ORIGINAL PAPER

Payment by results: validating care cluster allocation in the real world

Stavros Bekas,[1] Orlin Michev[2]

The Psychiatrist (2013), **37**, 349–355, doi: 10.1192/pb.bp.112.041780

[1]West London NHS Mental Health Trust; [2]Central and North West London NHS Foundation Trust

Correspondence to Stavros Bekas (stavros.bekas@nhs.net)

First received 15 Oct 2012, final revision 8 Mar 2013, accepted 15 Mar 2013

Aims and method To validate care cluster allocation for payment by results (PbR) in mental health and to evaluate clustering and auditing methodologies. We applied exclusion criteria to the patient population of a mental health trust. An automated validation compared cluster with expected ICD-10 codes or scores on the Health of the Nation Outcome Scales (HoNOS) and Mental Health Clustering Tool (MHCT). Six hundred 'mismatched' cases were reviewed in depth to better understand the reasons why these cases appeared misclustered.

Results There was a significant mismatch between ICD-10 codes, HoNOS and MHCT scores and allocated care cluster, with differences between services and localities. Some clusters appeared to be more accurately allocated. The 'deep dive' analysis indicated that most mismatches occurred because psychosis was allocated to a non-psychotic cluster and *vice versa*, but also as a result of inherent weaknesses of the MHCT.

Clinical implications High levels of inappropriate care cluster allocation highlight the need to improve practice. Weaknesses in the MHCT and ICD-10 coding mean that the final arbiter should be clinical judgement. Auditing will, by necessity, have a significant margin of error.

Declaration of interest S.B. has been involved in HoNOS and MHCT training with the Royal College of Psychiatrists and is currently on the PbR team in West London Mental Health NHS Trust.

The financial year 2012–13 witnessed a major change in the way that mental healthcare is funded, a shift from block contracts to payment by results (PbR) 'currencies'. These have been in use in acute care for years and are linked to ICD-10 codes and unit costs for procedures and treatments. Acute PbR was founded on the strategic priorities set by the National Service Framework policy launched in 2000 which included structural transformations in combination with financial 'levers for change'. Introducing PbR currencies was the Department of Health's chosen approach for the interface between commissioners and providers.[1]

Payment by results in mental health will be significantly different in that it will not use fixed tariffs for each condition; rather, payment will be linked with individual patient needs and care plans produced during their contact with mental health services. It was decided to produce currencies through the allocation of patients to 21 'clusters' of care using the Mental Health Clustering Tool (MHCT).[2,3] A cluster corresponds to a group of patients with similar clinical symptoms, needs and disabilities, with the idea that a single tariff on average will be sufficient to cover the cost of care for each patient allocated to the cluster. Mental health PbR has been enshrined as a commitment for the Department of Health and National Health Service (NHS) providers in the government's White Paper *Liberating the NHS*.[4] Current policy envisages the introduction of local prices which will form the basis of commissioning contracts in 2013–14, and the earliest date for introduction of national prices is set to be in 2015.[2] Guidance for mental health PbR recognises that local cluster allocation quality and quantity information for 2013–14 will continue to be variable and suggests that providers and commissioners should 'work together to mitigate the risks of financial instability'.[2]

West London Mental Health NHS Trust rolled out clustering in 2010 using the MHCT, in preparation of the implementation of PbR. All working-age adults, early intervention and older people service users should have been allocated to a cluster by 31 December 2011, but data have not been validated. Due to the fact that any gaps or inconsistencies in the implementation of clustering, along with substandard or inappropriate allocation to currencies, will potentially have implications in terms of probity and financial stability, it was important to develop a method of validation and to make recommendations for improvement before or in parallel with the next stages of PbR implementation in the forthcoming financial year. The aim of this study was to appraise the validity of cluster allocation and to evaluate clustering and auditing methodologies.

Method

The data-set was extracted from the data warehouse of the Trust's RiO electronic record system. The relevant sub-population was chosen from all active patients: those with no cluster and children and adolescent, forensic and gender identity service users were excluded, as at this stage these services are not within the scope of PbR.

The study was divided into two arms and each arm consisted of two phases (Fig. 1). In phase 1, we used an algorithmic, automated validation. A set of rules was designed to analyse the data using SQLServer 2008 for Windows XP. This provided an automated cross-check of the allocated cluster with a number of expected ICD-10 codes[5] for each cluster (e.g. 'Schizophrenia, code F20' was accepted in the 'psychotic' clusters 10–13, but not in the 'neurotic' clusters 1–8) and another to cross-check with a range of expected severity scores in the working-age adult HoNOS or the MHCT. These rules were taken from the MHCT booklet,[3] with some additional three-letter ICD-10 codes and other minor adjustments. Cases with no ICD-10 code were excluded from phase 1 and cases with no complete HoNOS or MHCT scores were excluded from phase 2.

In phase 2, we employed a 'deep dive' clinical analysis. We drew two samples of 300 cases each from the 'mismatching' cases of both arms. Sampling was random but proportional to the size of the case-loads of each service to control for variations between them. The sample size chosen was not based on statistical reasons, but on the largest size that could realistically be analysed within the resources provided. The samples were reviewed manually by two expert clinicians, based on clinical information (i.e. diagnoses, HoNOS scores and, if necessary, reading in more detail the records, using all available guidance). These in-depth analyses were conducted in order: to discover whether the algorithm indeed worked and the results were not spurious products of the software; to see whether apparent mismatches were justifiable on clinical grounds; and to record the reasons for confirmed erroneous allocations.

The results were analysed by locality, service and cluster. The services were grouped under 'Community' (e.g. all recovery teams), 'In-patients' and 'Older Adults', whereas services such as early intervention and assertive outreach were grouped under 'Specialist'. We considered separately the Hammersmith & Fulham Assessment Team, as at the time of the study it was the only service in the Trust exclusively focused on community intake and brief treatment, in order to see whether clustering was any more accurate soon after referral. We used proportions as descriptive statistics.

Ethics and governance

The study was commissioned and ethically approved by senior managers, including the medical director, providing exemption from further referral to an ethics committee. All efforts were made for complete anonymisation and safe-keeping of records. Neither the data-sets produced nor the results contained traceable information about patients or staff involved in the clustering process. Consent for data collection and processing was not sought, as the study used data routinely recorded as part of the Trust's business and quality improvement.

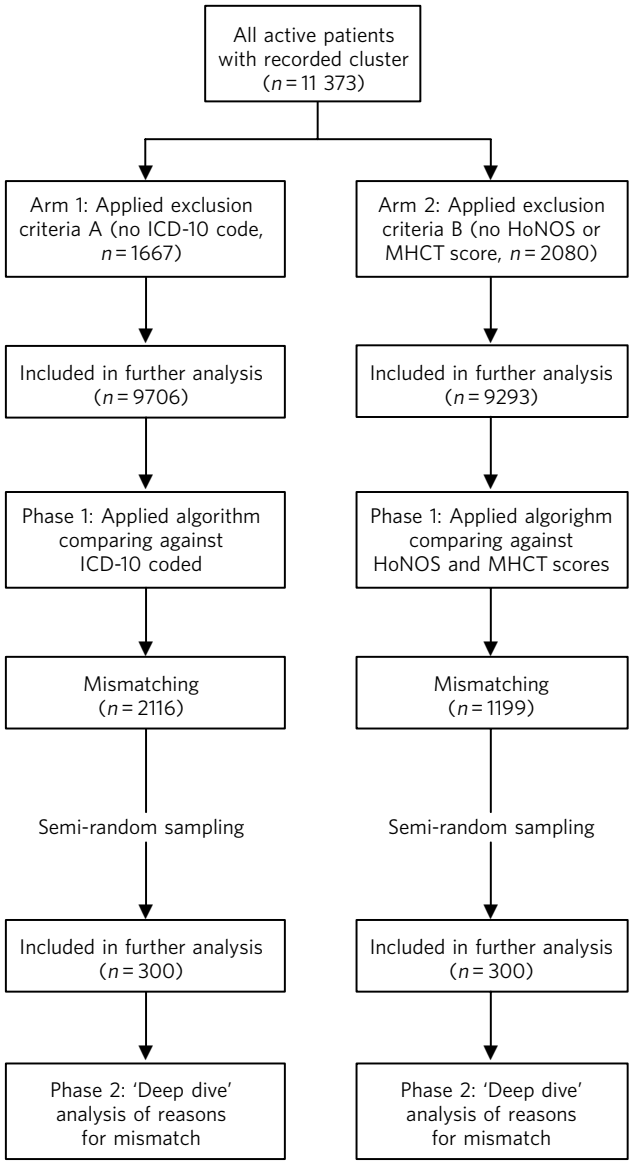

Fig 1 Outline of study process. HoNOS, Health of the Nation Outcome Scales; MHCT, Mental Health Clustering Tool.

Results

The results of phase 1 are shown in Table 1. A total of 1667 (14.7%) cases in arm 1 were excluded from phase 1 because they did not match the inclusion criteria, and 9706 were included. Compared against ICD-10 codes, the Trust average mismatch was 21.8%. We used this 21.8% as the benchmark to arbitrarily define 'underperforming' and 'overperforming' services and localities. It appeared that there were no major differences between locality averages, but some services were significantly less accurate than others, particularly in Hounslow in-patient and community teams, Ealing and Hammersmith & Fulham older adult services, the Hammersmith & Fulham Assessment Team and the Cassel Hospital.

https://doi.org/10.1192/pb.bp.112.041780 Published online by Cambridge University Press

Table 1 ICD-10 code and HoNOS/MHCT comparison

Service unit	Mismatch with ICD-10 ($n = 9706$), %	Mismatch with HoNOS/MHCT ($n = 9293$), %
Ealing		
Community	19.5	14.4
In-patients	15.0	6.4
Older adults	21.6	8.4
Specialist	19.5	11.1
Average	19.9	10.4
Cassel Hospital		
In-patients	0	25.0
Specialist	45.8	16.7
Average	37.9	20.0
Hammersmith & Fulham		
Community	15.7	8.3
Hammersmith & Fulham		
Assessment Team	31.3	15.3
In-patients	16.3	9.4
Older adults	26.2	12.5
Specialist	19.1	11.3
Average	22.9	12.0
Hounslow		
Community	26.9	15.2
In-patients	30.0	19.2
Older adults	18.7	12.2
Specialist	18.1	21.4
Average	23.1	16.9
No service recorded	22.2	10.0
Trust average	**21.8**	**12.9**

HoNOS, Health of the Nation Outcome Scales; MHCT, Mental Health Clustering Tool.

As seen in Fig. 2, some clusters appeared more accurately allocated than others. The 'psychotic' clusters seem to be more robust and presented lower rates of mismatch with diagnoses. Conversely, cluster 16 (dual diagnosis), cluster 15 (psychotic depression), cluster 8 (chaotic personality disorder) and the low-need 'non-psychotic' clusters 1–4 presented higher rates of mismatch with recorded diagnoses. All allocations to cluster 9 were considered wrong, as this is a 'blank' cluster which should not be used. The 'variance' cluster 0 should be reserved for those very few patients that cannot be allocated to any other cluster. According to our algorithm, about 70% of patients allocated to cluster 0 in our case-load had diagnoses that could inform allocation to another cluster.

A total of 2080 (18.2%) cases were excluded from phase 1 in arm 2, resulting in this phase being applied to a total of 9293 cases. Compared against HoNOS or MHCT scores, the Trust average mismatch was about 12.9%, significantly lower compared with the range of recorded diagnoses. Again, some services and localities seemed to be performing below the Trust average and the trends were similar to arm 1, with the exception of the Cassel Hospital. The Hammersmith & Fulham Assessment Team presented higher-than-average mismatches at 15.3% (Table 1). Some clusters appeared to be more accurately allocated using this methodology than others, but the level and distribution of mismatch was different from phase 1 in arm 1 (Fig. 3). Again, the 'psychotic' clusters seemed to be more robust. Conversely, allocations to cluster 8 (chaotic personality disorder), cluster 15 (psychotic depression), cluster 5

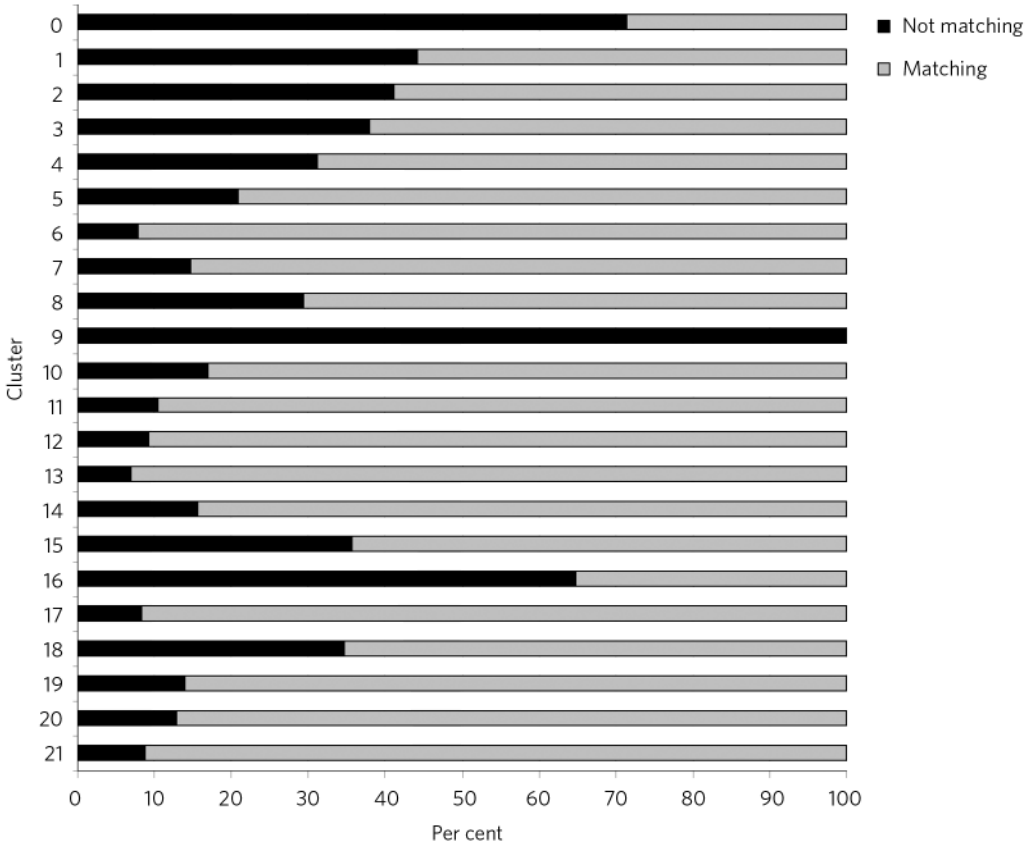

Fig 2 ICD-10 diagnosis comparison by cluster.

https://doi.org/10.1192/pb.bp.112.041780 Published online by Cambridge University Press

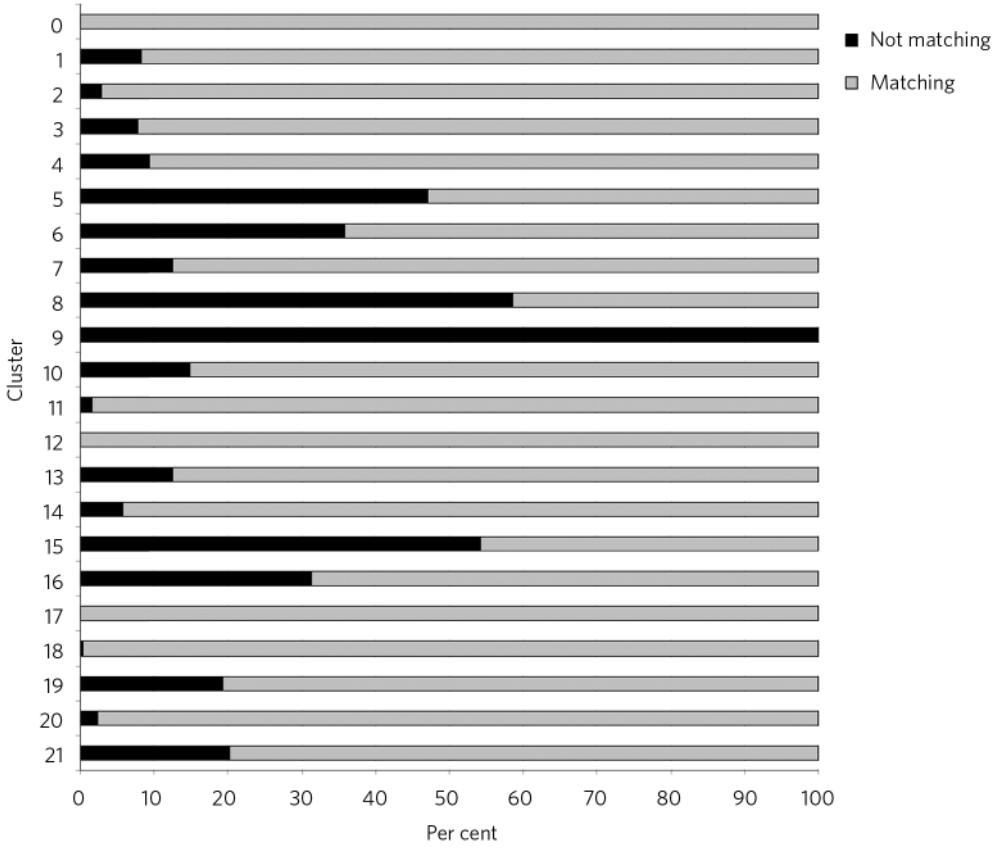

Fig 3 Comparison between the Health of the Nation Outcomes Scales and Mental Health Clustering Tool scores by cluster.

(severe non-psychotic) and cluster 16 (dual diagnosis) appeared particularly mismatched with HoNOS/MHCT scores. The low-need 'non-psychotic' clusters 1–4, however, showed lower rates of mismatch.

In our sample, the largest group of mismatches between ICD-10 code and cluster (Fig. 4) was due to the fact that some cases were allocated to a 'non-psychotic' cluster (34%) despite the presence of psychosis in recorded diagnoses. This was not justifiable on any clinical grounds – for example, the cluster included major psychosis such as schizophrenia (16.4%) and bipolar disorder (12%), for which the patients were receiving highly complex care packages under the care programme approach. Conversely, the second largest group of mismatches was due to cases being allocated to a 'psychotic' cluster despite having no psychotic condition recorded (16.7%). None of these was an obvious case of failure to record the correct diagnosis. The third most common reason (12.6%) for mismatch was the allocation of patients with recurrent depression to the low-need clusters 1–3. According to the MHCT, recurrent depression should generally be considered a higher-need condition (in our algorithm it was expected to be in clusters 4 or 5). Fourth in frequency (9.3%) were allocations to cluster 16 (dual diagnosis) in the absence of a recorded drug and alcohol use-related condition, which in all cases was as a result of failing to record this information in the first place, and a small percentage of patients in cluster 16 did not have a psychotic condition recorded. Finally, a significant 7% of the misallocations were as a result of chaotic and chronic personality disorders (primarily emotionally unstable) being allocated to the low-need, 'non-psychotic' clusters and 3.2% misallocations were due to the absence of a personality disorder diagnosis despite an allocation to cluster 8, which is for this purpose.

The vast majority (78%) of mismatch between HoNOS/MHCT severity and cluster allocation was because of lower-than-expected severity scores, and 11% was as a result of higher-than-expected severity scores in the cluster. Three per cent appeared mismatched because of a combination of problems that would not be expected in the cluster (e.g. high scores in hallucinations and delusions when the cluster was 'non-psychotic'). Finally, 8% of mismatched cases were deemed to be as a result of errors in the algorithm and the MHCT (Fig. 5).

Discussion

A report by the Sainsbury Centre for Mental Health raised concerns about the challenges that PbR presented in mental health, as this approach is characterised by long-term and episodic conditions, variability of services and the cost of care is influenced by a multitude of factors beyond diagnosis (i.e. multi-agency pathways and informal care).[6] Dangers of acute PbR have been established through many years of international experience – that providers may 'cherry pick' easier and cheaper cases, reduce quality of care by hastening turnover, manipulate patient coding into higher tariffs or increase activity to such an extent that commissioners

https://doi.org/10.1192/pb.bp.112.041780 Published online by Cambridge University Press

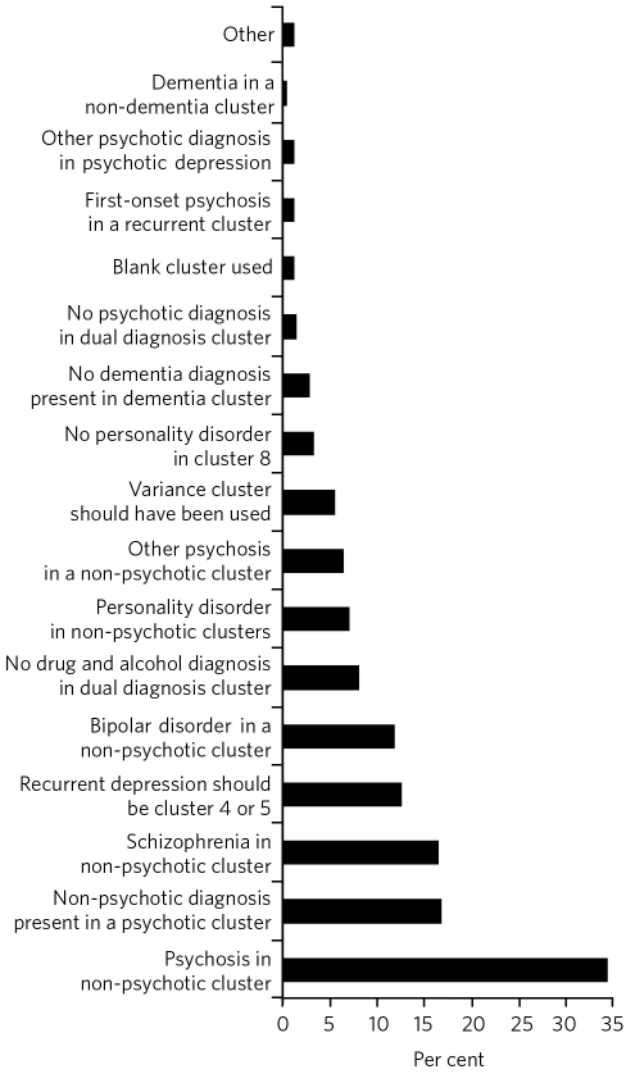

Fig 4 Reasons for mismatch between ICD-10 codes and cluster.

cannot afford the cost. As a result, it was suggested not to abandon PbR altogether, but to include safeguards that improve coding, prevent excessive hospital utilisation and promote quality of care.[6]

Similarly, problems identified by the developers of mental health PbR currencies included lack of a satisfactory classification system, large provider variations, case-mix of acute and chronic cases and variable care needs, in addition to major technical challenges in auditing and data collection.[1] The same developers later acknowledged the lack of robust evidence to further support the 'high face validity' of clusters and failed to demonstrate beyond doubt that they constitute a 'fit for purpose classification system', attributing the difficulty to data quality and methodological difficulties.[7]

Achieving homogeneity within clusters is considered essential to ensure that PbR will not introduce financial risk for providers.[8,9] This was the primary reason why similar approaches to payment were not implemented for mental health services in the USA, Australia or New Zealand.[9–11] Similarly, initial cost analyses in the UK demonstrated that homogeneity within clusters was 'unacceptably low' and that providers vary in their resource utilisation much more than what could be explained by differences in their case-load.[12,13]

Arguably, PbR challenges are multiplied by the fact that we are currently facing the most extensive financial retrenching since the introduction of the internal market in the NHS, in combination with a drive for quality improvement and patient choice. Payment by results was heralded as the mechanism by which payment is attached to patient choices, but the Department of Health also aimed to provide incentives to providers to achieve the lowest cost consistent with quality outcomes,[1] while the tools from which the MHCT is derived and the process of defining clusters did not take into account costs at all.[14] Later work replicated the significant problems in having quality data and in including outcomes as a means of incentivising quality improvement.[2,15] Nevertheless, PbR remains an opportunity to have more transparent commissioning and potentially can act as a strong incentive for the routine

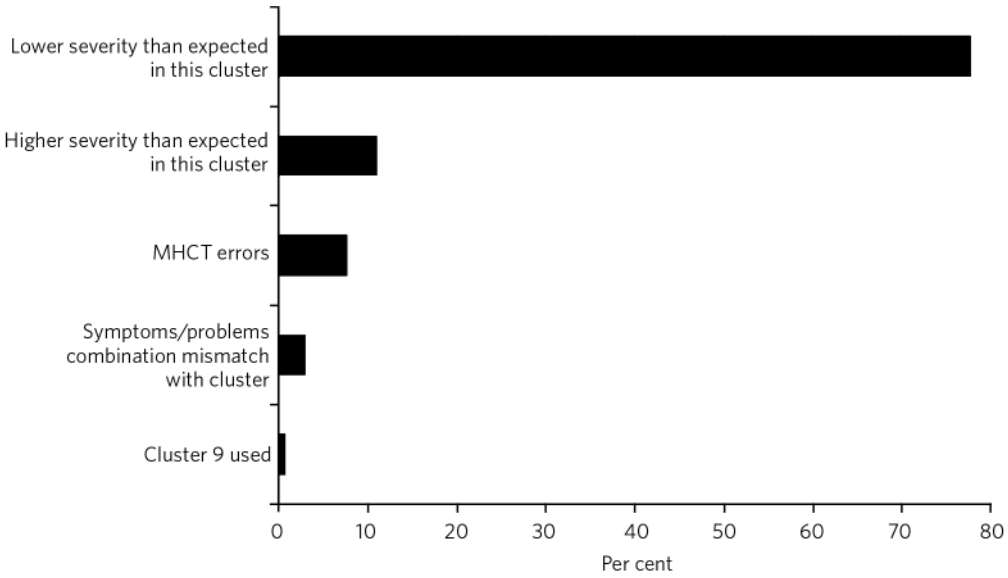

Fig 5 Reasons for mismatch between Health of the Nation Outcomes Scales and Mental Health Clustering Tool (MHCT) scores and cluster.

https://doi.org/10.1192/pb.bp.112.041780 Published online by Cambridge University Press

recording of clinical outcomes, and may encourage clinical involvement in financial management.[16]

Main findings

Some of the concerns mentioned have been replicated in our study, which provides evidence that in the real clinical world, the methodology of clustering as currently implemented presents significant weaknesses. The study provided evidence of potentially unacceptable levels of inappropriate allocation to care clusters, raising concerns about the readiness to fully implement the next stages of mental health PbR. The results are not significantly worse than those shared by other providers in London and the above concern is raised throughout the network of providers trying to implement PbR. As a result, we have implemented a series of recommendations for improved practice, including intensive clustering and care transition protocol training,[3] a dedicated clustering policy, guidance and clinical protocols, and improved IT systems and exception reports flagging breaches.

Differences between services can be partly explained by the nature of the work done and for which common disorders they are offered. The Hammersmith & Fulham Assessment Team presented very high mismatch, a finding that failed to support the hypothesis that clustering would be at its most precise nearer to the initial assessment of a patient, at least in the West London Mental Health NHS Trust. The Cassel Hospital is a service for patients with personality disorder and the inconsistent results reflect the difficulty in clustering these cases. Conversely, in-patient services appeared to be more accurate, but this could be explained by the fact that the majority of in-patients have psychotic disorders, which are more straightforward to allocate. Both the in-patient and Cassel Hospital results may also be less reliable because of low numbers.

This study provides evidence of how imprecise the allocation of patients with personality disorder is, but unfortunately limited access to detailed records prevented further explanations. It is possible that common practice is to reserve a diagnosis of personality disorder for the most extreme cases and instead give a diagnosis of affective disorder. Similar issues with failure to record all the applicable diagnoses explain most of the mismatch for the comorbid conditions.

The study also provides evidence that the results of the validation varied greatly depending on the methodology used. The vast majority of errors picked by the automated cross-check with ICD-10 codes proved to be genuine misallocations and were not justified on clinical grounds. When compared against HoNOS/MHCT scores, misallocations presented lower rates. This partly reflects the moderate discriminating properties of the MHCT in isolation, as the score profiles for many clusters are more or less identical. Also, many mismatches were because of lower-than-expected scores and could be due to fluctuation of severity during a period of care. False positives were also found; many of the apparent mismatches were not actually wrong allocations when examined clinically using diagnoses and more information. This finding reflected weaknesses in the algorithm and accordingly reflected inherent errors in the design of the MHCT. The latter includes incompatibilities with the ICD-10 diagnostic system, i.e. the requirement for delusions and hallucinations to be allocated to a 'psychotic' cluster and the allocation of bipolar disorder together with schizophrenia, which can be a source of confusion for practising clinicians. Some misallocations in cluster 16 appeared to be due to a similar confusion inherent in the concept of 'dual diagnosis' in the MHCT, which is narrowly defined as comorbidity of psychosis with drug and alcohol problems.

These results provide evidence that as a result of the inherent design of the MHCT, scores alone cannot be used for error-free, reliable allocation to clusters and corresponding care packages. Conversely, a more holistic and iterative, rather than a linear and algorithmic, approach would be more reliable. Lately, an increasing number of voices have called for the inclusion of diagnosis and care package descriptions as a more accurate approach to costing.[17] Diagnostic systems such as the ICD-10 have been developed after many years of rigorous research and international consensus, and have proven classification properties. Diagnostic labels tend to remain the same over time. They convey a great deal of clinically useful information and are already widely recorded in data-sets in trusts and provided to commissioners and the Department of Health. An additional argument for the inclusion of diagnosis in PbR is that current training, research, evidence-based guidelines, such as those by the National Institute for Health and Care Excellence, and service configurations so far have been primarily based on diagnoses. It is stressed within the principles of the clustering guidance that the final arbiter in allocating a PbR cluster should always be clinical judgement,[2,3] which can be better informed by a combination of diagnosis, MHCT and other rating tool scores, and the best match from an array of available packages of care.

This methodology or similar could be integrated in governance systems to provide automated, iterative validation audits and reports to track quality. However, it is apparent that both methods used in our study have limitations. The use of ICD-10 diagnosis appears more reliable than using the HoNOS/MHCT scores to audit practice. On the other hand, comparing against diagnostic codes is not validation in the classic sense, as there is no evidence that recorded ICD-10 codes are fully reliable. The absence of a true 'gold standard' in this case makes necessary the use of comparison and triangulation methods as those described here, preferably using the combination of both ICD-10 codes and HoNOS/MHCT scores as the most rigorous approach in auditing cluster allocation by clinical staff. For the same reasons, such auditing will, by necessity, have significant imprecision, and a margin of error for any automated, algorithmic methodology should be accepted and communicated to commissioners and other interested parties. Future work could potentially reveal where this acceptable margin of error lies.

Acknowledgements

We would like to thank Bijan Zainudini and Farinaz Mazhari for programming the data collection and analysis, and Dr Sam Nayrouz for his managerial support.

https://doi.org/10.1192/pb.bp.112.041780 Published online by Cambridge University Press

About the authors

Dr Stavros Bekas, MA, MSc, MRCPsych, FHEA, West London NHS Mental Health Trust, London, UK. **Dr Orlin Michev**, Central and North West London NHS Foundation Trust, London, UK.

References

1 Self R, Painter J, Davis R. *A Report on the Development of a Mental Health Currency Model.* Department of Health, 2008.

2 Department of Health. *Mental Health Payment by Results Guidance for 2013–14.* Department of Health, 2013.

3 Department of Health. *Mental Health Clustering Booklet v2.01 2011/12.* Department of Health, 2011.

4 Department of Health. *Equity and Excellence: Liberating the NHS.* Department of Health, 2010.

5 World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Description and Diagnostic Guidelines.* WHO, 1992.

6 The Sainsbury Centre for Mental Health. *Payment by Results: What Does it Mean for Mental Health? Policy Paper 4.* SCMH, 2004.

7 Self R, Painter J. *Study: To Improve and Demonstrate the Structural Properties of the Care clusters that form the basis of the PbR Currency Development Programme.* Care Pathways and Packages Project, 2009.

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

9 Tulloch AD. Care clusters and mental health Payment by Results. *Br J Psychiatry* 2012; **200**: 161.

10 Buckingham W, Burgess P, Solomon S, Pirkis J. Eagar K. *Developing a Casemix Classification for Mental Health Services. Volume 1: Main Report.* Commonwealth Department of Health and Family Services, 1998.

11 Cotterill PG, Thomas FG. Prospective payment for Medicare inpatient psychiatric care: assessing the alternatives. *Health Care Financ Rev* 2004; **26**: 85–101.

12 Health and Social Care Information Centre. *Casemix Service. Mental Health Casemix Classification Development: End Stage Report.* Health and Social Care Information Centre, 2006.

13 Audit Commission. *Maximising Resources in Adult Mental Health.* Audit Commission, 2010.

14 Self R, Rigby A, Leggett C, Paxton R. Clinical Decision Support Tool: a rational needs-based approach to making clinical decisions. *J Ment Health* 2008; **17**: 33–48.

15 Green C, Daniel D. *Payment by Results Quality and Outcomes Indicators: Report for Product Review Group Quality & Outcomes Sub Group.* Department of Health, 2011.

16 Macdonald AJD, Elphick M. Combining routine outcomes measurement and 'Payment by Results': will it work and is it worth it? *Br J Psychiatry* 2011; **199**: 178–9.

17 Kingdon DG, Solomka B, McAllister-Williams H, Turkington D, Gregoire A, Elnazer H, et al. Care clusters and mental health Payment by Results. *Br J Psychiatry* 2012; **200**: 162.

https://doi.org/10.1192/pb.bp.112.041780 Published online by Cambridge University Press