Report on the development of an algorithm to provide decision-support to clinicians when allocating to clusters.

Background

The Mental Health Payment by Results (PbR) system is centred on the clinical allocation of service users to a need-based classification system. Service users continue to be assessed according to their clinical presentation and treatment setting (i.e. according to local custom and practice) however this information is then summarised in a standardised format using the mental health Clustering Tool (MHCT). Severity ratings from 0-4 across the range of needs captured by the tool are then used to allocate service users to a single, best-fit cluster. Each cluster describes a group of patients with a particular combination and severity of needs.

This process has a high degree of transparency but professional judgement remains integral to the process and the final cluster allocation requires a clinical decision. The need for robust cluster allocation is important as the resources allocated to each cluster, will reflect the level of patient complexity and clinical input associated with each cluster. To support this process the Department of Health and the CPPP Consortium commissioned a piece of statistical analysis to produce an algorithm which describes at the first clustering assessment how well any combination of MHCT scores fit each viable cluster.

It is important to note that this development is **not** intended to automate the process of allocating a service user to a cluster, **or** to provide a definitive cluster for a particular set of MHCT scores. Instead, the algorithm should be seen primarily as decision-support for clinicians at the point of cluster allocation. There is the potential to apply the algorithm to data retrospectively, and this is likely to be of great interest to both commissioners and providers alike. This should however only be considered appropriate at an organisational level **not** at the individual patient level.

Overview of work to date

To obtain a high quality set of cluster data from which to develop the algorithm a multi-disciplinary group of staff from across the country attended a one-day workshop run jointly by the Royal College of Psychiatrists (RCPsych) and the Care Pathways and Packages Project (CPPP) in April 2011. The workshop was designed to improve the accuracy of this set of clinicians in both their rating of the individual scales in the tool, and the allocation to cluster. Following the workshops, first cluster assessments undertaken by the clinicians that attended have been submitted centrally and used to start developing the algorithm.

The development process has involved a number of stages:

1. Data cleansing.
   In order to increase confidence, only the ratings from new referrals into services were used. (NB There are some exceptions where, by definition allocation of new patients to particular clusters is highly unlikely. (See Appendix 1). Allocations not meeting the scoring range for the "must have" scales (as identified by red score ranges in the cluster booklet) were also excluded.

2. Feasibility study.
   Initial testing of the model was undertaken to ascertain the feasibility of producing a proportional membership model based on the mandated data items (i.e. MHCT scores). The initial analysis indicated that the clusters demonstrated robust properties that made the creation of a set of algorithm viable. This was significant as the clusters were not created through statistical methods alone, and the revisions made to ensure the clusters were also clinically meaningful could have adversely affected the model's statistical properties.

3. Refinement
   The initial algorithm was refined to take account of known clinical issues (e.g. the flooring effect of the tool, the marginal importance of certain MHCT scales to cluster membership for particular super classes etc.). Given the low prevalence of some clusters encountered by workshop attendees, this data was then augmented with data of an adequate quality from the CPPP data warehouse until sufficient data was captured for full analysis. At this point close attention was given to the scoring distributions for each cluster and a further 2% of assessments were deemed outliers and hence excluded. This resulted in an optimised level of agreement between algorithm and clinical cluster allocation on the original higher-quality block of data (See appendix 2).

4. Cross validation.
   Although retrospective use of the algorithm was not its primary purpose, it was also important to understand how the proposed algorithm performed on other blocks of data. Despite the obvious weaknesses in using data of variable quality from Q1, 2 and 3 2011/12 the Mental Health minimum Data Set (MHMDS), data held by the Information Centre (IC) was an obvious source of large volumes of multi-provider data. Appendix 3 sets out the analysis undertaken by the DH on MHMDS data and the findings. The results of this stage are difficult to interpret as low agreement rates could be a positive outcome if clustering is poor. However, by considering the correlations between the algorithm and clinical agreement rates, together with other variables (e.g. provider organisation, adherence to the cluster booklet's red rules) additional insight can be gained.

Outputs of the sub-group

Appendix 4 contains the algorithm produced through the process described above. This is in excel-format and is available for use by any organisation that is willing and able to do so. At this stage uptake may be limited to organisations able to make changes to their electronic patient record systems or wishing to use the algorithm retrospectively on data extracted from their systems and / or data warehouses. It will therefore also be made available via the CPPP website (www.cppconsortium.nhs.uk). This will allow clinicians to enter sets of scores and see the likelihood of cluster membership. It will also give system providers a better understanding of how the algorithms should behave in practice. A 'soft launch' will provide a useful road testing phase and allow further feedback to be gathered from staff using the algorithm in practice before any process of mandation is commenced.

Next Steps

1. Obtain approval from the Product Review group and the Project Board to make the algorithm available.
2. Build a version of the algorithm on the CPPP website that allows scores to be entered and results to be viewed without requiring any patient identifiable data to be entered.
3. Allow organisations to road test the algorithm and feedback on its use (particularly the strict application of the 'red rules'.
4. Undertake any amendments required following feedback from the roadtest period.
5. Commence the mandation process as and when clinical utility has been confirmed.

Jon Painter (CPPP)
Sue Nowak (DH PbR)
22nd Oct 2012

Appendix 1: Table showing exceptions to the new patient sampling

From the clustering data submitted only assessments carried out at certain points were used in developing the algorithm. This was to avoid some of the confounding variables that may be present when re-allocating patients to clusters at the point of clinical reviews. For most clusters only new referrals were used. However, there are a number of clusters where the allocation of new service users to them is highly unlikely.

The following table outlines which assessments were selected:

Clinical cluster allocation	New referral or First Allocation to this cluster at review	Clusters from which reviewed patients can transition
1	New referral	
2	New referral	
3	New referral	
4	New referral	
5	New referral	
6	New referral	
7	1st allocation*	1,2,3,4,5,6,
8	1st allocation*	1,2,3,4,5,6,7
10	New referral	
11	New referral	
12	New referral	
13	New referral	
14	New referral	
15	New referral	
16	1st allocation*	10-15, 17
17	1st allocation*	10, 11, 12, 13, 14, 15
18	New referral	
19	New referral	
20	New referral	
21	1st allocation*	18, 19, 20

Appendix 2: Box Clever statistical report

Our overarching analytical objective was to develop a best practice, statistically driven, Mental Health Clustering Framework that serves to consolidate the approach taken across Trusts and improves on the accuracy of existing cluster allocation tools.

In order to achieve this aim, we broke our analysis down into stages.

Stage 1

The first stage involved using exploratory statistical modelling to ascertain whether a suitably strong statistical relationship exists between the scores that clinicians gave patients using the MHCT handbook criteria and their ultimate cluster allocation. If we found no, or a very small statistical relationship then we could quickly conclude that the data is not robust enough for statistical modelling whereas if we found a strong statistical relationship we would then move forward and invest the analytical effort in developing a statistical algorithm.

The final sample of clinician records was collected from the period of April 2011 to August 2011from participating Trusts. In total, we collected 1241 patient assessments from our analysis as outlined in table 1 below.

Trust / Consortium	Frequency
A	64
B	84
C	1033
D	29
E	31
Total	**1241**

Table 1

We used a statistical modelling technique called Linear Discriminant Analysis to measure the accuracy with which we can predict patient cluster allocation if we know the scores they have been given for each of the MHCT items.

We made the reasonable assumption that we also know the super cluster to which the patient is to be allocated and constructed one Discriminant model per super cluster (Non-psychotic, Psychosis, Organic).

The overall accuracy of these Discriminant Models in predicting cluster membership for each of these super clusters are given in table 2 below. This

table quantifies the proportion of times our statistical models suggested patient allocation agrees with the expert clinicians judgement.

Super cluster	Statistical Model Accuracy
Non-psychotic	35%
Psychosis	56%
Organic	63%

Table 2

Whilst we agreed that the levels of accuracy achieved at this stage weren't sufficient, we did determine that a statistical relationship existed and that through further model refinement we could achieve a strong level of fit. This in turn provides evidence to support the validity of the clusters themselves.

Therefore we proceeded to stage 2 of the modelling process – model refinement.

Stage 2

In phase 2 we aimed to refine the basic algorithm developed in stage 1 in order to enhance its accuracy. We also aimed to overlay the 'red rules' outlined in the MHCT handbook that prevent patients from being allocated to a cluster if they don't possess certain scores on certain items to see if this enhanced accuracy.

The first step in model refinement involved exploring whether the inclusion of additional variables above and beyond the original 18 MHCT items could help to improve the accuracy of our algorithm. At this stage, we did not introduce any entirely new variables to our model. Rather we looked at whether any "derived" variables, calculated from our original 18 MHCT items could serve to increase algorithm accuracy. Specifically, we explored whether the following variables (which were deemed to be clinically meaningful) enhanced our algorithm:

- A variable that was calculated by taking the highest score given to items 7 or 8
- A variable that was calculated by taking the highest score of items 2, A & E
- A variable that was calculated as an average of items 5, 10, C, E & 9

We found that the incorporation of these additional variables marginally enhanced algorithm accuracy by an average of 3%.We didn't deem this step to be sufficient a refinement.

The second step we took in model refinement was to break each of the original 18 MHCT items into four separate dichotomous variables – one variable per point on the item rating scale where a value of "1" was coded if

the patient achieved that specific score and a value of “0” was coded if the patient was given another score. The rationale for doing this was based on a suggestion that a higher score on each of the rating scales indicated an increasing escalation of a patients symptoms. In statistical terms, we allowed the model greater flexibility to fit the data by including one variable for each score level on each MHCT item. This allowed our model to explore non-linear relationships between a patient’s MHCT score and their cluster membership.

We found that adjustment to the functional form of our Discriminant Model dramatically improved modelling accuracy as demonstrated in table 3 below:

Model	Accuracy after stage 1	Accuracy after non-linear scaling
Non-psychotic	35.4%	69.2%
Psychosis	55.6%	82.7%
Organic	62.7%	80.2%

Table 3

The final stage we took in our modelling refinement process was to overlay the existing MHCT Handbook “Red Rules”.

The existing MHCT handbook stipulates scoring patterns that must occur in order for a patient to be eligible for each of the clusters.By overlaying these rules onto the predictions made by the statistical algorithm and using them to “clean” the results we can ensure that the final algorithm never predicts a cluster that can’t occur (according to the MHCT handbook).This stage should also enhance the accuracy of the algorithm in correctly predicting cluster membership.

By overlaying the red rules, we found a number of clinician records were no longer valid due to the fact that the clusters they had allocated patients to broke the rules stipulated in the MHCT handbook. As such our overall sample size for analysis purposes was reduced to 919 as illustrated in table 4 below:

	Total Sample	Sample that adheres to rules	Sample that doesn't adhere to rules
1	30	18	12
2	33	18	15
3	119	83	36
4	152	118	34
5	71	38	33
6	37	16	21
7	38	7	31
8	56	14	42
10	41	39	2
11	54	40	14
12	45	22	23
13	31	25	6
14	40	32	8
15	44	43	1
16	45	42	3
17	54	51	3
18	104	99	5
19	141	134	7
20	41	33	8
21	48	47	1
Total	**1224**	**919**	**305**

Table 4

Having overlaid the Red Rules we used our sample to assess how well the new algorithm, including Red Rule, predicted segment membership. Tables 5, 6 and 7 below are the classification matrices that resulted from this analysis. They indicate the frequency and % of patient records that were correctly and incorrectly allocated and, if incorrectly allocated which segment the algorithm allocated them to.

Non-psychotic		Predicted Group Membership								Total
		1	2	3	4	5	6	7	8	
Count	1	**15**	3	0	0	0	0	0	0	**18**
	2	3	**15**	0	0	0	0	0	0	**18**
	3	0	0	**83**	0	0	0	0	0	**83**
	4	0	0	0	**114**	0	3	0	1	**118**
	5	0	0	0	0	**37**	1	0	0	**38**
	6	0	0	0	0	0	**16**	0	0	**16**
	7	0	0	0	0	0	1	**6**	0	**7**
	8	0	0	0	0	1	0	0	**13**	**14**
%	1	**83%**	17%	0%	0%	0%	0%	0%	0%	100%
	2	17%	**83%**	0%	0%	0%	0%	0%	0%	100%
	3	0%	0%	**100%**	0%	0%	0%	0%	0%	100%
	4	0%	0%	0%	**97%**	0%	3%	0%	1%	100%
	5	0%	0%	0%	0%	**97%**	3%	0%	0%	100%
	6	0%	0%	0%	0%	0%	**100%**	0%	0%	100%
	7	0%	0%	0%	0%	0%	14%	**86%**	0%	100%
	8	0%	0%	0%	0%	7%	0%	0%	**93%**	100%

Table 5

Psychosis		Predicted Group Membership								Total
		10	11	12	13	14	15	16	17	
Count	10	**33**	1	3	1	1	0	0	0	**39**
	11	0	**38**	1	0	0	0	0	1	**40**
	12	3	2	**15**	0	0	1	1	0	**22**
	13	1	0	0	**23**	0	1	0	0	**25**
	14	3	0	0	0	**25**	2	1	1	**32**
	15	0	0	0	0	0	**43**	0	0	**43**
	16	0	0	0	0	0	0	**42**	0	**42**
	17	1	0	1	1	1	0	1	**46**	**51**
%	10	**85%**	3%	8%	3%	3%	0%	0%	0%	100%
	11	0%	**95%**	3%	0%	0%	0%	0%	3%	100%
	12	14%	9%	**68%**	0%	0%	5%	5%	0%	100%
	13	4%	0%	0%	**92%**	0%	4%	0%	0%	100%
	14	9%	0%	0%	0%	**78%**	6%	3%	3%	100%
	15	0%	0%	0%	0%	0%	**100%**	0%	0%	100%
	16	0%	0%	0%	0%	0%	0%	**100%**	0%	100%
	17	2%	0%	2%	2%	2%	0%	2%	**90%**	100%

Table 6

Organic		Predicted Group Membership				Total
		18	19	20	21	
Count	18	**89**	10	0	0	99
	19	23	**102**	5	4	134
	20	0	5	**27**	1	33
	21	0	4	6	**37**	47
%	18	**90%**	10%	0%	0%	100%
	19	17%	**76%**	4%	3%	100%
	20	10%	15%	**82%**	3%	100%
	21	0%	9%	13%	**79%**	100%

Table 7

Overlaying the Red Rules resulted in an uplift in model accuracy as quantified in table 8 below.

Model	Accuracy after non-linear scaling	Accuracy after applying the Red Rules
Non-psychotic	69.2%	95.8%
Psychosis	82.7%	90.1%
Organic	80.2%	81.5%

Table 8

Stage 3

Upon review of the results from stage 2, although the model accuracy levels were high, it became apparent that the sample sizes we had used to construct the algorithms were not quite sufficient and that scope existed to use a larger data set to refine the algorithms further.

This larger data set was sourced from CPPP and we undertook the following steps to ensure that the sample provided was of a high enough quality to be used within the modelling process.

1. As with stage 2, we applied the “Red Rules” to ensure that all clinician records adhered to the rules stipulated within the MHCT handbook.
2. We conducted analysis exploring the distribution of scores given to patients within each cluster in order to identify and remove 2% of cases which were obvious outliers.

By applying the exclusion rules outlined above, we identified 14,842 cases to feed into stage 3 algorithm development as detailed in table 9 below:

Cluster	Total	Valid Cases	Excluded Cases
1	348	348	0
2	369	369	0
3	1778	1778	0
4	1186	1186	0
5	58	58	0
6	71	70	1
7	426	417	9
8	55	49	6
10	100	100	0
11	1376	1341	35
12	882	649	233
13	284	284	0
14	130	130	0
15	52	43	9
16	52	42	10
17	61	37	24
18	6645	6645	0
19	1202	1202	0
20	48	47	1
21	64	47	17

Table 9

We then applied linear Discriminant Analysis to this data set to build our predictive models of cluster membership. Tables 10, 11 and 12 below are the classification matrices that resulted from this analysis.

Non-psychotic		Predicted Group Membership								Total
		1	2	3	4	5	6	7	8	
Count	1	237	111	0	0	0	0	0	0	348
	2	151	218	0	0	0	0	0	0	369
	3	0	0	1774	0	0	0	4	0	1778
	4	0	0	0	1173	0	0	11	2	1186
	5	0	0	0	0	55	1	0	2	58
	6	0	0	0	0	5	58	6	1	70
	7	0	0	4	17	0	6	390	0	417
	8	0	0	0	0	0	0	0	49	49
%	1	**68.1%**	31.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100%
	2	40.9%	**59.1%**	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100%
	3	0.0%	0.0%	**99.8%**	0.0%	0.0%	0.0%	0.2%	0.0%	100%
	4	0.0%	0.0%	0.0%	**98.9%**	0.0%	0.0%	0.9%	0.2%	100%
	5	0.0%	0.0%	0.0%	0.0%	**94.8%**	1.7%	0.0%	3.4%	100%
	6	0.0%	0.0%	0.0%	0.0%	7.1%	**82.9%**	8.6%	1.4%	100%
	7	0.0%	0.0%	1.0%	4.1%	0.0%	1.4%	**93.5%**	0.0%	100%
	8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**100.0%**	100%

Table 10

Psychosis		Predicted Group Membership								Total
		10	11	12	13	14	15	16	17	
Count	10	**20**	1	27	29	8	4	9	2	100
	11	0	**1338**	3	0	0	0	0	0	1341
	12	0	28	**617**	0	0	1	3	0	649
	13	0	0	0	**266**	15	2	1	0	284
	14	2	0	0	40	**78**	1	1	8	130
	15	3	0	0	4	0	**36**	0	0	43
	16	0	0	2	0	0	0	**38**	2	42
	17	0	0	4	0	2	0	0	**31**	37
%	10	**20.0%**	1.0%	27.0%	29.0%	8.0%	4.0%	9.0%	2.0%	100%
	11	0.0%	**99.8%**	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	100%
	12	0.0%	4.3%	**95.1%**	0.0%	0.0%	0.2%	0.5%	0.0%	100%
	13	0.0%	0.0%	0.0%	**93.7%**	5.3%	0.7%	0.4%	0.0%	100%
	14	1.5%	0.0%	0.0%	30.8%	**60.0%**	0.8%	0.8%	6.2%	100%
	15	7.0%	0.0%	0.0%	9.3%	0.0%	**83.7%**	0.0%	0.0%	100%
	16	0.0%	0.0%	4.8%	0.0%	0.0%	0.0%	**90.5%**	4.8%	100%
	17	0.0%	0.0%	10.8%	0.0%	5.4%	0.0%	0.0%	**83.8%**	100%

Table 11

Organic		Predicted Group Membership				Total
		18	19	20	21	
Count	18	**6409**	236	0	0	6645
	19	52	**1141**	4	5	1202
	20	0	14	**31**	2	47
	21	0	1	11	**35**	47
%	18	**96.4%**	3.6%	0.0%	0.0%	100%
	19	4.3%	**94.9%**	0.3%	0.4%	100%
	20	0.0%	29.8%	**66.0%**	4.3%	100%
	21	0.0%	2.1%	23.4%	**74.5%**	100%

Table 12

Table 13 below summarises the overall accuracy of the algorithm at stage 3 compared to stage 2. Although the accuracy has fallen slightly for Non-psychotic it has risen slightly for Psychosis and Organic and as such we deem these algorithms to be more robust and accurate at an overall level.

Model	Algorithm accuracy after stage 2	Algorithm accuracy after stage 3
Non-psychotic	95.8%	92.5%
Psychosis	90.1%	92.3%
Organic	81.5%	95.9%

Table 13

Appendix 3: Results from the application of the algorithm to MHMDS data

Initial Data Selection 1

- The Mental Health Minimum Dataset (MHMDS) contained **493,000** records during the period for clustering assessments (there are actually nearer 740,000, but the figure above is those which upon submission have been flagged by the IC as containing all the required fields of information).

- Of the **493,000** only **58,037** were identified as initial clustering assessments with all appropriate data in place. The initial assessment element was identified by the date of the clustering event being within 30 days of the start date recorded for patient care in that spell.

Initial Data Selection 3

- The development of the clustering tool has been based on the assumption that the 'red rules' defined in the Mental Health Clustering booklet. As such it is only appropriate to test the accuracy of clustering assessment which adhere to these rules.

- In the **58,037** potential initial clustering assessments, **35,099** adhered to the rules.

Summary of Reductions in Clustering Assessments Used in Analysis

Clinical Cluster	Valid Initial Assessments	Passed Red Rules	% Retained for Analysis
1	4,941	1,817	37%
2	4,836	850	18%
3	9,582	4,484	47%
4	5,660	3,715	66%
5	1,571	800	51%
6	939	418	45%
7	1,126	398	35%
8	1,444	468	32%
10	2,518	2,196	87%
11	2,117	1,202	57%
12	1,189	600	50%
13	994	698	70%
14	1,377	958	70%
15	624	409	66%
16	666	287	43%
17	697	312	45%
18	6,969	5,803	83%
19	6,813	6,274	92%
20	2,585	2,209	85%
21	1,389	1,201	86%
Total	**58,037**	**35,099**	**60%**

Clustering Validation

Non-Psychotic		Predicted Cluster Membership							
		1	2	3	4	5	6	7	8
Clinical Cluster	1	77%	23%	0%	0%	0%	0%	0%	0%
	2	64%	36%	0%	0%	0%	0%	0%	0%
	3	0%	0%	96%	0%	0%	0%	4%	0%
	4	0%	0%	0%	82%	0%	0%	9%	9%
	5	0%	0%	0%	0%	88%	7%	0%	5%
	6	0%	0%	0%	0%	6%	43%	42%	9%
	7	0%	0%	1%	8%	0%	5%	77%	10%
	8	0%	0%	0%	3%	7%	0%	0%	90%

• On average, the tool predicts **82%** of the 12,590 non-psychotic clustering assessments accurately.

•The tool appears to have difficulty distinguishing between assessments for clusters 1 and 2. This is particularly seen in the allocation to cluster 1 by the tool of those clinically assessed as cluster 2.

•It also appears to have difficulty placing those clinically assessed as being in cluster 6, placing roughly equal amounts in cluster 6 and cluster 7.

Clustering Validation

Psychotic		Predicted Cluster Membership							
		10	11	12	13	14	15	16	17
Clinical Cluster	10	6%	6%	30%	35%	6%	9%	6%	2%
	11	1%	86%	12%	0%	0%	0%	0%	2%
	12	1%	8%	76%	0%	0%	4%	3%	7%
	13	2%	0%	0%	53%	21%	8%	7%	9%
	14	7%	0%	0%	50%	21%	8%	11%	3%
	15	6%	0%	2%	17%	6%	63%	3%	3%
	16	13%	0%	11%	5%	4%	4%	60%	4%
	17	2%	7%	23%	13%	20%	5%	7%	23%

• On average, the tool predicts **40%** of psychotic clustering assessments accurately.

•The results here somewhat mimic those experienced when the algorithm was run on the expert data. Clinically assigned cluster 10 patients were in both datasets predicted to be in a wide variety of clusters. There is though a far lower correct prediction rate for cluster 17, but the initial sample of expert data here was very small.

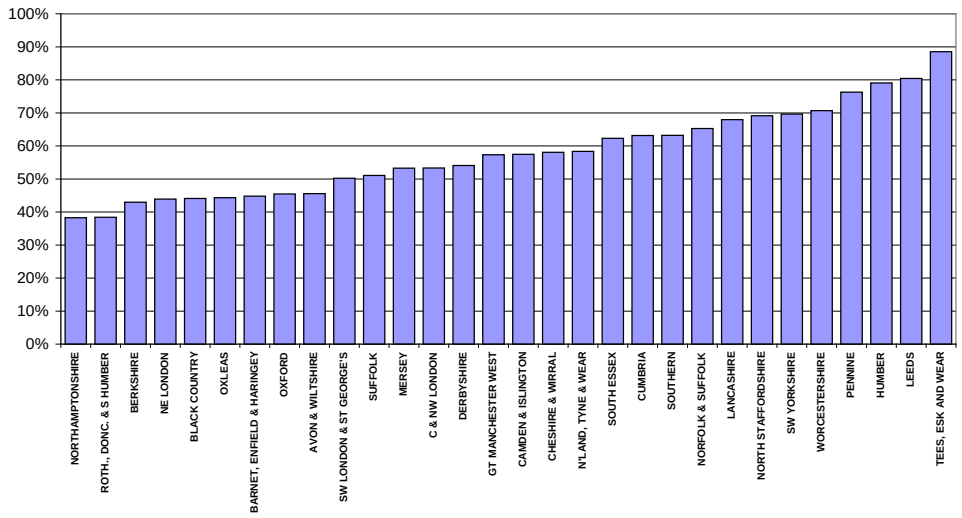
Clustering Validation

Organic		Predicted Cluster Membership			
		18	19	20	21
Clinical Cluster	18	86%	13%	0%	1%
	19	27%	63%	5%	4%
	20	0%	43%	40%	17%
	21	4%	26%	29%	41%

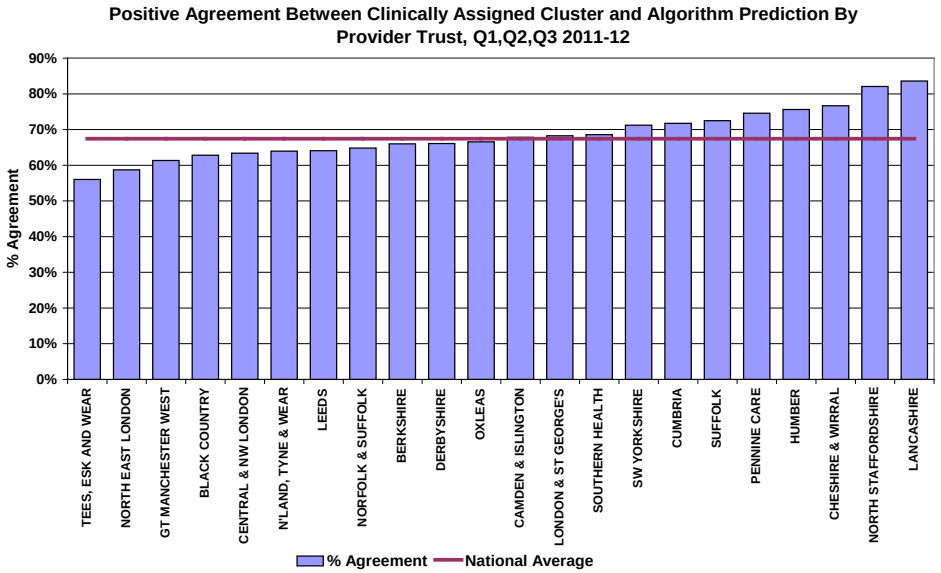
• On average, the tool predicts **67%** of organic clustering assessments accurately.

Red Rule Adherence

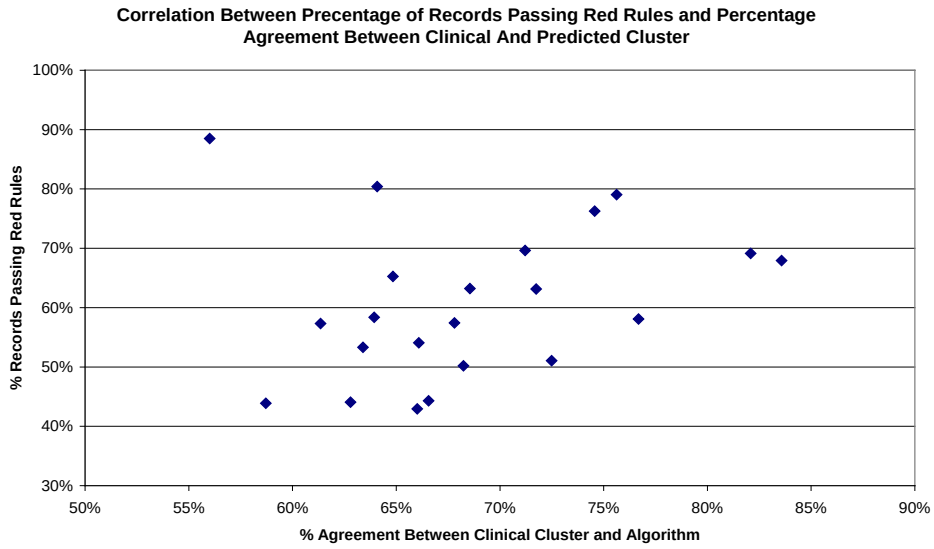
Percentage of Eligible Initial Clustering Episodes Which Agree With Red Rules, Q1 Q2 Q3 2011-12, By Provider



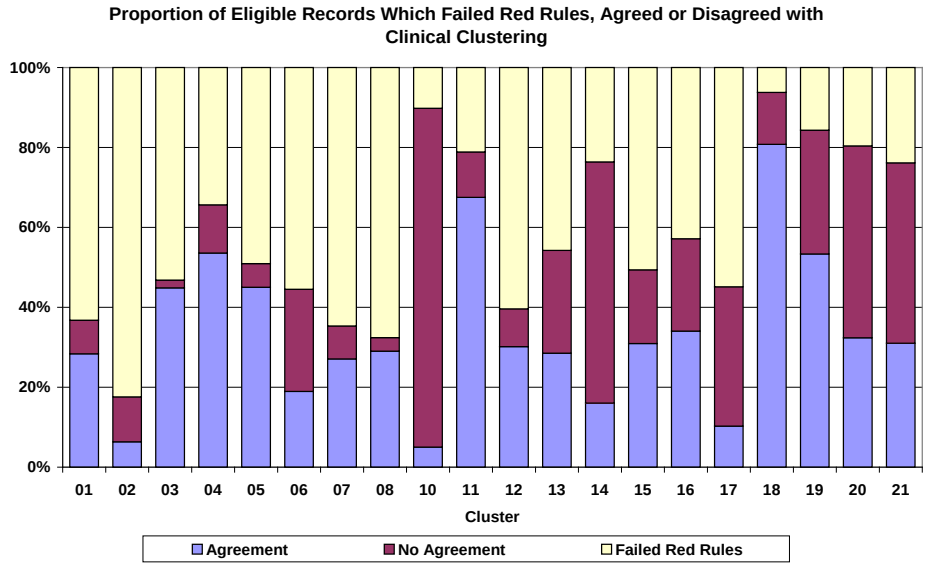
Agreement Rates



Comparison of Red Rules and Agreement Rates



Summary of Agreement with Red Rules and Clustering by Cluster



Appendix 4

Embedded below are two excel versions of the algorithm. One is in an older version of excel to make these accessible to as many organisations as possible.

MHCT algorithm V3.0.xlsx

MHCT algorithm V3.0.xls

Please note these are not intended for use in their current format by clinicians. Instead the CPPP website (www.cppconsortium.nhs.uk) is planning to host an example of how the tool might appear to front-line practitioners.