Centre for Chronic Diseases & Disorders

CHE
Centre For Health Economics

UNIVERSITY of York

Funding of Mental Health Services: Do Available Data Support Episodic Payment?

Rowena Jacobs, Martin Chalkley,
María José Aragón, Jan R. Böhnke,
Mike Clark, Valerie Moran,
Simon Gilbody

CHE Research Paper 137

Funding of mental health services: Do available data support episodic payment?

[1] Rowena Jacobs
[1]Martin Chalkley
[1]María José Aragón
[2]Jan R. Böhnke
[3]Mike Clark
[4]Valerie Moran
[2]Simon Gilbody

[1]Centre for Health Economics, University of York, York, UK
[2]HYMS and Department of Health Sciences, University of York, York, UK
[3]PSSRU, London School of Economics, London, UK
[4]Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

October 2016

Background to series

CHE Discussion Papers (DPs) began publication in 1983 as a means of making current research material more widely available to health economists and other potential users. So as to speed up the dissemination process, papers were originally published by CHE and distributed by post to a worldwide readership.

The CHE Research Paper series takes over that function and provides access to current research output via web-based publication, although hard copy will continue to be available (but subject to charge).

Acknowledgements

This work was part-funded by the Wellcome Trust [ref: 105624] through the Centre for Chronic Diseases and Disorders (C2D2) at the University of York. We would like to thank Phil Moore, lead for the NHS CC Mental Health Commissioners Network and participants at the workshop we ran for Commissioners for their helpful views and comments. We would like to thank David Yeomans from Leeds and York Partnership NHS Foundation Trust for providing an additional clinical view to our findings. Our thanks go to Jon Painter, Northumberland Tyne and Wear NHS Foundation Trust, and Carole Green, former lead of the Care Pathways and Packages Project (CPPP), for helping us with queries on specific Trusts and services. We would also like to thank policymakers at NHS England, NHS Improvement (formerly Monitor) and the Department of Health for useful discussions around mental health funding. The views expressed in this report are the authors' and not the funder and any errors or omissions that remain are our responsibility.

Further copies

Copies of this paper are freely available to download from the CHE website www.york.ac.uk/che/publications/ Access to downloaded material is provided on the understanding that it is intended for personal use. Copies of downloaded papers may be distributed to third-parties subject to the proviso that the CHE publication source is properly acknowledged and that such distribution is not subject to any payment.

Printed copies are available on request at a charge of £5.00 per copy. Please contact the CHE Publications Office, email che-pub@york.ac.uk, telephone 01904 321405 for further details.

Centre for Health Economics
Alcuin College
University of York
York, UK
www.york.ac.uk/che

Abstract

Background

The primary method of funding NHS mental health services in England has been block contracts between commissioners and providers, with negotiations based on historical expenditure. There has been an intention to change the funding method to make it similar to that used in acute hospitals (called the National Tariff Payment System or NTPS, formerly known as Payment by Results (PbR)) where fixed prices are paid for each completed treatment episode. Within the mental health context this funding approach is known as episodic payment. Patients are categorised into groups with similar levels of need, called clusters. The mental health clustering tool (MHCT) provides a guide for assignment of patients to clusters. Fixed prices could then be set for each cluster and providers would be paid for the services they deliver within each cluster based on these fixed prices, although the emphasis to date has been on local pricing. For this episodic payment system to work, the MHCT needs to assign patients to clusters, such that they are homogenous in terms of 1) patient need, and 2) resource use.

Objectives

We test whether the existing data collected on mental health activity amongst NHS providers would support this new payment system. Specifically we examine whether there is homogeneity within clusters in terms of 1) costs, and 2) activity/resource use, and 3) whether the MHCT effectively clusters people with similar levels of need.

Data

We use the Mental Health Services Dataset, a patient level dataset which records activity on the patient's classification into clusters, for the financial years 2012/13 and 2013/14. We link cluster activity within each provider to Reference Costs (for 2013/14), which provide information on the cost of the activity performed by each provider and facilitates the calculation of average costs for the different clusters both at provider and at national level.

Methods

We calculate a cost index to observe the variation in costs across providers. We run multilevel regressions of activity within clusters to test whether the observed differences in length of cluster-episodes translate into differences in the activity performed within them. We perform latent class analysis to determine whether the clusters as currently defined correspond to mutually exclusive groups of patients based on their answers to the MHCT. We ran a workshop for mental health commissioners to ascertain the key challenges of implementing this payment system.

Results

There is substantial variation in costs across providers. Considering all activity together, the ratio between the provider with the highest costs and the one with lowest is around two, but in some clusters this ratio can be as high as ten. Longer cluster-episodes do not translate into proportionally more activity. The existing 21 clusters do not correspond to 21 different groups of patients as defined by their answers to the MHCT. There is also great variation across commissioners in terms of capacity to contract for and implement this payment system.

Conclusions/Discussion

The results indicate that if a new cluster-based episodic payment system were introduced it would have different financial impacts across providers as there is variation in the activity they perform within clusters and the costs they report for them. However remaining under block contract incurs a risk of mental health services experiencing disinvestment. There is thus a need to invest in improving the quality of data collection, and refining the cluster classification system to facilitate sustainable payment system reform.

Key Words

Mental Health, Payment System, National Tariff Payment System (NTPS), episodic payment, Mental Health Clustering Tool (MHCT), Mental Health Services Dataset (MHSDS), Health of the Nation Outcome Scale (HoNOS).

Executive summary

Background

1. NHS mental health services in England have primarily been funded through block contracts agreed between commissioners and providers of care. This method of financing, offers little incentive for providers to efficiently meet health care need.

2. Recent policy guidance by NHS England and NHS Improvement (formerly Monitor) has set out two alternative approaches for contracting and paying for mental health services, the episodic payment approach, and the more recently introduced capitated payment approach. In this report, we have focused on the move towards an episodic payment approach.

3. Under episodic payment mental health services would adopt a payment system similar to that for acute hospital care from 2004, known as the National Tariff Payment System (NTPS) (formerly Payment by Results or PbR).

4. Under the episodic payment approach, patients are categorised into one of 21 *clusters* (analogous to Healthcare Resource Groups - HRGs under NTPS) according to their need and are reviewed at regular review intervals which have been recommended for each cluster. Providers of services would be paid for each period of care that the patient receives whilst assigned to that cluster for a defined review period or episode of care. Fixed prices may then be set for each cluster-episode. Patients are assigned to a cluster guided by the Mental Health Clustering Tool (MHCT). The MHCT consists of 18 items comprising the 13 items of the Health of the Nation Outcomes Scores (HoNOS) and the 5 items of the Summary Assessment of Risk and Need (SARN) to assess need.

5. For this episodic payment system to work there should not be too much variation in costs either *within* clusters, or *between* providers. The MHCT therefore needs to assign patients to clusters, such that they are homogenous in terms of 1) patient need, and 2) resource use. Numerous unintended consequences may arise if there is too much variation either *within* clusters, or *between* providers.

Our investigation

6. We evaluate whether the Mental Health Services Data Set (MHSDS) is of sufficient quality to provide accurate measures of activity and we undertake an examination of the variability that it reports. Specifically, we examine whether there is homogeneity within clusters in terms of 1) costs, and 2) activity/resource use, and 3) whether the MHCT effectively clusters people with similar levels of need.

7. We use two main sources of data, the Mental Health Services Data Set (MHSDS) for 2012/13 and 2013/14, and Reference Costs (RC) for 2013/14. The data covers nearly 2 million cluster-episodes.

8. Patients spend on average five months in a cluster, with three days admitted on a ward and have contact with a health care professional for seven days. Costs for inpatient care are around £360 per day whilst care in the community or as an outpatient is approximately £10 per day, though there is substantial variation between providers in terms of costs, activity and length of stay within clusters.

9. The largest clusters are 18 and 19 (cognitive impairment with low and moderate need) each of which accounts for around 11% of cluster-episodes, and clusters 3 and 4 (non-psychotic moderate and severe) each of which represents around 10% of cluster-episodes allocated to them respectively.

10. We observe that the provider with the highest cost has costs that are 55% higher than average and the provider with the lowest cost is 25% below average. Considering all activity together, the ratio between the provider with the highest costs and the one with the lowest is around two. Looking within clusters, one cluster that displays a large degree of variation in cost is cluster 0 but this is not surprising since it is the variance cluster and patients are assigned to this cluster when there is uncertainty about the correct allocation. Other clusters with large variability include clusters 1, 2, 15, 18, 19 and 21. Some clusters show a ten-fold variation between providers in terms of cost.

11. We produce plots for each of the 21 clusters showing the length of cluster episodes against the proportion of inpatient activity and the proportion of contact with healthcare professionals. Our results show that there is substantial variability across providers in the length of cluster episodes, and there is substantial variability within clusters in terms of the proportion of inpatient days and the proportion of contact with healthcare professionals.

12. We also look at the correlation between the length of the cluster episode and the activity performed within it to determine whether providers differ in the resources they devote to treating patients in any given cluster. We run multilevel regressions on the length of the cluster-episode using, as the explanatory variables respectively, the number of days admitted to hospital and the number of days with contact with a health care professional. The results show that longer cluster episodes do not translate into proportionally more activity of either type.

13. Taken together our results suggest that there is a substantial degree of variability within clusters in terms of activity and resource use.

14. To further address the question on how well the Mental Health Clustering Tool (MHCT) allocates patients with similar needs to clusters, which is a crucial underpinning of an episodic payment approach, we use Latent Class Analysis (LCA) to sort groups of patient-episodes into *classes* that are maximally homogeneous within and maximally heterogeneous between. Our approach is to examine whether the statistically defined *classes* correspond to the *clusters* determined by the MHCT.

15. Overall we find relatively little correspondence between statistical *classes* and *clusters*, which suggests that the MHCT may be capable of being refined in order to establish more homogeneous groupings of patients.

16. Alongside our empirical analysis, we ran a workshop for mental health commissioners to explore where commissioners are in terms of their developments and plans for funding and contracting models, the challenges, and the landscape in terms of moving forward. A number of key topics emerged.

    a. It was clear that there is a wide degree of variation between commissioners in terms of their thinking about and confidence in using care clusters and developing their plans for payment models. Some were only now beginning to think about how to move from a block contract model. A few had variations of block contracts, episodic models and embryonic capitation models operating locally; only one group had clear plans for taking the first steps towards a capitation model in the next financial year and evolving it over the next few years. Policy guidance on the choice between the episodic or capitated payment approach has caused confusion and anxiety amongst some of the commissioners. They were now more uncertain as to what they should be doing and how to do it. Capitation seems to be seen as a means to integration of services, particularly between physical and mental health services, and, in essence, is a means of pricing a form of grand 'block contract' with sophisticated plans for additional metrics of quality and performance.

b. Commissioners seemed to welcome the care cluster model. They appeared to be largely working towards using it as a framework to understand and discuss local patterns of care and variations, rather than as a categorical system to use as a threat against providers. The degree of understanding of MHSDS and its potential links to developing payment models varied substantially across the commissioners. There is a need to clarify the risks of using local data returns in place of national data for developing payment models. There also seemed to be a lack of expertise in being able to use data other than business and process data, such as epidemiological data, which may be needed for models of capitation payments.

Implications for payment reform

17. Our results suggest that there is substantial variation within the 21 clusters created by the MHCT in terms of the treatments that patients within a cluster receive and the cost of that treatment. This implies that an episodic-based payment system would result in large variation across providers in terms of their financial positions.

18. The significant heterogeneity we observe in terms of patient need, costs and resource use within clusters, does not bode well for an episodic payment approach which requires case-mix and resource homogeneity. We conclude however, that instead of abandoning the approach, a much clearer steer is needed from policymakers to support providers and commissioners to move towards refining and developing episodic payment as a viable payment option.

19. The clustering approach is already relatively well established in most providers and abandoning it would delay the development of a transparent funding system for mental health services.

20. Whilst mental health operates a less transparent funding system than that for physical health care, there is a risk of resources being directed away from it and towards acute providers. Commissioners need to have a clear sense of the value for money they are getting from investment in mental health care and in the absence of that they will be under pressure to disinvest.

21. The system also needs to implement change at a pace that does not risk destabilising local health economies.

22. The menu of options of payment approaches offered to commissioners is causing confusion and anxiety, and risks greater local variation and an overall lack of financial control.

23. There is only limited discussion around how any payment system will be linked to quality and outcome metrics. A much stronger policy steer and more evidence are urgently needed.

24. We encountered a number of data quality issues in the MHSDS. We conclude that it is not yet suitable for use as an information tool to accurately count activity which would be central to its use as a platform for the payment system.

25. Significant investment in information technology is required and improvement in data quality needs to be a priority in mental health services.

26. We urge commissioners to routinely use the MHSDS in their contracting and monitoring processes. This will facilitate a single consistent use of data across several commissioners with any given provider and prevent providers wasting resources filling in different dataset requirements for different commissioners. It will also incentivise rapid improvement in the data quality of the MHSDS. Improvements to Reference Cost data are also essential and introduction of Patient Level Information Costing Systems (PLICS) at provider level would be beneficial.

Table of contents

List of tables

List of figures

1. Introduction

NHS mental health services in England have primarily been funded through block contracts agreed between commissioners and providers of care, or on the basis of levels of existing 'inputs' such as the number of beds (Mason et al, 2011). These arrangements have typically been negotiated on the basis of historical expenditure. This method of financing offers little incentive for providers to deliver an efficient level of care (Mason et al, 2011).

Recent policy guidance (Monitor and NHS England, 2016) has set out two alternative approaches for contracting and paying for mental health services, the episodic payment approach, and the more recently introduced capitated payment approach. In this report, we have evaluated the move towards an episodic payment approach.

A mental health payment system for England (Khan et al, 2014), which remunerates providers according to the number and type of patients they treat, would move mental health services into a payment regime similar to that adopted for acute hospital care from 2004. That system was known until recently as Payment by Results (PbR) and is now termed the National Tariff Payment System (NTPS). Applied to mental health care, NTPS is a system in which patients are categorised into one of 21 *clusters* according to their need and the provider of services is paid for each period of care that the patient receives whilst assigned to that cluster. Patients are assigned to a cluster by a clinician or clinical team using the Mental Health Clustering Tool (MHCT). The MHCT consists of 18 items comprising the 13 items of the Health of the Nation Outcomes Scores (HoNOS) (Wing et al, 1994) and the five items of the Summary Assessment of Risk and Need (SARN) (Self, Painter et al, 2008) to assess need on both a current and historical basis. Clusters also define the relevant period of care and the system requires patients to be reviewed and assigned to clusters according to those periods. The clusters are mutually exclusive meaning that a patient should only be assigned to one cluster at any given time.

This approach to payment, also termed episodic payment, would link a provider's payment closely to the volume and type of mental health care *activity* that it supplies in such a way that it will know in advance how much each patient-cluster-period will yield it in terms of income. The payment would be independent of how much treatment any individual patient receives, or how that treatment is delivered. This *prospective fixed-pric*e mechanism has found favour in many health care systems especially in regard to paying for acute hospital care. A number of potential advantages and risks of such a system have been extensively discussed, and rather less-extensively evaluated (Charlesworth et al, 2012; Jacobs, 2014).

One obvious risk is in setting an inappropriate price - too high implies scarce financial resources are wasted and too low places providers of services in financial difficulties. However the literature has focused on the more subtle risks associated with how patients are assigned to a particular category (and hence price) (Papanicolas & McGuire, 2015). If a category is too broadly defined, a provider may be able to identify loss-making or profitable patients in advance of taking them on and will then have an incentive to re-assign some patients to other providers, or where the system is not so restrictive as to permit it, to reassign them to different treatment categories. Furthermore, if patients are not randomly assigned to providers, some providers may receive more than their fair share of difficult to treat patients and will therefore be faced with losses, whereas others may make large surpluses.

In essence, the prevalent discussion of NTPS suggests that a key to its viability and desirability is in limiting the variation in costs that is associated with the defined payment categories and that cost variation needs to be considered both across patients in a given category or cluster and across providers of care. For a given specification of categories, if either form of variation is *too great*, problems may ensue, albeit that there is no absolute measure of what constitutes acceptable variation.

The assignment of patients to one of the 21 clusters for episodic payment assumes that i) these clusters capture the variation in mental health symptoms and resulting needs, and that ii) the clusters do this in a meaningful way to allocate costs/payments, i.e. to determine prices or tariffs. Neither of these two assumptions may be true. The first is a question about the content validity of the clusters: Do the 21 clusters represent the different degrees of severity of needs, as well as the different qualities (e.g., psychotic vs. non-psychotic, depression, etc.) that are observed in mental health patients. The clusters should therefore represent empirical differences between patients.

The second question extends from the first one: Not only should the clusters represent individual differences between patients in an empirically adequate way, but also the tariff associated with each cluster should be at least reasonably close to costs that are currently occurring in a cluster as well as different to the costs occurring in other clusters. This second part therefore raises the question of whether the costs that occur within each of the 21 clusters are actually different enough to help determine tariffs or fixed prices. For the episodic payment system to work, therefore, the MHCT needs to assign patients to clusters, such that they are homogenous in terms of 1) patient need, and 2) resource use (Charlesworth et al, 2012), (Laudicella et al, 2013).

This gives the context for our investigation. Data collection to inform the mental health episodic payment approach has been underway since 2011/12 as part of the Mental Health Services Data Set (MHSDS)[1] and describes patient classification to clusters, and to a limited extent the treatments that are delivered. We also evaluate, using Reference Costs (RC), the average unit costs that providers associate with cluster episodes. We set out to first evaluate whether these data are useful in providing measures of variability that are central to the viability of the payment system and second to undertake a preliminary examination of that variability. We evaluate whether the MHSDS is of sufficient quality to provide accurate measures of activity which would be needed to underpin a funding system.

Specifically, we empirically examine three research questions; whether there is homogeneity within clusters and between providers in terms of
1) costs, and
2) activity/resource use, and
3) whether the MHCT effectively clusters people with similar levels of need.

Alongside the empirical analysis, we wanted to gain feedback on our results by running a workshop for mental health commissioners. This explored where commissioners are currently placed in terms of funding models and contracting for mental health services. Reflecting on the results we presented, we explored the key challenges in terms of variability in activity and costs, and challenges in setting prices, key issues around the classification of patients and the MHCT, data requirements to make a funding system work, and data quality.

In Section 2 we set out in more detail the conceptual basis for examining mental health episodic payment, in Section 3 we address the research questions around homogeneity of clusters in terms of cost and resource use. We provide details of the data sets we use, set out some preliminary description of the data, and provide results for our examination on variability in costs and activity.

[1] We use the Mental Health Minimum Dataset (MHMDS) for 2012/13 and 2013/14. This changed name to the Mental Health and Learning Disabilities Data Set (MHLDDS) in September 2014 and has been called the Mental Health Services Data Set (MHSDS) since January 2016. For simplicity, we refer to it by its current name the MHSDS.

Section 4 addresses our third research question on how well the MHCT classifies patients. Section 5 describes some of the issues around data quality which present challenges for underpinning a payment approach. Section 6 presents a description of the key themes from the workshop for commissioners while Section 7 concludes.

2. Economic framework for funding of mental health services

There are a number of motivations for moving away from block contracts. For example, these arrangements do not necessitate the collection of activity data and thus do not aid audit and accountability. One benefit of activity-based payments might therefore be simply that it increases transparency and accountability (The Mental Health Taskforce, 2016). Under block contracts each purchaser-commissioner pair can institute its own monitoring arrangements, but these do not need to be comparable across either providers or commissioners. Hence a benefit of a *nationally* agreed funding mechanism is the potential to compare performance across the health care system (Laudicella et al, 2013). Whilst these and other potential reasons for switching from block contracts for mental health care may be important, considerable emphasis has been placed on the economic implications of different purchasing arrangements.

Economists have typically focused on two types of consequences of different purchasing arrangements – the allocation of risk and incentives. Of these the second has been the primary focus, especially in publicly funded health care systems. The key issues can be understood by regarding a health care provider as incurring a cost from treating a number of patients with some level of *effort.* The term effort is used as a short-hand for almost any choices that the provider can make regarding what sorts of treatments to engage in. Hence effort may be related to treatment quality, or the complexity of treatment and can even include elements of efficiency savings if these require expending time and resources to achieve. The essential point is that greater effort will incur greater cost. In regard to the mental health care system, effort in treating a patient in a particular *cluster* might be reflected in the amount of time that the patient is treated as an inpatient, the number of health care professionals they see or the amount of time those professionals spend with them.

Under a block contract the provider receives fixed revenue, so that unless there are some alternative provisions and monitoring by the commissioner, they are free to choose effort. Depending on their intrinsic motivation, that might entail choosing very high effort but then turning away patients because the available budget has been consumed on the ‘lucky’ few, or it may entail treating as many patients as possible, but out of necessity of keeping within the fixed budget, doing so with a minimal effort. Put simply the provider has discretion how to utilize the fixed budget specified in the block contract and that discretion may result in different choices by different providers, according to their specific objectives (to be the ‘best’, to be the ‘busiest’ or to be the ‘biggest’). One of the biggest potential problems with this kind of arrangement is when the number of patients who will need treatment is uncertain and subject to variation, because then the only way of meeting demand if that turns out to be unexpectedly high is to reduce treatment. A block contract raises the threat of either unmet need, or low service quality or both.

In many other health care settings a standard method of paying providers has been to condition payment on the treatments they deliver. This is referred to as fee-for-service payment. This is equivalent to paying the provider for their effort. The more effort that is expended, the greater will be the budget and this sort of arrangement is therefore viewed as giving rise to a very strong incentive to increase effort. Whilst that is potentially a good thing from the point of view of service quality it implies that the costs of delivering services will be high. In a cash-constrained NHS, the risk here is that the overall health care budget will either have to be substantially increased or will face severe rationing. It is therefore clear why such an arrangement has not been countenanced either for acute hospital services or mental health services in the NHS. Nevertheless it serves to illustrate that a change in payment mechanism can plausibly shift the balance – in this case towards greater effort in delivering services.

The now most common payment mechanism for acute hospital care in many health care systems goes under a number of labels; prospective payment (US), Diagnosis Related Group payment (US and European) and Payment by Results (PbR) or the National Tariff Payment System (NTPS) (England). This payment mechanism faces the provider with a budget that varies according to the number of patients they treat, but not directly with the effort they expend. It therefore immediately gives a strong incentive to treat patients (reducing the risk of unmet need) provided the price is set high enough. There has been an enormous amount of attention given to the implication of this type of arrangement for effort and hence treatment quality (Chalkley & Malcomson, 2000; Sood et al, 2013). One potential benefit is that a provider certainly has an incentive to engage in cost savings or *efficiencies* under this system but an associated risk is that some of those savings might come from compromising service *quality* (Charlesworth et al, 2012; Jacobs, 2014). There are some potential mitigating effects for service quality. For example, if the fixed price per patient is attractive to the provider they will wish to expand the number of patients they attract and treat. If a reputation can be established for being the 'best' provider by increasing the quality of treatment then the fixed price system can be consistent with establishing an incentive towards effort and quality of care.

The rationale for adopting something analogous to NTPS in a mental health setting can therefore be summarized in terms of incentives. Whereas block contracts have indeterminate incentives for either effort (quality of care) or volume (meeting need) and fee-for-service arrangements may cause meeting need to be unaffordable, a fixed price NTPS type of mechanism potentially balances incentives to meet need and preserve service quality.

Being a potential late adopter of this kind of payment mechanism, mental health care can learn from previous experience and evaluation gained during the inception and operation of these kinds of systems in acute hospital settings. Since our focus is on data, it is useful to review some of that experience and evaluation as it applies to data requirements.

As the summary above highlights, one of the key requirements of a fixed price system is that the price should be chosen appropriately. In the NHS, that requirement has been interpreted as ensuring that prices reflect the costs of delivering treatments. A prerequisite for establishing the cost of treating a patient is measuring the resources that are devoted to their treatment. This is naturally very difficult to do even for direct resources – like clinician time – and requires often arbitrary attribution of indirect resources – like estates and administration costs. Nevertheless a building block for measuring costs is the measurement of direct resources and the data contained in the Mental Health Services Data Set (MHSDS) provides a starting point. Hence, our first investigations are directed at understanding the content and relevance of this data set, when combined with data relating to the costs of direct resources. This top level investigation is concerned with determining the extent to which the available data might support the determination of fixed prices for mental health care treatment episodes.

Much of the discussion and debate concerning the potential problems of fixed-price mechanisms for health care focuses on variability of costs. The reason is that the underlying assumption is that prices can be determined for a relatively homogeneous set of treatments. The usual means for trying to achieve this in acute hospital care is through the definition of Diagnosis Related Groups (DRGs) or in the English NHS, Healthcare Resources Groups (HRGs). Both of these constructs are intended to represent groups of patients for whom the treatments – and thus the associated costs – are similar. If an HRG in practice captures a very diverse group of patients, there are a number of risks associated with setting a single price. Patients within the group who are exceptionally costly to treat are loss-making to the provider. There is an incentive to avoid treating those patients ("cream skimming"), by referring them on ("dumping"), or to reduce the treatments they are given to try and contain their cost ("skimping") (Jacobs, 2014). Hence excessive variation of cost within a treatment group is concerning. It is perhaps worth noting that compared to more than 1200 acute care[2] HRGs (Department of Health, 2013), there are just 21 clusters for mental health, so that there would appear to be a real risk of excessive heterogeneity. A second element of our investigation is therefore concerned with the extent of heterogeneity of resource use across patients within clusters.

A second dimension of variability of cost is between providers. The impact of this dimension of variability is less clear. On the one hand a fixed price acts as an incentive for high cost providers to control their costs so that *ex ante* variability in costs may not be a great concern. Alternatively however, variation in costs across providers might indicate variation in their case mix and setting a uniform price based on average costs might drive high cost, high quality providers who treat difficult patients into financial distress. To begin informing this issue we conduct a preliminary evaluation of the variability of costs within clusters, across providers.

The finding that *a given* classification of mental health patients results in variation in costs that is problematic does not preclude the possibility of establishing a functioning fixed price system based on a different classification. Thus whilst clusters may have been defined in expectation of homogeneity there is the possibility of re-designing them based on the data that has been gathered to date. The final part of our investigation begins this process by first assessing the extent to which clusters reflect empirically justified groupings of patients.

[2] There were 1216 HRG tariffs in 2012/13.

3. Variability in costs and activity within clusters and between providers

As mentioned in Section 2, we are interested in three main research questions. The first examines variation in costs within clusters and between providers, the second examines variation in activity within clusters and between providers and the third explores the adequacy of the clustering process. In this section we focus on the first two research questions.

3.1 Data

We use two main sources of data, the Mental Health Services Dataset (MHSDS) (Health & Social Care Information Centre, 2016a; Health & Social Care Information Centre, 2016b) and the Reference Costs (RC) (Department of Health, 2014b).[3] The MHSDS is a patient-level dataset with national coverage for England, which was introduced in 2003. The variables included in the MHSDS have evolved over the years and information pertaining to the new payment system was first collected in 2011/12 coinciding with the commencement of the allocation of patients to clusters. The mandatory use of the clusters as the basis for contracting mental health services for working-age and older adults was introduced in 2012. From the MHSDS we use two financial years, 2012/13 and 2013/14, and from the RC only the last financial year 2013/14. This is because we require use of only one set of costs to calculate the cost index and this will generate variation from only one source (activity); if we use the RC from both years, the variation in the cost index may be due to intertemporal differences in costs between years for the same provider which we could not easily disentangle.

The MHSDS consists of three data files; Records, Events and Episodes. The *Records* data file contains information about the patient (e.g. age, gender, their Lower Layer Super Output Area (LSOA)[4]) and the Trust in charge of the treatment.[5] The *Events* data file provides information about different types of activity that occur during the care of a patient, such as evaluation through MHCT, contact with health care professionals, and update of information, and the date when that event took place. Contact with a health care professional may include, among others, contact with consultants, nurses, occupational therapists and social workers. The *Episodes* data file records information on periods of time spent in different circumstances, e.g. assigned to a given cluster, periods admitted as an inpatient within a Trust or under the care of a mental health team – these periods can overlap.

Table 1 shows the 21 care clusters, grouped into three *super-classes*: non-psychotic, psychotic and organic, plus a variance cluster (cluster zero). The variance cluster is used when clinicians are not able to readily assign patients to a particular cluster. Its use is intended to reduce over time. There is no cluster 9 since in the original version of clusters, it was related to Substance Abuse, but then subsequently dropped.

It is recommended that patients are assessed and allocated to a cluster at regular review intervals and maximum review periods have been recommended for each cluster (Monitor and NHS England, 2016). It is intended that providers will be paid on the basis of the number of patients in a given cluster for a defined review period or episode of care.

[3] The data we use is not available on the RC website, but is a version obtained directly from the HSCIC, which excludes one provider with reported costs and activity with data quality issues related to the split between admitted and non-admitted days.
[4] LSOAs represent a geographic area with a mean population of 1500.
[5] There were two PCTs in the data, Isle of Wight and Milton Keynes. Their activity was allocated to the Trusts that took over their activity after the abolition of PCTs, Isle of Wight and Central and North West London NHS Foundation Trust, respectively.

Table 1: Super-classes, clusters and review periods

Superclass	Cluster number	Cluster label	Cluster review period (maximum)
Variance Cluster	0	Variance	6 months
Non-psychotic	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
N/A	9	Blank cluster	N/A
Psychosis	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
Organic	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

The RC gives information regarding the cost of the activity performed by the different Trusts. The unit costs reported are for admitted days and non-admitted days in a given cluster. These two costs are available at provider level, i.e. we observe the cost that a provider reports for each cluster for which it has activity. We also observe the costs for admitted and non-admitted care at the national level, showing the (activity weighted) average of the costs reported by all providers.

In order to combine activity (MHSDS) and cost data (RC), it is necessary to focus on *cluster-episodes*, which record the time a patient spends assigned to the same cluster (even if MHCT assessments took place); and to know how many days a patient was assigned to a cluster when admitted to hospital. This can be done combining information from the *Episodes* data file, using the start and end dates of cluster and ward-stay episodes. Then, using the NTPS cluster number, we can match the corresponding (provider level and national average) costs, for admitted and non-admitted days.

We use Stata 13 to analyse the data (StataCorp, 2013).

3.2 Descriptive statistics

Table 2 shows the descriptive statistics of the variables we use in our analysis. Cluster-episodes can cover the period 2012/13 or 2013/14, and some cluster-episodes span both years. There are just under 2 million cluster-episodes.

Table 2: Descriptive statistics

Variable	Obs.	Mean	Std.Dev.	Min	Max
Number of days patient assigned to a cluster	1,965,582	151.87	143.93	1	730
Number of days spent admitted to hospital while assigned to a cluster	1,965,582	3.33	19.50	0	726
National average cost of an admitted day	1,965,582	360.00	21.67	336.53	404.75
National average cost of a non-admitted day	1,965,582	9.27	5.36	3.35	33.91
Provider level cost of an admitted day	1,783,479	363.64	74.75	160.23	1063.68
Provider level cost of a non-admitted day	1,827,615	10.03	8.42	0.58	134.21
Number of days with contact with a health care professional in the cluster episode	1,965,582	6.60	11.17	0	363

While we have added the number of days with contact with a health care professional in the cluster episode, this type of activity does not have a separate cost, but it represents most of the events reported in the data.

The number of observations for provider reported costs is smaller than that for the national average costs, this is because the provider level costs were not available for all providers but we can still match the national averages to their activity.

Patients spend on average five months in a cluster, with three days as an inpatient on a ward and have contact with a health care professional for around seven days. Costs for inpatient care are around £360 per day whilst care in the community or as an outpatient costs £10 per day, but there is a substantial amount of variation between providers in terms of costs, activity and length of stay.

Figure 1 shows the distribution of the cluster-episodes across the different clusters.

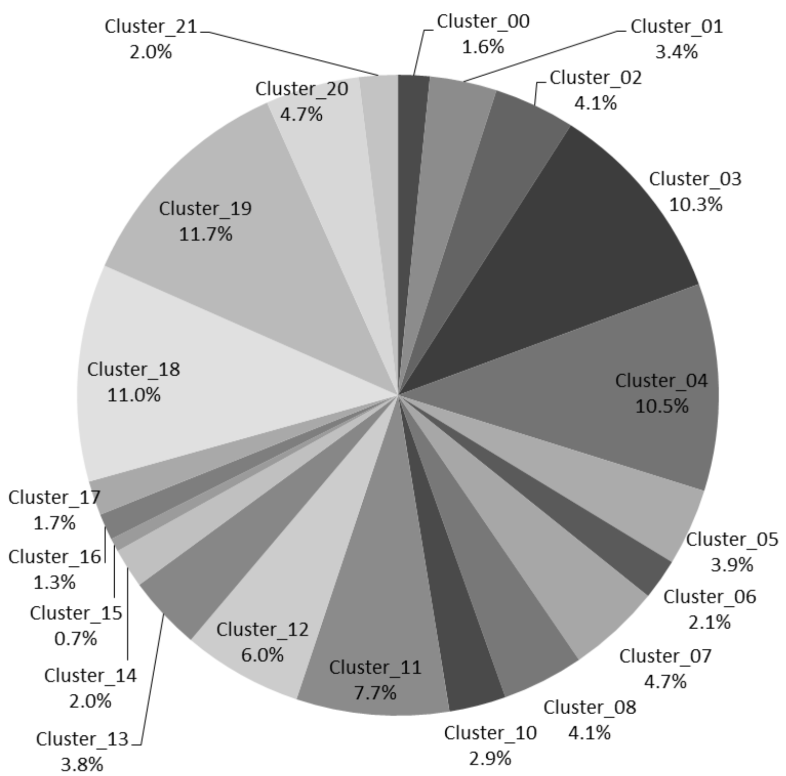


Figure 1: Volume of cluster-episodes by cluster

Note: Cluster labels are provided in Table 1 (page 8).

The largest are clusters 18 and 19 (cognitive impairment with low and moderate need) with around 11% of cluster-episodes, and clusters 3 and 4 (non-psychotic moderate and severe) with around 10% of cluster-episodes allocated to them respectively.

3.3 Analysis of costs

Using the two types of costs previously described (admitted and non-admitted care) for each cluster, we can calculate a cost index for each provider comparing its total cost with what this cost would have been if its cost for each cluster were equal to the national average. This index provides a simple way to identify providers that are outliers in terms of costs. Those providers with costs above the national average (the prospective tariff), would potentially lose out under an episodic payment approach, whereas providers with costs below the national average could potentially make a surplus (Dawson & Street, 1998).

The cost for each cluster can be calculated as the sum of costs for inpatient activity and for non-admitted days. We denote this cost

$$C_c = warddays_c * rc_{adm,c} + nonadm_c * rc_{nonadm,c},$$

where $warddays_c$ is the number of admitted days and $nonadm_c$[6] is the number of non-admitted days while assigned to cluster c, $rc_{adm,c}$ is the cost of an admitted day and $rc_{nonadm,c}$ is the cost of a non-admitted day in cluster c.

Since there are two types of costs, one at provider level and a national average, we can calculate two versions of this cost for each provider. These two costs in turn can be used to calculate a cost index for each provider.

Table 3 shows the cost index calculated for each provider as

$$\frac{\sum_{\forall c} C_c^p}{\sum_{\forall c} C_c^{NA}} = \frac{\sum_{\forall c}(warddays_c^p * rc_{adm,c}^p + nonadm_c^p * rc_{nonadm,c}^p)}{\sum_{\forall c}(warddays_c^{NA} * rc_{adm,c}^{NA} + nonadm_c^{NA} * rc_{nonadm,c}^{NA})}$$

where $warddays_c$ is the number of days spent as inpatient and $nonadm_c$ is the number of non-admitted days while assigned to cluster c. Costs differ in their superscript, p indicates provider level and NA national averages, $rc_{adm,c}$ is the cost of an admitted day and $rc_{nonadm,c}$ is the cost of a non-admitted day in cluster c.

We observe that the highest cost provider has costs 55% higher than average and that the lowest cost provider has costs 25% below average. The table includes all providers that reported costs.[7,8]

[6] $nonadm_c = epidays_c - warddays_c$

[7] Most providers report costs for all clusters, only 12 report costs for a subset of the clusters. The index is calculated using only the clusters for which the provider reported costs.

[8] Three providers did not report costs and one reported the same costs for admitted and non-admitted days.

Table 3: Cost index for each provider against the national average cost across all clusters

Provider	Cost in £000		Cost Index
	P	NA	
1	35,686	22,989	1.55
2	149,389	113,435	1.32
3	124,355	96,982	1.28
4	72,197	56,548	1.28
5	126,304	100,513	1.26
6	99,823	80,343	1.24
7	61,308	50,569	1.21
8	36,159	30,471	1.19
9	53,943	46,607	1.16
10	47,448	41,231	1.15
11	49,765	43,677	1.14
12	45,073	39,633	1.14
13	48,086	42,439	1.13
14	73,610	65,043	1.13
15	171,018	152,237	1.12
16	222,470	200,130	1.11
17	39,606	35,789	1.11
18	84,679	77,067	1.10
19	97,996	90,590	1.08
20	78,060	72,493	1.08
21	18,876	17,563	1.07
22	161,598	153,637	1.05
23	195,457	189,892	1.03
24	82,366	80,228	1.03
25	112,328	110,520	1.02
26	104,131	102,749	1.01
27	55,878	55,199	1.01
28	137,543	137,271	1.00
29	60,738	60,961	1.00
30	84,876	85,362	0.99
31	2,586	2,605	0.99
32	178,869	180,463	0.99
33	51,478	52,142	0.99
34	7,537	7,636	0.99
35	132,324	134,306	0.99
36	177,456	180,441	0.98
37	61,377	63,892	0.96
38	16,747	17,640	0.95
39	84,429	89,057	0.95
40	70,742	74,726	0.95
41	74,152	78,403	0.95
42	51,472	54,461	0.95
43	43,917	46,927	0.94
44	94,901	101,605	0.93
45	77,353	83,740	0.92
46	213,427	231,370	0.92
47	85,802	95,384	0.90
48	19,132	21,274	0.90
49	54,288	60,598	0.90
50	57,340	64,186	0.89
51	76,245	86,219	0.88

52	91,811	104,700	0.88
53	50,210	60,293	0.83
54	43,347	53,081	0.82
55	66,893	86,253	0.78
56	120,764	161,217	0.75

Table 4 shows the cost index for each cluster across all providers who have activity in that particular cluster. It shows the minimum cost as a proportion of the national average and the maximum cost as a proportion of the national average. It then shows the ratio of each of these percentages. Clusters which display a large degree of variation in cost are cluster 0. This is not surprising since it is the variance cluster and patients are assigned to this cluster, when there is uncertainty about the correct allocation. Other clusters with large variability include clusters 1, 2, 15, 18, 19 and 21. Cluster 1 for example shows a more than ten-fold variation between providers in terms of cost.

Table 4: Cost index for each cluster across all providers against the national average and the ratio of variability

	Numbers of providers	Minimum as % of national average	Maximum as % of national average	Maximum divided by minimum
Cluster 0	47	33%	337%	10.34
Cluster 1	50	39%	415%	10.61
Cluster 2	52	43%	349%	8.15
Cluster 3	52	46%	211%	4.60
Cluster 4	56	57%	171%	3.00
Cluster 5	56	53%	155%	2.90
Cluster 6	56	57%	183%	3.19
Cluster 7	56	40%	183%	4.53
Cluster 8	56	56%	150%	2.67
Cluster 10	56	53%	151%	2.87
Cluster 11	53	49%	140%	2.86
Cluster 12	56	64%	141%	2.22
Cluster 13	56	69%	155%	2.25
Cluster 14	56	72%	209%	2.91
Cluster 15	56	66%	421%	6.37
Cluster 16	55	68%	150%	2.21
Cluster 17	55	77%	145%	1.89
Cluster 18	53	40%	302%	7.51
Cluster 19	56	57%	297%	5.21
Cluster 20	56	57%	204%	3.60
Cluster 21	56	50%	283%	5.69

Note: Cluster labels are provided in Table 1 (page 8).

3.4 Analysis of activity

Using the information on activity, we can observe the variation across providers in the distribution of activity, i.e. how much activity is admitted and how much is non-admitted, in each cluster. Similarly, we can use the number of days with contacts with a health care professional (for which we do not have an associated cost) to find out whether there are differences between providers in terms of how often their patients from different clusters have access to a health care professional. This gives us a proxy of resource use.

In terms of activity, we observe differences between clusters and between providers; e.g. some clusters are more likely to require ward admissions, but not all providers will have their patients admitted for the same period.

One way to assess whether providers differ in how they treat their patients, and therefore in resource utilisation, is by comparing how long cluster episodes are, how many days are spent admitted to hospital and on how many days the patients within that cluster have contact with a health care professional. Figure 2 is the plot for Cluster 4 (Non-psychotic - severe) showing the total length of cluster episodes on the right axis. Providers are ordered in terms of lowest to highest length of cluster episode. The plot shows this varies from around 20 days for some providers to around 250 days for others. The left axis shows the percentage of the days spent admitted to hospital or having contact with a health care professional for this cluster. If we expect that within any given cluster, patients should spend a fixed proportion of time as an admitted patient or having contact with healthcare professionals, we would expect these lines to be flat, i.e. no matter how long the cluster-episode we anticipate the proportion of admitted days to be the same across providers.

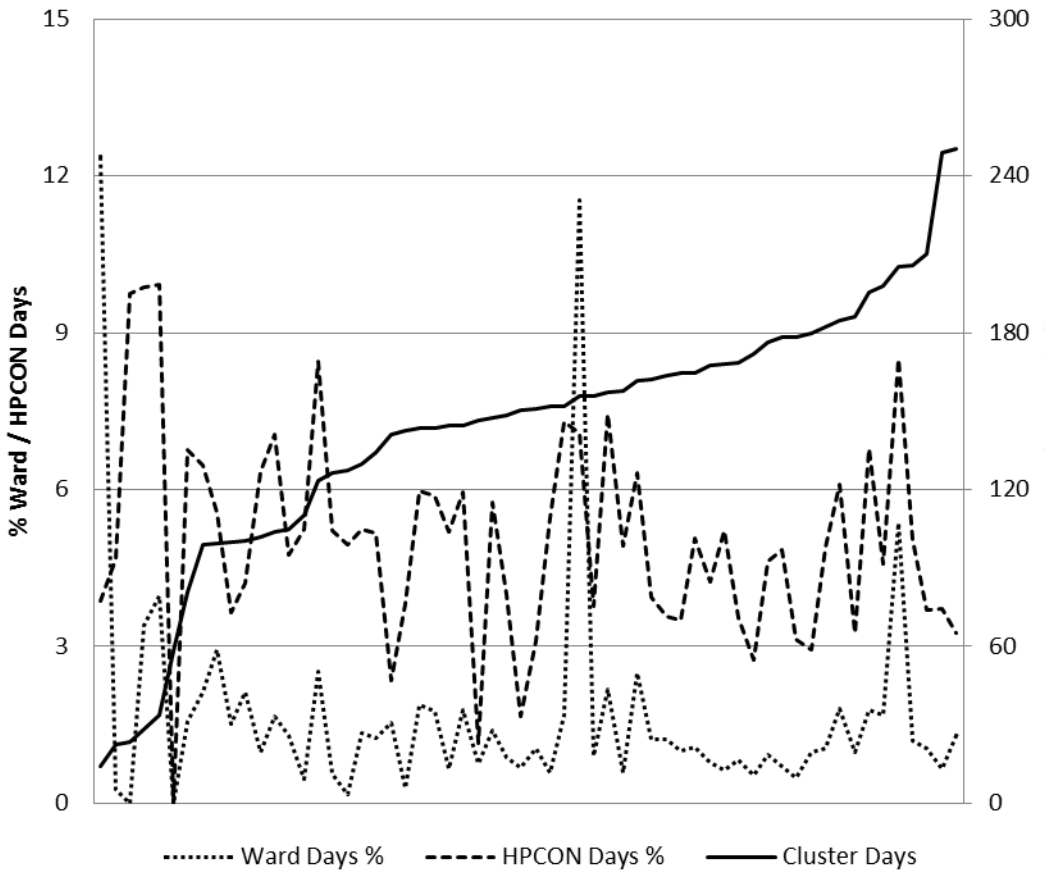


Figure 2: Cluster 4 - Non-psychotic (severe) - length of cluster episode, percentage of days spent as inpatient stay and percentage of days of contact with a health care professional

We see from Figure 2 that there is considerable variability across providers within Cluster 4 in terms of inpatient activity contact with healthcare professionals. Resource use is therefore not constant across providers within this classification. This implies that there is potentially considerable heterogeneity within clusters and across providers in terms of activity and resource use.

In Appendix A we show the plots for all 21 clusters and these confirm that providers have different patterns of activity; e.g. cluster 11 has greater variation in terms of its total length, and cluster 18 is the one with lowest average activity in terms of admissions and contact with health care professionals.

Another way to determine whether providers differ in resource utilisation in the treatment of patients is to look at the correlation between the length of the cluster-episode and the activity performed within it. If all providers in a cluster deliver the same services per period of time, but reported it at different intervals, we would expect the correlations to be around one, i.e. longer cluster-episodes translate into proportionally more services delivered in comparison to shorter ones.

Table 5 and Table 6 show the coefficient estimates of multilevel regression models (Rabe-Hesketh & Skrondal, 2008) expressed as elasticities (Castelli & Ukovich, 2003; Norton et al, 2002) for all clusters where the explanatory variable is the length of the cluster-episode and the dependent variable is either ward days (Table 5) or contact with health care professionals (Table 6). In both cases we consider a three level model (Rabe-Hesketh & Skrondal, 2008), where clusters are nested within patients and patients are nested within providers.[9] We use an elasticity because it is normalised to one for equi-proportional changes. We would therefore expect the estimate to be around 1.

The results show that longer cluster episodes do not translate into proportionally more days with activity of the types considered here and all elasticity estimates are below one. The results for ward-days (Table 5) are not significantly different from zero (p-value ≥ 0.10) for three clusters, 3, 4 and 6, while for days with contact with health care professional (Table6), they are significant in all clusters.

Table 5: Three-level regressions: ward days

	Elasticity	Std. Err.	p-value	N
Cluster 00	0.1281	0.0325	0.00	31,298
Cluster 01	0.1210	0.0214	0.00	66,394
Cluster 02	0.4058	0.0210	0.00	80,247
Cluster 03	0.0144	0.0156	0.36	202,626
Cluster 04	-0.0066	0.0127	0.61	205,716
Cluster 05	0.0341	0.0150	0.02	76,322
Cluster 06	0.0550	0.0330	0.10	40,649
Cluster 07	0.1652	0.0239	0.00	91,969
Cluster 08	0.2441	0.0209	0.00	80,725
Cluster 10	0.0735	0.0162	0.00	56,307
Cluster 11	0.2302	0.0205	0.00	151,287
Cluster 12	0.2677	0.0187	0.00	118,825
Cluster 13	0.3630	0.0214	0.00	73,832
Cluster 14	0.1851	0.0075	0.00	39,210
Cluster 15	0.1240	0.0133	0.00	13,180
Cluster 16	0.3173	0.0336	0.00	26,247
Cluster 17	0.4533	0.0311	0.00	34,056
Cluster 18	0.1388	0.0320	0.00	215,528
Cluster 19	0.0868	0.0191	0.00	229,717
Cluster 20	0.2406	0.0254	0.00	93,093
Cluster 21	0.3407	0.0397	0.00	38,354

Note: Cluster labels are provided in Table 1 (page 8).

[9] Using only two levels, patients and providers, gives similar results.

Table 6: Three-level regressions: days with contact with health care professional

	Elasticity	Std. Err.	p-value	N
Cluster 00	0.2217	0.0144	0.00	31,298
Cluster 01	0.1495	0.0074	0.00	66,394
Cluster 02	0.1779	0.0088	0.00	80,247
Cluster 03	0.2313	0.0116	0.00	202,626
Cluster 04	0.2757	0.0122	0.00	205,716
Cluster 05	0.2651	0.0107	0.00	76,322
Cluster 06	0.3455	0.0144	0.00	40,649
Cluster 07	0.3794	0.0157	0.00	91,969
Cluster 08	0.3843	0.0170	0.00	80,725
Cluster 10	0.3778	0.0167	0.00	56,307
Cluster 11	0.3949	0.0149	0.00	151,287
Cluster 12	0.3947	0.0153	0.00	118,825
Cluster 13	0.3839	0.0161	0.00	73,832
Cluster 14	0.1838	0.0087	0.00	39,210
Cluster 15	0.1910	0.0104	0.00	13,180
Cluster 16	0.3829	0.0208	0.00	26,247
Cluster 17	0.4757	0.0268	0.00	34,056
Cluster 18	0.2443	0.0102	0.00	215,528
Cluster 19	0.2225	0.0099	0.00	229,717
Cluster 20	0.2240	0.0126	0.00	93,093
Cluster 21	0.2267	0.0147	0.00	38,354

Note: Cluster labels are provided in Table 1 (page 8).

4. The assignment to clusters using the Mental Health Clustering Tool (MHCT)

In this section we consider how well the Mental Health Clustering Tool (MHCT) allocates patients with similar needs to clusters and the degree of homogeneity within the clusters. These are crucial underpinnings of an episodic payment approach. As set out in Section 1, the assignment of each episode to one of the 21 clusters assumes that these clusters capture the variation in mental health symptoms and resulting needs. We test whether the 21 clusters represent the empirical variation in needs.

To address this issue, we used statistical data-driven clustering to identify groups of cases that belong closer together based on similar patterns across a number of variables. The variables in this case were the 13-item HoNOS and 5-item SARN ratings comprising the MHCT, provided for each of the episodes in the MHSDS. The statistical algorithm identifies an optimal separation of all episodes into a number of 'classes'. This term is used to differentiate the results of our analysis from the *clusters* and because of the statistical technique used which is called latent class analysis (LCA). The statistical separation into these classes sorts the cluster-episodes in such a way that the patterns within one class are as similar as possible (i.e. maximally homogeneous within) and as different as possible between classes (i.e. maximally heterogeneous between). This ensures that for a given number of classes there is no data-driven way to separate the data into classes that are more different than the ones resulting from this analysis. This analysis of the HoNOS and SARN items with a given set of 21 classes will result in a differentiation of needs patterns that are maximally different, i.e. there would be no classification that describes the differences between patient episodes more optimally or efficiently for this number of classes. The larger the overlap between these *data-driven classes* and the *documented cluster* assignments, the more content-valid is the current practice of assigning patient-episodes to clusters.

We use the same data for this analysis as reported in Section 3.1. The analysis can only consider cluster-episodes with at least one valid recording of MHCT (HoNOS and SARN) items, so cluster-episodes without this must be excluded, but it does not require the cluster-episode to have a valid PbR cluster code (as the cost and activity analysis did), so cluster-episodes with valid MHCT but no PbR cluster code can be included. The sample is smaller than before because there are more cluster-episodes without a valid MHCT (around 847k) than without PbR cluster code (around 14k), it therefore includes 1,132,376 cluster-episodes.

4.1 The MHCT as a measurement tool

The MHCT was developed based on the HoNOS to cover relevant indicators of need for treatment planning in mental health care (Self, Rigby et al, 2008). The starting point of the development was the HoNOS (see Table 7), which 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 et al, 1994). HoNOS is comprised of 12 items and 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 (Wing et al, 1994).

The SARN (Self, Painter et al, 2008) (see Table 8) comprises five items and assesses historical problems that occur less frequently or sporadically and these five items (as well as the additional HoNOS item for the assesment of non-psychotic unreasonable beliefs) were a result of the development of additional items to cover relevant areas for a clinical decision tool that could be used across a broad range of services (Self, Rigby et al, 2008).

Both HONOS and SARN use the same rating scale (see Table 9). The MHCT is accompanied with detailed guidance that provides examples regarding how to differentiate between different levels of severity in order to increase inter-rater objectivity and reliability. For example for HONOS10 ("Problems with activities of daily living") guidance suggests "Self-care adequate, but major lack of performance of one or more complex skills" to be coded "2-Mild problem, but definitely present" and "Major problem in one or more areas of self-care (eating, washing, dressing, toilet) as well as major inability to perform several complex skills" as a "3-Moderately severe problem" (Department of Health, 2014a).

Table 7: Health of the Nation Outcome Scales (HoNOS), current status

Item Number	Description
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

Table 8: Summary of Assessments of Risk and Need (SARN)

Item Number	Description
1.	**Agitated behaviour/expansive mood (historical)** *Rate agitation and overactive behaviour causing disruption to social role functioning. Behaviour causing concern or harm to others.*
2.	**Repeat self-harm (historical)** *Rate repeat acts of self-harm with the intention of managing people, stressful situations, emotions or to produce mutilation for any reason.*
3.	**Safeguarding other children & vulnerable adults (historical)** *Rate the potential or actual impact of the patient's mental illness, or behaviour, on the safety and wellbeing of vulnerable people of any age.*
4.	**Engagement (historical)** *Rate the individual's motivation and understanding of their problems, acceptance of their care/treatment and ability to relate to care staff.*
5.	**Vulnerability (historical)** *Rate failure of an individual to protect themselves from risk of harm to their health and safety or wellbeing.*

Table 9: Rating scale used for HoNOS and SARN

Score	Codes used across all items
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

4.2 Latent class analysis

The statistical technique used for the following analyses is the pattern identification method called latent class analysis (LCA) (Lazarsfeld & Henry, 1968; McCutcheon, 1987). The advantage of the LCA over other techniques such as factor analyses or cluster analyses used before in this context (Leach et al, 2005; Salvi et al, 2005; Self, Rigby et al, 2008; Speak et al, 2012; Speak & Muncer, 2015) is that it assumes that there is a potential unknown causal factor that groups the available cluster-episodes into qualitatively different patterns that can be nominal (i.e. un-ordered), ordinal (some classes representing ‘more’ of the latent cause than others) or any mixture of these (Kempf, 2012).

The following aspects of the LCA model support the rationale for its use in clusters:

1. The assumption of an unknown underlying cause is mirrored i) in the clinical assumption that the care clusters should represent different categories or syndromes of mental illness that lead to these patterns; ii) the assumption that the clusters effectively should be caused by different combinations of need; and/or iii) the relevant question from a tariff system perspective, that the clusters represent patterns of needs-based resource use. While we won't be able to differentiate between these three causal assumptions in terms of appropriateness or plausibility, the LCA approach would be deemed a plausible statistical analysis approach under all three of them.

2. The fact that the resulting classes can represent both unordered patterns of needs (i.e. classes that are qualitatively different in their combination of symptoms and it cannot be said one class has a higher need level than another), as well as ordinal patterns of needs (i.e. for example more or less severe problems in the area of common mental disorders), makes this approach more flexible than dimensional approaches that have been used on the HoNOS before (Böhnke & Croudace, 2015; Speak & Muncer, 2015), which would assume that all patterns observed in the data could be ordered from a combination of symptoms that represents the lowest level of need to a pattern that represents the maximal need (within this sample). Additionally, it mirrors the assumption within the clusters of qualitatively different patterns (e.g., non-psychotic vs. psychotic) as well as ordinal variation within these (e.g., non-psychotic-moderate vs. non-psychotic-severe vs. non-psychotic-very severe).

LCA can be used as an exploratory (McCutcheon, 1987) as well as a confirmatory analytical strategy (Finch & Bronk, 2011). The following analysis sits in between these two positions: i) we are interested in the appropriateness of the 21 clusters in describing the empirical variation in patient need, therefore we set the number of *classes* to 21; ii) but we do not impose any statistical constraints on the structure of the patterns that should be identified by these classes.

After determining the structure of the latent classes (i.e., what the 21 patterns of need look like), we allocated all cluster-episodes to their most likely class and tested the overlap with the 21 care-clusters by determining *Cramer V* coefficients. These coefficients vary between 0 (no concordance between two nominal variables) and +1 (which would indicate perfect concordance). This coefficient is determined for the concordance between the LCA classes and the clusters as well as the LCA classes and the superclusters.

The program used to estimate the LCA was Mplus 7.11 (Muthén & Muthén, 1998).

4.3 Results for structure of the 21 classes

Overall, a substantial number of cluster-episodes were assigned to all 21 classes. Figure 3 displays the frequency distribution of the cluster episodes allocated to the 21 classes. Note that the letter assigned to each class is arbitrary and a result of the statistical program, it does not inform us about the importance or any other aspect of the estimation result. We deliberately chose to use letters for the classes, to differentiate them from clusters. Figure 3 shows variation in the frequency with which cluster-episodes were assigned to classes, ranging from $n_{\mathrm{cl=D}}$ = 5,882 (0.5%) to $n_{\mathrm{cl=Q}}$ = 96,928 (8.1%).

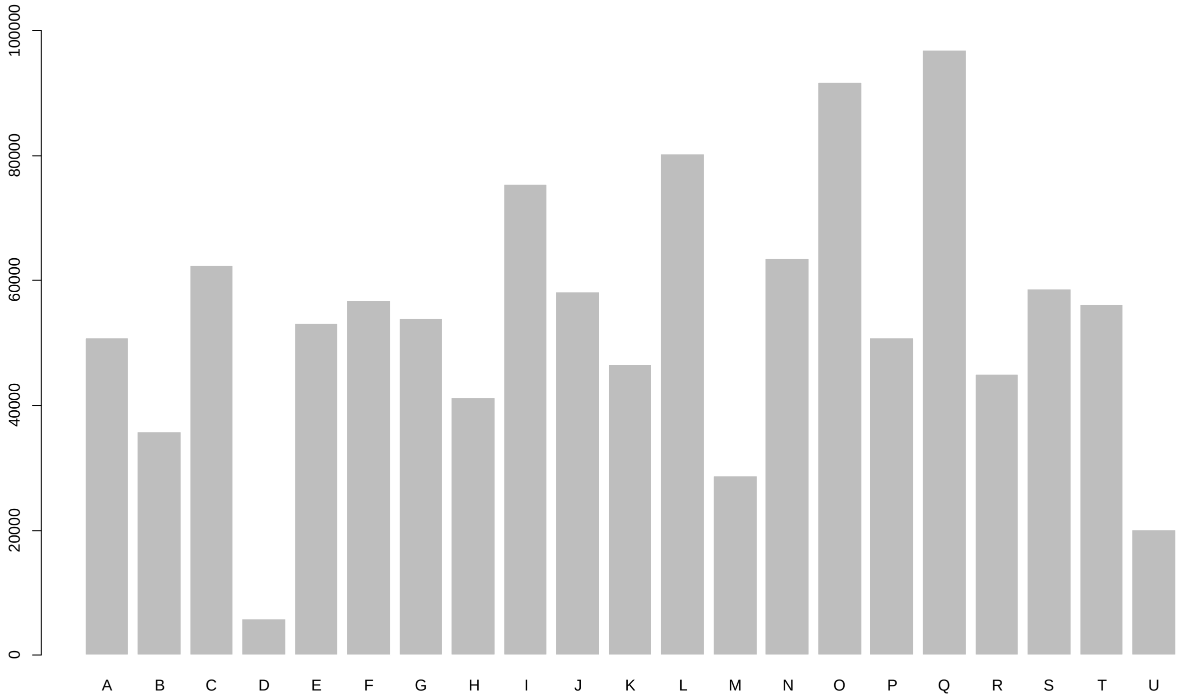


Figure 3: Frequency distribution of cluster episodes allocated to classes

To look at the result more closely, Figure 4 displays the pattern of probabilities for the largest of the estimated classes, i.e. the one with the highest prevalence in this data set (Q). The left panel of the figure presents the estimated category probabilities for all HoNOS (see Table 7) and SARN (see Table 8) items within this class. Grey represents the probability of "0 = no problems" indicated for that specific item; yellow a "1 = minor problem, not requiring action"; orange a "2 = mild problem, but definitely present"; red a "3 = moderately severe problem"; and dark-red a "4 = severe to very severe problem" (not present in Figure 4).

An episode allocated to class Q has a high probability of "no problems" on 16 of the 18 items (e.g., larger than 0.80 for item HoNOS1). The two items not in line with this trend are HoNOS7 (Problems with depressed mood) and HoNOS8 (Other mental and behavioural problems) where categories 1 ("minor problem, not requiring action", see Table 9) and 2 are the most likely response options.

The item HoNOS8 is accompanied by a qualifier that can be used to code more specifically which problem is rated. This information for class Q is presented in the right panel of Figure 4. With a probability p > 0.50 these cluster-episodes were accompanied by problems with anxiety; and to a lesser extent by problems with stress or sleep.

To give a more descriptive presentation, combining the item content, category labels and probabilities in a potentially more accessible vignette, cluster episodes allocated to class Q are characterised by at times gloomy mood and mild anxiety, a pattern typically associated with mild common mental disorders.

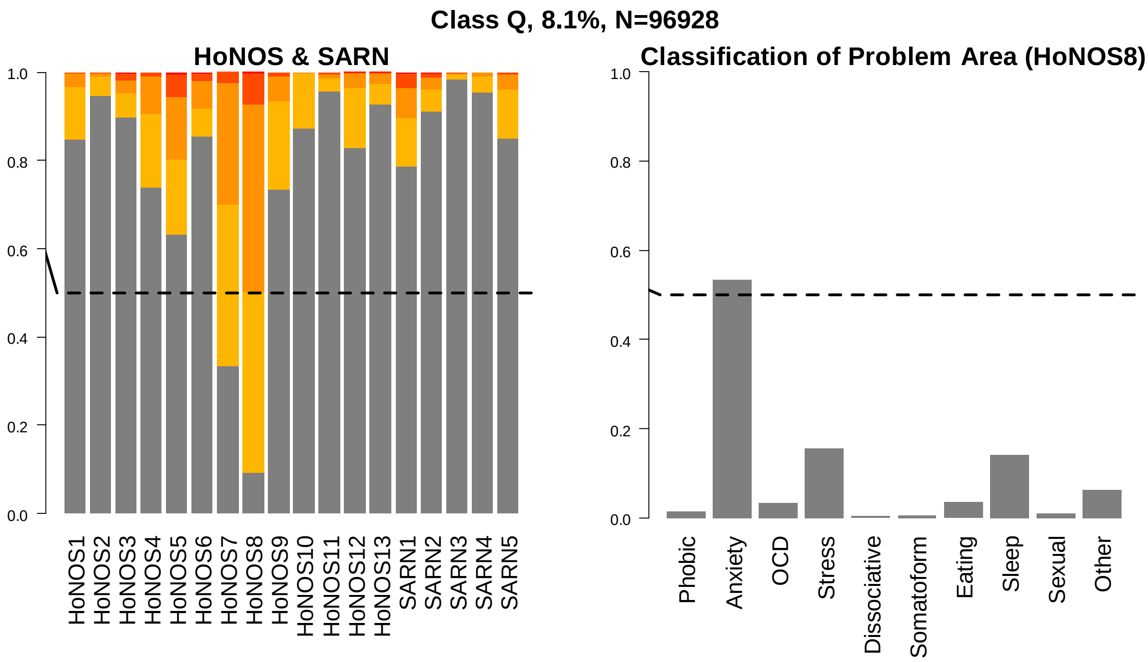


Figure 4: Category probability pattern for the largest identified class (Q)

In contrast to these results, Figure 5 presents the results for class M, which contains those cluster-episodes with the highest severity ratings. Cluster-episodes in this class display severe problems with anxiety (stress, sleep problems) dominating most activities (HoNOS8). Additionally, patients are very depressed (HoNOS7). Their relationships often cause persisting problems due to lack of support (HoNOS9). Beyond that, patients have major problems in several areas of daily living (HoNOS10) and lack opportunity of daytime activities to use their intact skills (HoNOS12). There is also a substantial probability that patients have been physically aggressive towards others / behaved in a threatening manner (HoNOS1) and caused harm to others in the past (SARN1). Finally, patients suffer from a breakdown in their ability to protect themselves (SARN5) and they tend to have little insight into their problems, engaging inappropriately with services (SARN4).

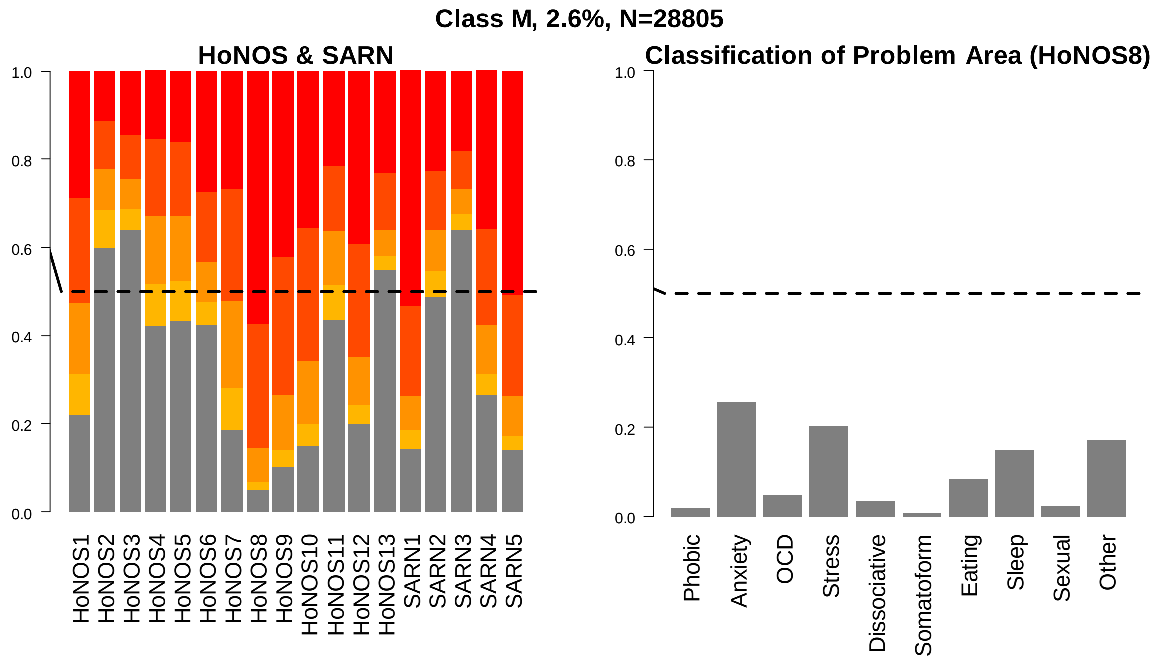


Figure 5: Category probability pattern for the most severe identified class (M)

The following sections will discuss further classes in more detail and are used to illustrate aspects of the overlap with super-classes and clusters of the MHCT. A graphical display of all classes can be found in Appendix B (page 60).

4.4 Results for overlap with super-classes

After allocating all observed cluster-episodes into LCA-classes, we looked at the overlap between our results and the overall categorisation that is provided by the three overarching categories of the MHCT that group the clusters into super-classes: "Non-psychotic", "Psychosis", and "Organic". The overlap between these two classifications is given by *Cramer V* = 0.48 (*p* < .001), which indicates that there is some overlap, but not a strong one.

Table 10 presents the cross-classification of these two variables, with column-percentages to identify within each clustering-tool super-class which is the most prevalent latent class pattern. If no relationship existed between these two variables, each of these cells would have a probability of (1/ 21) = 4.7% – highlighted in grey are therefore within each column those cells that show values > 10%, which is twice the expected probability under the assumption of no relation between these two classification systems.

Table 10: Cross-tabulation of LCA class results (rows) and super-class allocations (columns); column-wise percentages

	Non-Psychotic	Psychosis	Organic	Missing Superclass Allocation
A	1.31	13.69	0.56	2.49
B	4.17	3.99	0.46	1.89
C	6.19	8.74	0.89	3.28
D	0.41	0.48	1.15	0.05
E	2.71	11.45	0.37	4.08
F	1.14	1.07	21.22	3.19
G	8.89	1.92	0.37	2.52
H	1.1	5.59	8.26	1.95
I	11.48	3.21	0.19	6.23
J	7.25	3.16	1.33	7.2
K	0.57	0.84	14.83	7.69
L	11.91	2.42	0.63	9.08
M	2.26	4.38	1.31	1.58
N	3.8	11.83	2.65	3.54
O	11.54	4.18	3.11	11.2
P	3.74	8.11	1.83	3.56
Q	10.6	6.45	4.72	11.28
R	0.75	1.43	10.79	10.58
S	8.83	3.01	0.3	3.95
T	0.96	3.46	18.69	2.15
U	0.38	0.59	6.34	2.5
Total	100	100	100	100
n	495,192	289,305	204,844	143,035

The results show that the overlap is far from clearly identifying a reliable 1:1 mapping, for which we would expect large column percentages for few classes. But some convergence emerges that is also specific to the super-classes: the classes with probabilities > 10% in each column do not overlap across super-classes. For example, the "Non-Psychotic" super-class overlaps with classes I, L, O and Q from our analysis. The results for class Q were presented above (Figure 4) and episodes allocated in this class correspond strongly with non-psychotic experiences. As the following will illustrate, the empirically identified classes correspond to patterns that differ partly qualitatively, but mainly quantitatively in levels of need and severity.

Class L (Figure 6) is very similar in the overall pattern of probabilities as compared to class Q, only the probabilities for endorsing categories representing higher levels of severity are higher and the problems that accompany the HoNOS assessment are more diverse (HoNOS8).

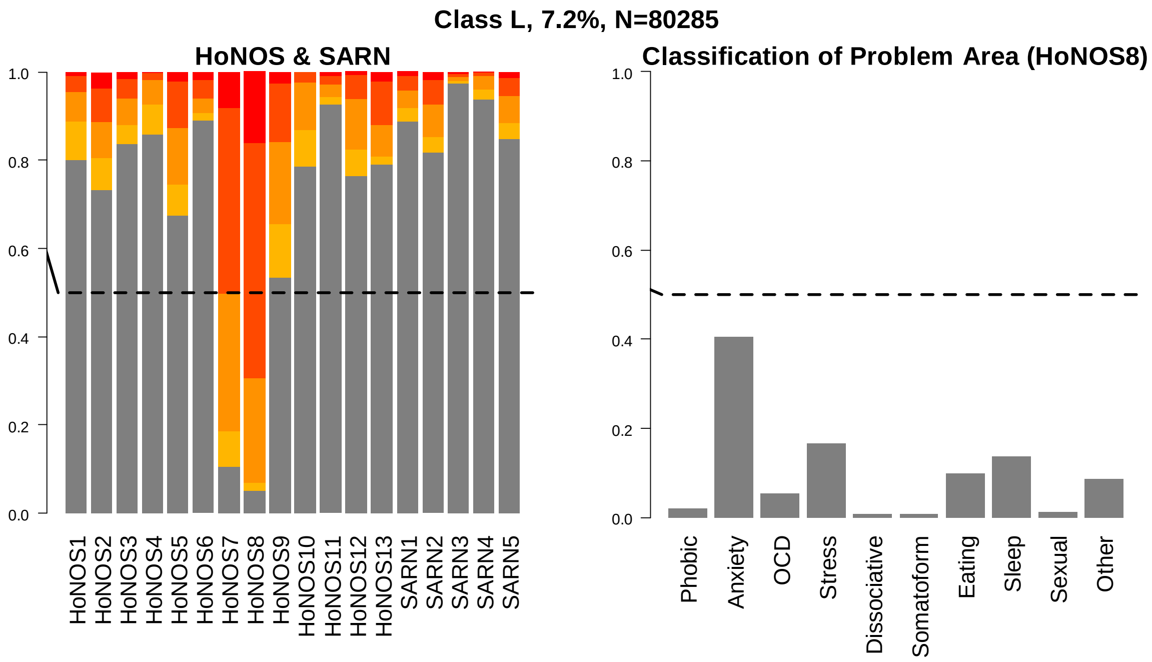


Figure 6: Category probability pattern for class L as a representative of the "Non-Psychotic" super-class

Class I (Figure 7) also shows a slight increase in the probability of higher problem severity on items HoNOS7 and HoNOS8 as compared to class Q (Figure 4). But this pattern goes together with an increase in the probability of at least mild problems for a range of other areas like overactive, aggressive, disruptive or agitated behaviour (current and historical: HoNOS1 + SARN1), non-accidental self-injury (current and historical, HoNOS2, SARN2) as well as problems with relationships (HoNOS9).

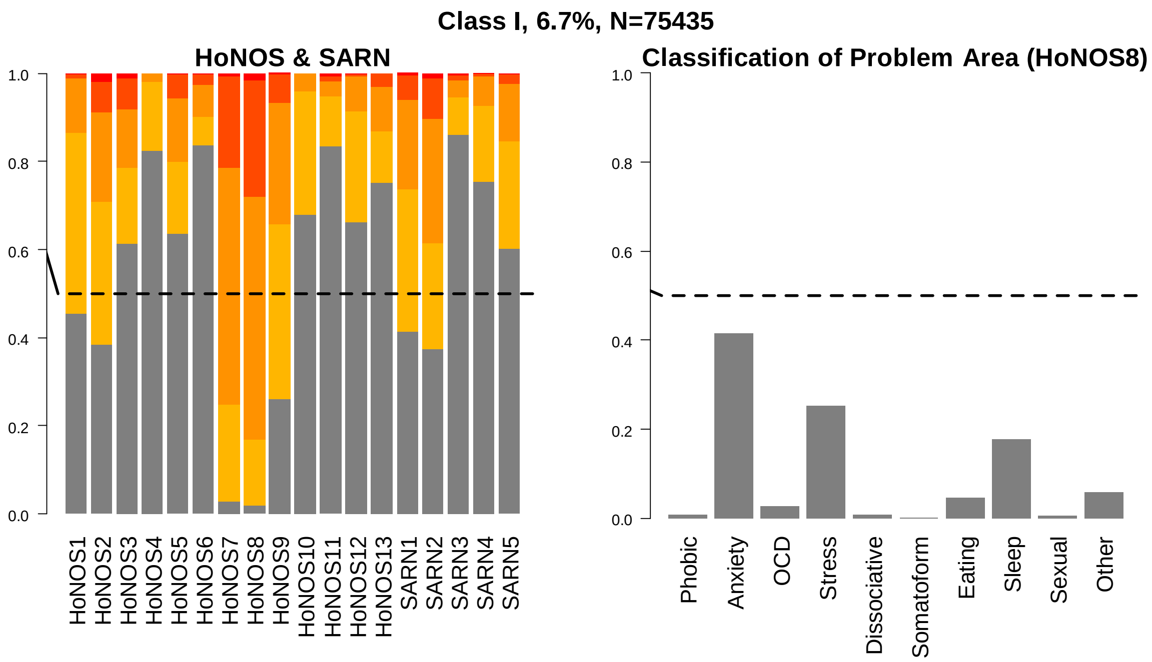


Figure 7: Category probability pattern for class I as a representative of the "Non-Psychotic" super-class

Class O (Figure 8) displays a slightly different pattern. The overall severity level of episodes allocated to this class is higher than in classes L and Q, but similar to class I with a qualitatively different element. While in class I, self-harm and aggressive behaviour were a potential problem, this is not as much the case for class O. Instead, episodes in class O are characterised by at least mild levels of problems due to physical illness or disability (HoNOS5), problems with activities of daily living (HoNOS10), and problems with occupation and activities (HoNOS12), which points at a very different mix of problems and needs compared to class I. Information on this is limited in the available data, but class O could relate to an overlap between common mental disorder and physical comorbidities.

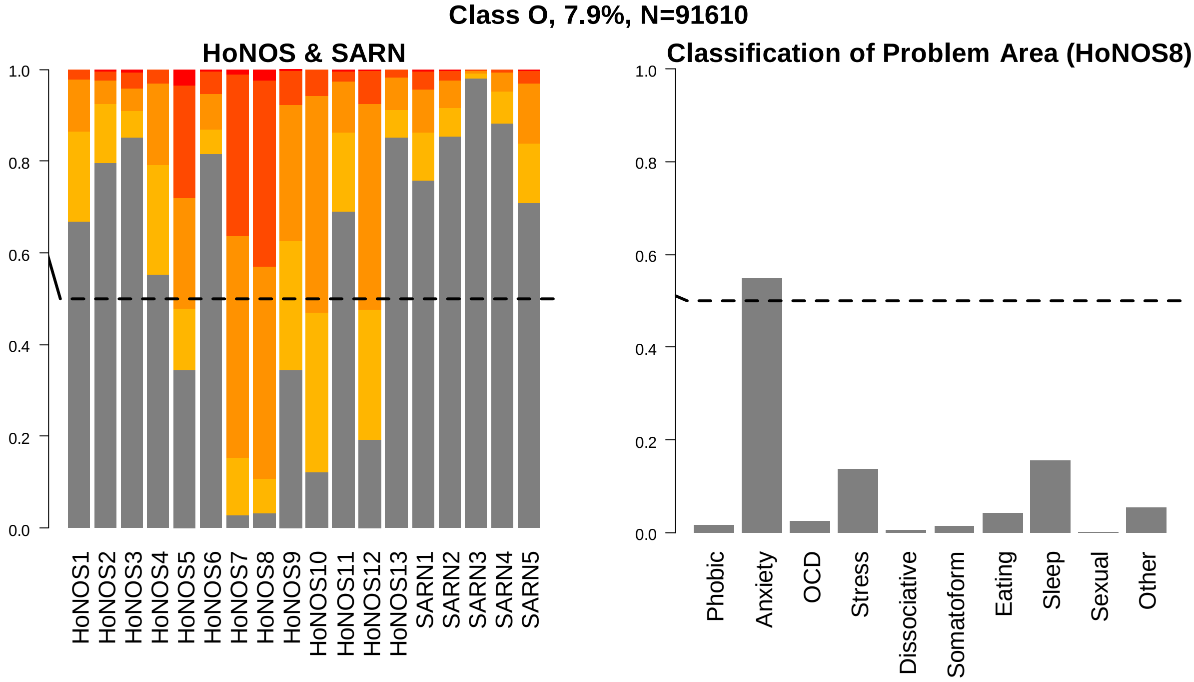


Figure 8: Category probability pattern for class O as a representative of the "Non-Psychotic" super-class

These examples illustrate that the empirical variation identified by the LCA makes clinical sense and identifies different patterns of severity and problem constellations. Although an assignment of severity levels to such classes is not straightforward, classes L and Q are representing clearly ordered levels of common mental disorder; both of which are probably lower in severity and needs than classes I and O which differ qualitatively in the problems that need to be addressed during treatment.

As can be seen in more detail in appendix B, in the Psychosis superclass, most of the classes that show strong overlap (A, N) have a distinct probability of hallucinations (HoNOS6; the only exception is class E) and the other classes where HoNOS6 has a high probability of occurring feature at least strongly (especially classes H & M) and are not common classes in any of the other super-classes. Cluster-episodes allocated to the four classes with the main overlap within the Organic superclass (F, K, R, T) all show heightened probabilities of cognitive disturbances (HoNOS4) and problems with daily living (HoNOS10).

While this is a positive result which indicates some correspondence between classes and super-classes, even aggregating on such a general level shows the mismatch between both classifications. For the non-psychotic superclass only 45% of episodes are recruited from the four main contributing latent classes, while 11% are drawn from classes that actually have the highest overlap with other super-classes. For the psychotic superclass these numbers are 37% from the main contributing classes and 23% for those that are main contributors for other super-classes. For the organic superclass these numbers are 66% from the main contributing classes and 12% for those that are main contributors for other super-classes.

We present the probability patterns for each of the 21 classes in Appendix B. The patterns of problems discovered by LCA make clinical sense and capture specific problem constellations regarding current and chronic difficulties. Looking at the most salient places of overlap in the cross-classification table also further shows that the overlap is at least to some extent due to common target problems and needs, but the extent to which this is true, differs across super-classes.

4.5 Results for overlap with MHCT cluster allocation

After allocating all observed cluster-episodes into LCA-classes, we looked at the overlap between our results and the overall categorisation that is provided by the 21 clusters. The overlap between these two classifications is *Cramer V* = 0.26 ($p < .001$), which indicates that there is some overlap, but not a strong one. Table 11 presents the cross-classification of these two variables, with row-percentages to identify within each clustering-tool cluster, which is the most prevalent latent class pattern. If no relationship existed between these two variables, each of these cells would have a probability of (1/21) = 4.7% – shown within each column therefore, are those cells that show values > 10%, which is twice the expected probability under the assumption of no relation between these two classification systems.

Similar trends as for the superclass concordance can be observed. For example class Q with its pattern of mild common mental distress, which was discussed in detail above (Figure 4) is a major contributing class for MHCT clusters 1-3 and since it is only covering mild mental health problems, its percentages decrease as the intensity of problems and need covered by the MHCT clusters increases. Finally, in cluster 3 classes I and O (see also Figure 7 and Figure 8 above) are more prevalent. Again, the overlap is not 1:1, but generally driven by quantitative (and some qualitative) differences in problems and needs.

Another good example of these trends is to look in more detail into the classes that make up MHCT cluster 5 (G, J, L). Class L was already described as a class with increased levels of common mental distress (Figure 6). These are also very pronounced in class G (HoNOS7, HoNOS8), but a broader range of problems adds to the severity, including problems associated with hallucinations/delusions (HoNOS6), problems with relationships (HoNOS9), and problems with daily living (HoNOS10).

Table 11: Concordance table comparing MHCT with results from LCA

CLASSES																					
MHCT	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**
0																	**16**				
1																	**37**				
2																	**26**				
3									**16**						**16**		**17**				
4												**18**			**16**						
5							**16**			**14**		**15**									
6							**24**														
7							**14**														
8																			**23**		
10														**12**		**10**	**10**				
11					**18**																
12	**15**													**15**							
13	**17**																				
14																**18**					
15										**13**											
16	**18**																				
17	**31**																				
18						**31**												**27**			
19						**22**					**20**									**25**	
20											**24**									**22**	
21								**21**													**34**

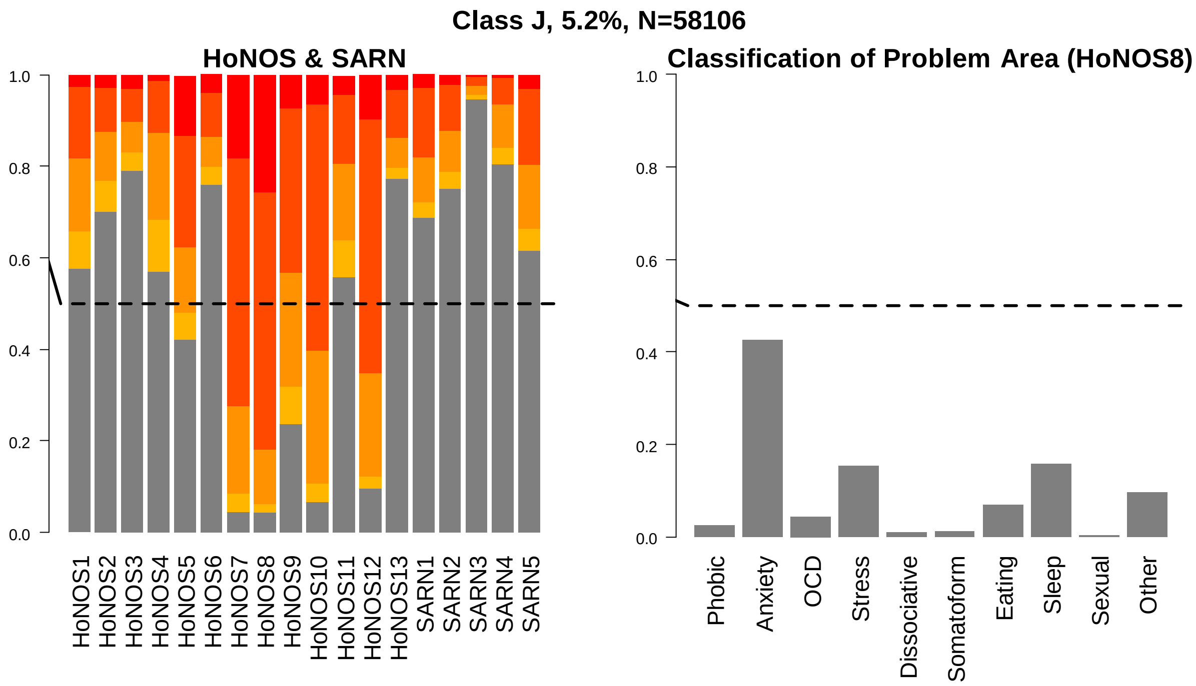


Figure 9: Category probability pattern for class J as a representative of the MHCT cluster 5 ("Non-psychotic (very severe)")

4.6 How well does the MHCT work?

Overall, the concordance between MHCT and the results from the LCA (assuming the same number of classes as MHCT clusters) is limited. The LCA classes separate the cluster-episodes into clinically meaningful patterns of problems and needs, but the 1:1 overlap between these two classification systems is low. Nevertheless, looking at the identified LCA patterns in detail reveals that the themes across both classification systems are broadly similar (separation between psychotic and non-psychotic episodes for example), but that the LCA classes still detect (mainly quantitative) differences in problems and needs between episodes that have been classified within the same super-class (Table 10) or the same cluster (Table 11). As such, the MHCT clusters seem to have a certain degree of validity with a view to identifying problem and need areas. But it is questionable whether the tool correctly quantifies the differences in problems and needs between patients/episodes.

Several limitations of the data quality will be discussed in the next section 5. Three points need a more thorough discussion with respect to this specific analysis. First, we did not set out to identify an 'optimal' number of classes representing the variation in the 18 items of HoNOS and SARN, the comparison between an existing theoretically built and an empirical system was the main point of this analysis. With regards to the statistical quality of the solution, the entropy of this specific solution is .77 and exploratory LCA runs so far have shown, that solutions between 10 and 25 classes all fall into the range of .77 to .80 and cannot be distinguished based on this criterion. The more commonly applied information criteria that quantify the relationship between overall model fit and the number of parameters needed to estimate this model (McCutcheon, 1987) are still decreasing up to 25 classes, which either indicates that still more heterogeneity is present in the data or that the huge volume of data would necessitate other approaches to select the final model.

Second, the inter-rater reliability of the original instruments provides important benchmarks for the concordance analysis presented here. Overall, the results for the HoNOS provide mixed evidence. Wing and colleagues (Wing et al, 1998) report a number of different reliability estimates, some focussed on individual items of the HoNOS, some on the whole score. They range from .49 to .97, and tend to be lower when more raters and more comparison instruments are used. In another validation study (Bebbington et al, 1999) the correlation between scores obtained from trained clinical staff and trained researchers ranged from .22 to .73. Based on these numbers and acknowledging that the comparison of multivariate patterns instead of scores is a much more stringent test than the comparison of scores, it seems unreasonable to expect the MHCT to work much better than it did. It may be that the HoNOS and therefore the MHCT performs within its range, but that its reliability is not particularly strong at the outset.

Third, the analysis undertaken in this project reflects both normative aspects of the system, as well as the 'pure' empirical variation. From a normative aspect we do not address whether other indicators would be more informative to separate the patient population, nor was the assumption of 21 clusters being optimal challenged. Although this might seem as a limiting factor, these are key aspects that allow the evaluation of the current practice that uses exactly these instruments and clusters. Another normative aspect that might influence the results is that clinicians might be aware of the functioning of the cluster assignment algorithm. This might therefore be reflected by the ratings that they provide on the HoNOS and SARN items, to make them 'more compatible' with the system. Although this cannot be ruled out, the little 1:1 overlap between assigned clusters and statistically determined clusters that was identified by the analyses makes strong influences of such behaviours unlikely.

5. Data issues and data quality

As previously mentioned, the data used in this report comes from the different data files drawn from the MHSDS. However we were not able to use all the variables we believed would be relevant to the analysis, as we encountered a number of data quality issues, which we describe below.

Probably the most common issue we found in the Events data file was the presence of observations that were identical except for the event identifier. Technically they were unique observations, but in practice they were duplicates as they gave the same information.

We found one type of event recorded several times in one day; this might be more or less plausible depending on the type of event. For example, having contact with different types of health care professionals on the same day might be more plausible than having several MHCT assessments on the same day. It is not impossible for a patient to have contact with more than one health care professional on the same day, but around 13% of observations correspond to patients with five or more contacts in one day, which is unlikely. An additional difficulty was the lack of information that could help determine which ones are actually different events, for example, if the variable that records the type of professional were better populated (the variable 'Job Role' has more than 65% missing values) it would be possible to see if the patient had contact with only one or more than one type of professional. Researchers and data users would then be able to assess the plausibility of the number of events recorded on the same day. In the case of the MHCT events, the same patient could have up to 18 responses[10] of the MHCT items on the same day.

In the Episodes data file the most common issue was the lack of end dates, and there were also observations that were identical except for the event identifier.

We found that both ward- and cluster-episodes did not record an end date. We could impute missing end dates using a variable calculated by the HSCIC (NHS Digital, 2013) that measures the length of the episode within the financial year; thus using the start date and this variable on the length of the episode it is possible to calculate the end date. For ward-episodes we also checked that the end date did not fall outside the cluster-episode within which the ward-episode started to make sure we were not counting in one cluster-episode, activity that corresponds to a different one.

Cluster-episodes without an end date in 2012/13 cannot be identified in 2013/14 using their episode-identifier and this is problematic when trying to use data from the two financial years as they would be duplicated. To overcome this issue we used only finished cluster-episodes from the 2012/13 data and all cluster-episodes from 2013/14.

Another issue with dates in the Episodes data file was the presence of overlapping cluster episodes, i.e. the end date of one cluster-episode was after the start of another cluster-episode for the same patient. This would mean double counting those days and all the activity performed during the overlapping period; to reduce the impact of this problem, we changed the end date of the cluster-episode starting earlier to match the start date of the following one.

Once the data was combined we excluded cluster-episodes for which the sum of the number of ward days exceeded the total number of days of the cluster, one possible cause for this is the way the length of the ward-episodes is calculated (length = end date minus start date plus one) which will count twice any days when a ward-episode ends and another one starts.

[10] Here we consider only valid (as indicated by a flag variable available in the data) full sets of replies, i.e. all HoNOS and SARN items have an answer. If we were to consider incomplete responses, the number of duplicates would increase, as only 77% of the valid MHCT events have a full set of responses.

In terms of evaluating whether the MHSDS is therefore able to be used as an information tool to count activity accurately which would be central to its use as a platform for the payment system, it would seem that the data is not yet of sufficient quality to support such a process.

6. Workshop for mental health commissioners

Alongside the empirical analysis, we wanted to obtain feedback on our results by running a one-day workshop for mental health commissioners. Twenty six commissioners attended from various CCGs across the country, as well as Councils and NHS England. Following presentations by the research team of our analytical work, there was a group discussion about the challenges for mental health commissioning arising from the presented analyses, where commissioners at the workshop are in terms of their developments and plans for funding models and contracting for services, and the landscape in terms of moving forward with new funding models. A number of key topics emerged at the workshop.

6.1 Episodic payment or capitated payment?

Recent guidance (Monitor and NHS England, 2016) has suggested that episodic payment models (which most commissioners seemed familiar with and to have done some level of thinking about) or a new approach of capitation (which most of the commissioners had done little to no thinking about) were acceptable alternatives to block contracts.

It was very clear that there is a wide degree of variation between commissioners in terms of their thinking about and confidence in using care clusters and developing their plans for payment models. Some were only now beginning to think about how to move from a block contract model. A few had variations of block contracts, episodic models and embryonic capitation models operating locally; only one group had clear plans for taking the first steps towards a capitation model in the next financial year and evolving it over the next few years.

Having the choice of options has caused much confusion and anxiety amongst some of the commissioners. They were now more uncertain as to what they should be doing and how to do it. Some had clearly felt, following the guidance from Monitor, that they ought to be developing capitation models to be keeping up with the latest thinking, but this felt like a huge step from their current practice with block contracts and/or seemed to be abandoning their work on episodic models before they had had chance to fully learn about this approach to commissioning. Clearer guidance on these matters from a national body would undoubtedly be helpful to many commissioners. This ought to help the commissioners to understand the journey that they are expected to be on, and provide navigational markers to help them on the way.

It was clear that the idea of capitation models of payment in mental health is very new and, hence, there were some quite diverse views as to what conceptually they might be and how to operationalize them. Overall, capitation seems to be seen as a means to integration of services, particularly between physical and mental health services, and, in essence, is a means of pricing a form of grand 'block contract' with sophisticated plans for additional metrics of quality and performance. Capitation can include episodic approaches, e.g. capitation for psychosis patients where they are classified into clusters, but the budget for that person is managed by one mental health Trust that pays for his/her acute and primary treatments.

The issue of integration of a person's care and its relationship to any payment model was worrying for the commissioners in a number of ways. Links across different health providers is one aspect, but then how to encourage better integration with social care and with housing were other challenges. The issue of linking payments to factors that may be beyond the control of mental health care providers, such as housing, was questioned.

Commissioners are keen to explore new ideas, such as accountable care organisations, outcomes-based commissioning, value-based commissioning, and place-based commissioning, but there seemed to be a need for more guidance from central organisations on these concepts and how they might be best operationalised.

Extrapolating from our understanding of the evidence base about forms of payments and issues of data quality, the idea of a capitation model in the forms discussed seems (at best) optimistic as a way forward. We recognise, though, that there is little in the way of robust evidence about the performance of capitation models in the context of mental health care in England chiefly because they are new, although there are some international examples (Monitor and NHS England, 2014).

6.2 Understanding data

Commissioners seemed to welcome the care cluster model. They appeared to be largely working towards using it as a framework to understand and discuss local patterns of care and variations, rather than as a categorical system to use as a threat against providers.

The degree of understanding of MHSDS and its potential links to developing payment models was, though, very varied amongst the commissioners. There may be an urgent need for more development opportunities for some commissioners to progress their thinking and understanding on these issues, if commissioning and payment models are to deliver improvements in mental health care systems. They would seem to want and need more help on how to understand existing data such as MHSDS. They struggle with the expertise and time to turn national data collections into helpful local insights.

It was recognised that there is much to do to improve data quality overall. Some commissioners have worked with their providers to develop incentives for them to improve their data returns. There was a mixed view as to whether this was now the best way forward, with some feeling it was still important to do so but others seeing that a good payment model would provide the incentives for providers to improve data quality. This, though, still leaves the question of whether the analysis can be good enough for the payment model to provide the right incentives.

Some commissioners discussed thinking about using local datasets to plan their commissioning and contracting work. This sometimes felt as if it was without full consideration of the MHSDS national dataset, and the pros and cons of using different datasets. Other instances were where commissioners were developing local data returns to supplement national data, which did not provide the data they needed for their commissioning plans. Some were very keen to use benchmarking within and across service providers to help focus quality improvement work and drive up value in their local systems. It was recognised that there is, though, some way to go to develop the local and national data and analyses to do this. Plus, once again we run in to the problem of a lack of expertise and capacity to achieve these ambitions for using data. There is a need to be clear of the risks of using local data returns in place of national data for developing payment models. There also seemed to be a lack of expertise in being able to use data other than business and process data, such as epidemiological data, which may be needed for models of capitation payments. The challenges involved in linking data to track people across systems of care to develop, manage and refine capitation models also seemed some way from being fully understood and answered in localities.

6.3 Other risks and challenges

There was a tension between wishing to move on quickly with developing payment models to improve care, whilst also wanting robust analysis to inform decisions, which takes time. Obviously, without good analyses to inform decisions there are risks of developing poor payment systems that introduce, for example, perverse incentives or unfair risks for some parties to the contracts.

Risk in the system was another point that was raised. Some participants felt that risk (most likely meaning financial risk), especially to providers, is necessary to provide incentives for change and to improve data collection. However, a tension was also discussed in the need to pay for existing services and even sometimes to make payments to providers to ‘shore them up’, even if services they provided were less than agreed in the quality goals. Data could play a crucial role in managing these issues as it could provide more clarity about the scale and nature of shortcomings and help to incentivise and monitor improvements in provision. It is unclear at this stage to what degree a payment system on its own can assist with this dilemma of supporting existing provision versus investing in new models.

Related to the point around investment in mental health, commissioners called for more clarity and stability with regard to levels of funding for mental health. The commissioners would welcome clear and consistent messages from the centre and practical tools to increase the level of investment in mental health.

There was mixed practice in terms of commissioning from the independent sector. Much seemed to be in the form of ‘spot purchases’. Some had framework or block contracts. It was recognised that contracting from the statutory, independent and third sectors needs to come closer together in terms of approaches being used locally.

Some commissioners face particular geographical challenges with commissioning and developing new models of services and contracts, for example where boundaries between commissioners and providers are not coterminous. Variation in practices arising from several commissioners working with one provider was one example. Some commissioners have worked together where they contract with the same provider.

Commissioners were keen that there were more opportunities such as this workshop to help them and to share experiences and learning. They especially welcomed the time at the workshop to discuss the issues raised with fellow commissioners. They were also keen on further input from academics on how to understand data and analyses.

Some additional, service-focused issues that the commissioners raised were:

- Need for a more coherent approach across the whole system of care, and less firefighting of problems for everyone (e.g. sending patients to A&E in crisis);
- Need to give thought to how to deal with ‘legal highs’.

Some additional points raised more specific to researchers and the analyses of the MHSDS data and care clusters:

- Linking of data sets for a more comprehensive analysis, such as HES and MHSDS;
- Analysis of activity and cost per person, especially considering different types of care/modes of delivery (e.g. GP, community mental health, acute mental health, acute physical health).

7. Discussion and conclusions

This report has explored the proposed episodic payment approach for mental health services whereby clinicians allocate patients into one of 21 clusters on the basis of similar levels of need using the MHCT. For this episodic payment system to effectively work, we have argued that there should not be too much variation in costs either *within* clusters, or *between* providers. The MHCT therefore plays a crucial role in that it needs to assign patients to clusters, such that they are homogenous in terms of 1) patient need, and 2) resource use.

We test whether the existing data collected on mental health activity amongst NHS providers within the MHSDS would support this new payment system. Specifically we examine whether there is homogeneity within clusters in terms of 1) costs, and 2) activity/resource use, and 3) whether the MHCT effectively clusters people with similar levels of need.

In this report we have been concerned with examining the relative variation across providers in terms of activity rates and costs. Our results suggest a large amount of variation between providers in terms of costs, activity rates and length of stay within clusters. Our results show that there is substantial variability across providers in the length of cluster episodes, and there is huge variability within clusters in terms of the proportion of inpatient days and the proportion of contact with healthcare professionals. Longer cluster episodes do not translate into proportionally more activity in terms of either inpatient days or contacts with healthcare professionals. With high levels of variation within clusters, accurate baseline activity rates cannot be determined for commissioning.

Variation in activity rates means that providers see different numbers of patients, have different treatment approaches, levels of productivity, and put different care pathways and packages of care in place for patients within each cluster. This could lead to differences in care quality and outcomes across providers, generating potential geographic inequalities for patients. We are unable to say from current data which of these pathways are associated with better outcomes.

At present, there is enormous variation within clusters in terms of costs. Variations in costs mean that patients with similar levels of need may be using different levels of resource, leading to a potential waste of scarce resources. The cost data cannot therefore be used at present to identify a reliable pricing system.

We also found that there is not a great degree of overlap between the MHCT and 'statistical classes', generated by sorting patients into classes which are maximally homogeneous within and maximally heterogeneous between. This suggests that there is variation within the 21 clusters created by the MHCT as a classification tool in terms of patient need. The key challenge for the classification system would be to refine the MHCT tool such that it fairly captures similarities and differences between patients. The categories of the classification system need to be casemix homogenous, that is patients within a given care cluster have similar needs profiles.

These key conclusions: 1) significant heterogeneity in patient need, 2) significant heterogeneity in costs, and 3) significant heterogeneity in terms of resource use within clusters, does not bode very well for an episodic payment approach which requires casemix and resource homogeneity within clusters. The reduction of variations in care, activity levels and costs is therefore pivotal to the establishment of a well-designed classification and payment system. In addition to provider variation, we also observed significant variation between commissioners in terms of their abilities and confidence in developing payment models. Furthermore, we encountered a number of data quality issues in the MHSDS. We found it is not yet able to be used as an information tool to accurately count activity which would be central to its use as a platform for the payment system. There are therefore significant challenges facing the system.

We would argue however, that instead of abandoning the episodic payment approach and clustering altogether, a much clearer steer is needed from policymakers to support providers and commissioners to move towards refining and developing episodic payment as a viable payment option. There is the possibility of re-designing or refining the clusters to improve homogeneity, using the data that has been gathered to date. It has taken more than a decade since implementation in the acute sector to refine and develop the PbR approach. There has been significant investment in information technology and collection of cost data over time which the sector has benefited from. Similar investment in information technology and improvement in data quality needs to be a priority in mental health services. The system also needs to implement change at a pace that does not risk destabilising local health economies. And more research is needed to support evidence-based policy-making and guidance.

We highlight three particular areas of concern and priority:

First, the current policy framework for the funding of mental health care is not providing a clear steer to commissioners and providers of care. Offering the service a choice of payment approach, is causing confusion and anxiety. The current policy proposals offer a lack of clarity, risk further fragmentation, greater local variation and an overall lack of financial control. Providers and commissioners should not be offered a menu of options to payment approaches. This generates further variation at local levels. Once providers and commissioners embark on a particular payment approach, it will be very difficult to change course and re-establish a common set of incentives that can reduce variations. The reality is, that commissioners feel they need to implement the latest 'trend' in funding approach, so as not to be left behind. There needs to be a much stronger policy directive on a single best mental health funding approach.

We would argue that the episodic payment approach has several significant advantages over the capitated payment approach and has stronger incentives than the capitated payment approach to increase activity rates and control costs. It may also be simpler to implement, and given capacity constraints within commissioners, may be more pragmatic. The choice of payment approach also has implications for risk attribution between providers and commissioners, with the capitated approach using an Accountable Care Organisation model, shifting risk onto providers. Episodic payment is a more transparent funding approach than the capitated payment approach. Therefore the episodic payment approach has the potential to establish greater parity of esteem between mental and physical health.

The arguments of the 'institutional bias' towards acute providers in the funding system is well rehearsed, with larger cuts in tariff prices seen in 2014/15 for mental health services (-1.8%) compared to acute services (-1.5%) (Monitor & NHS England, 2013). Given the current and future projected financial position of providers, with mental health providers delivering overall surpluses year on year, compared to huge increases in deficits year on year for acute providers (Dunn et al, 2016), mental health will continue to be at risk of having their resources diverted away towards acute providers. The disparity in payment systems between mental and physical health care is a major risk factor for mental health services. As long as there are parallel funding systems operating, where in one, better quality activity data and a more transparent system make the return on investment of limited budgets more obvious, it will always win out. Thus as long as mental health operates a block contract system, or an opaque 'capitation' contract or variants thereof, commissioners will not have a clear sense of the value for money they are getting from investment in these services and mental health care will lose out.

We would argue that the episodic approach is better developed and has a limited but robust, evidence-base supporting it. The clustering approach is already relatively well established in most providers and could be refined and improved upon. Scrapping it all and starting from scratch risks putting mental health services back almost a decade in terms of developing a more transparent and fair funding system.

With further financial pressures looming, mental health services need a stable funding environment and sustainable commissioning, based on an evidence-informed payment system that generates the right incentives and reduces variations in care.

Second, a fundamental aspect which seems to be missing from the proposed funding models and from discussions with commissioners in any meaningful way, is how any payment system will be linked to quality and outcome metrics. This should be based on sound evidence and not left to local health economies to try to develop. We would argue that it is imperative that the classification and payment system puts incentive structures in place that are able to drive improvements in efficiency of care delivery, without compromising service quality. Again, a stronger policy steer and more evidence is urgently needed and would be found useful.

Third, our research shows that it will be difficult to create a classification and payment system with the currently available data. For the development of any payment system, high quality activity and cost data would be a key requisite. Data quality is a significant challenge with any payment system, but it is at least underway for the episodic payment approach using clustering, collected routinely and we find evidence of some improvements in data quality over time in its collection. We would urge commissioners to routinely use the MHSDS in their contracting and monitoring processes. This is the only way a single consistent use of data can be achieved across several commissioners with any given provider. This prevents providers wasting precious resources filling in different datasets for different commissioners and will incentivise rapid improvement in the data quality of the MHSDS which will facilitate national benchmarking and performance improvement. Commissioners should be offered support to use and understand the MHSDS for local decision-making. Improvements to Reference Cost data are also essential and introduction of Patient Level Information Costing Systems (PLICS) at provider level can support the process of generating this.

In summary: we would call for a clear policy steer to implement and improve the episodic payment approach, underpinned by an imperative to collect high quality activity and cost data that can underpin the classification and payment system. Effort should be put into refining the approach and generating evidence to link appropriate quality and outcome metrics.

8. References

Bebbington, P., Brugha, T., Hill, T., Marsden, L. & Window, S. (1999) Validation of the health of the nation outcome scales. *The British Journal of Psychiatry*, 174(5), 389-394.

Böhnke, J. R. & Croudace, T. J. (2015) Factors of psychological distress: clinical value, measurement substance, and methodological artefacts. *Social Psychiatry and Psychiatric Epidemiology*, 50(4), 515-524.

Castelli, L. & Ukovich, W. (2003) *An Airline-Based Multilevel Analysis of Airfare Elasticity for Passenger Demand*. NASA; Washington, DC, United States: Nasa Technical Reports Server (NTRS), (NTRS), N. T. R. S.

Chalkley, M. & Malcomson, J. M. (2000) Government purchasing of health services, *Handbook of Health Economics* Elsevier, 847-890.

Charlesworth, A., Davies, A. & Dixon, J. (2012) *Reforming Payment for Healthcare in Europe to Achieve Better Value*

Dawson, D. & Street, A. (1998) *Reference Costs and the Pursuit of Efficiency in the 'New' NHS* Centre for Health Economics Discussion Paper Series, York, U. o.

Department of Health (2013) *Payment by Results in the NHS: tariff for 2012 to 2013; part of: Payment by Results 2012-13, Payment by Results (PbR) in the NHS and NHS efficiency* England: Gov.UK. Available online: https://www.gov.uk/government/publications/confirmation-of-payment-by-results-pbr-arrangements-for-2012-13 [Accessed added GR 280416].

Department of Health (2014a) *Mental Health Clustering Booklet (V3.0) (2013-14)*. London: Gateway ref. 18540, Health, D. o.

Department of Health (2014b) *NHS Reference Costs 2013 to 2014*, 2014b. Available online: https://www.gov.uk/government/publications/nhs-reference-costs-2013-to-2014 [Accessed.

Dunn, P., KcKenna, H. & Murray, R. (2016) *Deficits in the NHS 2016*. London: Briefing, K Fund,london.

Finch, W. H. & Bronk, K. C. (2011) Conducting confirmatory latent class analysis using M plus. *Structural Equation Modeling*, 18(1), 132-151.

Health & Social Care Information Centre, a. (2016a) *Mental Health Services Data Set Specifications and Guidance*, 2016a. Available online: http://www.hscic.gov.uk/mhsds/spec [Accessed.

Health & Social Care Information Centre, b. (2016b) *Mental Health Minimum Dataset (MHMDS) v4.0 – ISB 0011, Amd 22/2011*, 2016b. Available online: http://webarchive.nationalarchives.gov.uk/20120802111034/http:/www.ic.nhs.uk/services/mental-health/using-the-service/datasets-databases-and-data-collections/mental-health-minimum-dataset-mhmds/specifications-and-guidance [Accessed.

Jacobs, R. (2014) Payment by Results for mental health services: Economic considerations of casemix funding. *Advances in Psychiatric Treatment*, 20, 155-64.

Kempf, W. (2012) *A Pragmatic Approach to Rasch-Modeling: The Loss of Information Index*, *74*. Berlin: Irena Regener.

Khan, S., Nowak, S. & NHS England (2014) *Background to the 2015/16 proposals for the mental health payment system*.London: England, N.

Laudicella, M., Donni, P. L. & Smith, P. C. (2013) Hospital readmission rates: signal of failure or success? *Journal of Health Economics*, 32(5), 909-921.

Lazarsfeld, P. F. & Henry, N. W. (1968) *Latent Structure Analysis*. Boston, MA: Houghton Mifflin.

Leach, C., Lucock, M., Barkham, M., Noble, R., Clarke, L. & Iveson, S. (2005) Assessing risk and emotional disturbance using the CORE-OM and HoNOS outcome measures at the interface between primary and secondary mental healthcare. *The Psychiatrist*, 29(11), 419-422.

Mason, A., Goddard, M., Myers, L. & Verzulli, R. (2011) Navigating uncharted waters? How international experience can inform the funding of mental health care in England. *Journal of Mental Health*, 20(3), 234-248.

McCutcheon, A. L. (1987) *Latent Class Analysis*. Newbury Park, CA: Sage.

Monitor & NHS England (2013) *2014/15 National Tariff Payment System*.London.

Monitor and NHS England (2014) *Capitation: International Examples.* 26 November 2014. London: Monitor. Available online: https://www.gov.uk/government/publications/capitation-international-examples [Accessed

Monitor and NHS England (2016) *Guidance on mental health currencies and payment. A supporting document for the 2016/17 National Tariff Payment System: A consultation notice.*, 2016. Available online: https://www.gov.uk/government/publications/mental-health-currencies-and-payment-guidance [Accessed.

Muthén, L. K. & Muthén, B. O. (1998) *Mplus user's guide* 7th edition edition. Los Angeles, CA: Muhén & Muthén.

NHS Digital (2013) *Mental Health Bulletin, Annual Report From MHMDS Returns, England - 2012-13: National Reference Tables*

Norton, E. C., Van Houtven, C. H., Lindrooth, R. C., Normand, S. T. & Dickey, B. (2002) Does prospective payment reduce inpatient length of stay? *Health Economics*, 11(5), 377-87.

Papanicolas, I. & McGuire, A. (2015) Do financial incentives trump clinical guidance? Hip Replacement in England and Scotland. *Journal of Health Economics*, 44, 25-36.

Rabe-Hesketh, S. & Skrondal, A. (2008) *Multilevel and Longitudinal Modeling Using Stata, Second Edition*, Second Edition edition. Stata Press.

Salvi, G., Leese, M. & Slade, M. (2005) Routine use of mental health outcome assessments: choosing the measure. *The British Journal of Psychiatry*, 186(2), 146-152.

Self, R., Painter, J. & Davis, R. (2008) *A report on the development of a Mental Health Currency Model (including suggestions on the development and testing of PbR tariffs)* London.

Self, R., Rigby, A., Leggett., C, Paxton., R. (2008) Clinical Decision Support Tool: A rational needs-based approach to making clinical decisions. *Journal of Mental Health*, 17(1), 33-48.

Sood, N., Huckfeldt, P. J., Grabowski, D. C., Newhouse, J. P. & Escarce, J. J. (2013) The effect of prospective payment on admission and treatment policy: Evidence from inpatient rehabilitation facilities. *Journal of Health Economics*, 32(5), 965-979.

Speak, B., Hay, P., Muncer, S., Craig, B., Green, C., Rigby, A., Painter, J. & James, M. (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: Final Report. *Care Pathways and Packages Project Consortium*.

Speak, B. & Muncer, S. (2015) The structure and reliability of the Health of the Nation Outcome Scales. *Australasian Psychiatry*, 23(1), 66-68.

StataCorp (2013) *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.

The Mental Health Taskforce (2016) *The Five Year Forward View for Mental Health: A report from the independent Mental Health Taskforce to the NHS in England.*

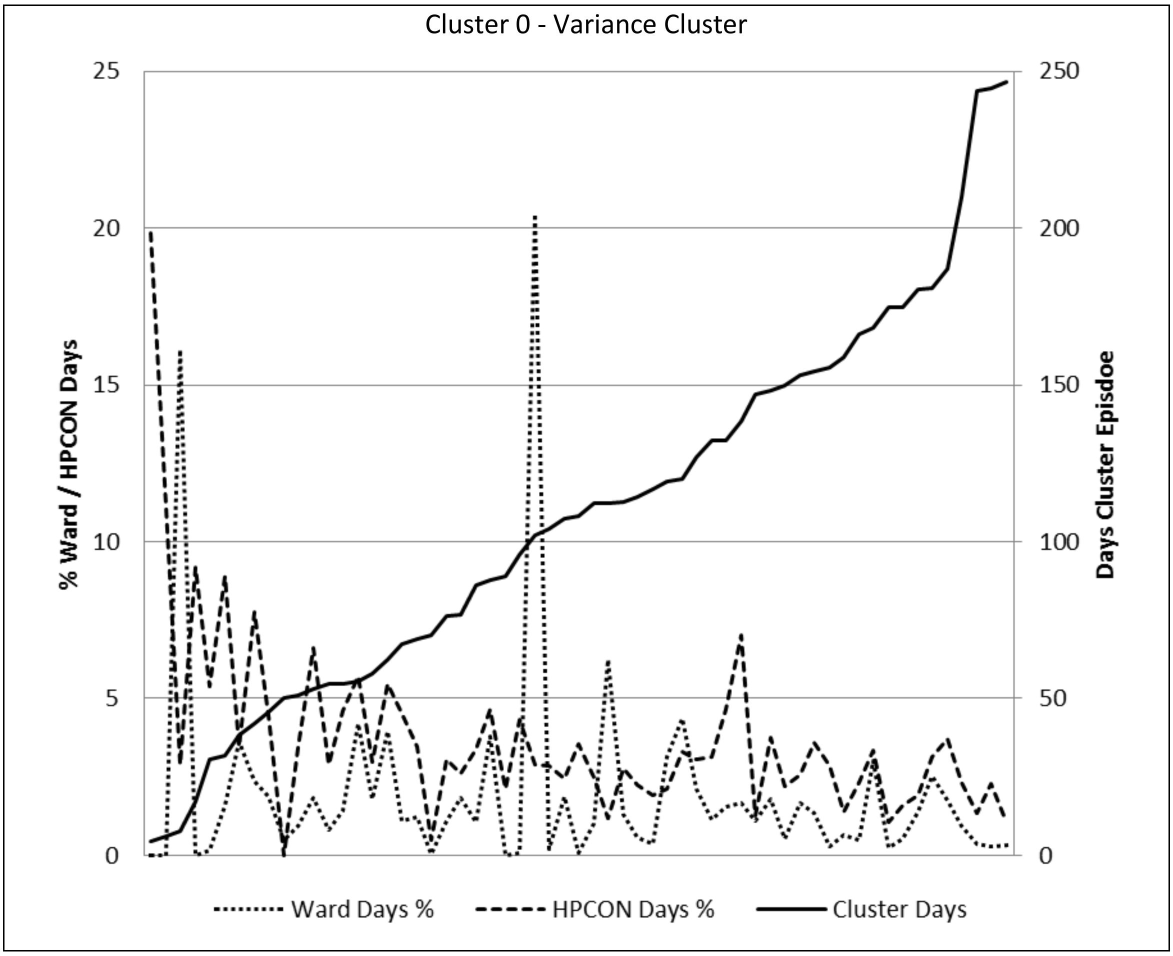
Wing, J., Curtis, R. & Beevor, A. (1994) 'Health of the Nation': measuring mental health outcomes *Psychiatric Bulletin*, 18(11), 690-91.

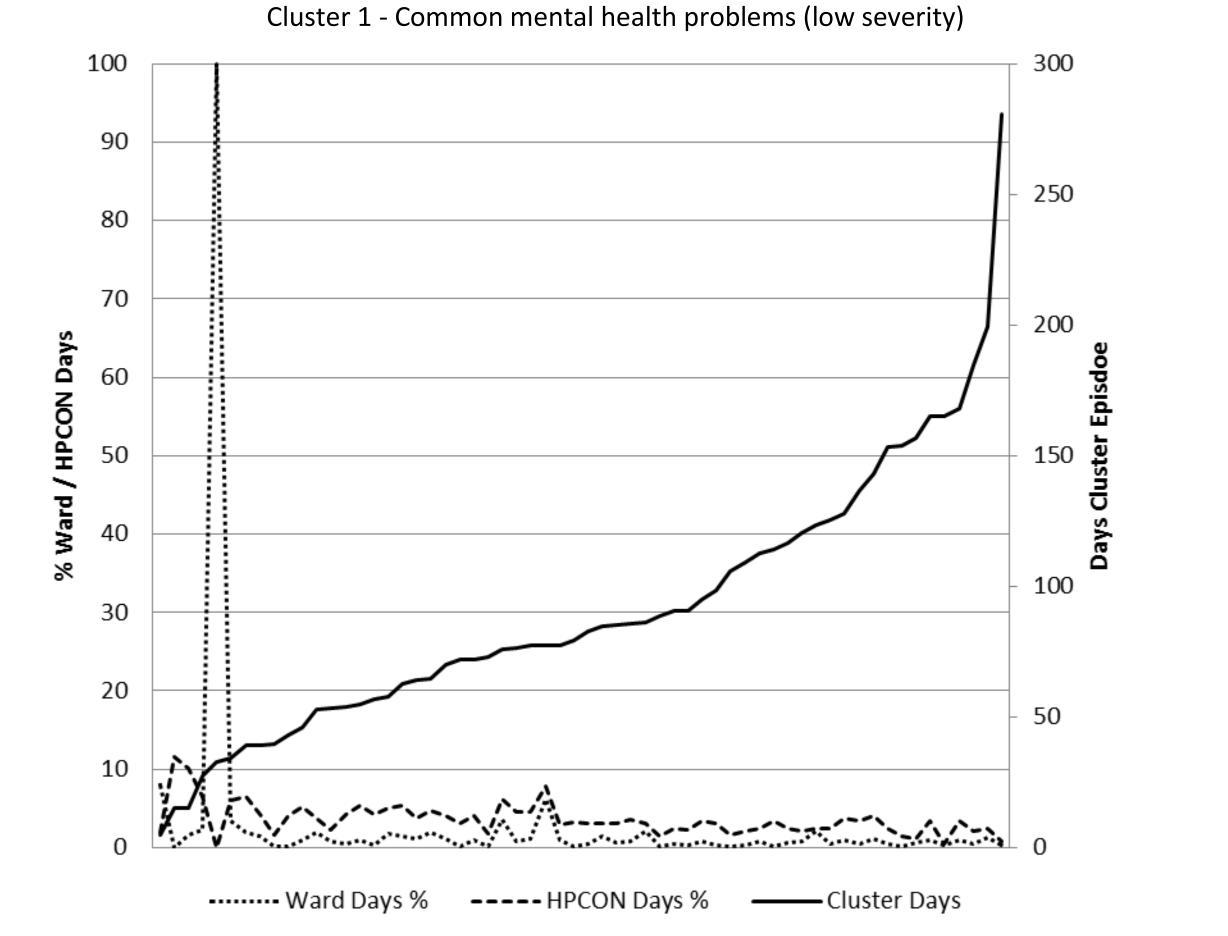
Wing, J. K., Beevor, A. S., Curtis, R. H., Park, S. B., Hadden, S. & Burns, A. (1998) Health of the Nation Outcome Scales (HoNOS). Research and development. *The British Journal of Psychiatry*, 172(1), 11-18.

Appendices

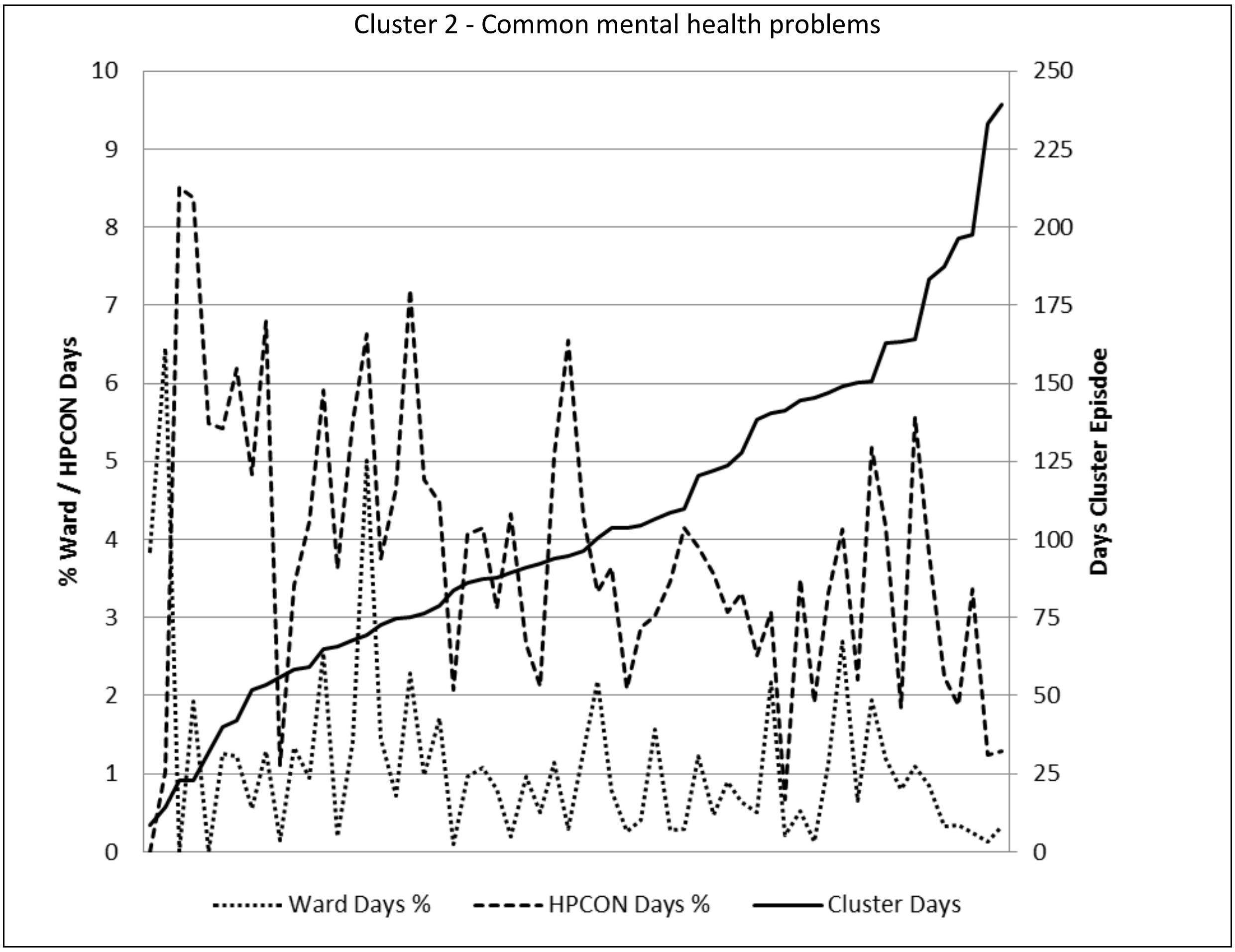
A. Activity Plots

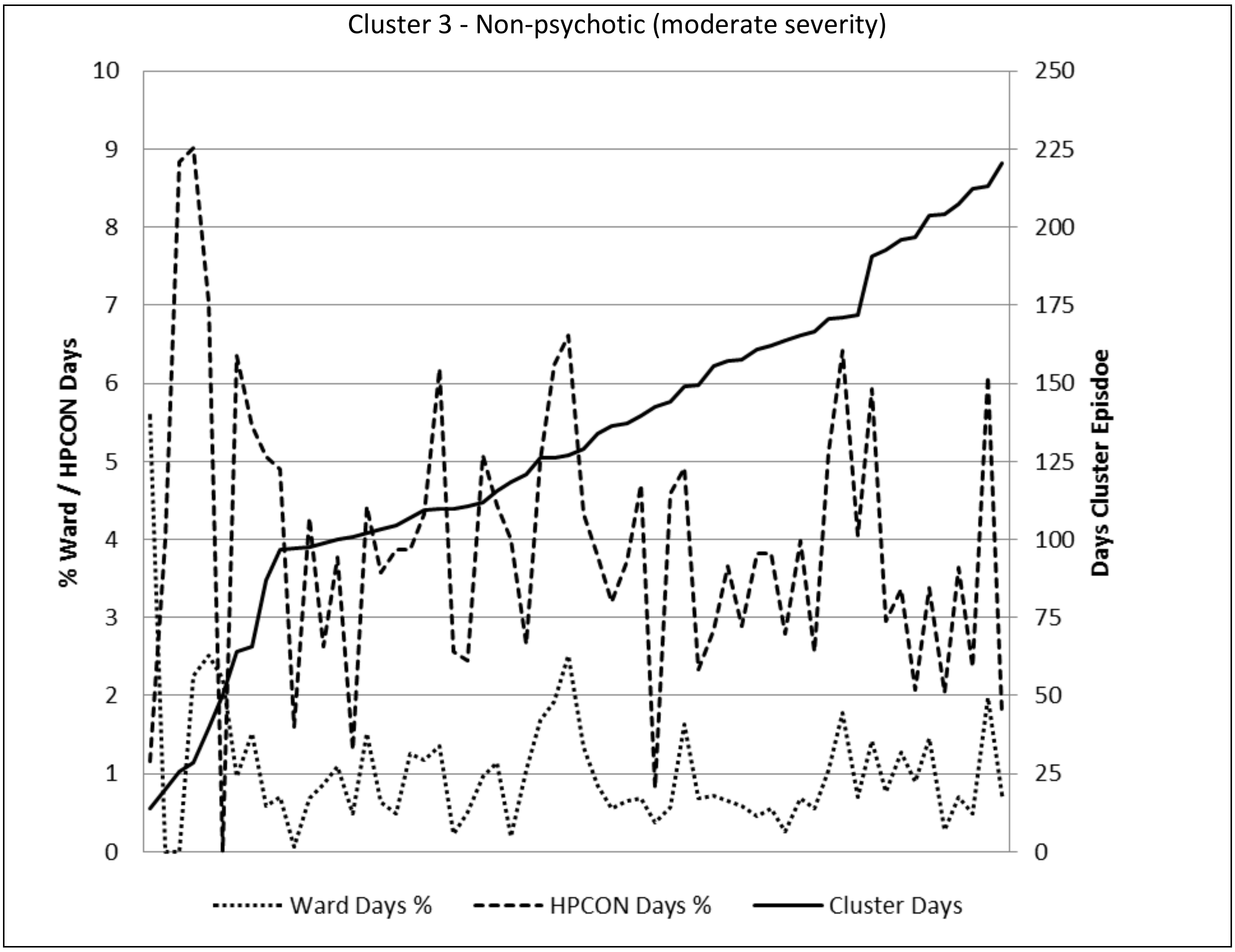
The following plots show for each cluster the average total length of cluster episodes (continuous line), the percentage of that time spent as an inpatient (dotted line) and the percentage of those days in which the patient had contact with a health care professional (dashed line).

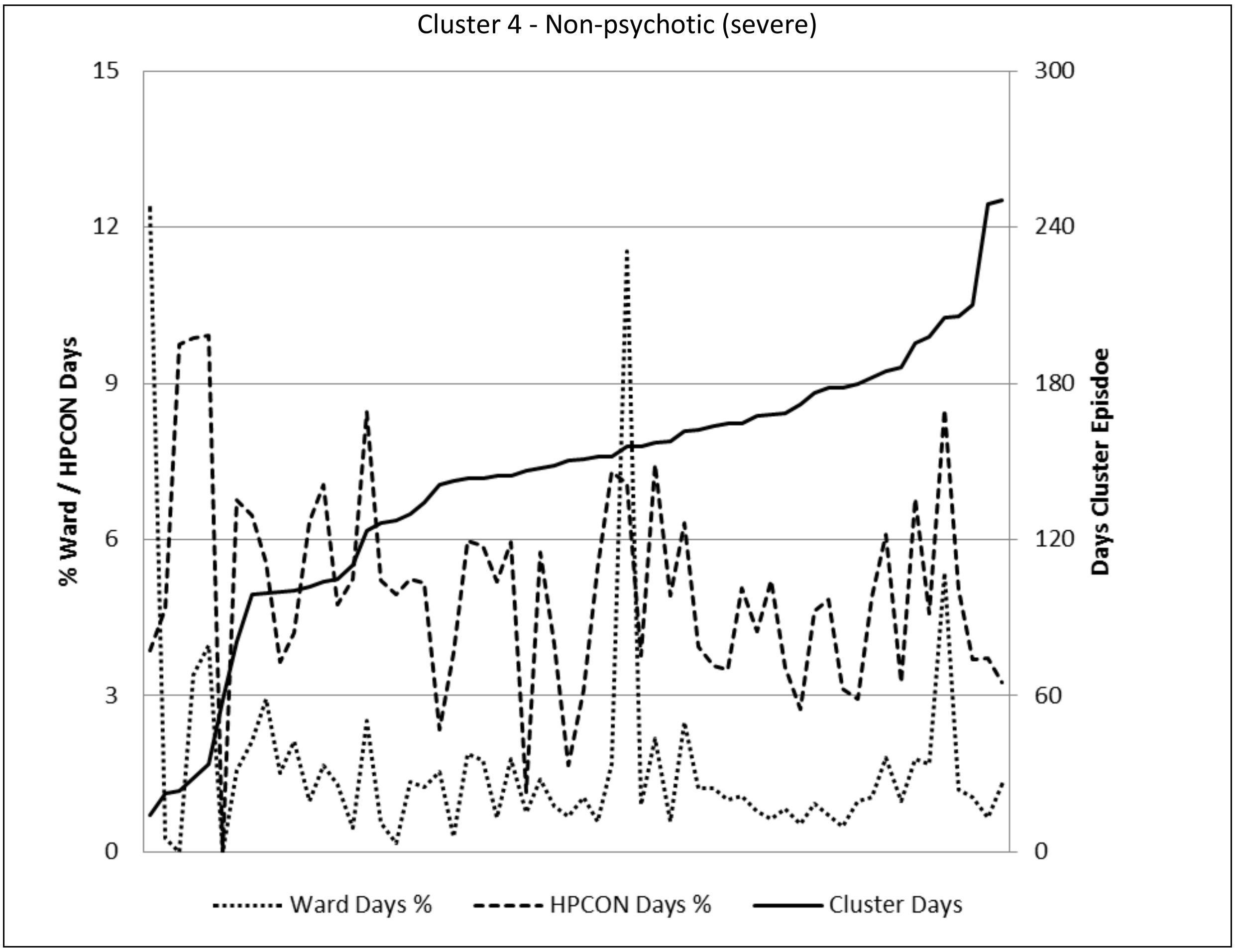


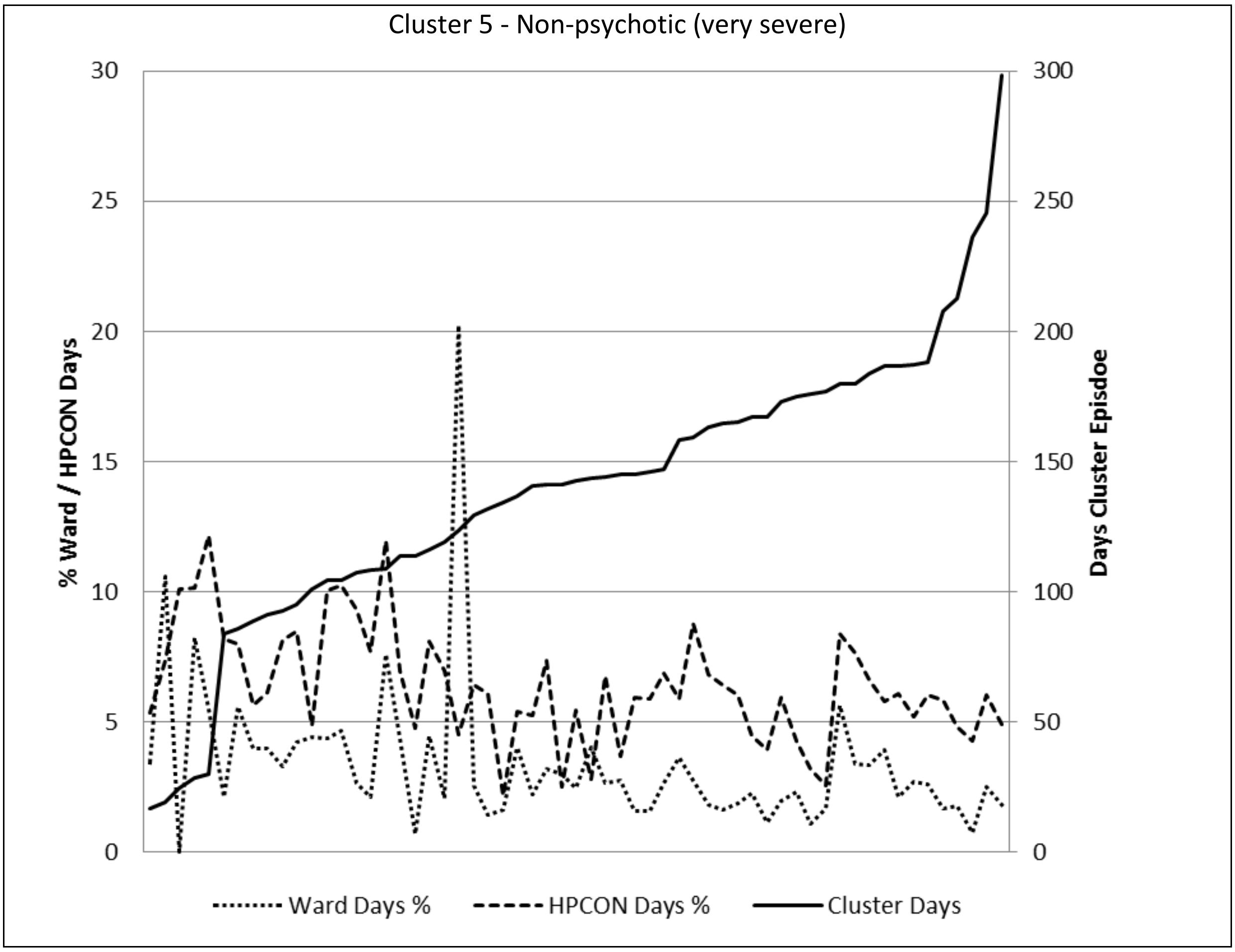


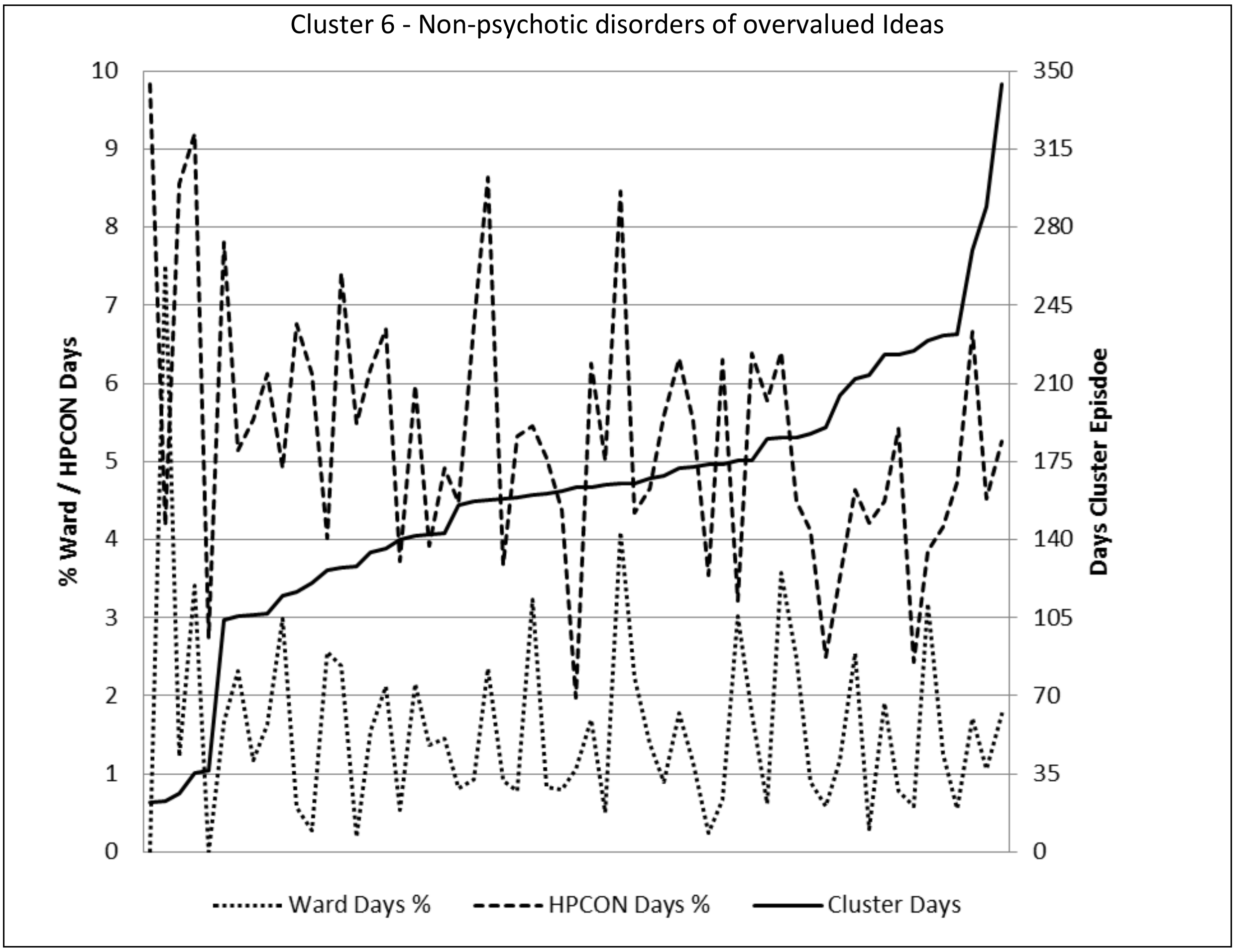
There is one provider that records all days in the cluster as admitted days.

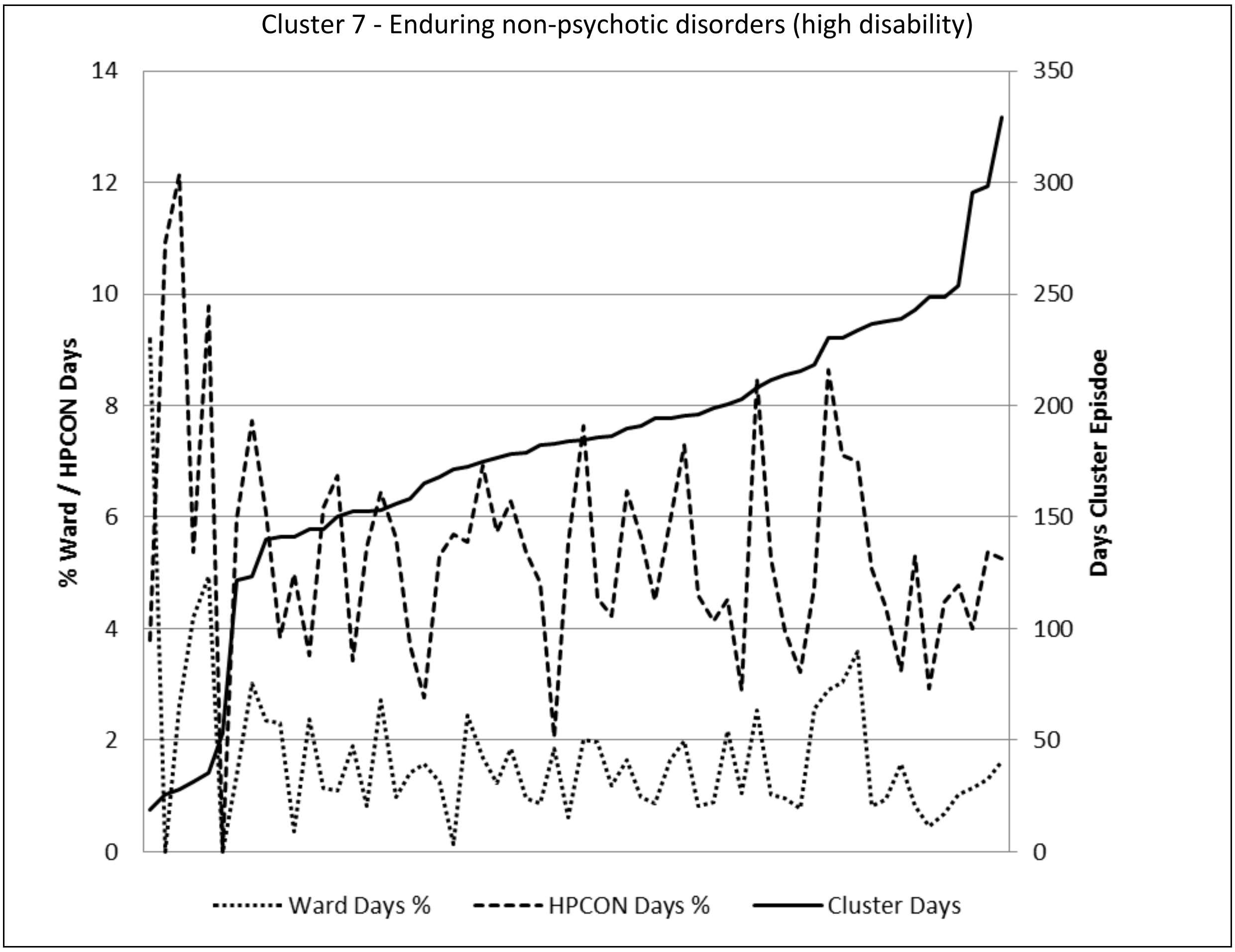

Cluster 2 - Common mental health problems
% Ward / HPCON Days
Days Cluster Episdoe
10
9
8
7
6
5
4
3
2
1
0
250
225
200
175
150
125
100
75
50
25
0
Ward Days %
HPCON Days %
Cluster Days

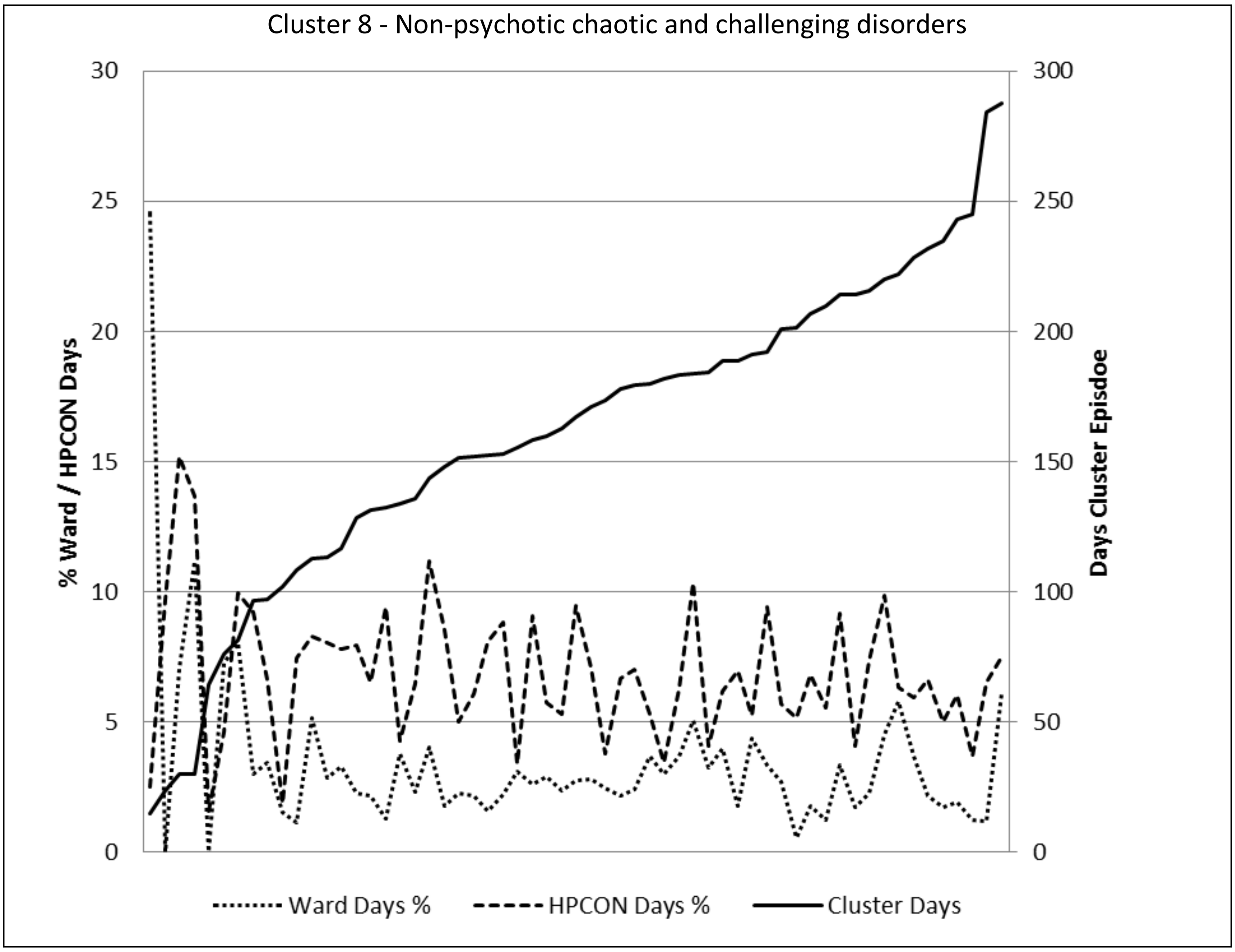

Cluster 3 - Non-psychotic (moderate severity)
% Ward / HPCON Days
Days Cluster Episdoe
0
1
2
3
4
5
6
7
8
9
10
0
25
50
75
100
125
150
175
200
225
250
Ward Days %
HPCON Days %
Cluster Days

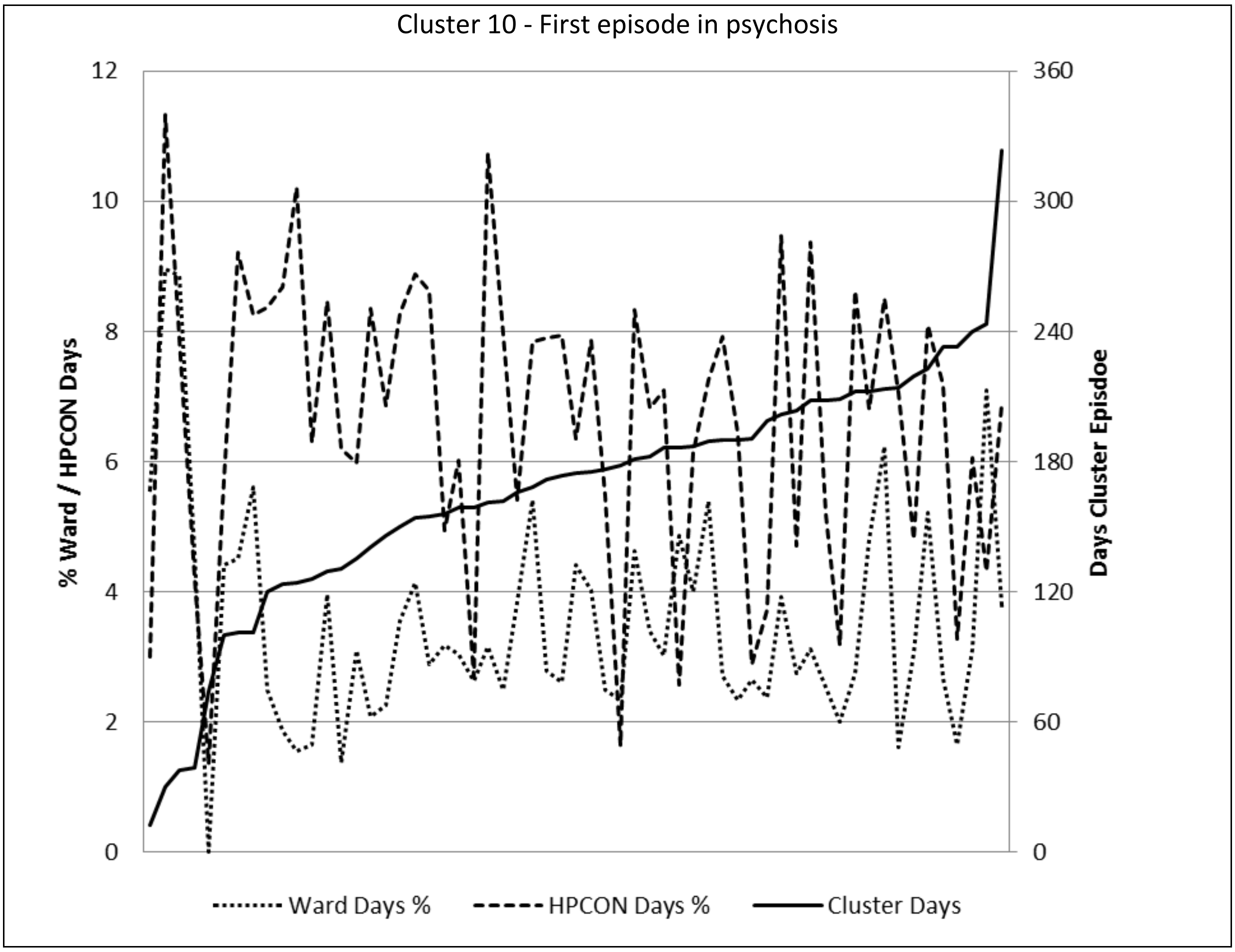

Cluster 4 - Non-psychotic (severe)
15
12
9
6
3
0
300
240
180
120
60
0
% Ward / HPCON Days
Days Cluster Episdoe
Ward Days %
HPCON Days %
Cluster Days


Cluster 5 - Non-psychotic (very severe)
% Ward / HPCON Days
Days Cluster Episdoe
0
5
10
15
20
25
30
0
50
100
150
200
250
300
Ward Days %
HPCON Days %
Cluster Days

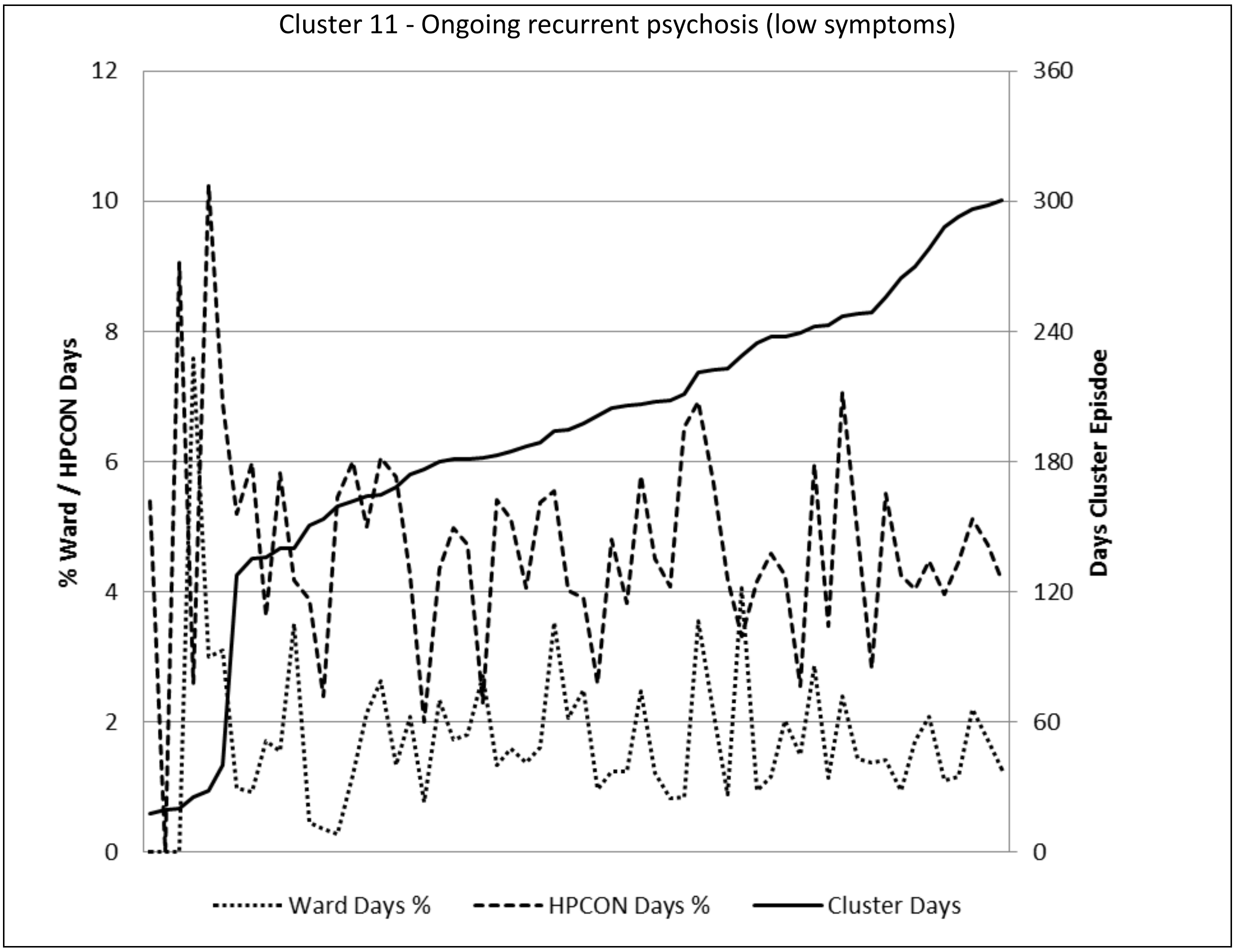

Cluster 6 - Non-psychotic disorders of overvalued Ideas
% Ward / HPCON Days
Days Cluster Episdoe
0
1
2
3
4
5
6
7
8
9
10
0
35
70
105
140
175
210
245
280
315
350
Ward Days %
HPCON Days %
Cluster Days

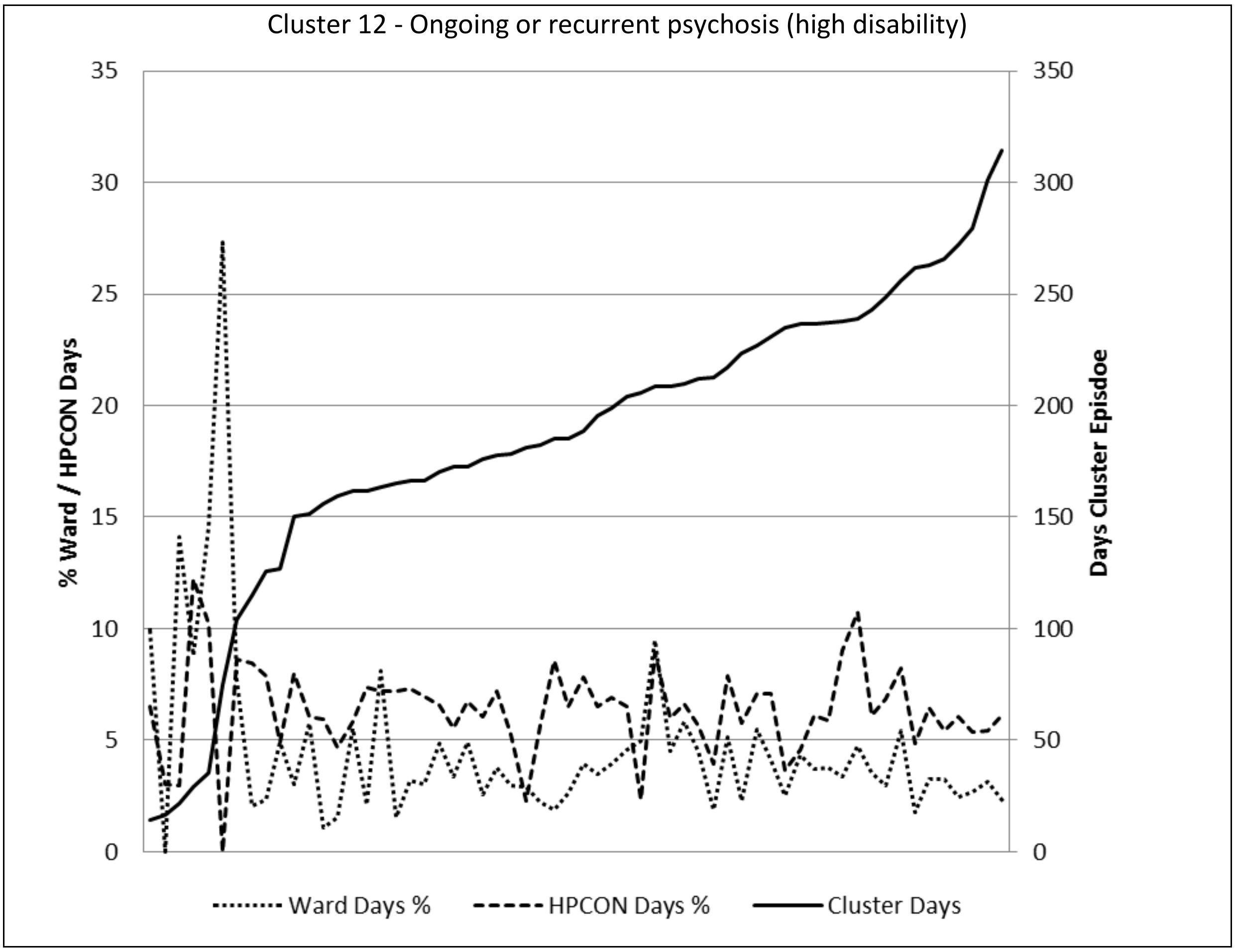

Cluster 7 - Enduring non-psychotic disorders (high disability)
% Ward / HPCON Days
Days Cluster Episdoe
0
2
4
6
8
10
12
14
0
50
100
150
200
250
300
350
Ward Days %
HPCON Days %
Cluster Days

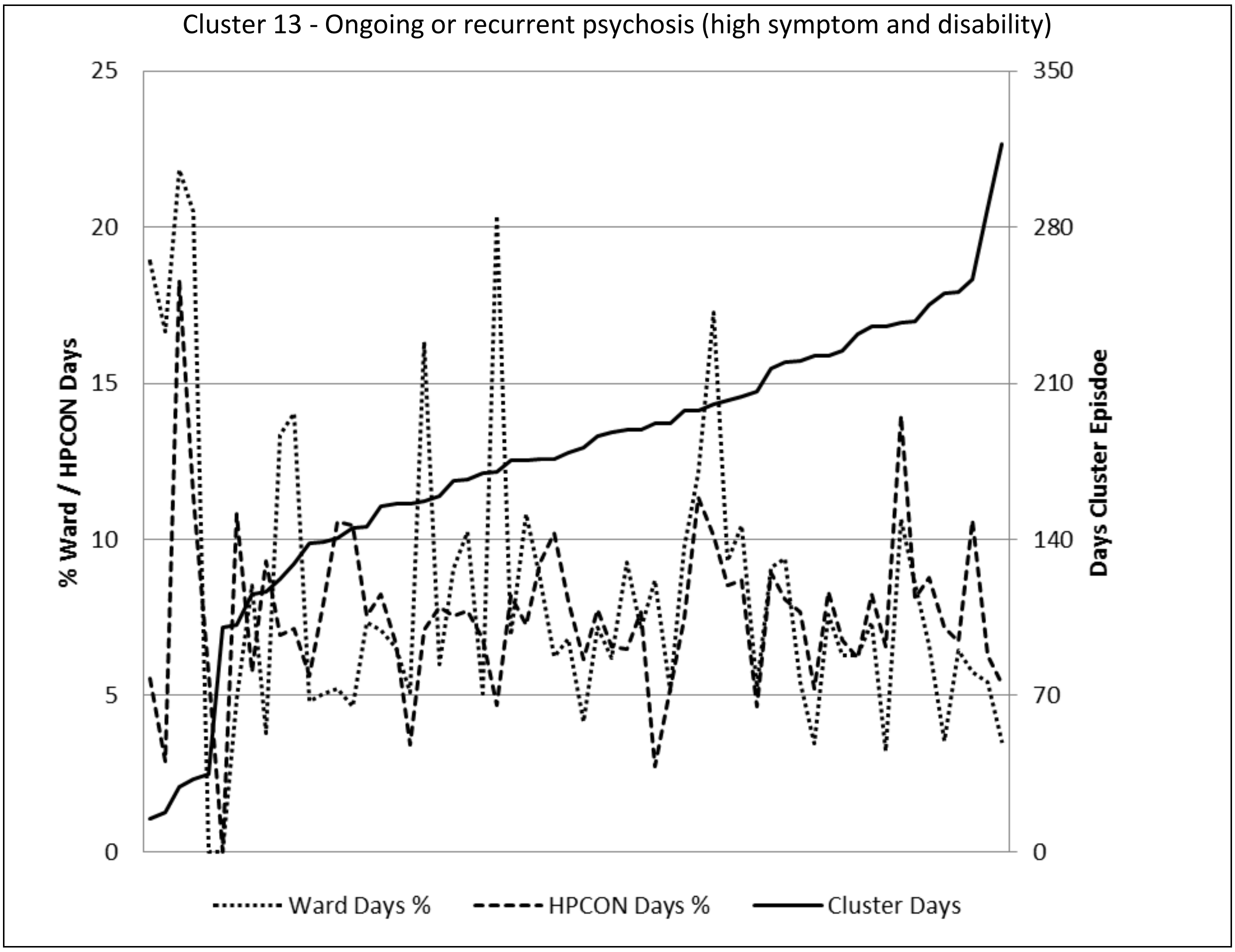

Cluster 8 - Non-psychotic chaotic and challenging disorders
% Ward / HPCON Days
Days Cluster Episdoe
0
5
10
15
20
25
30
0
50
100
150
200
250
300
Ward Days %
HPCON Days %
Cluster Days

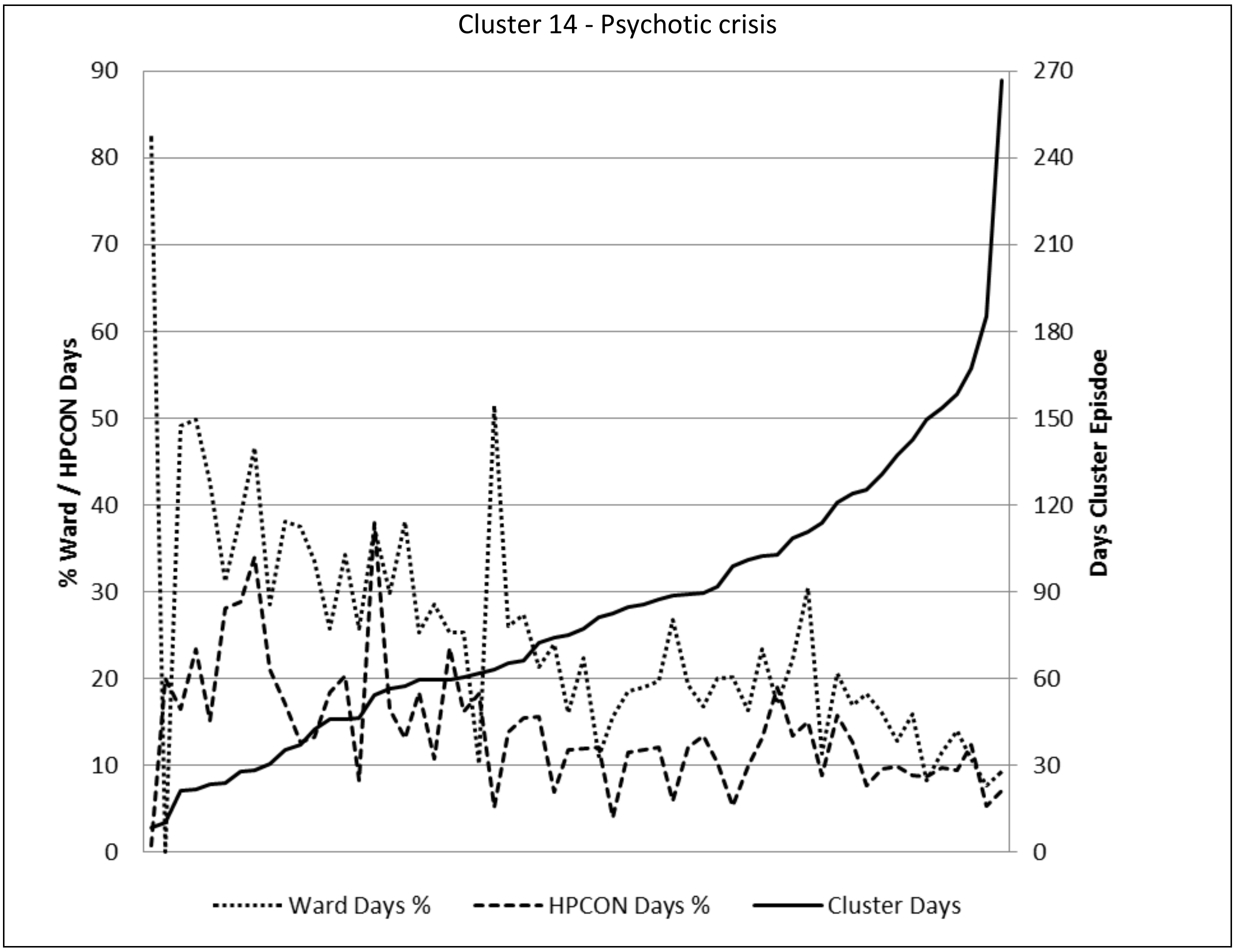


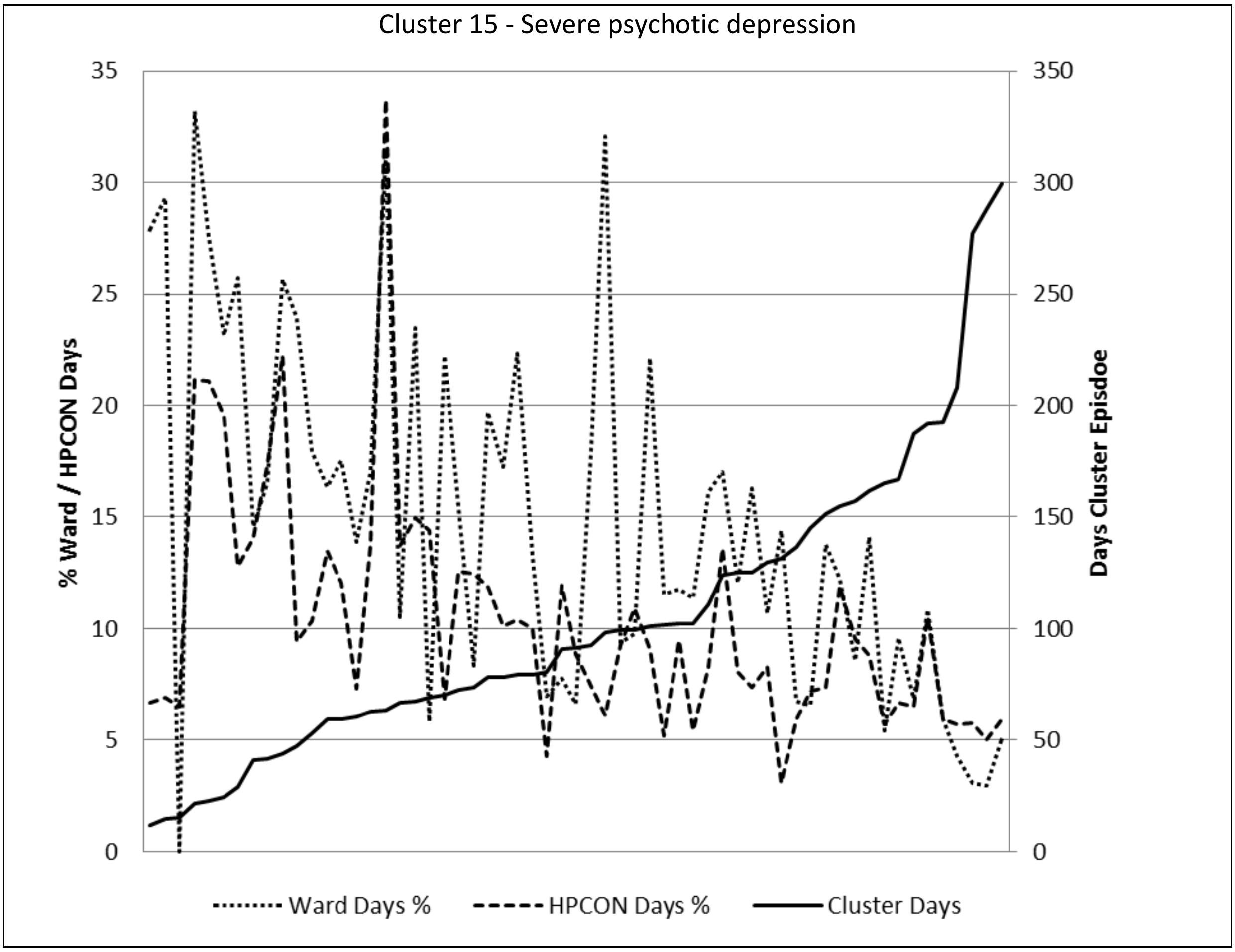
Cluster 10 - First episode in psychosis
% Ward / HPCON Days
Days Cluster Episdoe
0
2
4
6
8
10
12
0
60
120
180
240
300
360
Ward Days %
HPCON Days %
Cluster Days

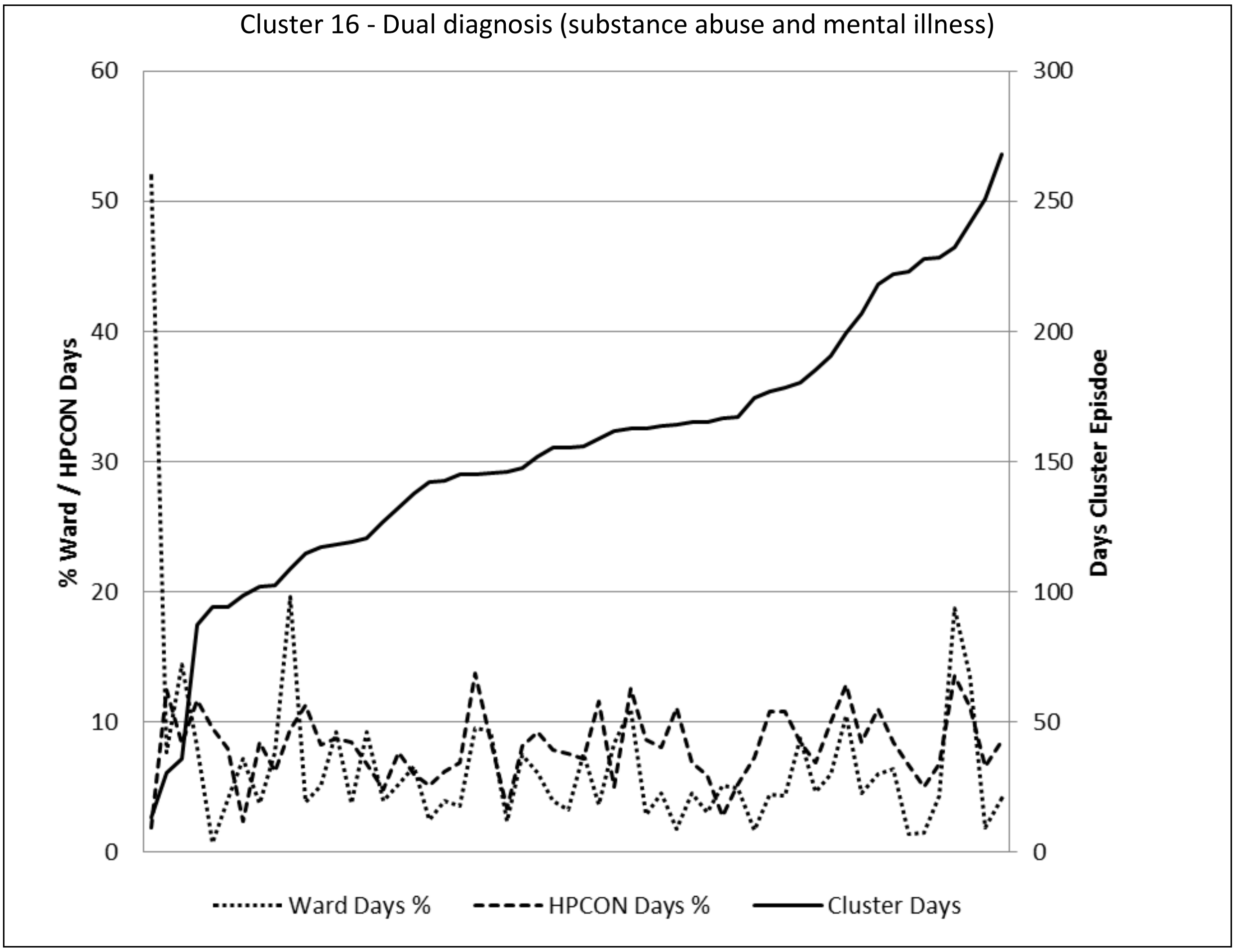

Cluster 11 - Ongoing recurrent psychosis (low symptoms)
% Ward / HPCON Days
Days Cluster Episdoe
0
2
4
6
8
10
12
0
60
120
180
240
300
360
Ward Days %
HPCON Days %
Cluster Days

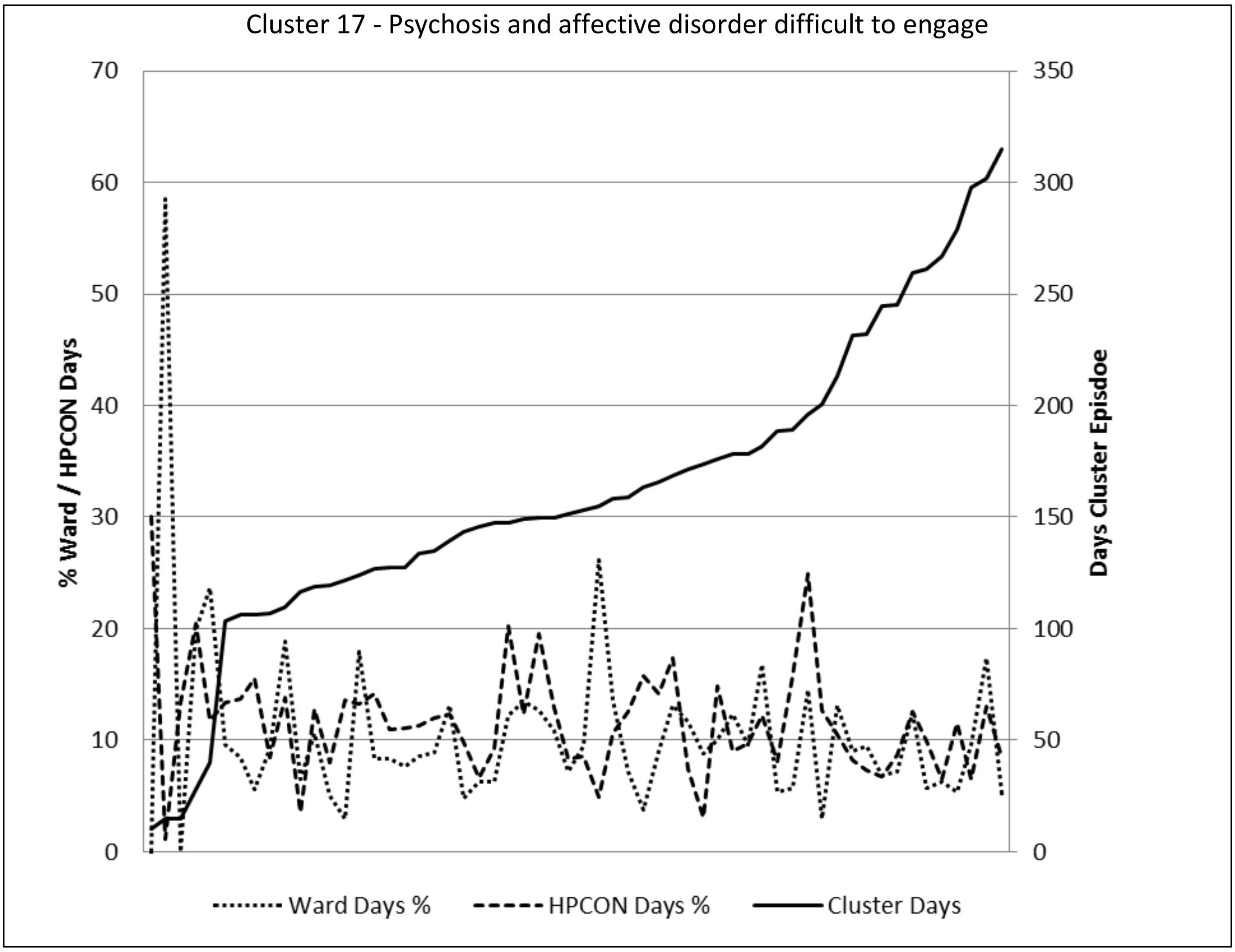

Cluster 12 - Ongoing or recurrent psychosis (high disability)
% Ward / HPCON Days
Days Cluster Episdoe
35
30
25
20
15
10
5
0
350
300
250
200
150
100
50
0
Ward Days %
HPCON Days %
Cluster Days

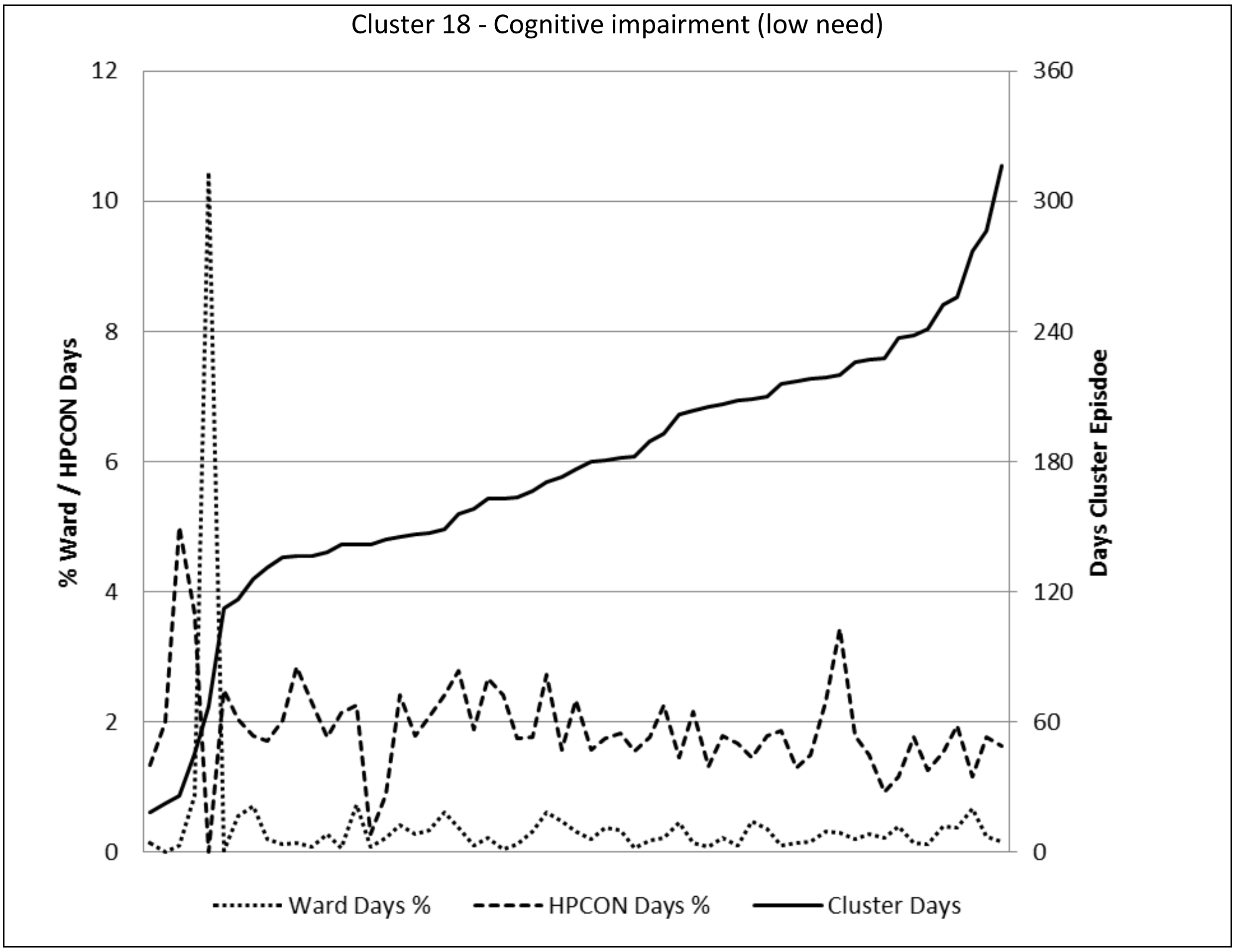

Cluster 13 - Ongoing or recurrent psychosis (high symptom and disability)
% Ward / HPCON Days
Days Cluster Episdoe
25
20
15
10
5
0
350
280
210
140
70
0
Ward Days %
HPCON Days %
Cluster Days

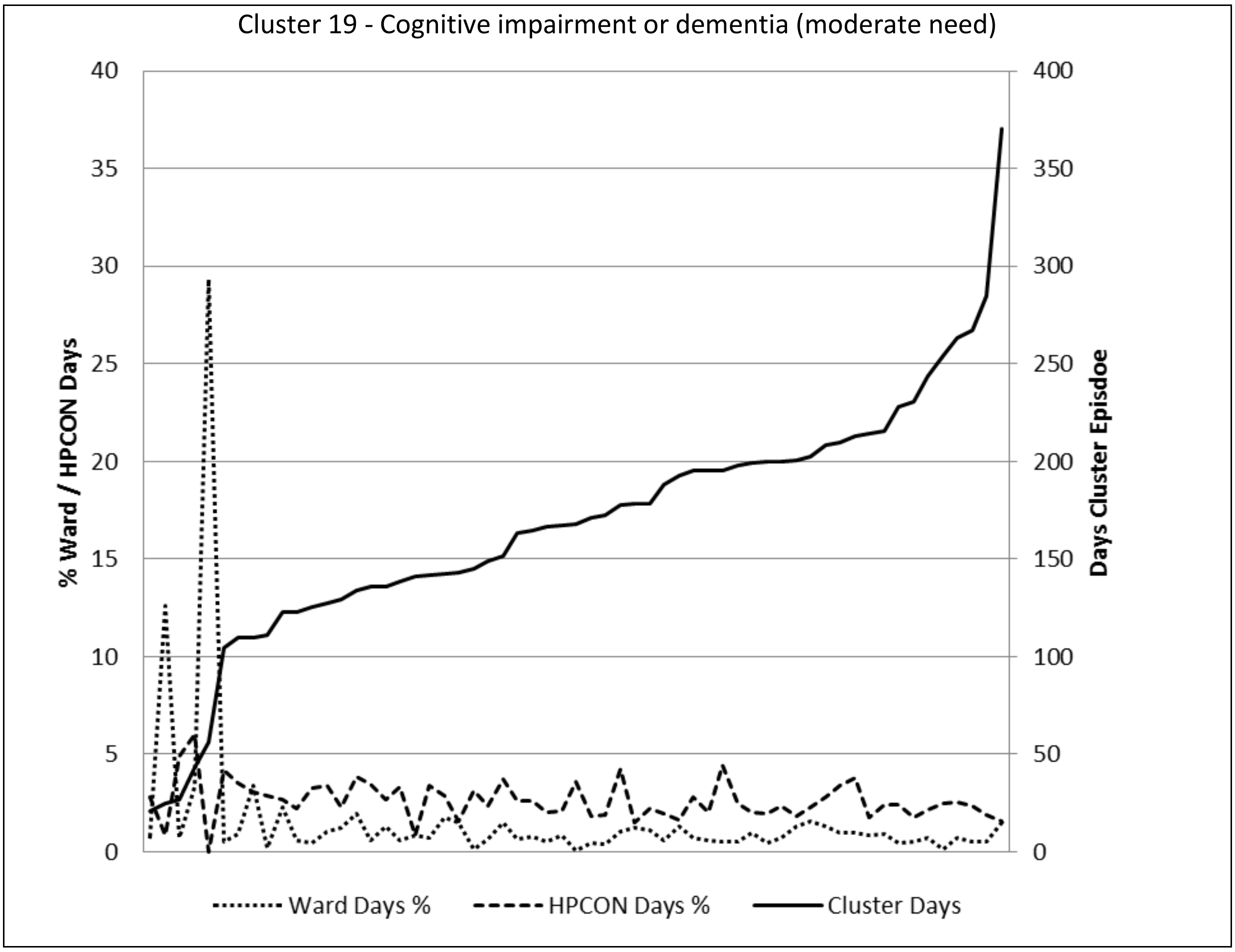

Cluster 14 - Psychotic crisis
% Ward / HPCON Days
Days Cluster Episdoe
0
10
20
30
40
50
60
70
80
90
0
30
60
90
120
150
180
210
240
270
Ward Days %
HPCON Days %
Cluster Days

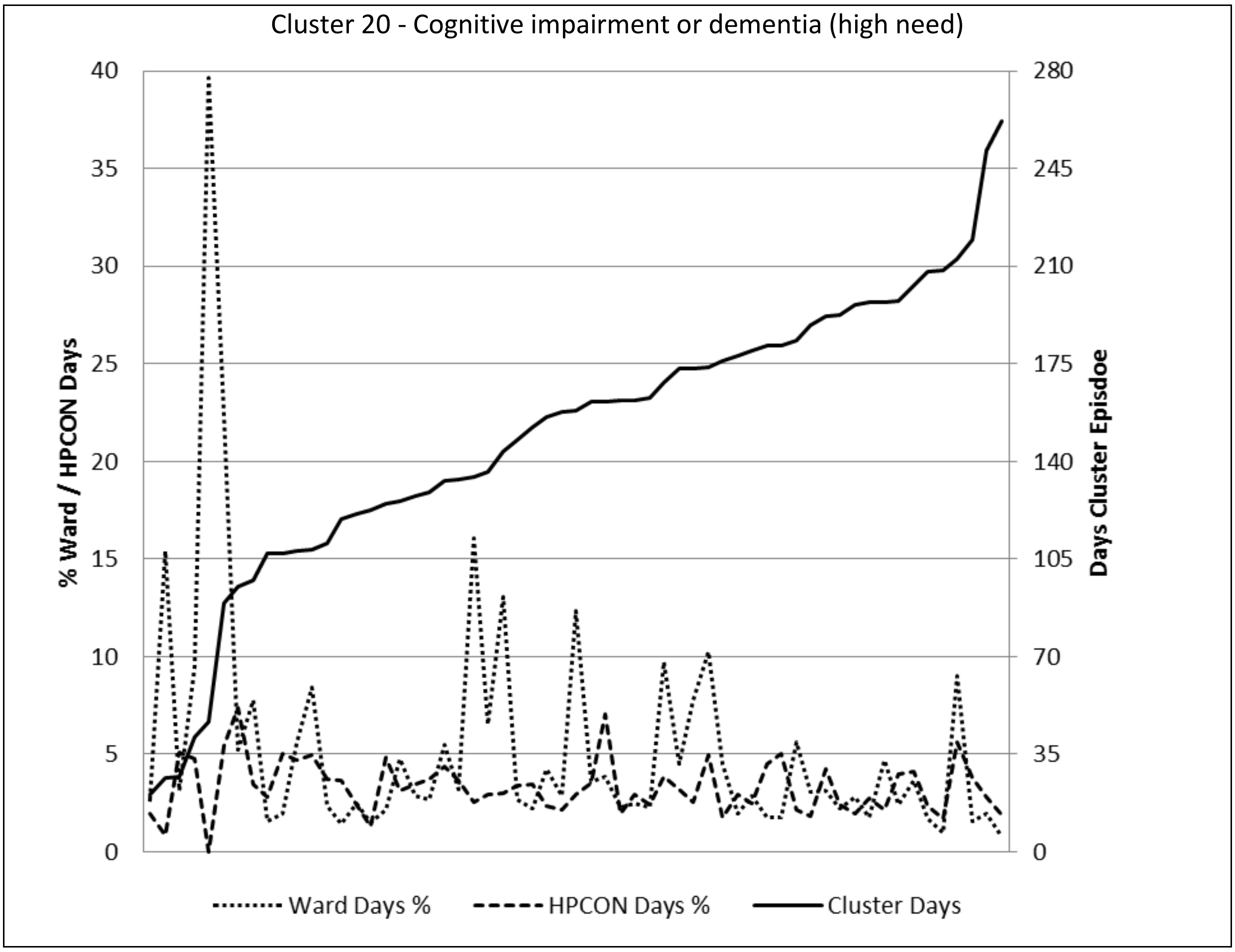

Cluster 15 - Severe psychotic depression
% Ward / HPCON Days
Days Cluster Episdoe
0
5
10
15
20
25
30
35
0
50
100
150
200
250
300
350
Ward Days %
HPCON Days %
Cluster Days

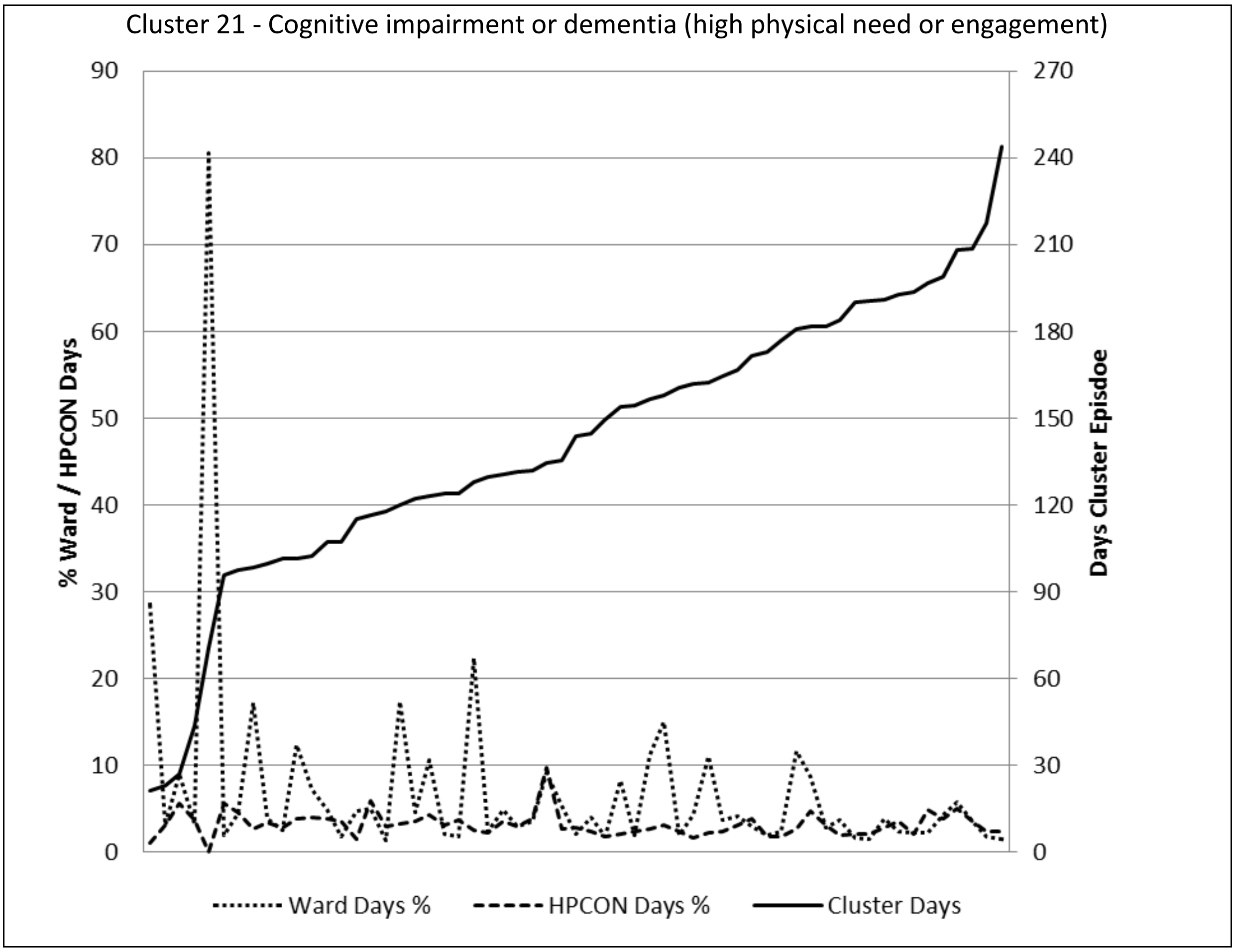

Cluster 16 - Dual diagnosis (substance abuse and mental illness)
% Ward / HPCON Days
Days Cluster Episdoe
60
50
40
30
20
10
0
300
250
200
150
100
50
0
Ward Days %
HPCON Days %
Cluster Days


Cluster 17 - Psychosis and affective disorder difficult to engage
% Ward / HPCON Days
Days Cluster Episdoe
70
60
50
40
30
20
10
0
350
300
250
200
150
100
50
0
Ward Days %
HPCON Days %
Cluster Days


Cluster 18 - Cognitive impairment (low need)
% Ward / HPCON Days
Days Cluster Episdoe
12
10
8
6
4
2
0
360
300
240
180
120
60
0
Ward Days %
HPCON Days %
Cluster Days


Cluster 19 - Cognitive impairment or dementia (moderate need)
% Ward / HPCON Days
Days Cluster Episdoe
40
35
30
25
20
15
10
5
0
400
350
300
250
200
150
100
50
0
Ward Days %
HPCON Days %
Cluster Days


Cluster 20 - Cognitive impairment or dementia (high need)
% Ward / HPCON Days
Days Cluster Episdoe
40
35
30
25
20
15
10
5
0
280
245
210
175
140
105
70
35
0
Ward Days %
HPCON Days %
Cluster Days


Cluster 21 - Cognitive impairment or dementia (high physical need or engagement)
% Ward / HPCON Days
Days Cluster Episdoe
0
10
20
30
40
50
60
70
80
90
0
30
60
90
120
150
180
210
240
270
Ward Days %
HPCON Days %
Cluster Days

B. Probability patterns for latent classes

This appendix presents the probability patterns for all 21 classes in the order of the estimation results. For further information regarding how to interpret these, please refer to Section 4.3.

To guide browsing through these results, a number of potential groups across these classes will be highlighted in the following. This section is not meant to give a higher level classification or grouping of these classes, instead, it is to guide interested readers to classes with specific problem constellations they may be interested in. Although all but one of the classes appear in at least one category, some appear in multiple ones, since they fulfil multiple criteria. The only class occurring in none of the groups is class N, which is characterised by some probability of mild problems in several areas.

The first group to note is those that cover *common mental disorders* of varying degrees of severity. They are mainly characterised by problems with depression (HoNOS7) and other mental and behavioural problems (HoNOS8) and relatively small probabilities of other problems (current and historic). As discussed in Section 4.3, these are classes I, L, O, Q and class G is an additional candidate (see overlap with next group).

The second group, *severe mental disorders* with low probabilities of both cognitive problems (HoNOS4) and psychotic experience (HoNOS6) is represented by classes G and J.

For the following suggested groups, for a specific problem to be counted as likely to be present in episodes of a specific class the sum of the probabilities for the categories 2-4 should be larger than 50%, since it should indicate actionable levels of distress or need (see Table 9).

Classes with a probability of some *cognitive problems* (indicative of potential dementia involvement, HoNOS4) are classes D, F, H, K, R, T, and U.

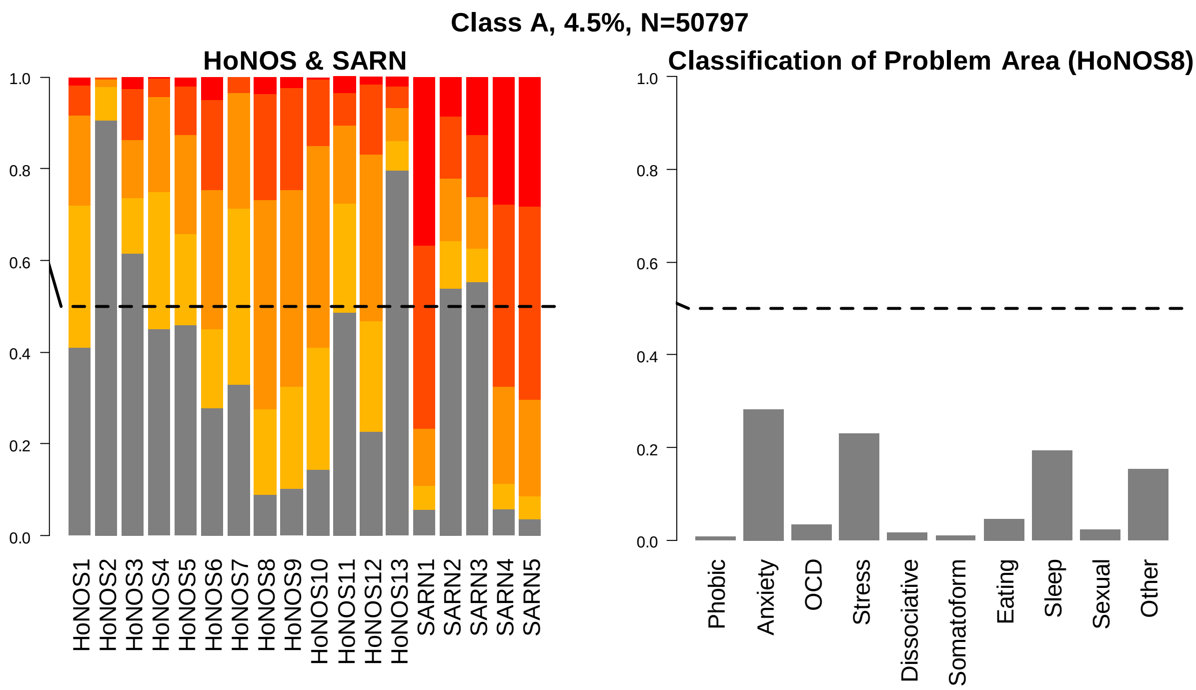
Classes with some probability of *psychotic experiences* (indicative of psychotic episodes, HoNOS6) are A, H, and M.

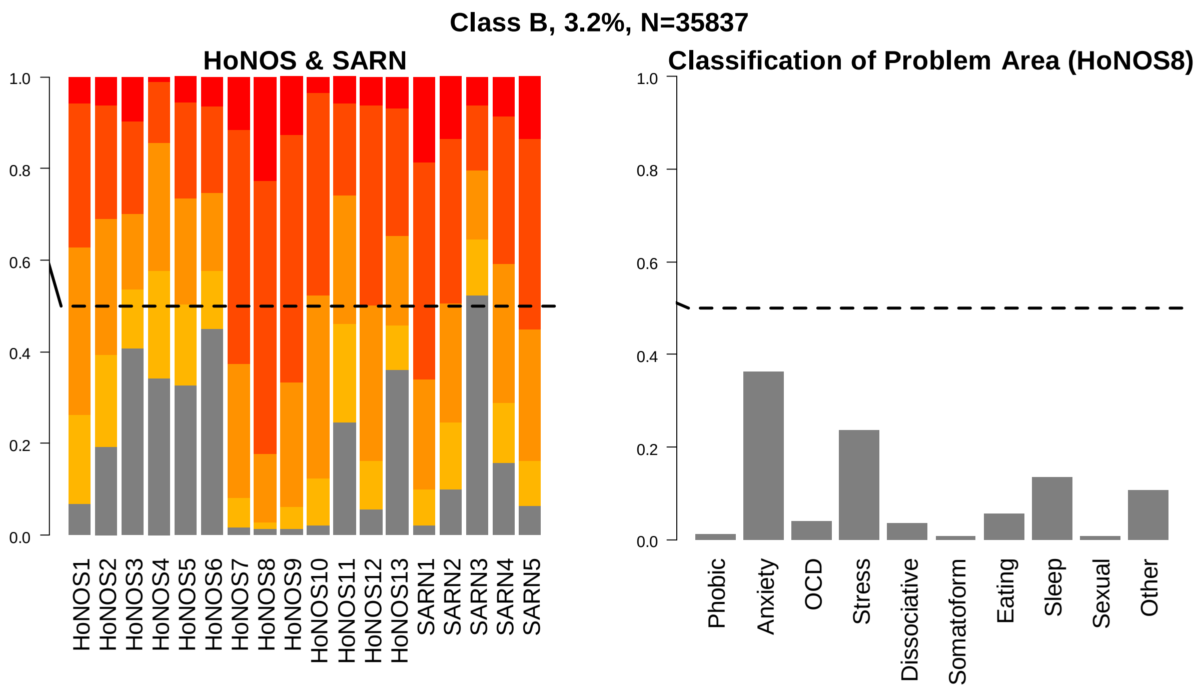
Classes with *comorbid physical* problems (HoNOS5) are D, F, H, J, K, O, T and U.

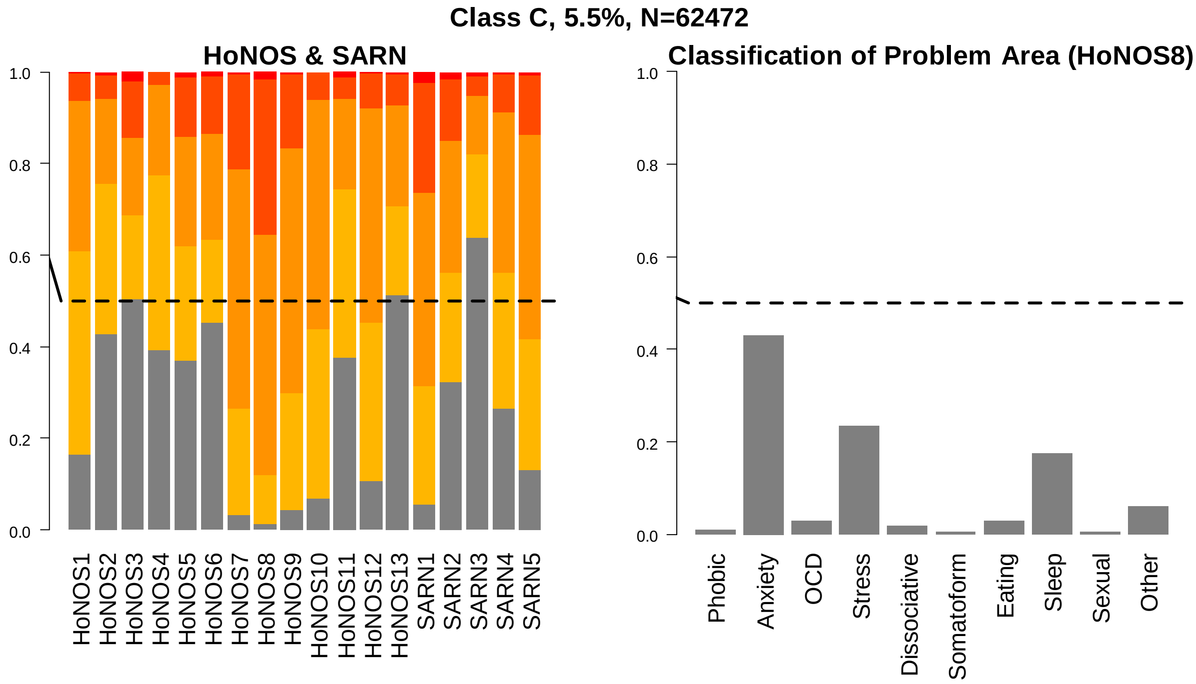
Classes with increased risk of current *self-harm* (HoNOS2) or historical self-harm (SARN2) are B, H, M, P and S.

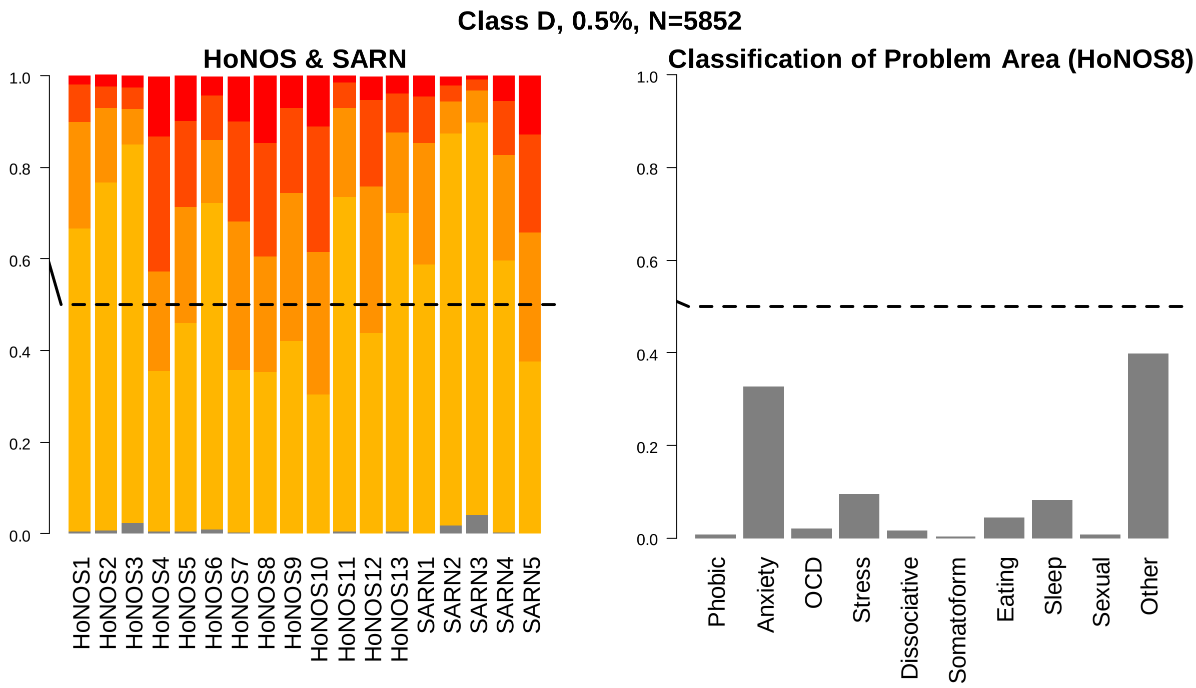
A couple of classes are characterised by a high probability of observing only other mental or behavioural problems (HoNOS8) and these are E, P, and Q.

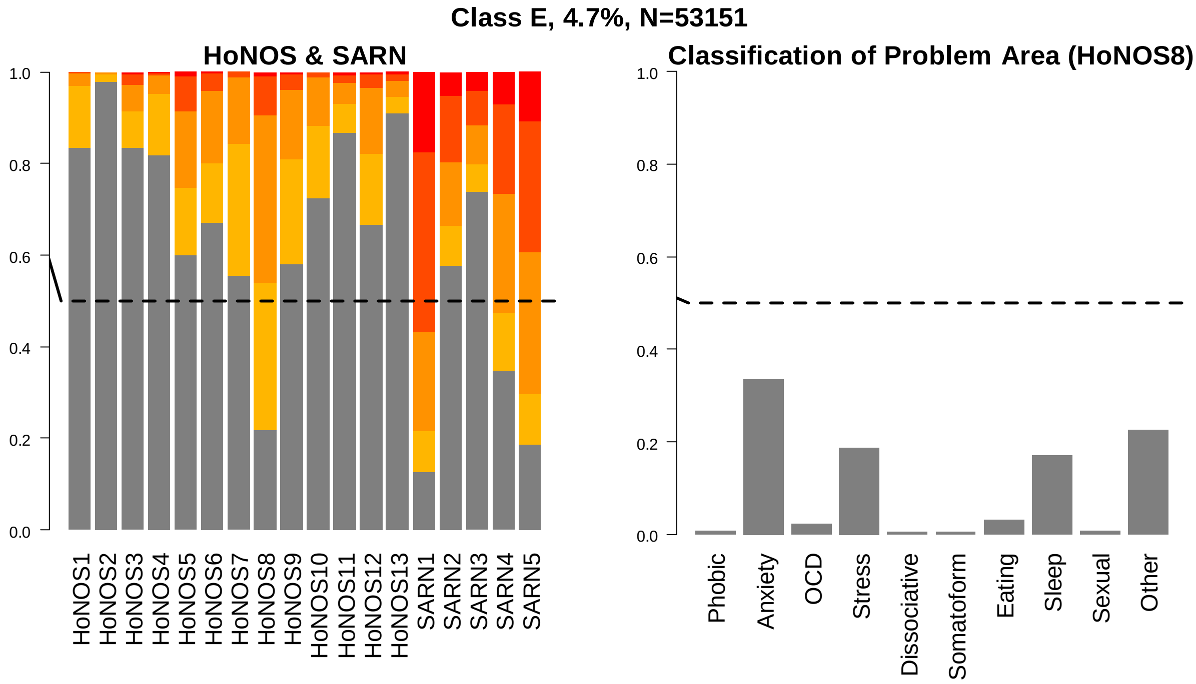
Classes with a high probability of at least two chronic or historical problems (i.e. any SARN item), are classes A, B, C, D, E, H, M and S.

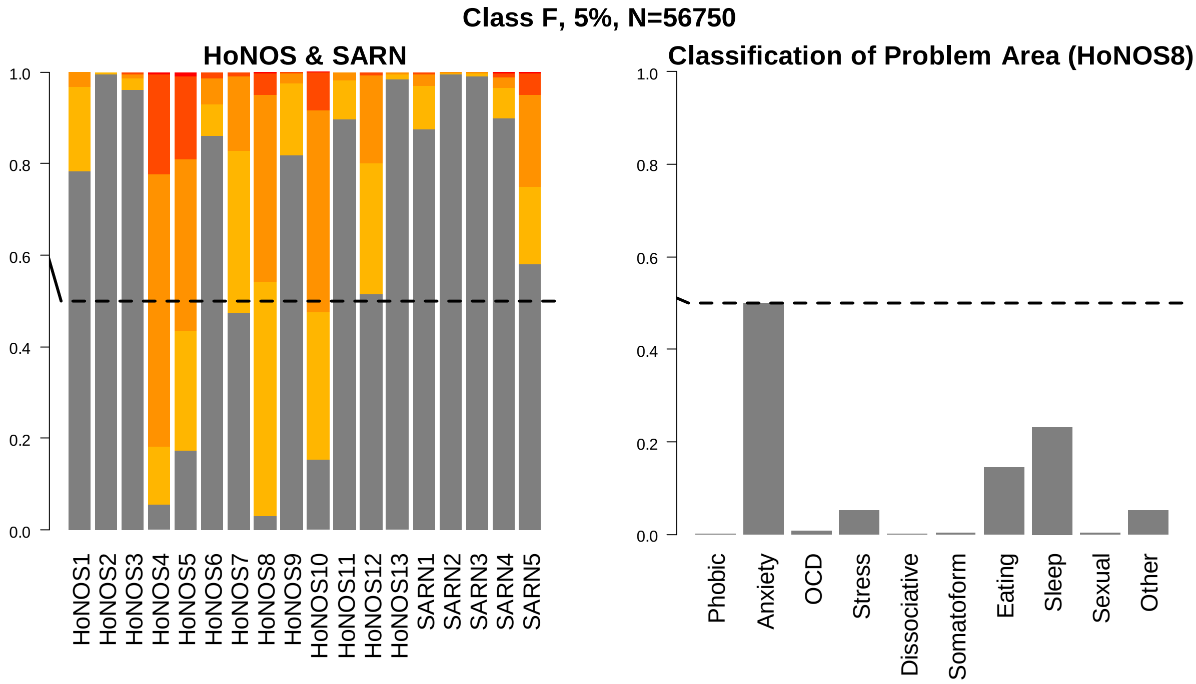

Class A, 4.5%, N=50797
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other

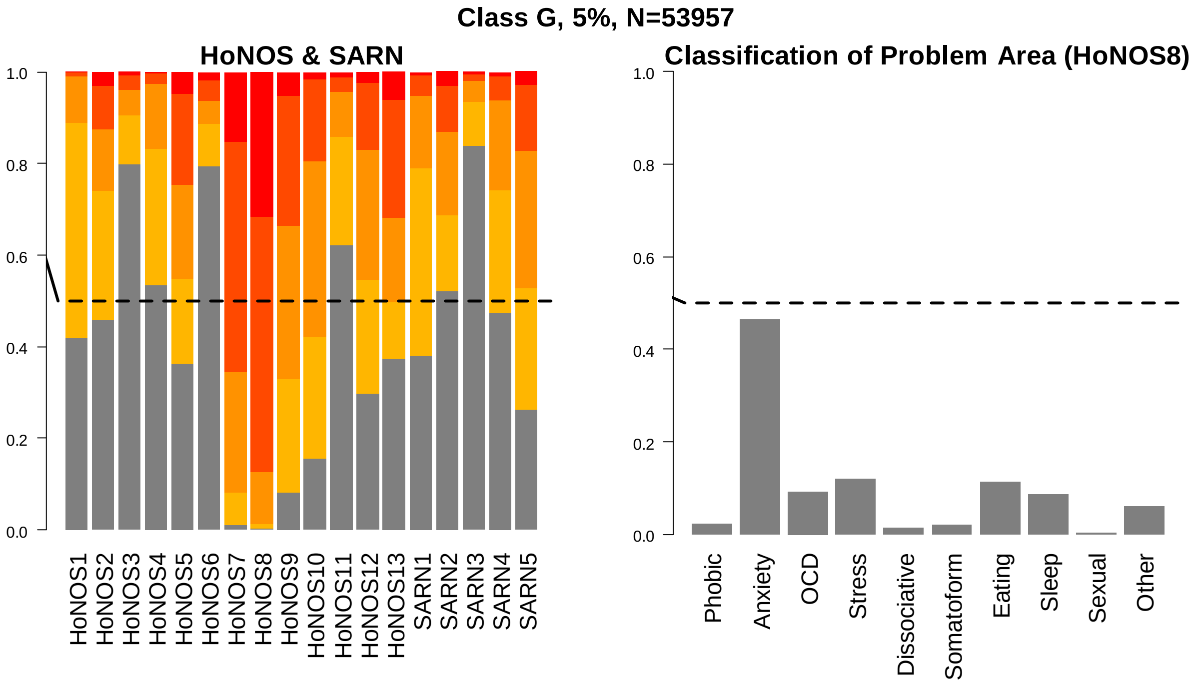

Class B, 3.2%, N=35837
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other

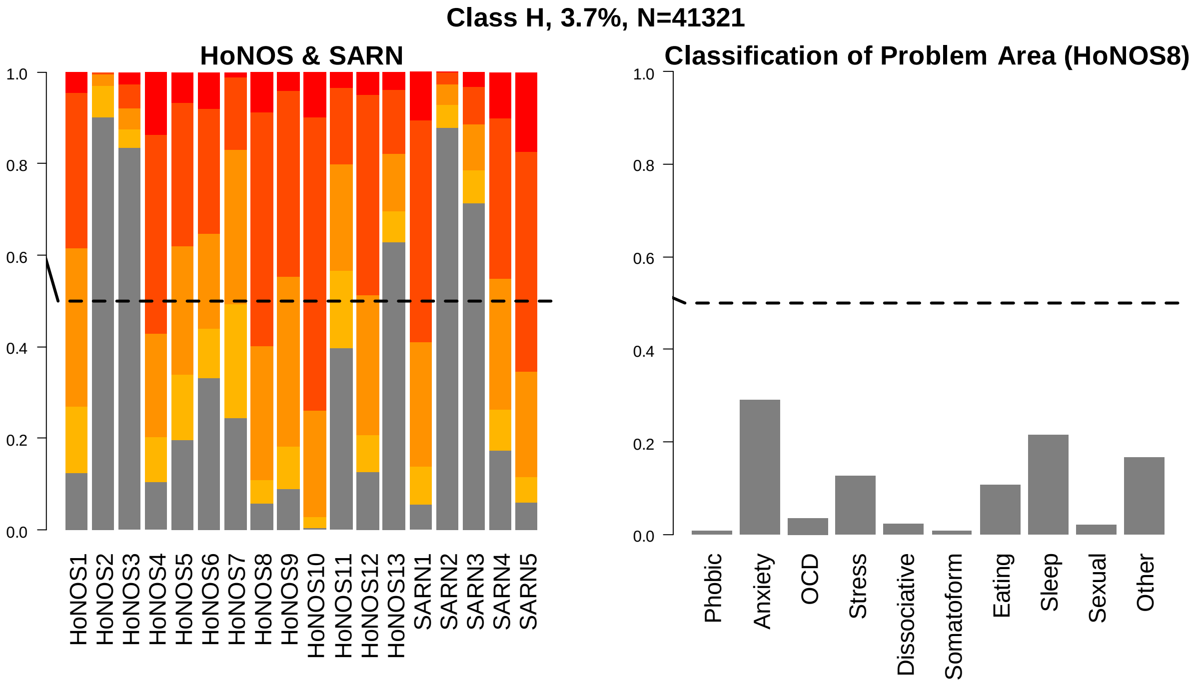

Class C, 5.5%, N=62472
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other


Class D, 0.5%, N=5852
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other

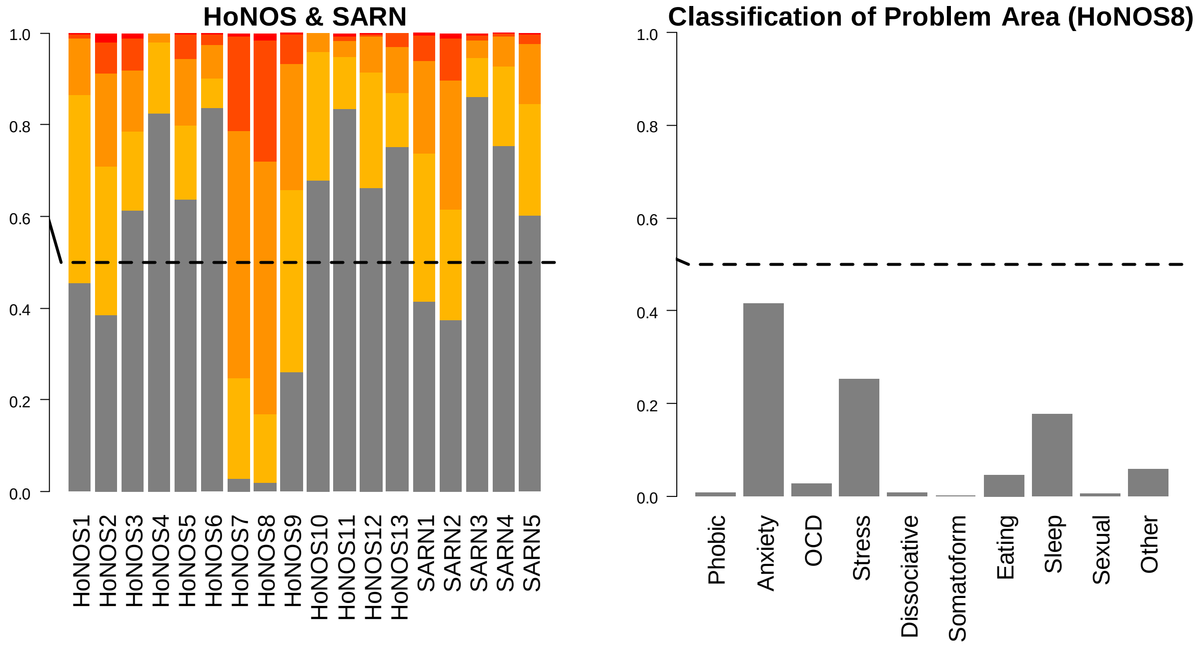

Class E, 4.7%, N=53151
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other


Class F, 5%, N=56750
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other

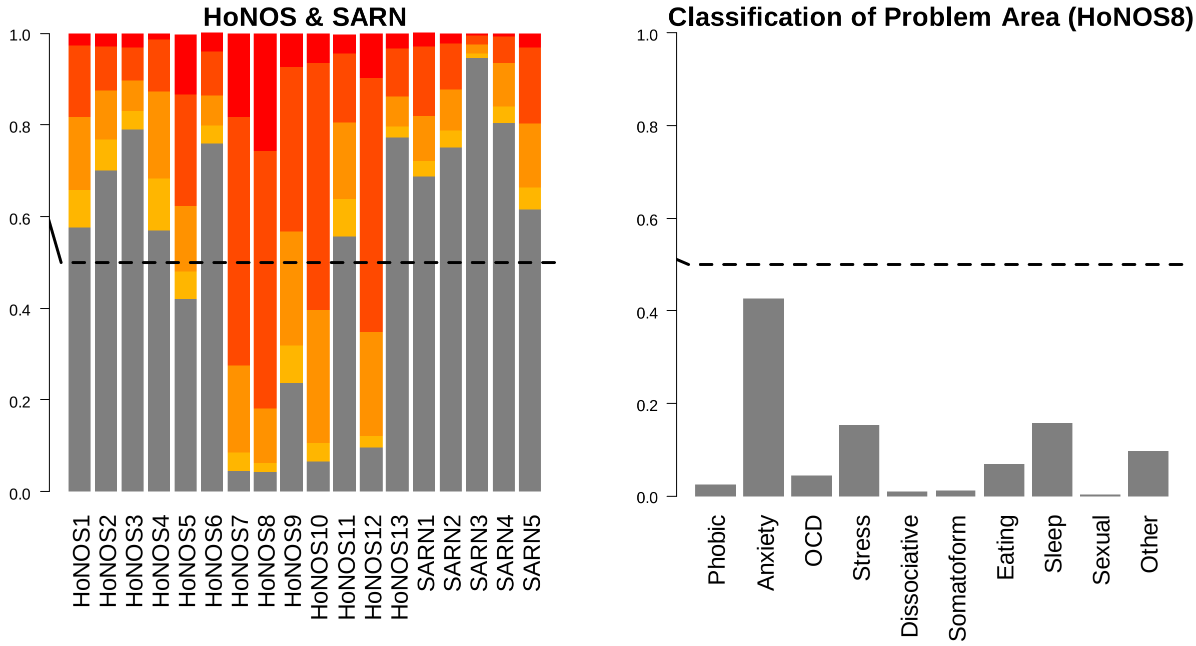

Class G, 5%, N=53957
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other

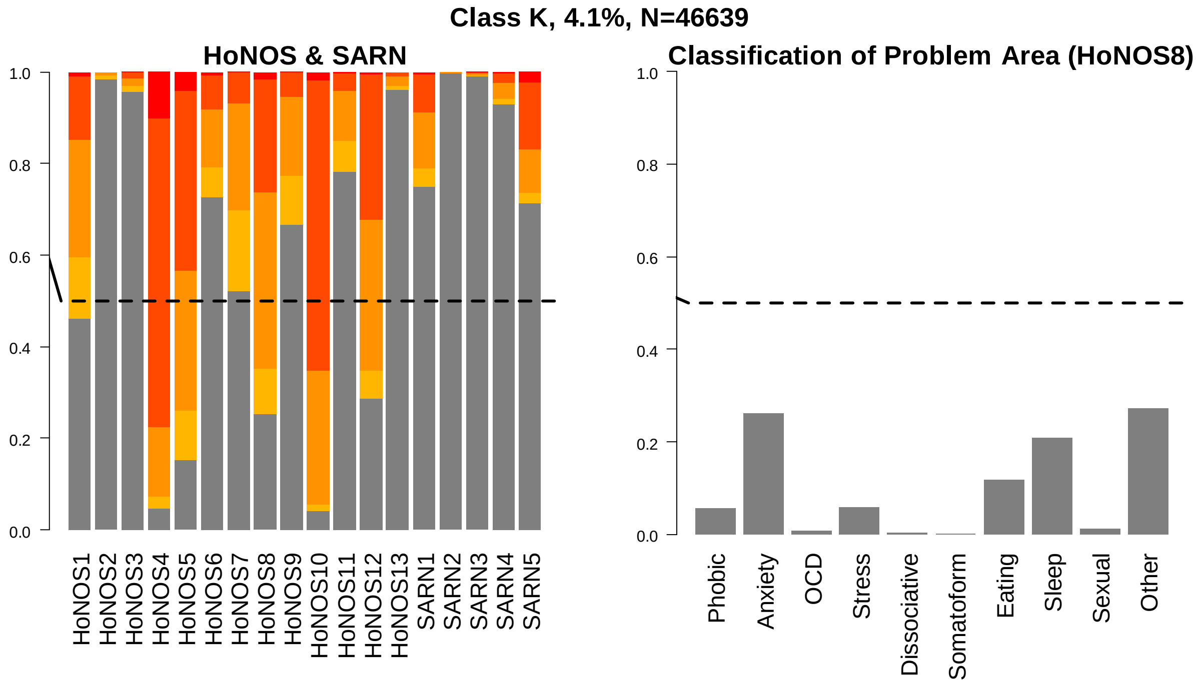

Class H, 3.7%, N=41321
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other

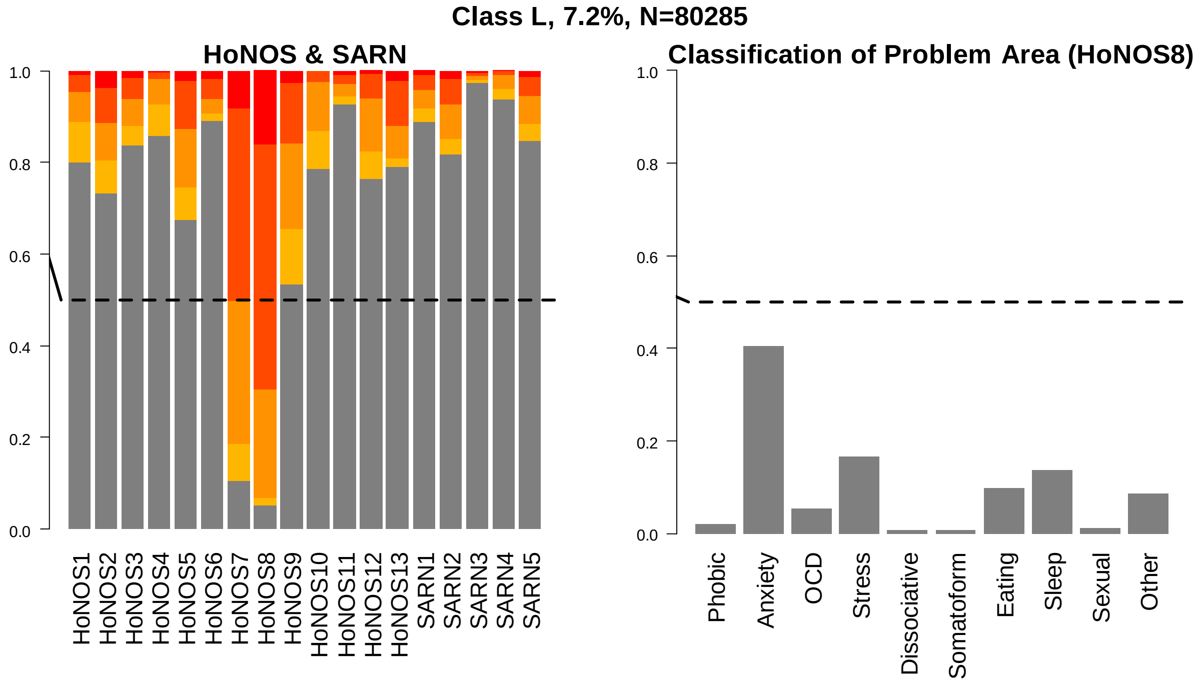
Class I, 6.7%, N=75435

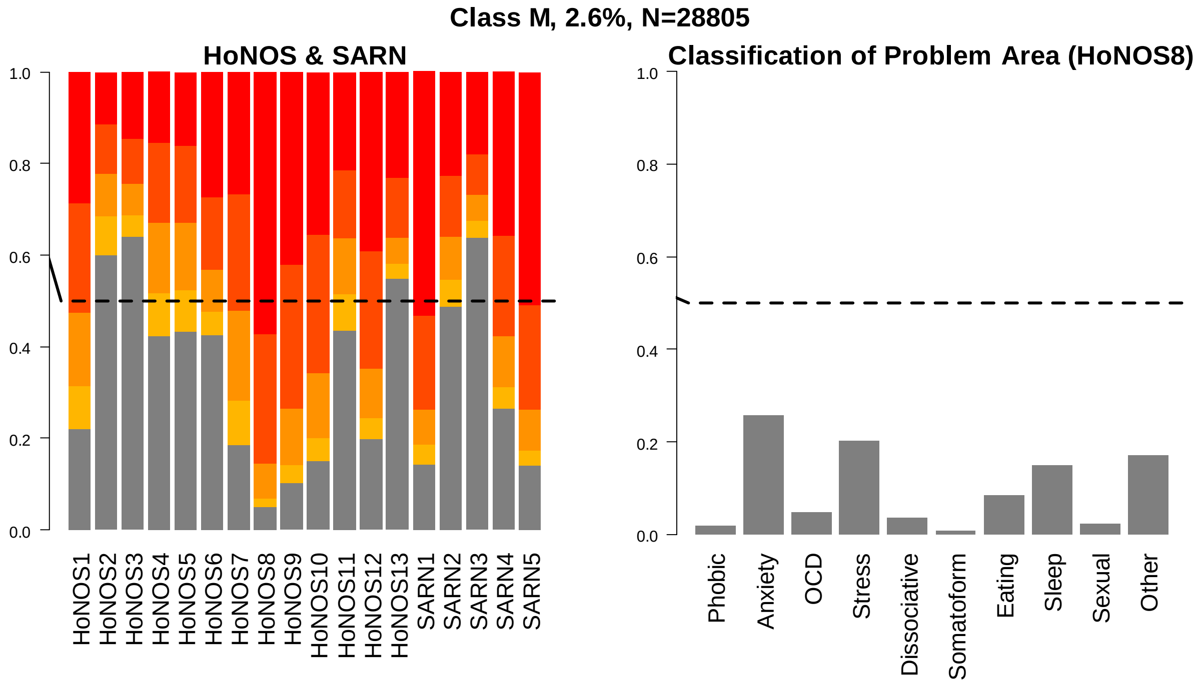


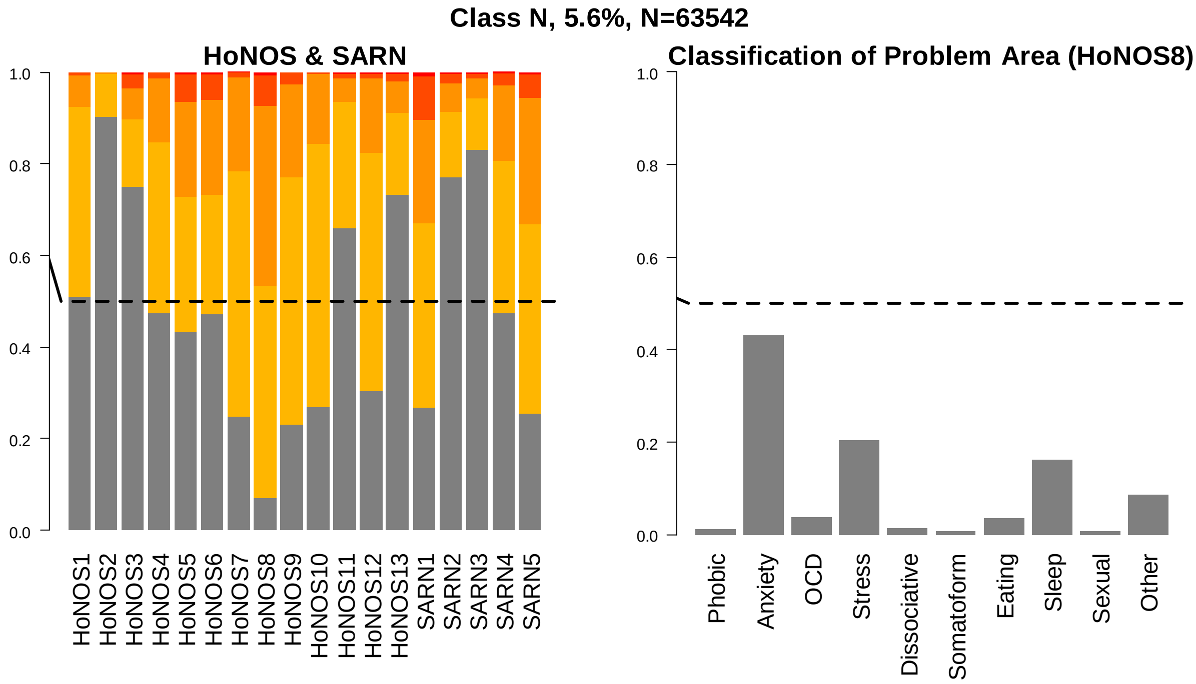
Class J, 5.2%, N=58106

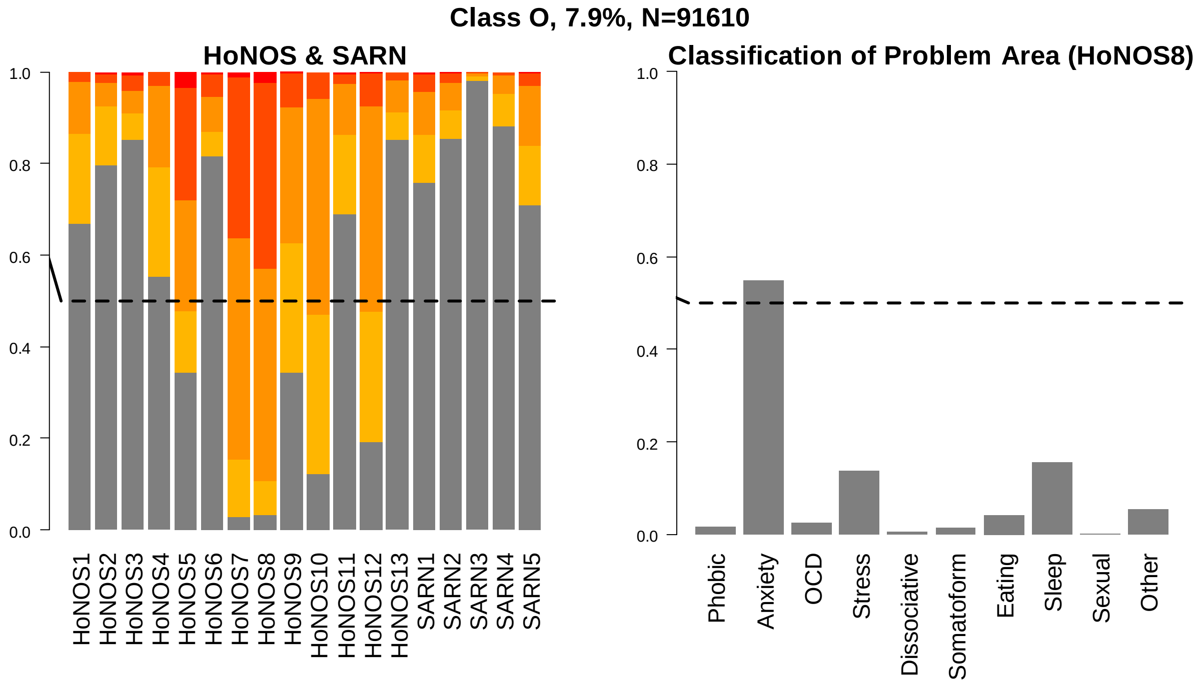


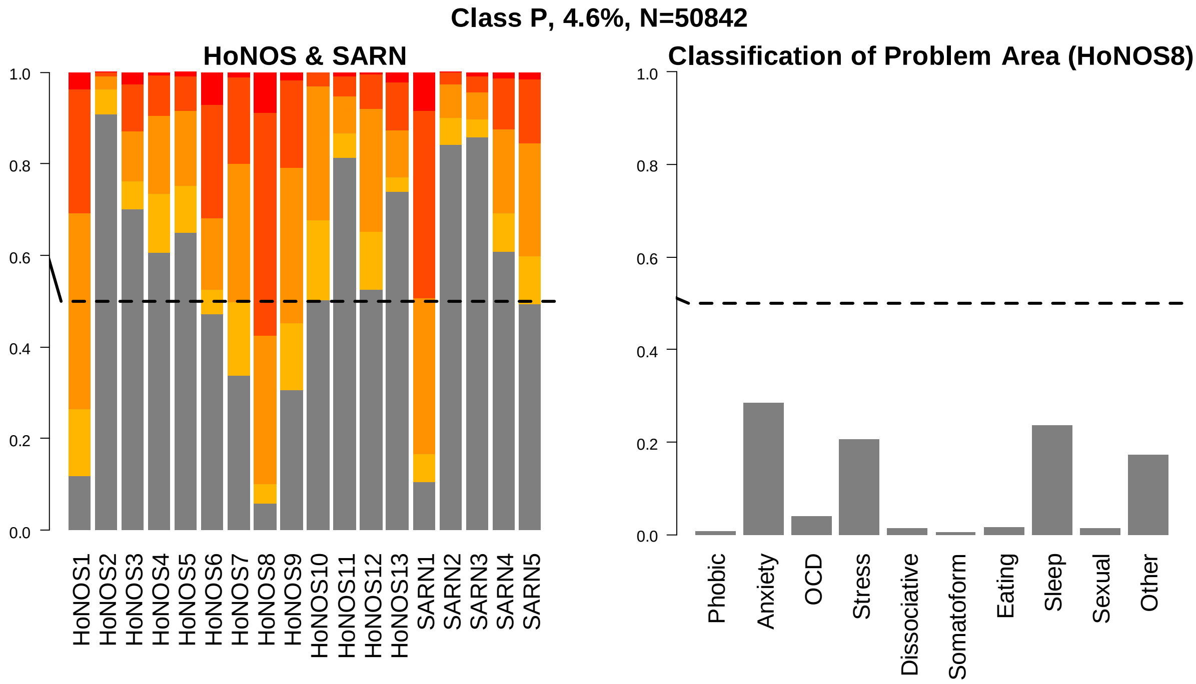

Class K, 4.1%, N=46639
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other

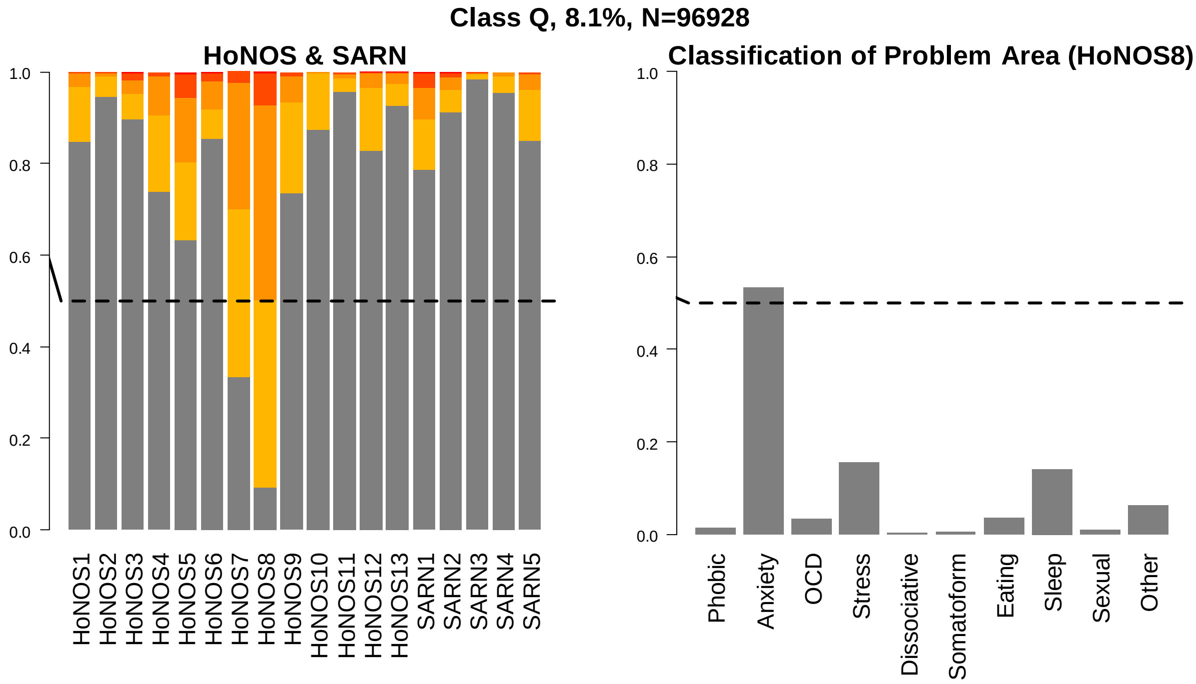

Class L, 7.2%, N=80285
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other

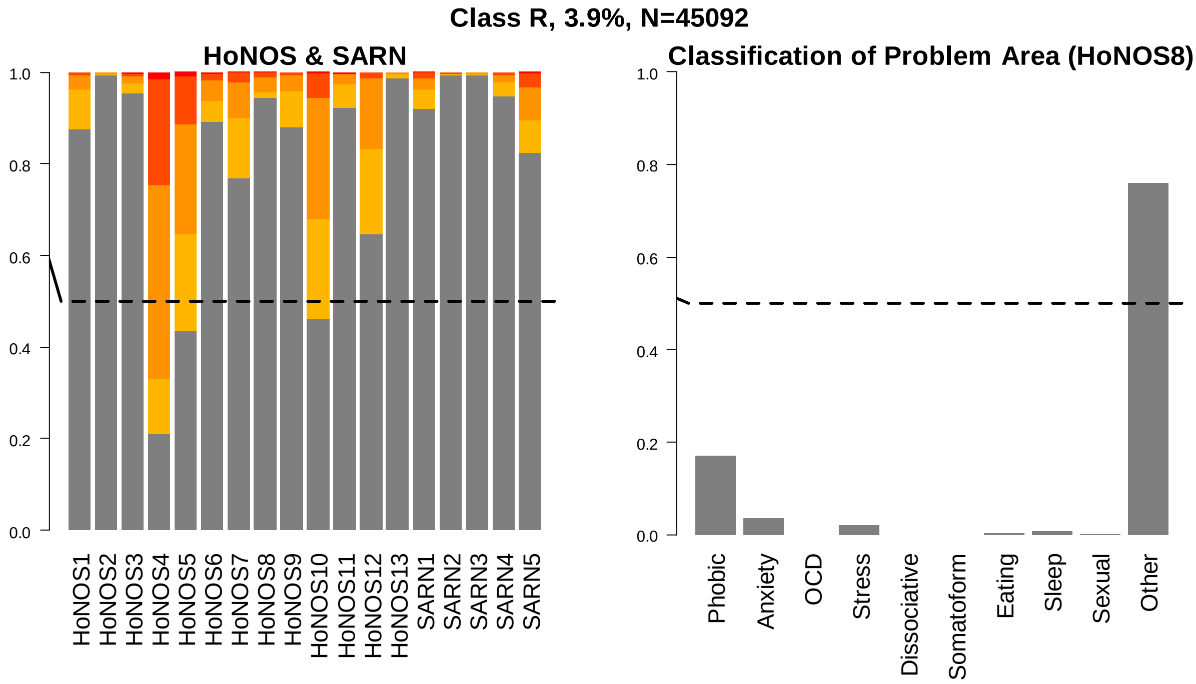

Class M, 2.6%, N=28805
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other

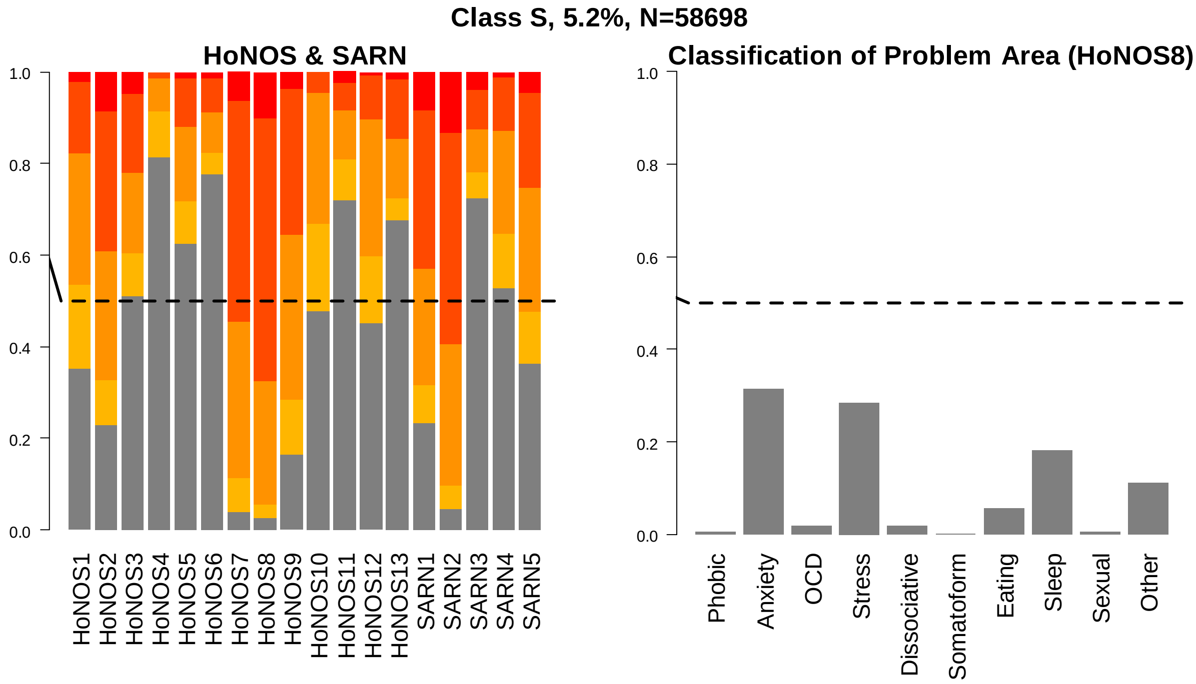

Class N, 5.6%, N=63542
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other

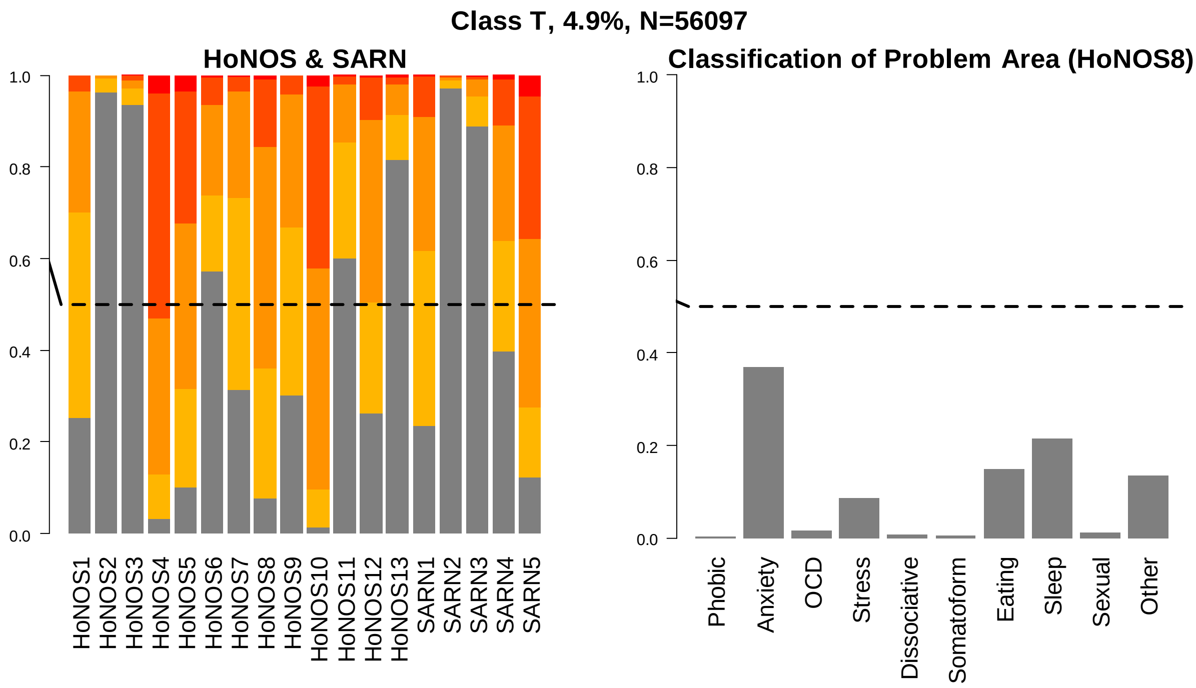

Class O, 7.9%, N=91610
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other

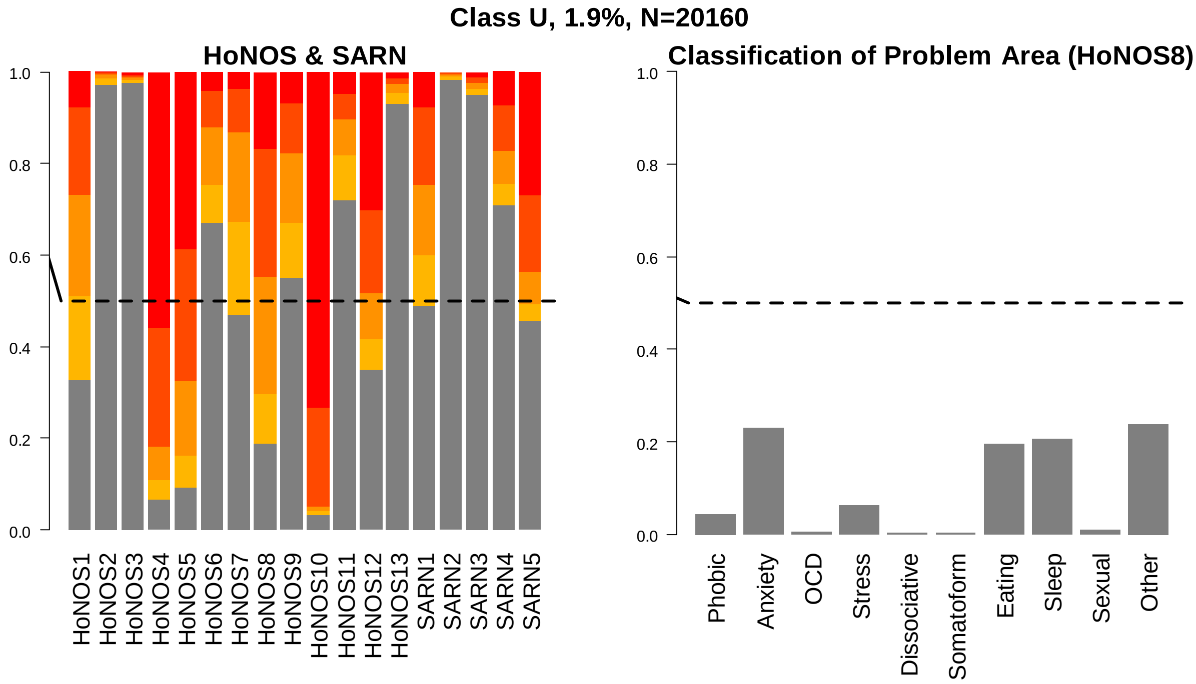

Class P, 4.6%, N=50842
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other


Class Q, 8.1%, N=96928
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other


Class R, 3.9%, N=45092
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other


Class S, 5.2%, N=58698
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other


Class T, 4.9%, N=56097
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other


Class U, 1.9%, N=20160
HoNOS & SARN
Classification of Problem Area (HoNOS8)
1.0
0.8
0.6
0.4
0.2
0.0
HoNOS1
HoNOS2
HoNOS3
HoNOS4
HoNOS5
HoNOS6
HoNOS7
HoNOS8
HoNOS9
HoNOS10
HoNOS11
HoNOS12
HoNOS13
SARN1
SARN2
SARN3
SARN4
SARN5
Phobic
Anxiety
OCD
Stress
Dissociative
Somatoform
Eating
Sleep
Sexual
Other