Commentary on the proposed adult mental health PbR cluster decision support algorithm (internal group discussion note)

Dr Andy Fugard[1], Research Fellow
CAMHS Evidence Based Practice Unit
University College London and Anna Freud Centre
Version of 7th November 2012

This is a short note – originally for discussion in the CAMHS PbR group on Monday 5th November – which summarises my understanding of the algorithm, identifies areas where more information would be helpful to support clearer understanding of the tool, notes possible challenges raised by the approach taken, and presents ideas for possible areas for future exploration. This tool is clearly important for Adult PbR and provides stimulating input to the developing CAMHS approach.

I understand that the team working on the adult mental health PbR cluster decision support algorithm are inviting rapid feedback and I am therefore sharing this internal document in the spirit of wanting to share learning. *These are my own personal views of the approach and aren't necessarily shared by others in the CAMHS PbR group.*

Brief summary of algorithm

The algorithm was developed using a dataset consisting of

1. Clinicians' Mental Health Clustering Tool (MHCT) responses.
2. Clinicians' cluster choice.

Linear Discriminant Analysis was used to produce a set of equations predicting the probability of cluster membership as a function of the MHCT responses.

A predictor was added for each possible response for each item, i.e., for each of the 18 items of the MHCT, 5 predictors were added for the five levels of severity:

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

For each cluster there are therefore 18 x 5 = 90 predictors, plus a constant term giving 91 predictors per cluster. There are 20 clusters (cluster 9 was dropped), giving 1,820 predictors, of which 1,204 have non-zero coefficients.

[1] Email: andy.fugard@annafreud.org

See Appendix 1 for an example of one of the equations and an illustration of how the equations are used to predict cluster membership (I didn't use the red-rule constraints as I wanted to understand how the underlying statistical model is working).

Areas where more information would be helpful about the approach taken

Specification of the statistical model

1. There are different versions of Linear Discriminant Analysis. What was the exact form of the model fitted and method used to estimate the coefficients? This explanation could take the form of a formula, and a sentence or two summarising the approach.

2. What was the precision of the estimates, e.g., are standard errors available?

3. Why, before computing the exponential, are the Discriminant Fisher Scores divided by 10 for the first two super-clusters and by 100 for the third Organic cluster? Presumably this is to ensure that the numbers are computable in Excel. These seem commonly to be very large numbers in the model. The example in the Appendix computed for cluster 21 is approximately 616,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000, which is unusually large for computational implementations of statistical models.

4. Why are 616 of the coefficients exactly 0? Were they fixed to zero as a result of model simplification (in effect the predictor is dropped from the model) or did they come out that way? If the former then how was model selection done?

5. No interactions were allowed between the predictors. It seems possible that this type of non-linearity would be crucial to allow the probability of cluster membership to decrease as well as increase, as more information is added. Details of any attempts to test for interactions would be helpful.

Information used to input to algorithm

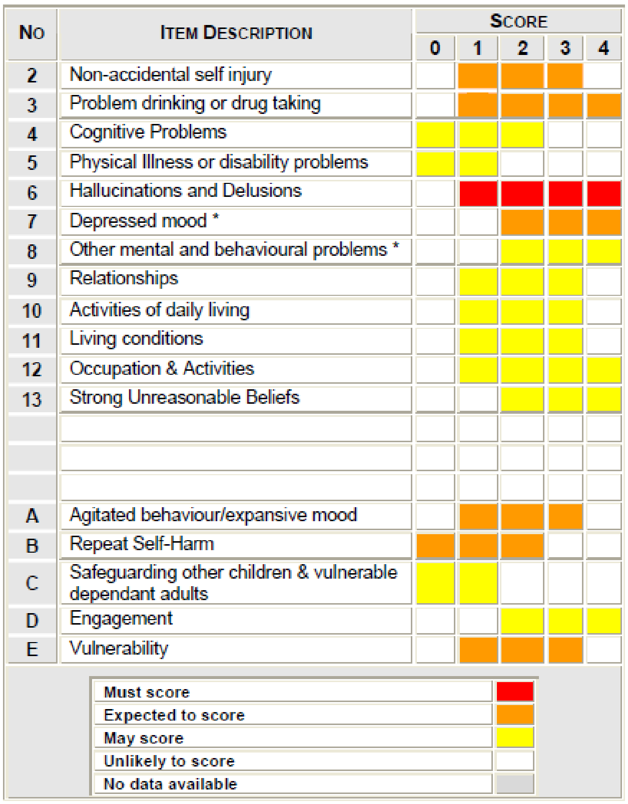
That clinicians can disagree with the red-rule guidance suggests that there's additional information not included in the MCHT which is necessary to guide cluster selection. For example diagnostic codes are included, "Likely to include F60 Personality disorder", "Likely to include (F20-F29) Schizophrenia, schizotypal and delusional disorders, Bi-polar disorder". It would be helpful to have more detail on how this information does and should guide decisions.

How the issue of overfitting was or could be addressed

The very large number of predictors (1,820) means there is a danger of overfitting, i.e., that the model will not generalise beyond the sample. This will be certainly true of the first dataset with 919 cases as the authors acknowledge, but also for the second dataset – especially for cluster B which consists of only ~1.5 cases per predictor in the model and to a lesser extent for cluster A with 2.4 cases per predictor.

Possible circularity of the approach

1. Clinicians were trained in a partial description of an algorithm, specified in a colour-coded booklet. See the image below for an example for Cluster 10. Red denotes that the score is required, orange denotes that it is likely, yellow denotes that it is possible, and white denotes that it is unlikely.



No	Item Description	Score 0	1	2	3	4
2	Non-accidental self injury					
3	Problem drinking or drug taking					
4	Cognitive Problems					
5	Physical Illness or disability problems					
6	Hallucinations and Delusions					
7	Depressed mood *					
8	Other mental and behavioural problems *					
9	Relationships					
10	Activities of daily living					
11	Living conditions					
12	Occupation & Activities					
13	Strong Unreasonable Beliefs					
A	Agitated behaviour/expansive mood					
B	Repeat Self-Harm					
C	Safeguarding other children & vulnerable dependant adults					
D	Engagement					
E	Vulnerability					

Must score	
Expected to score	
May score	
Unlikely to score	
No data available	

2. Clinicians attempted to use this to generate a cluster, using item responses in the MHCT booklet.

3. Responses were excluded from analysis if they didn't satisfy red-rules – as defined in the booklet. (From the first dataset, 25% of cases were excluded however from the second sourced from CPPP only 2% had to be excluded.)

4. A statistical model was used to predict clinicians' cluster choice from their MHCT selections – in order to generate an algorithm.

5. The end result is a statistical incarnation of the original guidance, giving most likely clusters chosen.

It would be extremely helpful to have more detail on where the initial cluster-selection guidance comes from, and justification as to why it is suitable as a basis for the present modelling efforts. Clearly it will also be crucial going forward to understand how these clusters relate to resource use.

Possible issues with the sample

One important application of the algorithm is to indicate the “prevalence” of clusters in adult mental health, conditional on the MHCT scores, so the representativeness of the sample is important. The first dataset used to fit the models had only 919 cases. A larger sample of 14,842 cases was sourced, however there is a clear over-representation of one cluster – Cognitive Impairment (Low Need) – so it is unclear how representative the sample is (see table below).

	Cluster	*% of cases*
1	Common Mental Health Problems (Low Severity)	2.3
2	Common Mental Health Problems (Low Severity with greater need)	2.5
3	Non Psychotic (Moderate Severity)	12.0
4	Non-psychotic (Severe)	8.0
5	Non-psychotic Disorders (Very Severe)	0.4
6	Non-psychotic Disorder of Over-valued Ideas	0.5
7	Enduring Non-psychotic Disorders (High Disability)	2.8
8	Non-Psychotic Chaotic and Challenging Disorders	0.3
10	First Episode Psychosis	0.7
11	Ongoing Recurrent Psychosis (Low Symptoms)	9.0
12	Ongoing or recurrent Psychosis (High Disability)	4.4
13	Ongoing or Recurrent Psychosis (High Symptom & Disability)	1.9
14	Psychotic Crisis	0.9
15	Severe Psychotic Depression	0.3
16	Dual Diagnosis	0.3
17	Psychosis and Affective Disorder – Difficult to Engage	0.2
18	Cognitive Impairment (Low Need)	44.8
19	Cognitive Impairment or Dementia Complicated (Moderate Need)	8.1
20	Cognitive Impairment or Dementia Complicated (High Need)	0.3
21	Cognitive Impairment or Dementia (High Physical or Engagement)	0.3

Ideas for possible areas to explore

1. Consider exploring, perhaps through interviews, what information clinicians are using to assign clusters. It seems plausible that there is something beyond the MCHT scores – but what exactly is influencing decisions?

2. Try multinomial logistic regression. This will also provide a set of equations which can predict cluster membership, and which could be embedded in a spread sheet. One advantage is that the response will be easier to convert to a probability. Of course there’s no guarantee that this will do any better, but it seems worth a try.

3. Allow interactions between predictors. It seems possible that this will be necessary so that adding additional information can also reduce the probability of cluster membership, depending on what other MHCT responses have been chosen.

4. Consider the application of this model to resource use.

Appendix 1: example of computation

The formula below shows how to compute (a value proportional to) the probability of membership in cluster 1. Each of the variables in the formula below is either 0 or 1. For instance *Item1_Score_0* is set to 1 if "no problem" is ticked for the item 1 on the MHCT.

$$
\begin{aligned}
&\exp((274.1261\times \mathrm{Item1\_Score\_0} + 231.374\times \mathrm{Item2\_Score\_0} + 74.79094\times \mathrm{Item3\_Score\_0} +\\
&1116.246\times \mathrm{Item4\_Score\_0} + 0\times \mathrm{Item5\_Score\_0} + 643.8571\times \mathrm{Item6\_Score\_0} + 135.7461\times \mathrm{Item7\_Score\_0}\\
&+ 219.093\times \mathrm{Item8\_Score\_0} + 429.2262\times \mathrm{Item9\_Score\_0} + 69.18559\times \mathrm{Item10\_Score\_0} +\\
&0\times \mathrm{Item11\_Score\_0} + 232.9812\times \mathrm{Item12\_Score\_0} + 42.61629\times \mathrm{Item13\_Score\_0} +\\
&43.61732\times \mathrm{ItemA\_Score\_0} + 334.1883\times \mathrm{ItemB\_Score\_0} + 0\times \mathrm{ItemC\_Score\_0} +\\
&122.8488\times \mathrm{ItemD\_Score\_0} + 203.1964\times \mathrm{ItemE\_Score\_0} + 278.323\times \mathrm{Item1\_Score\_1} +\\
&230.8871\times \mathrm{Item2\_Score\_1} + 68.66777\times \mathrm{Item3\_Score\_1} + 1118.597\times \mathrm{Item4\_Score\_1} + 0\times \mathrm{Item5\_Score\_1}\\
&+ 655.0367\times \mathrm{Item6\_Score\_1} + 282.4226\times \mathrm{Item7\_Score\_1} + 467.9079\times \mathrm{Item8\_Score\_1} +\\
&429.2671\times \mathrm{Item9\_Score\_1} + 72.16214\times \mathrm{Item10\_Score\_1} + 0\times \mathrm{Item11\_Score\_1} +\\
&232.2064\times \mathrm{Item12\_Score\_1} + 40.46893\times \mathrm{Item13\_Score\_1} + 39.20682\times \mathrm{ItemA\_Score\_1} +\\
&337.9674\times \mathrm{ItemB\_Score\_1} + 0\times \mathrm{ItemC\_Score\_1} + 103.0121\times \mathrm{ItemD\_Score\_1} + 201.9949\times \mathrm{ItemE\_Score\_1}\\
&+ 276.9076\times \mathrm{Item1\_Score\_2} + 233.1237\times \mathrm{Item2\_Score\_2} + 76.34718\times \mathrm{Item3\_Score\_2} +\\
&1120.471\times \mathrm{Item4\_Score\_2} + 0\times \mathrm{Item5\_Score\_2} + 587.0671\times \mathrm{Item6\_Score\_2} + 51.84121\times \mathrm{Item7\_Score\_2}\\
&+ 90.04503\times \mathrm{Item8\_Score\_2} + 431.6853\times \mathrm{Item9\_Score\_2} + 66.49924\times \mathrm{Item10\_Score\_2} +\\
&0\times \mathrm{Item11\_Score\_2} + 232.6356\times \mathrm{Item12\_Score\_2} + 14.69476\times \mathrm{Item13\_Score\_2} +\\
&39.19009\times \mathrm{ItemA\_Score\_2} + 349.5469\times \mathrm{ItemB\_Score\_2} + 0\times \mathrm{ItemC\_Score\_2} +\\
&113.5403\times \mathrm{ItemD\_Score\_2} + 201.7448\times \mathrm{ItemE\_Score\_2} + 310.5489\times \mathrm{Item1\_Score\_3} +\\
&242.7606\times \mathrm{Item2\_Score\_3} + 69.52968\times \mathrm{Item3\_Score\_3} + 1133.2\times \mathrm{Item4\_Score\_3} + 0\times \mathrm{Item5\_Score\_3} +\\
&0\times \mathrm{Item6\_Score\_3} + 56.5532\times \mathrm{Item7\_Score\_3} + 66.51422\times \mathrm{Item8\_Score\_3} + 425.317\times \mathrm{Item9\_Score\_3} +\\
&87.08731\times \mathrm{Item10\_Score\_3} + 0\times \mathrm{Item11\_Score\_3} + 234.038\times \mathrm{Item12\_Score\_3} +\\
&23.89605\times \mathrm{Item13\_Score\_3} + 18.36489\times \mathrm{ItemA\_Score\_3} + 13.40368\times \mathrm{ItemB\_Score\_3} +\\
&0\times \mathrm{ItemC\_Score\_3} + -20.17667\times \mathrm{ItemD\_Score\_3} + 252.9833\times \mathrm{ItemE\_Score\_3} + 0\times \mathrm{Item1\_Score\_4} +\\
&0\times \mathrm{Item2\_Score\_4} + 0\times \mathrm{Item3\_Score\_4} + 0\times \mathrm{Item4\_Score\_4} + 0\times \mathrm{Item5\_Score\_4} + 0\times \mathrm{Item6\_Score\_4} +\\
&0\times \mathrm{Item7\_Score\_4} + 0\times \mathrm{Item8\_Score\_4} + 0\times \mathrm{Item9\_Score\_4} + 0\times \mathrm{Item10\_Score\_4} + 0\times \mathrm{Item11\_Score\_4} +\\
&0\times \mathrm{Item12\_Score\_4} + 0\times \mathrm{Item13\_Score\_4} + 0\times \mathrm{ItemA\_Score\_4} + 0\times \mathrm{ItemB\_Score\_4} + 0\times \mathrm{ItemC\_Score\_4} +\\
&0\times \mathrm{ItemD\_Score\_4} + 0\times \mathrm{ItemE\_Score\_4} - 2282.15)/10)
\end{aligned}
$$

The final probability is then computed relative to the reference class which is defined by the clinician-chosen super-cluster and also clusters which are consistent with the red-rules.

This is illustrated in the table below (ignoring the red rules). The probabilities within a super-cluster sum to 100. Note that when computing *exp(DFS)* for the first two clusters, the Discriminant Fisher Score is first divided by 10; for the third Organic cluster it is first divided by 100.

The items were selected as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	A	B	C	D	E
0	0	0	4	0	0	0	0	0	0	0	4	0	0	0	0	4	4

			Probability		
Cluster	*Discriminant Fisher Score(DFS)*	*exp(DFS)*	*Non-psychotic*	*Psychosis*	*Organic*
1	215.6709	2.33E+09	0	-	-
2	210.3489	1.37E+09	0	-	-
3	346.8847	1.16E+15	20	-	-
4	354.1459	2.40E+15	42	-	-
5	324.8617	1.28E+14	2	-	-
6	350.712	1.70E+15	30	-	-
7	335.6802	3.79E+14	7	-	-
8	268.9057	4.77E+11	0	-	-
10	620.479	8.85E+26	-	0	-
11	650.2272	1.73E+28	-	6	-
12	643.8599	9.17E+27	-	3	-
13	589.954	4.18E+25	-	0	-
14	623.7976	1.23E+27	-	0	-
15	607.5817	2.44E+26	-	0	-
16	590.555	4.44E+25	-	0	-
17	678.1406	2.83E+29	-	91	-
18	11639.144	3.53E+50	-	-	24
19	11634.664	3.38E+50	-	-	23
20	11565.296	1.69E+50	-	-	11
21	11694.812	6.16E+50	-	-	42