity of esteem. The funding of mental health providers using block contracts, combined with a lack of good cost and activity data has left mental health vulnerable to underfunding. The use of the NTPS for acute physical healthcare has made costs and activity more transparent and meant it was easier for commissioners to evaluate what budgets were being spent on. The measurement and costing of mental health activity using the care clusters means that commissioners will have a more tangible sense of what they are purchasing in mental healthcare. This will make it more challenging for commissioners to arbitrarily shift resources away from mental health, which is an all too realistic prospect given the financial difficulties of acute physical healthcare providers. A fairer allocation of resources to mental health according to patient need (and potentially outcomes) and provider activity is clearly warranted given the high disease burden and economic cost associated with mental illness. This will help to improve patient care in order to deliver better physical and mental health outcomes for users of mental health services.

6.3. *Limitations of work and areas for future research*

There are a number of limitations to the work presented in this thesis, which could be addressed in future research. In Chapter 2, we only consider the inpatient part of the mental healthcare pathway due to the unavailability of the MHMDS at the time of the research. Inpatient mental healthcare warrants particular attention given that it is a resource intensive care setting that has been singled out for productivity improvements. However, only a minority of mental healthcare is provided in inpatient settings (Health and Social Care Information Centre 2014). A more complete picture of the impact of provider readmission rates on LOS is provided by linking HES to the MHMDS. This would allow the inclusion of additional admission-level variables on community-based care such as CPA, accommodation status, and type and intensity of community-based care received (e.g. contacts with different healthcare professionals and types of mental health teams and other forms of care such as outpatient or day care) to be added to the model. This would provide greater insights into the effect of community-based care and support on inpatient LOS. It is worth noting the limitations of the primary diagnosis data recorded in HES for our sample, in particular the relatively high percentage (around 20%) of observations coded as "Unknown and unspecified causes of morbidity" (ICD-10 R69X). This may be why we observe fewer observations with diagnoses such as psychosis than we would expect given high admission thresholds for acute inpatient mental health care.

Mental health costs are reported in the Reference Cost (Department of Health 2012) data as per diem admitted (and non-admitted) cost per cluster by provider. By linking HES to the MHMDS we would also gain information on the care cluster assigned to the patient which would allow us to measure resource use in terms of cost in order to replicate the analysis. Future research could also consider alternative ways of modelling the data to investigate the relationship between LOS and readmission rates, such as using a bivariate model such as that employed in Chapter 5. Another limitation of the analysis in Chapter 2 is that we only estimate the association between LOS and readmission rates. In order to make inferences about a causal relationship we would need an instrumental variable – a variable correlated with readmission rates but not with LOS. To our knowledge, such a variable is not available in either HES or MHMDS.

A limitation of the research presented in Chapter 3 is the inability to include data on diagnosis due to the poor coding of the variables measuring primary and secondary diagnoses in the MHMDS. Previous research (Buckingham et al. 1998; Drozd et al. 2006; Kapur, Young and Murata 2000; Robst 2009) has shown that diagnosis is predictive of cost with more severe diagnoses associated with higher costs (Buckingham et al. 1998; Drozd et al. 2006). While the care clusters utilise diagnostic labels there is a distinct difference between the care cluster classification system and diagnostic systems (Yeomans 2014). Diagnosis is not considered when a patient is allocated to a care cluster and it has been suggested that the clusters should be viewed as complementary to diagnosis (Trevithick, Painter and Keown 2014). Primary diagnosis is relatively well recorded in HES so future work could link HES and MHMDS and restrict the analysis to patients with an inpatient stay only. While this would lead to a considerable reduction in the sample size and would limit the analysis to high severity patients with an inpatient stay, it would nevertheless provide an insight into the effect of different primary diagnoses on cost. If the coding of diagnosis in the MHMDS improves over time then it would be worthwhile replicating the analysis with information on diagnosis included. The introduction of a new variable measuring provisional diagnosis in the new Mental Health Services Data Set (MHSDS) (which replaces the MHMDS) may help to improve coding of diagnosis. A potential reason for poor coding of primary diagnosis is that it can take time for a mental disorder to present itself and clinicians are often reluctant to prematurely code diagnosis in order to avoid the problem of false positive diagnoses (Wakefield 2010; Wykes and Callard 2010). The use of a provisional diagnosis variable may address this issue to a certain extent.

As we use data from the initial years of the implementation of the NTPS, we can expect data coding to improve over time, in particular for the cost and care cluster variables. An improvement in the coding of the cost data from 2011/12 to 2012/13 is evident (see Figures 3.1 and 3.2 in Chapter 3). With the development and subsequent roll-out of PLICS for mental health providers commencing in 2016 we may expect cost data to improve considerably. A further limitation of the cost data used in Chapter 3 is that it is essentially provider-level cost data that underpins our dependent cost variable. While variation in the dependent cost variable will arise for patients in different clusters with different care patterns (in terms of LoS and admitted and non-admitted care), a greater level of variation would be observed if we had access to data on the actual costs incurred by individual patients, rather than the provider average. Cluster data may also be expected to improve once it forms the basis for provider payment. Therefore, it would be worthwhile replicating the analysis in future with additional years of data.

A limitation of Chapter 4 is the large number of observations with missing follow-up HoNOS scores. If the data is Missing Not at Random (MNAR) then this can result in biased estimates of provider performance. Future work could use multiple imputation to address the problem of missing HoNOS data when making inferences about the performance of mental health providers on patient outcomes. Multiple imputation requires the generation of multiple imputed data sets which, conditional on the missingness assumption, correctly reflect the distribution of the missing data given the observed data. The model of interest is then fitted to each imputed data set and the results of all the iterations are averaged – for example using Rubin's rules (Rubin 1987) – for the final inference. The use of multiple imputation might lead to alternative inferences about provider performance.

It is important to recognise the limitations of RCSC as applied to our data. The RCI of 10 and clinical severity cut-off of 9 imply that reliable and clinically significant improvements or deteriorations will be evident only for observations at the extreme ends of the severity distribution in terms of HoNOS scores. Moreover, it is not possible to distinguish reliable and clinically significant improvements or deteriorations measured using RCSC from changes that would have occurred anyway due to the tendency for a large measurement to be followed by a measurement closer to the average value, a phenomenon commonly known as "regression to the mean" (Barnett, van der Pols and Dobson 2005). For example, to the extent that regression to the mean applies, patients who enter mental health services in a crisis or distressed state (rather than in a non-distressed state) and subsequently record high HoNOS scores, will be more likely to record a lower subsequent rating regardless of the treatment received or quality of provider care (Evans, Margison and Barkham 1998). It has also been suggested that a cut-off point will not be accurately estimated when distributions have different variances and skew (Evans, Margison and Barkham 1998). Bearing these caveats in mind, the application of RCSC to a more homogenous patient population e.g. patients allocated to the same cluster or attending the same provider or service type may make the measurement more sensitive to change. This would mean a smaller change in HoNOS scores needed to establish a reliable change (RC) and a higher change in HoNOS scores for the clinically significant (CS) cut-off determining an improvement or deterioration in outcome (Parabiaghi et al. 2005). As in Chapter 3, better quality data on diagnosis would allow refinement of the models used in Chapter 4 and also aid interpretation of the ordered outcome dependent variable. The majority (88%) of patients were classified as having a "stable" outcome, which can be viewed as a favourable outcome for severe mental illness such as psychosis. Better data on diagnosis would also allow the calculation of RCSC in HoNOS scores for different diagnostic groups, given sufficient sample sizes.

As additional HoNOS measurements become available for the same patients, these could be analysed using a repeated measures model. This would allow the analysis of multiple outcome measures taken at different points in time and take into account the correlation between subsequent measures (Steele 2014). A repeated measures model would provide insights into the change in outcomes over time and whether different patterns are associated with different providers, for example do patients experience more rapid improvements or deteriorations when treated by certain providers? As mental illness is a chronic disease, modelling provider performance based on more than two measurements could provide inferences that are more robust.

Future work could also consider the causal relationship between costs and outcomes if data on a suitable instrumental variable correlated with outcomes but not with costs should become available.

We find variations in performance across providers for both resource use and quality. It is not within the scope of this thesis to investigate the factors driving these differences in performance beyond the variables considered in this research. Future work could further investigate what features of providers can explain differences in performance to inform future developments in the way mental health providers are funded so the reimbursement system can incentivise efficiency and quality without creating perverse incentives for providers.

Glossary

A&E	Accident and Emergency
ADL	Activities of Daily Living
AIC	Akaike’s Information Criterion
BIC	Bayesian Information Criterion
CAMHS	Child and Adolescent Mental Health Services
CCGs	Clinical Commissioning Groups
CGIS	Clinical Global Impression Scale
CPA	Care Programme Approach
CQUIN	Commissioning for Quality and Innovation
CRHT	Crisis Resolution and Home Treatment
CROM	Clinician Rated Outcome Measure
CRP	Cluster Review Period
CS	Clinically Significant
CV	Coefficient of Variation
DRGs	Diagnosis Related Groups
EB	Empirical Bayes
ECT	Electroconvulsive Therapy
FMM	Finite Mixture Models
FT	Foundation Trust
FTE	Full Time Equivalent
GLM	Generalized Linear Model
HES	Hospital Episode Statistics
HoNOS	Health of the Nation Outcome Scales
HRGs	Healthcare Resource Groups
HSCIC	Health and Social Care Information Centre
IAPT	Improving Access to Psychological Therapies
IMD	Index of Multiple Deprivation
IRRs	Incidence Rate Ratios
LOS	Length Of Stay
LSOA	Lower Layer Super Output Area
MAPE	Mean Absolute Prediction Error
MAR	Missing At Random
MCAR	Missing Completely At Random
MCMC	Monte Carlo Markov Chain
MFF	Market Forces Factor
MHA	Mental Health Act
MHCT	Mental Health Clustering Tool
MHMDS	Mental Health Minimum Data Set
MHSDS	Mental Health Services Data Set
MNAR	Missing Not At Random
NHS	National Health Service
NTPS	National Tariff Payment System
PbR	Payment by Results

PCTs	Primary Care Trusts
PLICS	Patient Level Information Costing Systems
PRs	Predictive Ratios
PREMS	Patient Related Experience Measures
PROMs	Patient Reported Outcome Measures
RC	Reliable Change
RCSC	Reliable and Clinically Significant Change
RIGLS	Restricted Iterative Generalized Least Squares
RMSE	Root Mean Square Error
SARN	Summary of Assessments of Risk and Need
VHA	Veterans Health Administration

Bibliography

Abas, M. A., J. Vanderpyl, and E. Robinson. 2008. "Socioeconomic deprivation and extended hospitalization in severe mental disorder: a two-year follow-up study." *Psychiatr Serv* 59(3):322-5.

Ahmed, K., N. Iwegbu, M.I. Husain, W.A. Addae, and M. Baggaley. 2015. "Put a lid on patient overspill to private care." *Health Service Journal*.

Appleby, J., J. Thompson, and J. Jabbal. 2015. "The King's Fund Quarterly Monitoring Report 16." London: The King's Fund.

Appleby, L., P. N. Desai, D. J. Luchins, R. D. Gibbons, and D. R. Hedeker. 1993. "Length of stay and recidivism in schizophrenia: a study of public psychiatric hospital patients." *Am J Psychiatry* 150(1):72-6.

Audin, K., F. R. Margison, J. M. Clark, and M. Barkham. 2001. "Value of HoNOS in assessing patient change in NHS psychotherapy and psychological treatment services." *Br J Psychiatry* 178:561-6.

Barnett, Adrian G, Jolieke C van der Pols, and Annette J Dobson. 2005. "Regression to the mean: what it is and how to deal with it." *International Journal of Epidemiology* 34(1):215-20.

Bartlett, J. , and J. Carpenter. 2013. "Module 14: Missing Data Concepts." in *LEMMA (Learning Environment for Multilevel Models and Applications)*: Centre for Multilevel Modelling, University of Bristol.

Ben-Zeev, D., M.A. Young, and P.W. Corrigan. 2010. "DSM-V and the stigma of mental illness." *Journal of Mental Health* 19(4):318-27.

Boden, R., L. Brandt, H. Kieler, M. Andersen, and J. Reutfors. 2011. "Early non-adherence to medication and other risk factors for rehospitalization in schizophrenia and schizoaffective disorder." *Schizophr Res* 133(1-3):36-41.

Browne, William J. 2012. "MCMC estimation in MLwiN Version 2.26." Bristol: Centre for Multilevel Modelling University of Bristol.

Bryan, Karen. 2010. "Policies for reducing delayed discharge from hospital." *British Medical Bulletin* 95(1):33-46.

Buckingham, B., P. Burgess, S. Solomon, J. Pirkis, and K. Eagar. 1998. "Developing a Casemix Classification for Mental Health Services: Summary." Canberra: Commonwealth Department of Health and Family Services.

Burgess, J., A. Street, J. Burgess, and A. Street. 2011. "Measuring organisational performance." in *The Oxford Handbook of Health Economics*, edited by P. C. Smith, S. A. Glied, P. C. Smith, and S. A. Glied. Oxford: Oxford University Press.

Burgess Jr., J.F., and J.M. Hockenberry. 2014. "Can all cause readmission policy improve quality or lower expenditures? A historical perspective on current initiatives." *Health Economics, Policy and Law* 9(2):193-213.

Cameron, C.A., and P.K. Trivedi. 2011. *Microeconometrics Using Stata (Revised Edition)*. Texas: Stata Press.

Canadian Institute for Health Information. 2008. "Hospital Length of Stay and Readmission for Individuals Diagnosed With Schizophrenia: Are They Related?,." Ottawa: Canadian Institute for Health Information.

Capdevielle, D., and K. Ritchie. 2008. "The long and the short of it: are shorter periods of hospitalisation beneficial?" *The British Journal of Psychiatry* 192:164-65.

Capita. 2013. "Payment by Results Data Assurance Framework. Report on Assuring Mental Health Currencies.". England: Capita Health and Wellbeing Limited.

Care Quality Commission. 2009. "Performance Assessment 2008-09." London: Care Quality Commission.
—. 2010. "2009 Inpatient survey “Patient experience” performance indicator calculation." London: Care Quality Commission.
Carey, K. 2000. "A multilevel modelling approach to analysis of patient costs under managed care." *Health Econ* 9(5):435-46.
Carey, K., and J. F. Burgess, Jr. 1999. "On measuring the hospital cost/quality trade-off." *Health Econ* 8(6):509-20.
Christensen, Larry, and Jorge Mendoza. 1986. "A method of assessing change in a single subject: An alteration of the RC index." *Behavior Therapy* 17(3):305-08.
Chung, W., H. S. Chang, S. M. Oh, and C. W. Yoon. 2013. "Factors associated with long-stay status in patients with schizophrenia: an analysis of national databases covering the entire Korean population." *Int J Soc Psychiatry* 59(3):207-16.
Chung, W., S. M. Oh, T. Suh, Y. M. Lee, B. H. Oh, and C. W. Yoon. 2010. "Determinants of length of stay for psychiatric inpatients: analysis of a national database covering the entire Korean elderly population." *Health Policy* 94(2):120-8.
Compton, M. T., J. Craw, and B. E. Rudisch. 2006. "Determinants of inpatient psychiatric length of stay in an urban county hospital." *Psychiatr Q* 77(2):173-88.
Cooper, N. J., P. C. Lambert, K. R. Abrams, and A. J. Sutton. 2007. "Predicting costs over time using Bayesian Markov chain Monte Carlo methods: an application to early inflammatory polyarthritis." *Health Econ* 16(1):37-56.
Cromwell, J., E. M. Drozd, B. Gage, J. Maier, E. Richter, and H. H. Goldman. 2005. "Variation in patient routine costliness in U.S. psychiatric facilities." *J Ment Health Policy Econ* 8(1):15-28.
Csipke, E., C. Flach, P. McCrone, D. Rose, J. Tilley, T. Wykes, and T. Craig. 2014. "Inpatient care 50 years after the process of deinstitutionalisation." *Soc Psychiatry Psychiatr Epidemiol* 49(4):665-71.
Daidone, S., and A. Street. 2013. "How much should be paid for specialised treatment?" *Soc Sci Med* 84:110-8.
Dausey, D. J., R. A. Rosenheck, and A. F. Lehman. 2002. "Preadmission care as a new mental health performance indicator." *Psychiatr Serv* 53(11):1451-5.
Dekker, J., J. Peen, A. Goris, H. Heijnen, and H. Kwakman. 1997. "Social deprivation and psychiatric admission rates in Amsterdam." *Soc Psychiatry Psychiatr Epidemiol* 32(8):485-92.
Department of Health. 2010. "Practical Guide to Preparing for Mental Health Payment by Results." London: Department of Health PbR Development Team.
—. 2011. "No Health Without Mental Health: A Cross-Government Mental Health Outcomes Strategy for People of All Ages.". London: Department of Health.
—. 2012. "Reference costs 2011-12." Leeds: Department of Health.
—. 2014. "Closing the gap: priorities for essential change in mental health." London: Department of Health.
Department of Health and Human Services Agency for Healthcare Research and Quality. 2002. "Guide to Inpatient Quality Indicators: Quality of Care in Hospitals – Volume, Mortality, and Utilisation." Washington, D.C.

Department of Health and NHS England. 2014. "Achieving Better Access to Mental Health Services by 2020." London: Department of Health.
Department of Health Payment by Results team. 2013a. "Key steps for successful implementation of Mental Health Payment by Results." Leeds: Department of Health.
—. 2013b. "Mental Health Payment by Results Guidance for 2013-14." Leeds: Department of Health.
Dickey, B., and S. L. Normand. 2004. "Toward a model for testing the relationship between quality of care and costs." *J Ment Health Policy Econ* 7(1):15-21.
Douzenis, A., D. Seretis, S. Nika, P. Nikolaidou, A. Papadopoulou, E. N. Rizos, C. Christodoulou, C. Tsopelas, D. Mitchell, and L. Lykouras. 2012. "Factors affecting hospital stay in psychiatric patients: the role of active comorbidity." *BMC Health Serv Res* 12:166.
Dow, M. G., T. L. Boaz, and D. Thornton. 2001. "Risk adjustment of Florida mental health outcomes data: concepts, methods, and results." *J Behav Health Serv Res* 28(3):258-72.
Drozd, E. M., J. Cromwell, B. Gage, J. Maier, L. M. Greenwald, and H. H. Goldman. 2006. "Patient casemix classification for medicare psychiatric prospective payment." *Am J Psychiatry* 163(4):724-32.
Druss, B.G., C.L. Miller, H.A. Pincus, and S. Shih. 2004. "The volume-quality relationship of mental health care: does practice make perfect?" *American Journal of Psychiatry* 161(12):2282-6.
Durbin, J., E. Lin, C. Layne, and M. Teed. 2007. "Is readmission a valid indicator of the quality of inpatient psychiatric care?" *J Behav Health Serv Res* 34(2):137-50.
Eagar, K., P. Gaines, P. Burgess, J. Green, A. Bower, B. Buckingham, and G. Mellsop. 2004. "Developing a New Zealand casemix classification for mental health services." *World Psychiatry* 3(3):172-7.
Edwards, N. 2014. "The squeeze is on." *Health Services Journal*.
Eisen, S. V., G. Ranganathan, P. Seal, and A. Spiro, 3rd. 2007. "Measuring clinically meaningful change following mental health treatment." *J Behav Health Serv Res* 34(3):272-89.
Ettner, S. L., R. G. Frank, T. G. McGuire, and R. C. Hermann. 2001. "Risk adjustment alternatives in paying for behavioral health care under Medicaid." *Health Serv Res* 36(4):793-811.
Ettner, S. L., R. G. Frank, T. G. McGuire, J. P. Newhouse, and E. H. Notman. 1998. "Risk adjustment of mental health and substance abuse payments." *Inquiry* 35(2):223-39.
Ettner, S. L., and E. H. Notman. 1997. "How well do ambulatory care groups predict expenditures on mental health and substance abuse patients?" *Adm Policy Ment Health* 24(4):339-57.
Evans-Lacko, S. E., M. Jarrett, P. McCrone, and G. Thornicroft. 2008. "Clinical pathways in psychiatry." Pp. 4-5 in *Br J Psychiatry*. England.
Evans, Chris, Frank Margison, and Michael Barkham. 1998. "The contribution of reliable and clinically significant change methods to evidence-based mental health." *Evidence Based Mental Health* 1(3):70-72.
Fairbairn, Andrew. 2007. "Payment by results in mental health: the current state of play in England." *Advances in Psychiatric Treatment* 13(1):3-6.

Figueroa, R., J. Harman, and J. Engberg. 2004. "Use of claims data to examine the impact of length of inpatient psychiatric stay on readmission rate." *Psychiatr Serv* 55(5):560-5.

Fong Chan, L., and O. Lieh Yan. 2010. "Factors influencing inpatient duration among insanity acquittees in a Malaysian mental institution." *ASEAN Journal of Psychiatry* 11(1):25-35.

Forti, A., C. Nas, A. van Geldrop, G. Franx, I. Petrea, Y. van Strien, and P. Jeurissen. 2014. "Mental Health Analysis Profiles (MhAPs): Netherlands." in *OECD Health Working Papers No. 71*. Paris: OECD Publishing.

Fries, Brant E., Paul W. Durance, David R. Nerenz, and Marie L. F. Ashcraft. 1993. "A Comprehensive Payment Model for Short- and Long-Stay Psychiatric Patients." *Health Care Financing Review* 15(2):31-50.

Gaines, P., A. Bower, B. Buckingham, K. Eagar, P. Burgess, and J. Green. 2003. "New Zealand Mental HealthClassification and Outcomes Study: Final Report." New Zealand: Health Research Council of New Zealand.

Gaughan, J., A. Mason, A. Street, and P. Ward. 2012. "English hospitals can improve their use of resources: an analysis of costs and length of stay for ten treatments." in *CHE Research Paper 78*. York: Centre for Health Economics, University of York.

Goldstein, H. 1989. "Restricted unbiased iterative generalized least-squares estimation." *Biometrika* 76(3):622-23.

Goldstein, H., and M.J.R. Healy. 1995. "The Graphical Presentation of a Collection of Means." *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 158(1):175-77.

Gomes, M., N. Gutacker, C. Bojke, and A. Street. 2015. "Addressing Missing Data in Patient-Reported Outcome Measures (PROMS): Implications for the Use of PROMS for Comparing Provider Performance." *Health Econ*.

Gutacker, N., C. Bojke, S. Daidone, N. J. Devlin, D. Parkin, and A. Street. 2013a. "Truly inefficient or providing better quality of care? Analysing the relationship between risk-adjusted hospital costs and patients' health outcomes." *Health Econ* 22(8):931-47.

Gutacker, N., C. Bojke, S. Daidone, N. Devlin, and A. Street. 2013b. "Hospital variation in patient-reported outcomes at the level of EQ-5D dimensions: evidence from England." *Med Decis Making* 33(6):804-18.

Gutacker, N., and A. Street. 2015. "Multidimensional performance assessment using dominance criteria." in *CHE Research Paper 115*. York: Centre for Health Economics, University of York.

Haas, L., T. Stargardt, J. Schreyoegg, R. Schlosser, B. F. Klapp, and G. Danzer. 2013. "The trade-off between costs and quality of care in the treatment of psychosomatic patients with somatoform pain disorder." *Appl Health Econ Health Policy* 11(4):359-68.

Harman, J. S., B. J. Cuffel, and K. J. Kelleher. 2004. "Profiling hospitals for length of stay for treatment of psychiatric disorders." *J Behav Health Serv Res* 31(1):66-74.

Hauck, K., and A. Street. 2006. "Performance assessment in the context of multiple objectives: a multivariate multilevel analysis." *J Health Econ* 25(6):1029-48.

Health and Social Care Information Centre. 2011. "Emergency readmissions to hospital within 28 days of discharge." Leeds UK,: Health and Social Care Information Centre.

—. 2013. "“Lower Layer Super Output Area”." in *NHS Data Dictionary*. Leeds: Health and Social Care Information Centre.
—. 2014. "Mental Health Bulletin. Annual Report from MHMDS Returns 2013/14.". Leeds: Health and Social Care Information Centre.
Health and Social Care Information Centre Clinical Indicators Team. 2015. "Indicator Specification: CCG OIS 3.16 Unplanned readmissions to mental health services within 30 days of a mental health inpatient discharge in people aged 17 and over.". Leeds: Health and Social Care Information Centre.
Hendryx, M. 2008. "State mental health funding and mental health system performance." *J Ment Health Policy Econ* 11(1):17-25.
Hendryx, M. S., A. Beigel, and A. Doucette. 2001. "Introduction: risk-adjustment issues in mental health services." *J Behav Health Serv Res* 28(3):225-34.
Hendryx, M. S., D. G. Dyck, and D. Srebnik. 1999. "Risk-adjusted outcome models for public mental health outpatient programs." *Health Serv Res* 34(1 Pt 1):171-95.
Hendryx, M. S., and G. B. Teague. 2001. "Comparing alternative risk-adjustment models." *J Behav Health Serv Res* 28(3):247-57.
Hermann, R. C., C. K. Rollins, and J. A. Chan. 2007. "Risk-adjusting outcomes of mental health and substance-related care: a review of the literature." *Harv Rev Psychiatry* 15(2):52-69.
Hodgson, R.E., M. Lewis, and J. Boardman. 2000. "The prediction of in-patient length of stay for acute psychiatric admissions." *Journal of Mental Health* 9(2):145-53.
Horgan, C. , and S.F Jencks. 1987. "Research on psychiatric classification and payment systems." *Med Care* 87(25):22-36.
House of Commons Health Committee. 2013. "Post-legislative scrutiny of the Mental Health Act 2007 First Report of Session 2013-14." London: House of Commons.
Huntley, D. A., D. W. Cho, J. Christman, and J. G. Csernansky. 1998. "Predicting length of stay in an acute psychiatric hospital." *Psychiatr Serv* 49(8):1049-53.
Hvenegaard, A., J. N. Arendt, A. Street, and D. Gyrd-Hansen. 2011. "Exploring the relationship between costs and quality: does the joint evaluation of costs and quality alter the ranking of Danish hospital departments?" *Eur J Health Econ* 12(6):541-51.
Hvenegaard, A., A. Street, T. H. Sorensen, and D. Gyrd-Hansen. 2009. "Comparing hospital costs: what is gained by accounting for more than a case-mix index?" *Soc Sci Med* 69(4):640-7.
Imai, H., J. Hosomi, H. Nakao, H. Tsukino, T. Katoh, T. Itoh, and T. Yoshida. 2005. "Characteristics of psychiatric hospitals associated with length of stay in Japan." *Health Policy* 74(2):115-21.
Jacobs, R. 2009. "Investigating Patient Outcome Measures in Mental Health." in *CHE Research Paper 48*. York: Centre for Health Economics, University of York.
—. 2010. "Some economic and policy considerations for outcome measurement." in *Outcome Measurement in Mental Health Theory and Practice*, edited by T. Trauer. UK: Cambridge University Press.
—. 2014. "Payment by results for mental health services: economic considerations of case-mix funding." *Advances in Psychiatric Treatment* 20:155 - 64.
Jacobs, R., N. Gutacker, A. Mason, A. Goddard, H. Gravelle, T. Kendrick, and S. Gilbody. 2015. "Determinants of hospital length of stay for people with

serious mental illness in England and implications for payment systems: a regression analysis." *BMC health services research* 15(439).
Jacobson, Neil S., William C. Follette, and Dirk Revenstorf. 1984. "Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance." *Behavior Therapy* 15(4):336-52.
Jencks, S. F., C. Horgan, H. H. Goldman, and C. A. Taube. 1987. "Bringing excluded psychiatric facilities under the Medicare Prospective Payment System. A review of research evidence and policy options." *Med Care* 25(9 Suppl):S1-51.
Jones, A.M. 2005. *Applied Econometrics for Health Economists: A Practical Guide*. London: Office of Health Economics.
—. 2010. "Models for Health Care." *HEDG Working Paper* 10(01).
Kallert, T. W., M. Glockner, and M. Schutzwohl. 2008. "Involuntary vs. voluntary hospital admission. A systematic literature review on outcome diversity." *Eur Arch Psychiatry Clin Neurosci* 258(4):195-209.
Kapur, K., A. S. Young, and D. Murata. 2000. "Risk adjustment for high utilizers of public mental health care." *J Ment Health Policy Econ* 3(3):129-37.
Kobel, C., J. Thuilliez, M. Bellanger, and K-P. Pfeiffer. 2011. "DRG systems and similar patient classification systems in Europe." in *Diagnosis-Related Groups in Europe Moving towards transparency, efficiency and quality in hospitals*, edited by R. Busse, A. Geissler, W. Quentin, and M. Wiley. England: Open University Press.
Korkeila, J. A., V. Lehtinen, T. Tuori, and H. Helenius. 1998. "Frequently hospitalised psychiatric patients: a study of predictive factors." *Soc Psychiatry Psychiatr Epidemiol* 33(11):528-34.
Kramer, TeresaL, RichardB Evans, Reid Landes, Michael Mancino, BrendaM Booth, and G. Richard Smith. 2001. "Comparing outcomes of routine care for depression: The dilemma of case-mix adjustment." *J Behav Health Serv Res* 28(3):287-300.
Kristensen, T., M. Laudicella, C. Ejersted, and A. Street. 2010. "Cost variation in diabetes care delivered in English hospitals." *Diabet Med* 27(8):949-57.
Lakhani, A., J. Coles, D. Eayres, C. Spence, and C. Sanderson. 2005. "Creative use of existing clinical and health outcomes data to assess NHS performance in England: Part 2 - more challenging aspects of monitoring." *BMJ* 330:1486-92.
Laudicella, Mauro, Paolo Li Donni, and Peter C. Smith. 2013. "Hospital readmission rates: Signal of failure or success?" *Journal of Health Economics* 32(5):909-21.
Laudicella, Mauro, Kim Rose Olsen, and Andrew Street. 2010. "Examining cost variation across hospital departments–a two-stage multi-level approach using patient-level data." *Social Science & Medicine* 71(10):1872-81.
Lay, B., C. Lauber, and W. Rossler. 2006. "Prediction of in-patient use in first-admitted patients with psychosis." *Eur Psychiatry* 21(6):401-9.
Leckie, G. 2013. "Module 12: Cross-Classified Multilevel Models." in *LEMMA (Learning Environment for Multilevel Methodology and Applications)*. Centre for Multilevel Modelling, University of Bristol.
Leckie, G., and C. Charlton. 2012. "runmlwin: A Program to Run the MLwiN Multilevel Modeling Software from within Stata." *Journal of Statistical Software* 52(11).

Lerner, Y., and N. Zilber. 2010. "Predictors of cumulative length of psychiatric inpatient stay over one year: a national case register study." *Isr J Psychiatry Relat Sci* 47(4):304-7.

Lin, H. C., W. H. Tian, C. S. Chen, T. C. Liu, S. Y. Tsai, and H. C. Lee. 2006. "The association between readmission rates and length of stay for schizophrenia: a 3-year population-based study." *Schizophr Res* 83(2-3):211-4.

Lintern, S. 2014a. "Analysis reveals mental health trust funding cuts." *Health Service Journal*.

—. 2014b. "Providers allege 'institutional bias'." *Health Services Journal*.

—. 2014c. "Robert Francis joins critics in mental health funding row." *Health Services Journal*.

—. 2015a. "Trusts accuse CCGs of mental health squeeze." *Health Service Journal*.

—. 2015b. "Trusts fear CCGs will withhold funds." *Health Services Journal*.

Martin, S., and P. Smith. 1996. "Explaining variations in inpatient length of stay in the National Health Service." *J Health Econ* 15(3):279-304.

Mason, A., and M. Goddard. 2009. "Payment by results in mental health: a review of the international literature and an economic assessment of the approach in the English NHS." in *CHE Research Paper 50*. York: Centre for Health Economics, University of York.

Mason, A., M. Goddard, L. Myers, and R. Verzulli. 2011. "Navigating uncharted waters? How international experience can inform the funding of mental health care in England." *J Ment Health* 20(3):234-48.

Mason, A., P. Ward, and A. Street. 2011. "England: The Healthcare Resource Group system." in *Diagnosis-Related Groups in Europe Moving towards transparency, efficiency and quality in hospitals*, edited by R. Busse, A. Geissler, W. Quentin, and M. Wiley. England: Open University Press.

McCrone, P. 1994. "Allocating resources in psychiatric hospitals according to casemix." *The Psychiatric Bulletin (United Kingdom)* 18:212-13.

McCrone, P., and P. Lorusso. 1999. "A comparison of psychiatric casemix in the UK and Italy." *Soc Psychiatry Psychiatr Epidemiol* 34(8):432-6.

McCrone, P., and M. Phelan. 1994. "Diagnosis and length of psychiatric in-patient stay." *Psychol Med* 24(4):1025-30.

McCrone, P., G. Thornicroft, S. Parkman, D. Nathaniel-James, and W. Ojurongbe. 1998. "Predictors of mental health service costs for representative cases of psychosis in south London." *Psychol Med* 28(1):159-64.

McCrone, Paul. 1995. "Predicting mental health service use: Diagnosis based systems and alternatives." *Journal of Mental Health* 4(1):31-40.

McCrone, Paul, Sonia Johnson, and Graham Thornicroft. 2001. "Predicting the Costs of Community Care for Individuals With Severe Mental Illness in South London." *Schizophrenia Bulletin* 27(4):653-60.

McGuire, T. G., B. Dickey, G. E. Shively, and I. Strumwasser. 1987. "Differences in resource use and cost among facilities treating alcohol, drug abuse, and mental disorders: implications for design of a prospective payment system." *Am J Psychiatry* 144(5):616-20.

Miles, Ashley A., R. Walter Heinrichs, Narmeen Ammari, Leah Hartman, Stephanie McDermid Vaz, and Eva Muharib. 2014. "Stability and change in symptoms, cognition, and community outcome in schizophrenia." *Schizophr Res* 152(2–3):435-39.

Mitchell, J. B., B. Dickey, B. Liptzin, and L. I. Sederer. 1987. "Bringing psychiatric patients into the Medicare prospective payment system: alternatives to DRGs." *Am J Psychiatry* 144(5):610-5.

Monitor. 2013. "Closing the NHS funding gap: how to get better value health care for patients." London: Monitor.

—. 2015. "Improving the costing of NHS services: proposals for 2015 to 2021 Our response to feedback.". London: Monitor.

Monitor, and NHS England. 2014. "2015/16 National Tariff Payment System: A consultation notice." London: Monitor.

Monitor and NHS England. 2013a. "2014/15 National Tariff Payment System." London: Monitor.

—. 2013b. "Guidance on mental health currencies and payment." London: Monitor.

—. 2013c. "A guide to the Market Forces Factor." London: Monitor.

—. 2013d. "National Tariff 2014/15 Payment System Annex 7C Mental health clustering tool booklet." London: Monitor.

—. 2013e. "National Tariff 2014/15: A Consultation Notice.". London: Monitor.

Montez-Rath, Maria, CindyL Christiansen, SusanL Ettner, Susan Loveland, and AmyK Rosen. 2006. "Performance of statistical models to predict mental health and substance abuse cost." *BMC Medical Research Methodology* 6(1):1-11.

NHS Choices. 2015a. "NHS authorities and trusts: what is a foundation trust?" Pp. http://www.nhs.uk/NHSEngland/thenhs/about/Pages/authoritiesandtrusts.aspx.

—. 2015b. "NHS authorities and trusts: What is a mental health trust." Pp. http://www.nhs.uk/ServiceDirectories/Pages/MentalHealthTrustListing.aspx.

NHS England. 2014a. "Choice in Mental Health Care." NHS England.

—. 2014b. "Five Year Forward View." NHS England.

—. 2014c. "The Forward View into Action: Planning for 2015/16." NHS England.

—. 2015. "Commissioning for Quality and Innovation (CQUIN) Guidance for 2015/16.". Leeds: NHS England.

Noble, M., McLennan, D., Wilkinson, K., Whitworth, A., Barnes, H. and Dibben, C. 2008. "The English Indices of Deprivation 2007." London: Communities and Local Government.

Nuttall, D., D. Parkin, and N. Devlin. 2015. "Inter-provider comparison of patient-reported outcomes: developing an adjustment to account for differences in patient case mix." *Health Econ* 24(1):41-54.

OECD. 2015. "OECD Health Statistics 2015." Paris: OECD Publishing.

Oiesvold, T., O. Saarento, S. Sytema, L. Christiansen, G. Gostas, O. Lonnerberg, S. Muus, M. Sandlund, and L. Hansson. 1999. "The Nordic Comparative Study on Sectorized Psychiatry--length of in-patient stay." *Acta Psychiatr Scand* 100(3):220-8.

Olsen, K. R., and A. Street. 2008. "The analysis of efficiency among a small number of organisations: how inferences can be improved by exploiting patient-level data." *Health Econ* 17(6):671-81.

Or, Z., and U. Hakkinen. 2011. "DRGs and quality: For better or worse?" in *Diganosis-Related Groups in Europe Moving towards transparency, efficiency and quality in hospitals*, edited by R. Busse, A. Geissler, W. Quentin, and M. Wiley. England: Open University Press.

Oyebode, Femi. 2007. "Payment by volume (not results): Invited commentary on… Payment by results in mental health." *Advances in Psychiatric Treatment* 13(1):7-9.

Padgett, D. K., C. Patrick, B. J. Burns, and H. J. Schlesinger. 1994. "Ethnic differences in use of inpatient mental health services by blacks, whites, and Hispanics in a national insured population." *Health Serv Res* 29(2):135-53.

Parabiaghi, A., A. Barbato, B. D'Avanzo, A. Erlicher, and A. Lora. 2005. "Assessing reliable and clinically significant change on Health of the Nation Outcome Scales: method for displaying longitudinal data." *Aust N Z J Psychiatry* 39(8):719-25.

Parabiaghi, A., F. Rapisarda, B. D'Avanzo, A. Erlicher, A. Lora, and A. Barbato. 2011. "Measuring clinical change in routine mental health care: differences between first time and longer term service users." *Aust N Z J Psychiatry* 45(7):558-68.

Peiro, S., G. Gomez, M. Navarro, I. Guadarrama, and J. Rejas. 2004. "Length of stay and antipsychotic treatment costs of patients with acute psychosis admitted to hospital in Spain. Description and associated factors. The Psychosp study." *Soc Psychiatry Psychiatr Epidemiol* 39(7):507-13.

Pertile, R., V. Donisi, L. Grigoletti, A. Angelozzi, G. Zamengo, G. Zulian, and F. Amaddeo. 2011. "DRGs and other patient-, service- and area-level factors influencing length of stay in acute psychiatric wards: the Veneto Region experience." *Soc Psychiatry Psychiatr Epidemiol* 46(7):651-60.

PriceWaterhouseCoopers. 2012. "An Evaluation of the Reimbursement System for NHS-Funded Care: Report for Monitor." UK: PriceWaterhouseCoopers LLP.

Rabash, J., C. Charlton, W.J. Browne, M. Healy, and B. Cameron. 2009. *MLwiN Version 2.1.* Centre for Multilevel Modelling: University of Bristol.

Rabash, J., F. Steele, W.J. Browne, and H. Goldstein. 2012. "A User's Guide to MLwiN Version 2.26." Bristol: Centre for Multilevel Modelling, University of Bristol.

Rabe-Hesketh, S., A. Skrondal, and A. Pickle. 2004. "GLLAMM Manual." *U.C. Berkeley Division of Biostatistics Working Paper Series* 160.

Robst, J. 2009. "Development of a Medicaid behavioral health case-mix model." *Eval Rev* 33(6):519-38.

Rosen, A. K., S. Chatterjee, M. E. Glickman, A. Spiro, 3rd, P. Seal, and S. V. Eisen. 2010. "Improving risk adjustment of self-reported mental health outcomes." *J Behav Health Serv Res* 37(3):291-306.

Rothbard, Aileen B., and Arie P. Schinnar. 1996. "Community determinants of psychiatric hospitalization and length of stay." *Socio-Economic Planning Sciences* 30(1):27-38.

Royal College of Psychiatrists. 2011. "Do the right thing: how to judge a good ward. Ten standards for adult in-patient mental health care." London: Royal College of Psychiatrists.

Rubin, D.B. 1987. *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.

Sabes-Figuera, R., P. McCrone, J. Sharac, E. Csipke, T. K. Craig, D. S. Rose, D. Pearman, and T. Wykes. 2012. "Developing a tool for collecting and costing activity data on psychiatric inpatient wards." *Epidemiol Psychiatr Sci* 21(4):393-9.

Sartorius, N. 2002. "Iatrogenic stigma of mental illness." *British Medical Journal* 324:1470-71.

Schreyogg, J., and T. Stargardt. 2010. "The trade-off between costs and outcomes: the case of acute myocardial infarction." *Health Serv Res* 45(6 Pt 1):1585-601.
Schulz, R. I., J. R. Greenley, and R. W. Peterson. 1983. "Management, cost, and quality of acute inpatient psychiatric services." *Med Care* 21(9):911-28.
Self R., J. Painter and R. Davis. 2008. "A report on the development of a Mental Health Currency Model (including suggestions on the development and testing of PbR tariffs)." London: Department of Health.
Shleifer, A. 1985. "A theory of yardstick competition." *RAND Journal of Economics* 16(3):319-27.
Skrondal, A., and S. Rabe-Hesketh. 2009. "Prediction in multilevel generalized linear models." *Journal of the Royal Statistical Society* 172:659-87.
Slade, M. 2010. "Outcome measurement in England." in *Outcome Measurement in Mental Health Theory and Practice*, edited by T. Trauer. UK: Cambridge University Press.
Slade, M., P. McCrone, E. Kuipers, M. Leese, S. Cahill, A. Parabiaghi, S. Priebe, and G. Thornicroft. 2006. "Use of standardised outcome measures in adult mental health services: Randomised controlled trial." *British Journal of Psychiatry* 189:330-36.
Sloan, K. L., M. E. Montez-Rath, A. Spiro, 3rd, C. L. Christiansen, S. Loveland, P. Shokeen, L. Herz, S. Eisen, J. N. Breckenridge, and A. K. Rosen. 2006. "Development and validation of a psychiatric case-mix system." *Med Care* 44(6):568-80.
Smith, G., K. Nicholson, C. Fitch, and L. Mynors-Wallis. 2015. "The Commission to Review the Provision of Acute Inpatient Psychiatric Care for Adults in England, Wales and Northern Ireland. Background briefing paper.": The Commission to review the provision of acute inpatient psychiatric care for adults.
Smith, P. C., A. Street, P. C. Smith, and A. Street. 2012. "Concepts and challenges in measuring the performance of health care organizations." Pp. 341-49 in *The Elgar Companion to Health Economics*, edited by A. Jones and A. Jones. Cheltenham: Edward Elgar.
Smith, P.C., and A.D. Street. 2013. "On the uses of routine patient-reported health outcome data." *Health Econ* 22(2):119-31.
Speak, B., and P. Hay. 2012. "The utility of the Mental Health Clustering Tool (MHCT) as a Generic Outcome Measure for the Payment by Results Quality and Outcomes Framework in Working Age and Older Adult Mental Health Services." Teeside University Tees, Esk and Wear Valleys NHS Foundation Trust.
StataCorp. 2013. "Stata: Release 13." College Station Texas,: StataCorp L.P.
Steele, F. 2008. "Module 5: Introduction to Multilevel Modelling Concepts." in *LEMMA (Learning Environment for Multilevel Models and Applications)*. Bristol: Centre for Multilevel Modelling, University of Bristol.
—. 2014. "Module 15: Multilevel Modelling of Repeated Measures Data." in *LEMMA (Learning Environment for Multilevel Models and Applications)*. Bristol: Centre for Multilevel Modelling, University of Bristol.
Stevens, A., K. Hammer, and G. Buchkremer. 2001. "A statistical model for length of psychiatric in-patient treatment and an analysis of contributing factors." *Acta Psychiatr Scand* 103(3):203-11.
Steyerberg, E.W. 2009. *Clinical Prediction Models. A Practical Approach to Development, Validation and Updating.*: Springer.

Street, A., J. O'Reilly, P. Ward, and A. Mason. 2011. "DRG-based hospital payment and efficiency: Theory, evidence and challenges." in *Diagnosis-Related Groups in Europe Moving towards transparency, efficiency and quality in hospitals*, edited by R. Busse, A. Geissler, W. Quentin, and M. Wiley. England: Open University Press.

Sutton, M. , S.R. Kristensen, Y-S. Lau, G. Glover, W. Whittaker, J. Wildman, H. Gravelle, and P.C. Smith. 2012. "Estimation of a formula for mental health services based on person-level data." London: Department of Health.

Swan-Tan, S., L. Serden, A. Geissler, M. van Ineveld, K. Redekop, M. Heurgren, and L. Hakkaart-van Roijen. 2011. "DRGs and cost accounting: Which is driving which?" in *Diagnosis-Related Groups in Europe Moving towards transparency, efficiency and quality in hospitals*, edited by R. Busse, A. Geissler, W. Quentin, and M. Wiley. England: Open University Press.

Sytema, S., and P. Burgess. 1999. "Continuity of care and readmission in two service systems: a comparative Victorian and Groningen case-register study." *Acta Psychiatr Scand* 100(3):212-9.

Sytema, S., P. Burgess, and M. Tansella. 2002. "Does Community Care Decrease Length of Stay and Risk of Rehospitalization in New Patients With Schizophrenic Disorders? A Comparative Case Register Study in Groningen, the Netherlands; Victoria, Australia; and South-Verona, Italy." *Schizophrenia Bulletin* 28(2):273-81.

The Centre for Economic Performance's Mental Health Policy Group. 2012. "How Mental Illness Loses Out in the NHS." London: London School of Economics and Political Science.

The Commission to review the provision of acute inpatient psychiatric care for adults. 2015. "Improving acute inpatient psychiatric care for adults in England. Interim report.": The Commission to review the provision of acute inpatient psychiatric care for adults.

Timimi, S. 2014. "No more psychiatric labels: why formal psychiatric diagnostic systems should be abolished." *International Journal of Clinical and Health Psychology*.

Trauer, T. 2010a. "Future directions." in *Outcome Measurement in Mental Health Theory and Practice*, edited by T. Trauer. UK: Cambridge University Press.

—. 2010b. "Outcome measurement - applications and utility." in *Outcome Measurement in Mental Health Theory and Practice*, edited by T. Trauer. UK: Cambridge University Press.

—. 2010c. *Outcome Measurement in Mental Health Theory and Practice*. UK: Cambridge University Press.

Trevithick, Liam, Jon Painter, and Patrick Keown. 2014. "Mental health clustering and diagnosis in psychiatric in-patients." *The Psychiatrist*.

Tulloch, A. D., A. S. David, and G. Thornicroft. 2015. "Exploring the predictors of early readmission to psychiatric hospital." *Epidemiol Psychiatr Sci*:1-13.

Tulloch, A. D., P. Fearon, and A. S. David. 2011. "Length of stay of general psychiatric inpatients in the United States: systematic review." *Adm Policy Ment Health* 38(3):155-68.

Verzulli, R., R. Jacobs, and M Goddard. 2011. "Do Hospitals Respond to Greater Autonomy? Evidence from the English NHS." in *CHE Research Paper 64*. York: Centre for Health Economics, University of York.

Wakefield, J.C. 2010. "Misdiagnosing normality: Psychiatry's failure to address the problem of false positive diagnoses of mental disorder in a changing professional environment." *Journal of Mental Health* 19(4):337-51.

Weech-Maldonado, R., D. Shea, and V. Mor. 2006. "The relationship between quality of care and costs in nursing homes." *Am J Med Qual* 21(1):40-8.

Weich, S. , O. McBride, L. Twigg, P. Keown, E. Cyhlarova, D. Crepaz-Keay, H. Parsons, J. Scott, and K. Bhui. 2014. "Variation in compulsory psychiatric inpatient admission in England: a cross-sectional, multilevel analysis." *Health Services and Delivery Research* 2(49).

Williams, P., E. Csipke, D. Rose, L. Koeser, P. McCrone, A. D. Tulloch, G. Salaminios, T. Wykes, and T. Craig. 2014. "Efficacy of a triage system to reduce length of hospital stay." *Br J Psychiatry* 204(6):480-5.

Wing, J.K., R.M. Curtis, and A.S. Beevor. 1996. "HoNOS - Health of the Nation Outcome Scales: Report of Research and Development." London: Royal College of Psychiatrists Research Unit.

Wing, John, Roy Curtis, and Anne Beevor. 1994. "'Health of the Nation': measuring mental health outcomes." *Psychiatric Bulletin* 18(11):690-91.

Wolff, J., P. McCrone, L. Koeser, C. Normann, and A. Patel. 2015a. "Cost drivers of inpatient mental health care: a systematic review." *Epidemiol Psychiatr Sci* 24(01):78-89.

Wolff, Jan, Paul McCrone, Anita Patel, Klaus Kaier, and Claus Normann. 2015b. "Predictors of length of stay in psychiatry: analyses of electronic medical records." *BMC Psychiatry* 15(1):1-7.

Wolpert, M. 2014. "Uses and abuses of patient reported outcome measures (PROMs): potential iatrogenic impact of PROMs implementation and how it can be mitigated." *Adm Policy Ment Health* 41(2):141-5.

Woolridge, J.M. 2010. *Econometric analysis of cross section and panel data*. Cambridge Massachusetts: MIT Press.

Wykes, T., and F. Callard. 2010. "Diagnosis, diagnosis, diagnosis: towards DSM-5." *Journal of Mental Health* 19(4):301-04.

Yeomans, D. 2014. "Clustering in mental health payment by results: a critical summary for the clinician." *Advances in Psychiatric Treatment* (20):227-34.

Zhang, J., C. Harvey, and C. Andrew. 2011. "Factors associated with length of stay and the risk of readmission in an acute psychiatric inpatient facility: a retrospective study." *Aust N Z J Psychiatry* 45(7):578-85.