Article:

Moran, Valerie and Jacobs, Rowena orcid.org/0000-0001-5225-6321 (2017) Costs and performance of English mental health providers. Journal of Mental Health Policy and Economics. pp. 83-94. ISSN 1099-176X

Investigating variations in cost performance of English mental health care providers

Authors: Moran, Valerie; Jacobs, Rowena.

Author accepted manuscript version; accepted by: The Journal of Mental Health Policy and Economics, 3 April 2017.

Abstract

Background

Despite limited resources in mental health care, there is little research exploring variations in cost performance across mental health care providers. In England, a prospective payment system for mental health care based on patient needs has been introduced with the potential to incentivise providers to control costs. The units of payment under the new system are 21 care clusters. Patients are allocated to a cluster by clinicians, and each cluster has a maximum review period.

Aims of the Study

The aim of this research is to explain variations in cluster costs between mental health providers using observable patient demographic, need, social and treatment variables. We also investigate if provider-level variables explain differences in costs. The residual variation in cluster costs is compared across providers to provide insights into which providers may gain or lose under the new financial regime.

Methods

The main data source is the Mental Health Minimum Data Set (MHMDS) for England for the years 2011/12 and 2012/13. Our unit of observation is the period of time spent in a care cluster and costs associated with the cluster review period are calculated from NHS Reference

Cost data. Costs are modelled using multi-level log-linear and generalised linear models. The residual variation in costs at the provider level is quantified using Empirical Bayes estimates and comparative standard errors used to rank and compare providers.

Results

There are wide variations in costs across providers. Characteristics associated with higher costs include older age, black ethnicity, admission under the Mental Health Act, and higher need as reflected in the care clusters. Provider type, size, occupancy and the proportion of formal admissions at the provider-level are also found to be significantly associated with costs. After controlling for patient- and provider-level variables, significant residual variation in costs remains at the provider level.

Discussion and Limitations

The results suggest that some providers may have to increase efficiency in order to remain financially viable if providers are paid national fixed prices (tariffs) under the new payment system. Although the classification system for payment is not based on diagnosis, a limitation of the study is the inability to explore the effect of diagnosis due to poor coding in the MHMDS.

Implications for Health Care Provision and Use

We find that some mental health care providers in England are associated with higher costs of provision after controlling for characteristics of service users and providers. These higher costs may be associated with higher quality care or with inefficient provision of care.

Implications for Health Policies

The introduction of a national tariff is likely to provide a strong incentive to reduce costs. Policies may need to consider safe-guarding local health economies if some providers make substantial losses under the new payment regime.

Implications for Further Research

Future research should consider the relationship between costs and quality to ascertain whether reducing costs may potentially negatively impact patient outcomes.

INTRODUCTION

Internationally, there is little research to date exploring variations in cost performance across mental health care providers. Nevertheless this is an important issue that warrants investigation as resources are scarce in mental health care; in Europe the median percentage of the health budget allocated to mental health is only 5% [1]. In the UK, mental health problems contribute to 23% of the total burden of illness, yet mental health receives only 13% of the NHS budget [2]. Therefore it is imperative that the scarce resources available for mental health care are used efficiently and that value-for-money is achieved. Provider payment is a key policy lever that can be used to achieve these objectives. A number of countries including Germany, the Netherlands, Switzerland and the US have implemented prospective payment systems for mental health providers based on activity and/or treatment duration in order to incentivise efficiency [3-6].

In England, mental health care is undergoing a substantial process of reform in the way providers are reimbursed with a move from a system of block contracts towards a prospective payment system based on patient episodes known as the National Tariff Payment System (NTPS). A block contract is akin to a global budget i.e. a provider is paid a fixed amount of money for a defined period of time – usually a financial year [7]. The block contract or budget is agreed between commissioners (who purchase and organise care) and providers and is usually negotiated on the basis of previous expenditure [8]. The use of a block contract can place constraints on the amount a provider can spend [7] and therefore limits the financial risk faced by the commissioner or purchaser of health care. Recent policy has indicated a desire to move away from block contracts for mental health [9,10]. A block contract or global budget does not incentivise the provision of higher quality care or continuity of care [9] and does not incentivise providers to increase activity [11]. Moreover, block contracts inhibit transparency as it is difficult to identify the costs and outcomes of services delivered [9].

Under the episodic payment method, providers will be paid a price agreed prospectively for all mental health related care delivered during a defined time period or episode of care [12]. The move towards an episodic payment system increases transparency around funding as providers will be reimbursed based on the amount and quality of care they deliver [12]. However, this payment method shifts risk to the commissioner or purchaser of health care.

Cots et al. [13] describe the incentives associated with case-based or episodic payment systems as 1) reducing costs per treated patients; 2) increasing revenues per patient; and 3) increasing the number of patients treated. Providers can attempt to reduce costs by reducing length-of-stay or the intensity of services provided or by selecting low-cost patients. Revenues may be increased by providers manipulating the coding of patients to achieve higher payment. The number of patients treated may be increased by splitting the treatment of patients into multiple episodes, for example by discharging and subsequently re-admitting patients [13]. An evaluation of the introduction of the NTPS (formerly known as Payment by Results (PbR)) to physical acute care in England found that the payment system was associated with a reduction in unit costs as evidenced by a fall in length-of-stay and an increase in the proportion of day cases provided [14]. There was also some evidence of an association between the new payment method and growth in acute hospital activity. There was little change in quality of care as measured by in-patient mortality, mortality within 30 days following surgery, or emergency readmission after treatment for hip fracture.

As part of the payment reform, a completely new classification system specific to mental health has been developed. The units of activity 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, psychotic and organic mental illness (Figure 1). Cluster 0 corresponds to a variance cluster which is used only if a patient cannot be allocated to any other cluster while Cluster 9 is currently not in use.

Insert Figure 1 around here

Users of mental health care services are allocated to a cluster by clinicians or a clinical team, using a tool specifically developed as part of the new payment system: the Mental Health Clustering Tool (MHCT) [15]. The MHCT is an 18-item tool which measures patient need in terms of both current problems (experienced in the previous two weeks) and historical problems across multiple domains including behaviour, functioning, symptoms and social. It is recommended that service users be assessed and allocated to a cluster at regular intervals and maximum review periods have been recommended for each cluster. Therefore, under this payment approach, providers will be paid on the basis of the number of patients in a given cluster for a defined period of time or episode of care.

Treatment reimbursed under the clusters can be provided in any setting (inpatient or community) in order to encourage care that is clinically appropriate, least restrictive, and cost-effective. It is intended that a fixed price (tariff) will be attached to each cluster [8,16], which can create an incentive for providers to control costs and increase efficiency as providers with costs above the national price (tariff) potentially face financial losses.

There is a marked absence of literature investigating the performance of mental health care providers in relation to costs. Rather, previous studies have investigated mental health costs in the context of provider payment. Research has broadly followed two strands: 1) the application of classification systems used in physical health care to mental health care [17-21] and 2) the development of new classification or resource allocation systems specifically for mental health care [22-26]. Results reveal that DRGs or other classification systems designed for use in physical health care have not been successful in explaining a substantial part of variation in mental health costs [17-21]. The use of additional control variables such as demographics, co-morbidities [19-21,27] as well as social (homelessness), functioning, previous costs [27], severity, and treatment [22] improved the predictive ability of models. Nevertheless, despite additional control variables, the predictive power of physical health classification systems applied to mental health remains weak. A number of studies [17,22,24,25] also found that provider factors influenced costs. Classification systems developed specifically for mental health explained up to 78% of the variation in episode costs [24,25]. However costs driven by patient casemix were correlated with costs associated with provider factors, which impeded the use of these classification systems to inform provider payment [24].

Recently, a body of literature [28-34] examining provider performance in relation to costs for physical health care has emerged. These studies have used multi-level models with fixed [28-32] or random [28,33,34] effects, which were subsequently used to rank providers in order to assess performance. After controlling for a range of patient-level demographic, case-mix, health and treatment, socio-economic, and quality variables, residual variation in costs remained at the provider-level. The ranking of providers in terms of fixed or random effects revealed significant differences in cost-containment performance [28,29,31-33].

The aim of this research is to examine the performance of mental health providers in relation to cost efficiency in the context of the introduction of the NTPS to mental health in England.

We investigate variations in costs across providers due to observable patient risk-factors. We compare residual variation in costs across providers to gain insights into provider performance on cost control. The motivation for the study is to investigate the extent of variation in costs unattributable to patient casemix or provider governance or capacity constraints. This unobservable cost variation will become more pertinent if provider payment is linked to national average costs as is current practice for acute physical health care. In particular, providers with above-average costs may be motivated to engage in behaviours that reduce costs to the detriment of patient outcomes and quality of care.

We add to the existing literature on mental health costs in several ways. We go beyond the remit of using risk-adjustment to explain variations in mental health costs by explicitly comparing the performance of mental health providers in terms of residual cost variation. This complements recent literature in the physical acute sector by extending similar methodologies to mental health care. Additionally, we improve upon existing studies in mental health by using a large, nationally representative patient-level dataset and exploit the richness of this dataset by using multi-level models. We control for variations in costs across providers using a classification system developed specifically for funding mental health care and supplement this with a range of demographic, treatment and socio-economic variables as well as a number of provider-level variables.

METHODS

Construction of the dependent cost variable

Each provider submits central returns (Reference Costs) annually on their average costs for both admitted and non-admitted care within each of the 21 clusters associated with a given Cluster Review Period (CRP). We compared Reference Cost (RC) data for 2011/12 and 2012/13 by provider for both admitted and non-admitted care and omitted data for outliers (n=99,232) defined as greater than 4 times the cost reported in the previous (for 2012/12 data) or following (for 2011/12 data) year. This meant that one provider with consistently high costs for all clusters across both years was dropped from the analysis, resulting in 55 providers included in the estimation sample. We calculated the total length of stay (days) in admitted and non-admitted care during a CRP. We then applied the per diem unit costs for admitted and non-admitted care for the particular cluster and provider to this activity in order to construct a variable reflecting the total cost associated with a CRP. We applied the 2011/12 RC data to CRPs occurring between 1 April 2011 and 31 March 2012 and the 2012/13 RC data to CRPs falling between 1 April 2012 and 31 March 2013. Due to the chronic nature of mental illness, some CRPs started during 2011/12 and ended during 2012/13 so a weighted average cost reflecting the number of days during a CRP in each year was calculated for these observations.

Multi-level model

The unit of observation is the CRP – the period of time between two MHCT assessments. A patient can have more than one CRP and the maximum number of CRPs per patient in our dataset is 43. This means that our data is characterised by a multi-level structure with three levels: CRPs nested in patients nested in providers (see Figure 2).

Insert Figure 2 around here

Data Analytic Procedures

We adopt two estimation approaches: 1) a linear model with the log of total cost as the dependent variable, and 2) a multi-level generalised linear model (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 Empirical Bayes 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(Y)} does not necessarily equal ln{E(Y)}[35] . The use of a multi-level 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 re-transformation of the dependent variable.

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

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

where y_{ijk} is the dependent cost variable, X_{ijk} represents a vector of risk-adjustment covariates at the cluster-review- and patient-levels, Z_k a vector of provider-level constraints, 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. 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 which can be calculated as (exp(β) – 1)*100. 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 multi-level GLM for CRP *i* in patient *j* in provider *k*:

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

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. $X'_{ijk}\beta + u_k + v_{jk}$ 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 can be interpreted as average marginal effects. All but one of our independent variables are dummy variables so coefficients can be interpreted in terms of average marginal 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 continuous IMD Income Deprivation 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.

In order to compare the residual variation across providers we predict the random effect u_k from the log-linear model using Empirical Bayes estimates with comparative standard errors. This approach has previously been used to measure the cost performance of providers in acute physical health care[36]. Empirical Bayes predictions combine the prior distribution with the likelihood to obtain the posterior distribution given the observed responses. The Empirical Bayes prediction is the mean of the posterior distribution with parameter estimates taken as the true values [37]. An attractive feature of Empirical Bayes estimates is that the Empirical Bayes prediction for a particular provider is shrunken toward zero (the mean of the prior) due to a shrinkage factor that lies between zero and one. The shrinkage factor will be closer to zero when group (i.e. provider) sizes are small or there is high within-group variability. In both cases there is relatively little information about the group so the group mean is shrunken towards the overall mean of zero [38]. The posterior standard deviation is used as the standard error of the Empirical Bayes predictions and this allows comparison of differences between individual providers and the group mean of zero [39]. We calculate the percentage difference in the EB estimates of provider-level residual variation for the best and worst performing providers compared to the average performing provider as $(\exp(u\mathrm{k} - u0) - 1)*100$ where $u0$ refers to the average provider.

The models are estimated in Stata 13.0 [40] using the *meglm, margins*, and *predict* commands and in MLwiN 2.29 [41] using the *runmlwin* command [42] in Stata 13.0 [40].

DATA

We use RC data published by the Department of Health to construct our dependent cost variable. RC data is collected on publicly owned providers and gives an indication of the costs of providing mental health services. While the RC data for the care clusters covers 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 are reimbursed under separate non-cluster units of activity.

Risk-adjustment variables are sourced from the Mental Health Minimum Data Set (MHMDS), a patient-level data set that describes specialist mental health care services. The MHMDS was introduced in 2000 and since 2003, all National Health Service (NHS) funded providers of specialist adult, including elderly, mental health services are required to submit central MHMDS returns on a quarterly and annual basis. We use Version 4.0 of the dataset which covers 2011/12 and 2012/13. We cleaned the MHMDS data to remove observations that: are duplicates; have age coded as less than 18 years or greater than 110 years, and are treated by private providers. We also dropped observations (n=833) with inpatient days in the 99th percentile (>=48 days for Cluster 1 and >=74 days for Cluster 2) for clusters covering common mental health problems as we would not expect patients in these clusters to receive long periods of inpatient treatment. Additionally, we only considered activity reimbursed under the care clusters.

Demographic variables include age, gender, ethnicity and marital status. We categorise age in order to capture any non-linearities in the relationship between age and cost. Ethnicity is also categorised into White, Black, Asian and Other categories to represent the various ethnic groups in the data. 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 had 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 - or has been admitted to hospital under the Mental Health Act (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 likely been subject to the MHA or under CPA (as these events are well regulated and documented) in order to preserve sample size. We include dummy variables for each of the care clusters to investigate the extent to which these explain variations in cost. We use the cluster with the lowest cost as the reference category. The MHMDS also contains a geographic marker for each individual at small area level or Lower Layer Super Output Area (LSOA). LSOAs are a geographic hierarchy with a minimum population of 1000 and a mean of 1500 [43]. The LSOA codes can be matched to data on the Index of Multiple Deprivation (IMD) [44] in order to enable variables reflecting various domains of deprivation to be used in the analysis. The IMD has seven domains, of which we use the IMD Income Domain to capture the proportions of the population experiencing income deprivation in an area. Observations include those with an actual 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 inflation with 2011/12 used as the reference category.

We include a number of provider-level variables reflecting provider governance and capacity constraints, sourced from the website of NHS Digital – the national provider of information, data and IT systems for health and social care in England. These include provider size as measured by the number of available mental health beds, percentage occupancy of mental health beds, and whether the provider has Foundation Trust (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%) [45] may have higher costs if high rates result in patients being discharged early and subsequently re-admitted. Providers with FT status are distinguished from other NHS providers as they have more autonomy and control over their finances so can be expected to be associated with higher financial performance [46] and hence lower costs. We also include a variable measuring the proportion of admissions under the Mental Health Act (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 [47] and we expect compulsory admission to be positively associated with cost.

RESULTS

Dependent variables

Figure 3 shows our untransformed dependent variable – total cost for CRPs. The graph shows that there is considerable variation both within and between providers.

Insert Figure 3 around here

Figure 4 displays the log transformation of the total cost per CRP variable, which approximates a normal distribution.

Insert Figure 4 around here

Descriptive statistics

Table 1 displays the descriptive statistics for our dependent and independent variables for the estimation sample of 681,027 observations with reference categories in brackets.

Insert Table 1 around here

Estimation results

A Hausman test confirmed our preference for the random-effects model (chi-squared (33) = 36.86, Prob>chi-squared = 0.2950).

Table 2 displays the estimation results for the three-level log-linear model and GLM.

Insert Table 2 around here

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 Asian and Other ethnicity, Cluster 18 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. These findings echo those of previous studies [26,48,49]

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 357% (Cluster 16) to 644% (Cluster 14).

For the GLM, Black ethnicity is associated with an increased cost of a CRP of £185 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 £613 compared to the age 18-34. Admission under the MHA is associated with an increase in costs of £2,276. The care clusters are broadly increasing in cost within the broad diagnostic groupings shown in Figure 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 £3,510 and Cluster 17 is associated with a higher cost of £3,828 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 £881 in the GLM.

None of the provider-level variables are statistically significant in the log-linear model but they are all significant in the GLM 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.

Provider-level residual variation

Around 8% of the residual variation in log of Total Cost is at the provider-level.

Figure 5 displays the Empirical Bayes predictions of the provider level random effects for the log-linear model. The graph shows 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 71% below the average while the worst performing provider has residual costs 181% above the average performing provider.

Insert Figure 5 around here

DISCUSSION

This paper has compared costs across mental health providers in England and attempted to explain variations in these costs due to observable patient and provider factors. Furthermore, we provide insight into the extent to which the classification system for a new mental health payment system explains variation in costs. After controlling for a wide range of risk-adjustment variables, we find substantial residual variation in costs across providers which we interpret as differences in performance. Our results show that a number of providers have above average residual costs and this indicates these providers may face financial instability if national fixed prices (tariffs) are introduced.

Our research 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. Nevertheless, the direction of the effects of the cluster variables does appear intuitive, with the clusters reflecting higher severity and need associated with higher costs. 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 a number of countries have developed psychiatric classification systems but have not implemented these in a provider payment system [4].

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 relatively large 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 [24]. This may be a legitimate concern in England as despite tight regulation of the MHA, it has been suggested that the MHA is used to acquire access to an inpatient bed due to high demand pressures on beds [50]. Moreover, research has found that some variation in formal admission rates is attributable to unobservable provider factors [47], which it may not be legitimate to reward. Ideally, the clinical or patient factors driving formal admission should be adequately reflected in the classification system [24]. Our findings suggest that these clinical factors are not adequately captured by the care clusters in themselves. Research has suggested that patient factors such as ethnicity and age are associated with compulsory admission [47]. 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 drivers of cost and formal admissions (e.g. Black ethnicity).

There are a number of possible reasons why some providers have higher residual costs that we are unable to address in this analysis. Firstly, those providers with higher residual costs may be providing better quality care. 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 [16] so that quality of care will not be sacrificed in the drive to increase activity and contain costs. The MHMDS contains data on a clinician-reported outcome measure – the Health of the Nation Outcome Scales, which is one of the measures recommended for use in contracts [16]. However, modelling the relationship between quality and costs poses challenges due to endogeneity and the lack of suitable instrumental variables. Secondly, providers with higher residual costs may be treating a certain case-mix 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, as a result of poor coding of diagnoses in the MHMDS. Previous studies have found that diagnosis can explain some variation in costs with more severe diagnoses such as psychoses being associated with higher costs [22,24]. 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 case mix [8,51].

Thirdly, 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 [52,53]. We find that CRPs that started in 2012/13 are associated with lower costs compared to those that started in 2011/12, which may reflect improved coding of the cost data in 2012/13. There is a low implementation rate of Patient Level Information Costing Systems (PLICS) in mental healthcare compared to acute physical care [53]. 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. 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 [54].

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. There may be a greater response to incentives on the part of providers in mental health compared to physical health care [3]. These could include "dumping' more expensive patients and "selecting" 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 ("cluster creep") and it could be argued that this may be relatively easier in mental health care where clinicians themselves will be the coders as opposed to acute physical care 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 [8,51]. 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 which may have a deleterious effect on the local healthcare provision system.

REFERENCES

1. World Health Organisation. Mental Health Atlas 2011. Geneva: WHO Press; 2011.
2. All Party Parliamentary Group on Mental Health. Parity in Progress? The All Party Parliamentary Group on Mental Health's inquiry into parity of esteem for mental health2015.
3. Douven R, Remmerswaal M, Mosca I. Unintended effects of reimbursement schedules in mental health care. J Health Econ 2015;42:139-50.
4. Mason A, Goddard M. Payment by results in mental health: a review of the international literature and an economic assessment of the approach in the English NHS. York: Centre for Health Economics, University of York; 2009.
5. Hubener F. Financial reforms could worsen German psychiatric care. Lancet (London, England) 2013;381:440.
6. Pletscher M. Marginal revenue and length of stay in inpatient psychiatry. The European journal of health economics : HEPAC : health economics in prevention and care 2015.
7. Berenson R, Upadhyay D, Delbanco S, Murray R. Payment Methods and Benefit Designs: How They Work and How They Work Together to Improve Health Care. Global Budgets for Hospitals. . Washington DC: Urban Institute; 2016.
8. Jacobs R. Payment by results for mental health services: economic considerations of case-mix funding. Advances in psychiatric treatment 2014;20:155 - 64.
9. Monitor and NHS England. Guidance on mental health currencies and payment. A supporting document for the 2016/17 National Tariff Payment System: A consultation notice2016.
10. The Mental Health Taskforce. The Five Year Forward View for Mental Health A report from the independent Mental Health Taskforce to the NHS in England,2016.
11. Paris V, Devaux M, Wei L. Health Systems Institutional Characteristics: A Survey of 29 OECD Countries. Paris: OECD; 2010.
12. Monitor and NHS England. Developing an episodic payment approach for mental health: Monitor and NHS England; 2015.
13. Cots F, Chiarello P, Salvador X, Castells X, Quentin W. DRG-based hospital payment: Intended and unintended consequences. In: Busse R, Geissler A, Quentin W, Wiley M, eds. Diagnosis-Related Groups in Europe Moving towards transparency, effi ciency and quality in hospitals. England: Open University Press; 2011.
14. Farrar S, Yi D, Sutton M, Chalkley M, Sussex J, Scott A. Has payment by results affected the way that English hospitals provide care? Difference-in-differences analysis. BMJ 2009;339.
15. Monitor and NHS England. National Tariff 2014/15 Payment System Annex 7C Mental health clustering tool booklet. London: Monitor; 2013.
16. Department of Health Payment by Results team. Mental Health Payment by Results Guidance for 2013-14. Leeds: Department of Health; 2013.
17. McGuire TG, Dickey B, Shively GE, Strumwasser I. Differences in resource use and cost among facilities treating alcohol, drug abuse, and mental disorders: implications for design of a prospective payment system. The American journal of psychiatry 1987;144:616-20.

18. Mitchell JB, Dickey B, Liptzin B, Sederer LI. Bringing psychiatric patients into the Medicare prospective payment system: alternatives to DRGs. The American journal of psychiatry 1987;144:610-5.
19. Ettner SL, Notman EH. How well do ambulatory care groups predict expenditures on mental health and substance abuse patients? Administration and policy in mental health 1997;24:339-57.
20. Ettner SL, Frank RG, McGuire TG, Newhouse JP, Notman EH. Risk adjustment of mental health and substance abuse payments. Inquiry : a journal of medical care organization, provision and financing 1998;35:223-39.
21. Ettner SL, Frank RG, McGuire TG, Hermann RC. Risk adjustment alternatives in paying for behavioral health care under Medicaid. Health services research 2001;36:793-811.
22. Drozd EM, Cromwell J, Gage B, Maier J, Greenwald LM, Goldman HH. Patient casemix classification for medicare psychiatric prospective payment. The American journal of psychiatry 2006;163:724-32.
23. Sloan KL, Montez-Rath ME, Spiro A, 3rd, et al. Development and validation of a psychiatric case-mix system. Medical care 2006;44:568-80.
24. Buckingham B, Burgess P, Solomon S, Pirkis J, Eagar K. Developing a Casemix Classification for Mental Health Services: Summary. Canberra: Commonwealth Department of Health and Family Services; 1998.
25. Gaines P, Bower A, Buckingham B, Eagar K, Burgess P, Green J. New Zealand Mental HealthClassification and Outcomes Study: Final Report. New Zealand: Health Research Council of New Zealand; 2003.
26. Sutton M, Kristensen SR, Lau Y-S, et al. Estimation of a formula for mental health services based on person-level data. London: Department of Health; 2012.
27. Kapur K, Young AS, Murata D. Risk adjustment for high utilizers of public mental health care. The journal of mental health policy and economics 2000;3:129-37.
28. Olsen KR, Street A. The analysis of efficiency among a small number of organisations: how inferences can be improved by exploiting patient-level data. Health economics 2008;17:671-81.
29. Hvenegaard A, Street A, Sorensen TH, Gyrd-Hansen D. Comparing hospital costs: what is gained by accounting for more than a case-mix index? Social science & medicine (1982) 2009;69:640-7.
30. Kristensen T, Laudicella M, Ejersted C, Street A. Cost variation in diabetes care delivered in English hospitals. Diabetic medicine : a journal of the British Diabetic Association 2010;27:949-57.
31. Laudicella M, Olsen KR, Street A. Examining cost variation across hospital departments–a two-stage multi-level approach using patient-level data. Social Science & Medicine 2010;71:1872-81.
32. Gaughan J, Mason A, Street A, Ward P. English hospitals can improve their use of resources: an analysis of costs and length of stay for ten treatments. York: Centre for Health Economics, University of York; 2012.
33. Carey K. A multilevel modelling approach to analysis of patient costs under managed care. Health economics 2000;9:435-46.
34. Cooper NJ, Lambert PC, Abrams KR, Sutton AJ. Predicting costs over time using Bayesian Markov chain Monte Carlo methods: an application to early inflammatory polyarthritis. Health economics 2007;16:37-56.
35. Montez-Rath M, Christiansen C, Ettner S, Loveland S, Rosen A. Performance of statistical models to predict mental health and substance abuse cost. BMC Me Res Methodol 2006;6:1-11.

36. Gutacker N, Bojke C, Daidone S, Devlin NJ, Parkin D, Street A. Truly inefficient or providing better quality of care? Analysing the relationship between risk-adjusted hospital costs and patients' health outcomes. Health Econ 2013;22:931-47.
37. Skrondal A, Rabe-Hesketh S. Prediction in multilevel generalized linear models. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2009;172:659-87.
38. Steele F. Module 5: Introduction to Multilevel Modelling Concepts. Bristol: Centre for Multilevel Modelling, University of Bristol; 2008.
39. Stata Press. Stata Multilevel Mixed-Effects Reference Manual Release 13. Texas: StataCorp L.P.; 2013.
40. StataCorp. Stata: Release 13. College Station Texas,: StataCorp L.P.; 2013.
41. Rabash J, Charlton C, Browne WJ, Healy M, Cameron B. MLwiN Version 2.1. Centre for Multilevel Modelling: University of Bristol; 2009.
42. Leckie G, Charlton C. runmlwin: A Program to Run the MLwiN Multilevel Modeling Software from within Stata. Journal of Statistical Software 2012;52.
43. Lower Layer Super Output Area. 2013.
44. Noble M, McLennan, D., Wilkinson, K., Whitworth, A., Barnes, H. and Dibben, C. The English Indices of Deprivation 2007. London: Communities and Local Government; 2008.
45. Royal College of Psychiatrists. Do the right thing: how to judge a good ward. Ten standards for adult in-patient mental health care. London: Royal College of Psychiatrists; 2011.
46. Verzulli R, Jacobs R, Goddard M. Do Hospitals Respond to Greater Autonomy? Evidence from the English NHS. York: Centre for Health Economics, University of York; 2011.
47. Weich S, McBride O, Twigg L, et al. Variation in compulsory psychiatric inpatient admission in England: a cross-sectional, multilevel analysis. Health Services and Delivery Research 2014;2.
48. Robst J. Development of a Medicaid behavioral health case-mix model. Evaluation review 2009;33:519-38.
49. Eagar K, Gaines P, Burgess P, et al. Developing a New Zealand casemix classification for mental health services. World psychiatry : official journal of the World Psychiatric Association (WPA) 2004;3:172-7.
50. House of Commons Health Committee. Post-legislative scrutiny of the Mental Health Act 2007 First Report of Session 2013-14. London: House of Commons; 2013.
51. Yeomans D. Clustering in mental health payment by results: a critical summary for the clinician. Advances in psychiatric treatment 2014:227-34.
52. Capita. Payment by Results Data Assurance Framework. Report on Assuring Mental Health Currencies. England: Capita Health and Wellbeing Limited; 2013.
53. PriceWaterhouseCoopers. An Evaluation of the Reimbursement System for NHS-Funded Care: Report for Monitor. UK: PriceWaterhouseCoopers LLP; 2012.
54. Monitor. Improving the costing of NHS services: proposals for 2015 to 2021 Our response to feedback. London: Monitor; 2015.

Figure 1. Care clusters decision tree

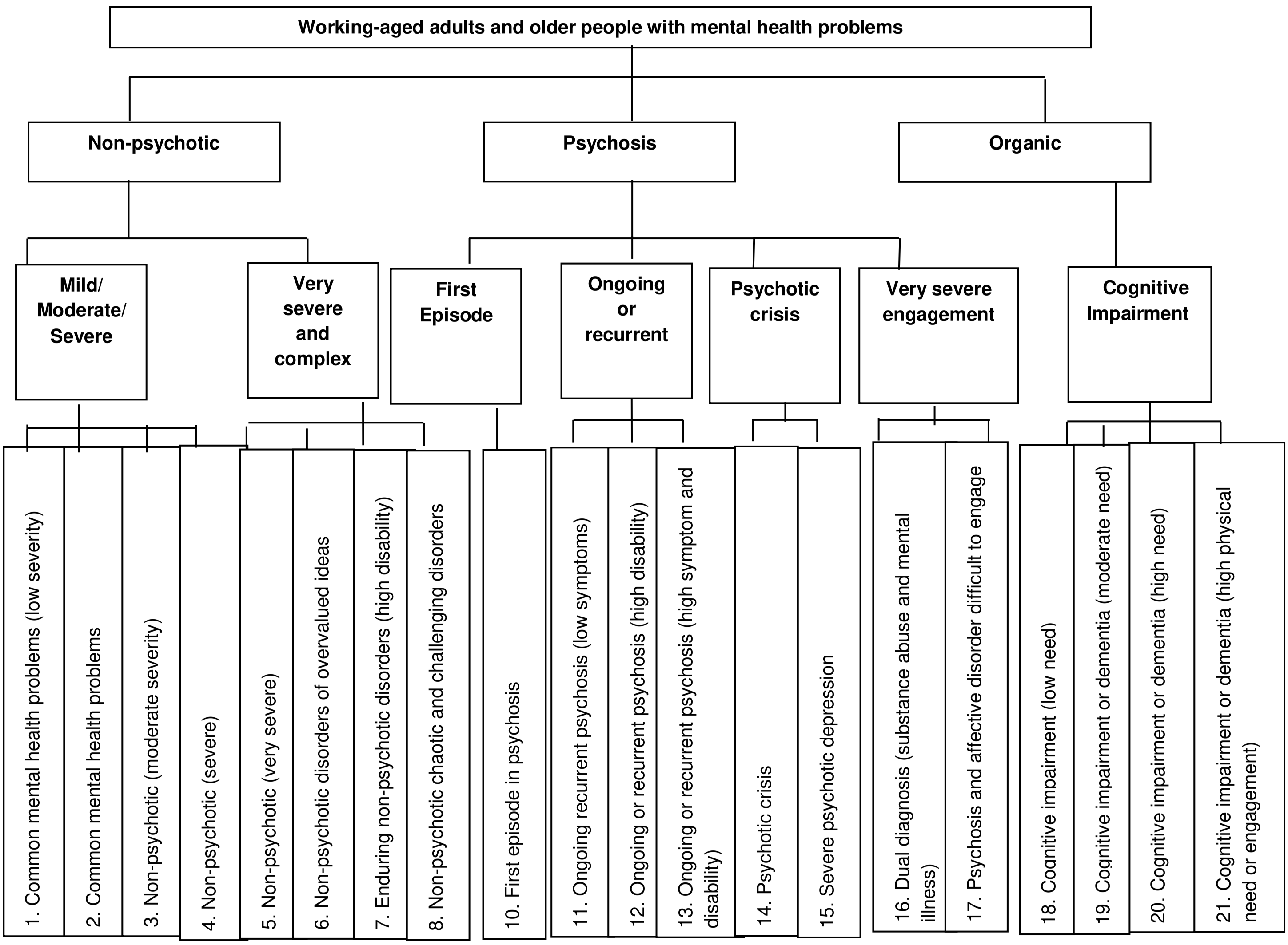

Figure 2. Multi-level data structure

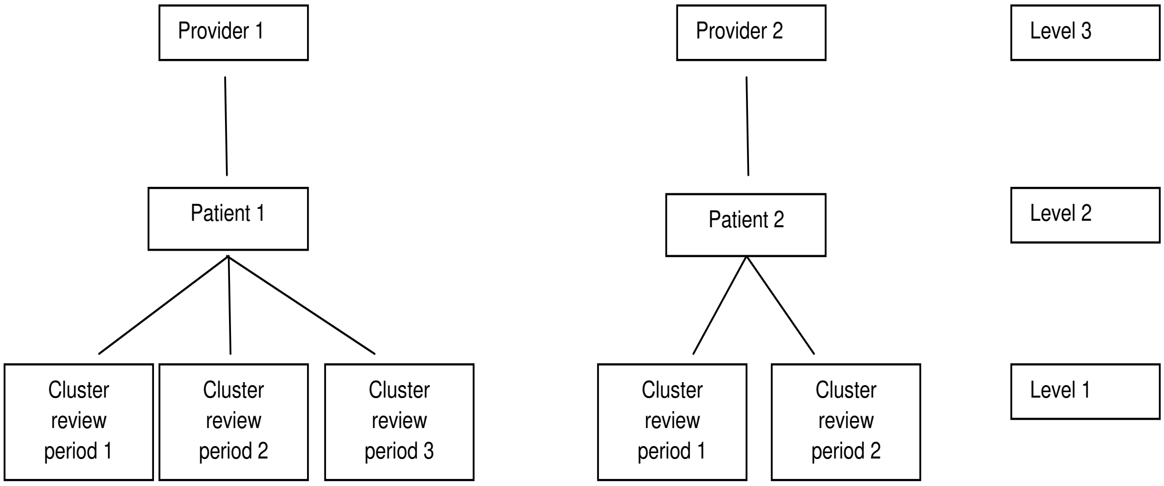

Figure 3. Dependent variable, total cost per Cluster Review Period (CRP) by provider, n=51

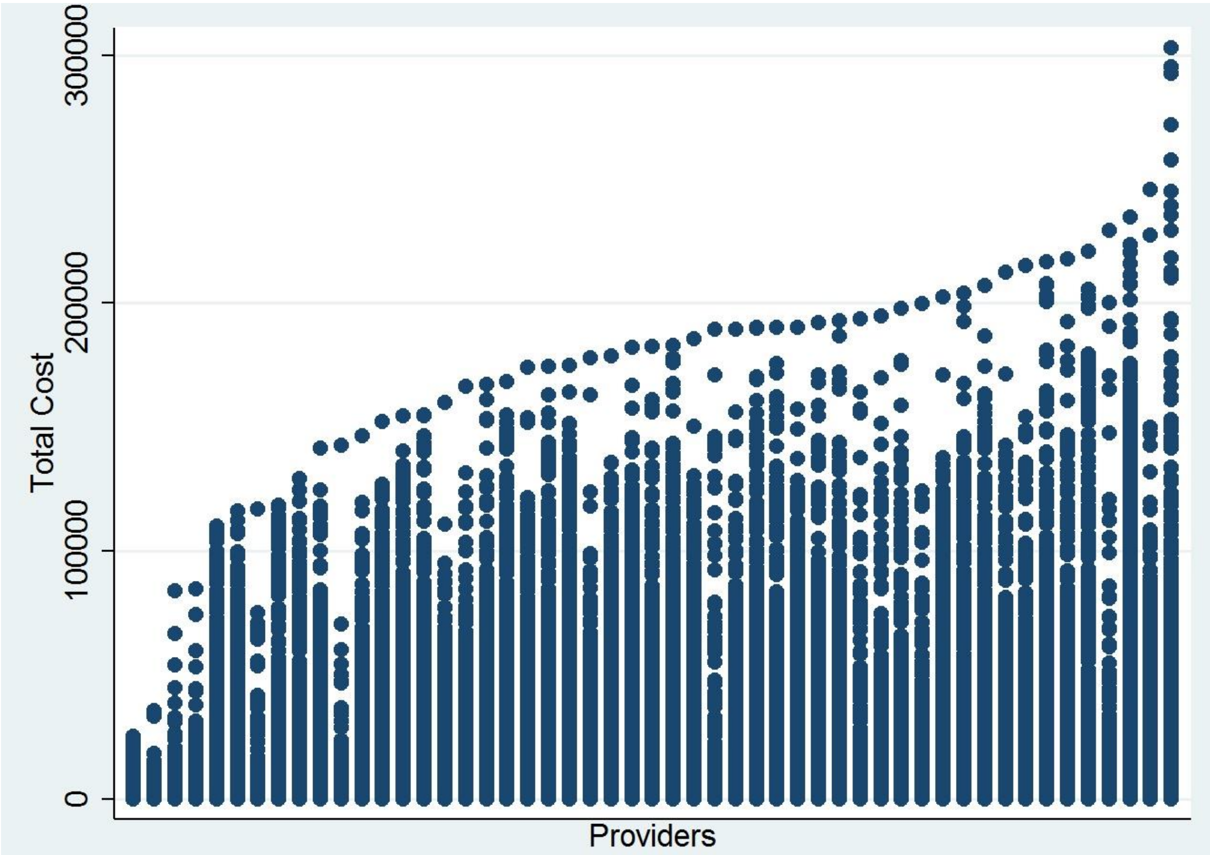

Figure 4. Log of Total Cost per CRP dependent variable

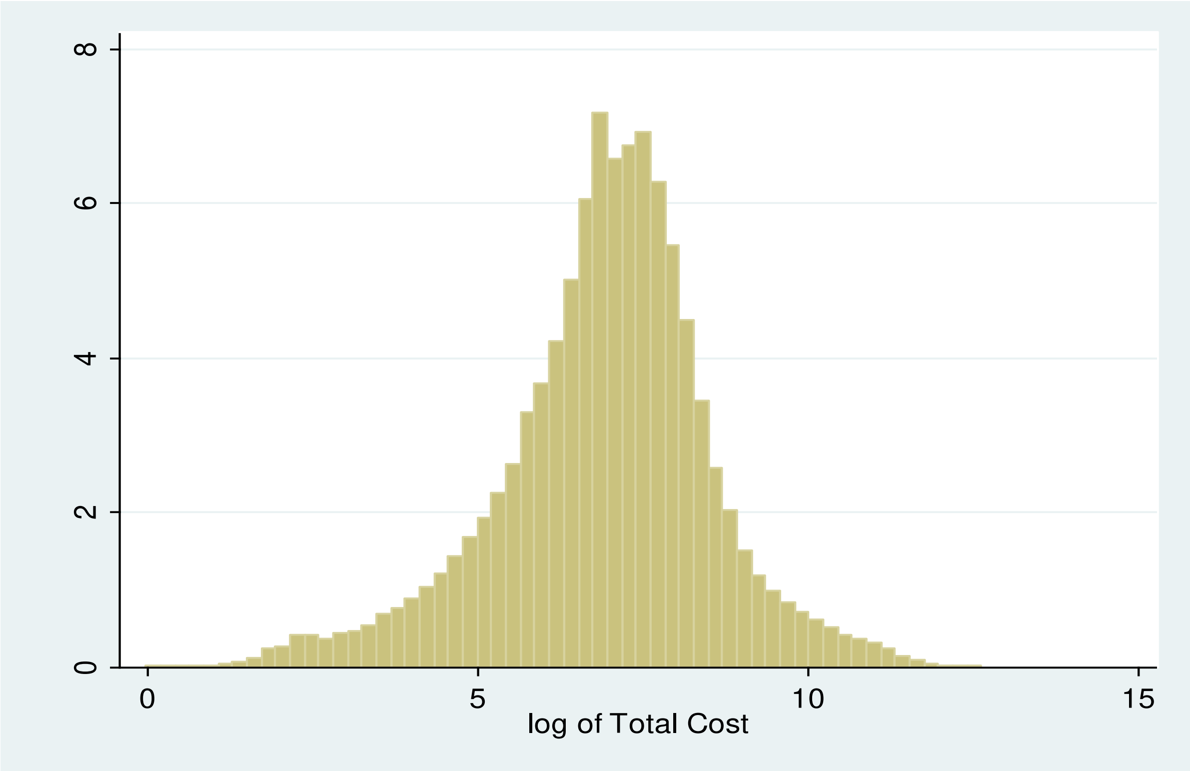

Figure 5. Variation in provider-level residual variation

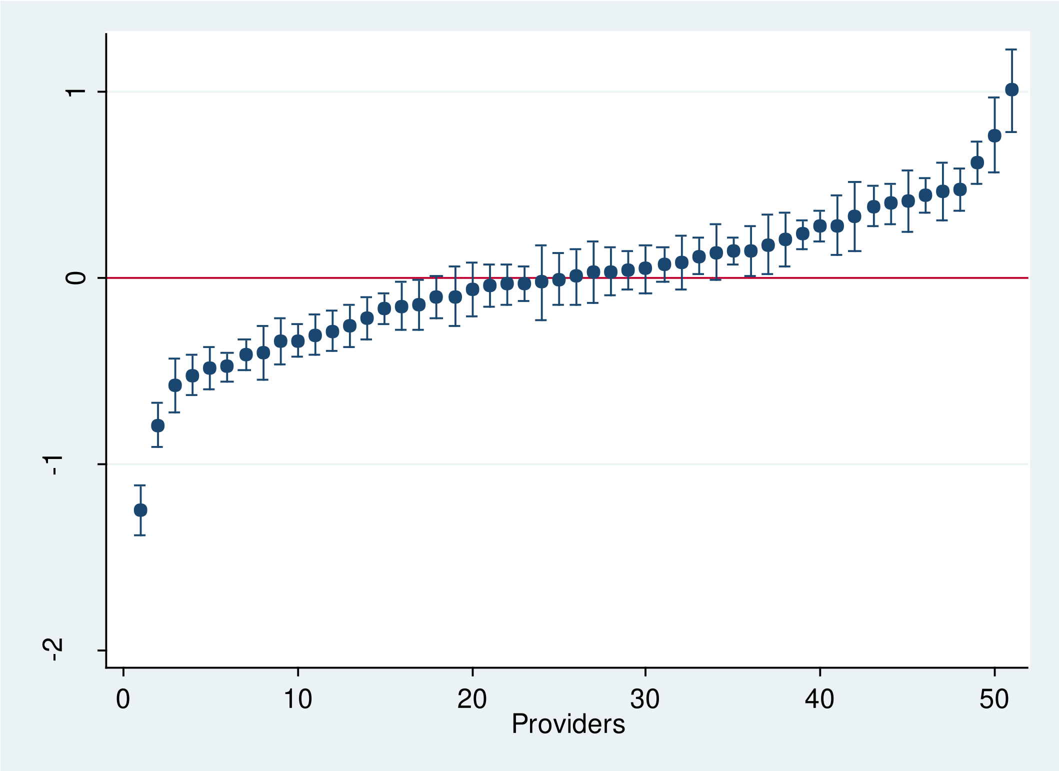

Table 1. Descriptive statistics (n=681,027)

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
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

Table 2. Model estimates of three-level log-linear and generalised linear models

	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 ideas	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 symptoms	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 crisis	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**