Developing a case mix classification for child and adolescent mental health services: the influence of presenting problems, complexity factors, and service providers on number of appointments

Authors: Peter Martin[1,5], Roger Davies[2], Amy Macdougall[1], Benjamin Ritchie[1], Panos Vostanis[3, 4], Andy Whale[1], and Miranda Wolpert[1]

[1] Evidence Based Practice Unit, University College London and Anna Freud Centre for Children and Families, [2] City and Hackney CAMHS, East London Foundation Trust, [3] Department of Neuroscience, Psychology and Behaviour, University of Leicester, [4] Department of Clinical, Educational and Health Psychology, University College London, [5] Department of Applied Health Research, University College London.

Corresponding author:

Peter Martin, Department of Applied Health Research, University College London, Gower Street, London WC1E 6BT, United Kingdom.

Email: peter.martin@ucl.ac.uk .

1

Developing a case mix classification for child and adolescent mental health services: the influence of presenting problems, complexity factors, and service providers on number of appointments

Abstract

Background: Case-mix classification is a focus of international attention in considering how best to manage and fund services, by providing a basis for fairer comparison of resource utilization. Yet there is little evidence of the best ways to establish case mix for child and adolescent mental health services (CAMHS).

Aims: To develop a case mix classification for CAMHS that is clinically meaningful and predictive of number of appointments attended, and to investigate the influence of presenting problems, context and complexity factors, and provider variation.

Method: We analysed 4573 completed episodes of outpatient care from 11 English CAMHS. Cluster analysis, regression trees, and a conceptual classification based on clinical best practice guidelines were compared regarding their ability to predict number of appointments, using mixed effects negative binomial regression.

Results: The conceptual classification is clinically meaningful and did as well as data-driven classifications in accounting for number of appointments. There was little evidence for effects of complexity or context factors, with the possible exception of school attendance problems. Substantial variation in resource provision between providers was not explained well by case mix.

Conclusions: The conceptually-derived classification merits further testing and development in the context of collaborative decision making.

Introduction

Determining what characteristics of patients and services best predict resource use is crucial for developing efficient financial allocations across service providers and for informing clinical care. Case-mix classification underpins considerations how best to manage, compare and pay health care services (Costa, Poss, & McKillop, 2015; Halsteinli, Kittelsen, & Magnussen, 2010; Hornbrook, 1980; Ogles, Carlson, Hatfield, & Karpenko, 2008; Phillips, Kramer, Compton, Burns, & Robbins, 2003).

Previous studies indicate that for children and young people, age, functioning (including school attendance) and symptom severity (including hallucinations, emotional and behavioural) affect the cost of mental health service use in the community setting (Buckingham et al. 1998, Gaines et al. 2003). These factors appear within a list of a dozen or so patient-level factors suggested to be associated with mental health service cost, in a recent literature review that looked across settings and ages (Harris et al., 2013). The review noted the relative paucity of research into children and community mental health care (in contrast to working age adults and inpatient care). Studies that have included the community setting report difficulty in predicting costs from assessment information about service-user characteristics (Buckingham, Burgess, Solomon, Pirkis, & Eagar, 1998; Health and Social Care Information Centre, 2006).

It is therefore unsurprising that where countries have tried to develop innovative systems to apply case-mix measurement to the funding of mental health service providers, implementation has been challenging and/or slow, particularly in the community context (Independent Hospital Pricing Authority, 2015b; Mason,

Goddard, Myers, & Verzulli, 2011). In England, the implementation of a clustering approach to underpin payment for adult mental health services has attracted a mixture of optimism and criticism (Care Pathways and Packages Project, 2015; Denham-Vaughan & Clark, 2012; Jones et al., 2013; Royal College of Psychiatrists, 2014; Yeomans, 2014). There is a lack of evidence about the relationship between cluster membership and resource use among adult patients. Consideration of this relationship is crucial for upholding the principle of fairness of provider payment (Vostanis et al., 2015).

This paper presents the development of a classification to inform the processes of contracting with and paying providers of child and adolescent mental health services (CAMHS), focusing on outpatient treatment, and specifically modelling the relationship between cluster membership and number of appointments. The principles of our approach were explained in Vostanis et al. (2015). Extensive stakeholder consultation identified factors that CAMHS clinicians, service managers and commissioners thought most likely to be relevant to clinical decision making and number of appointments. These factors included the nature and severity of presenting problems, a range of contextual and complexity factors, and education, employment and training issues (Jones, et al., 2013; Vostanis, et al., 2015). Our aim was to develop a classification of children and young people accessing CAMHS that satisfied four criteria derived from the case-mix literature and stakeholder requirements: (i) ability to account for variation in episode costs, (ii) clinical meaningfulness, (iii) assignment of service users to groupings at the beginning of episodes, and (iv) reliability of assignment (Fetter, Shin, Freeman, Averill, & Thompson, 1980; NHS England, 2015). Since previous evidence suggested large

variability in resource provision between providers (Vostanis, et al., 2015), we also investigated whether this can be explained by casemix. In contrast to studies in Australia (Buckingham, et al., 1998) and New Zealand (Gaines et al., 2003), we focus exclusively on CAMHS outpatient episodes.

Method

Sample

The CAMHS Payment System Project collected clinical records from 20 CAMH service providers that responded to an open call for participation. Information about children and young people seen between 1 September 2012 and 30 June 2014 was recorded in compliance with the Children and Young People's Improving Access to Psychological Therapies data set (Version 3) (CYP IAPT, 2013). We performed data quality checks on each of the 20 providers, using a combination of data inspection and communication with provider representatives. Eleven providers had sufficiently high data quality to be included in the analysis. These came from a variety of English regions, both urban and rural. All were outpatient services, ranging from uni-disciplinary community services to multidisciplinary teams dealing with complex problems. Specialist teams within one or more providers included: learning disability, neurodevelopmental disorders, looked after and vulnerable children, outreach and intensive community engagement, eating disorders, and substance misuse.

The eleven providers submitted 11352 outpatient periods of contact that included information on presenting problems at assessment, and that had at least one

appointment recorded as having been attended by the child or young person. We call this the Full Sample. Of these, 4573 cases were closed or dormant (without activity for at least six months), so that number of appointments until case closure could be established. These 4573 cases constitute our Analysis Sample, which was used for development of the classification.

Measures

Presenting Information

The Current View Tool (Jones, et al., 2013), a one-page form completed at assessment, was used to capture clinician view at the start of treatment. It records 30 presenting problems, 14 complexity factors, as well as six contextual problems and issues in education, employment or training (EET). Lists of the problems measured are given in Figures 1b – 1d. Ratings need not imply a diagnosis.

Complexity factors are rated on a three-point scale with response categories "Yes" (= present), "No", and "Not known". All other problems are on a five-point scale with the response categories "None", "Mild", "Moderate", "Severe", and "Not known". In all measures, ratings of "None" and "Not known" were combined with missing ratings to form a common category that indicates no evidence for the presence of a problem or factor. In contextual problems and EET issues, ratings were coded into a numeric scale as follows: "None/Not Known" = 0, Mild = .33, Moderate = 0.67, Severe = 1. For complexity factors, ratings were coded "No/Not known" = 0, "Yes" = 1.

Number of appointments. Our primary indicator of resource use was "Number of

Appointments", i.e. the number of direct contacts (face-to-face or telephone) that the service user (child or young person) had with the provider. Recording of indirect activity – work other than direct contact carried out to benefit the service user – was not of sufficient quality in our data set to be usable for analysis.

Demographic information. We used age (in years) and gender (male/female) to investigate the demographic composition of our sample. We did not have access to information on socio-economic status.

Data analysis

The rationale underlying our analysis is to compare data-driven approaches, which explicitly maximize the fit of the classification to the data (without consideration of clinical plausibility), with a conceptually guided approach based on our interpretation of best practice guidelines. We compared three methods of classification: k-medoids cluster analysis (Kaufman & Rousseeuw, 2009), regression trees (specifically, the recursive partitioning algorithm; Hothorn, Hornik, & Zeileis, 2006), and a conceptual approach. We considered mixture models such as latent class analysis. However, when using a large number of latent class indicators (in our case, 30 presenting problems), a mixture model requires many parameters per latent class. Given our sample size, as the number of latent classes rises, the estimation quickly becomes unstable, and eventually the model becomes unidentified.

K-medoids cluster analysis

K-medoids cluster analysis aims to cluster cases on the basis of a given set of

characteristics, without reference to a dependent variable. Clusters are developed so as to maximize between-cluster differences and minimize within-cluster differences. We used the Gower distance measure (Gower, 1971) to define a distance between all pairs of cases based on the 30 presenting problems, and analysed the distance matrix using the k-medoids algorithm. Well-fitting clusters consist of individuals who are more similar to each other than to individuals in other clusters. Fit is assessed by the average silhouette statistic (Rousseeuw, 1987).

Regression Trees

Regression trees cluster cases on the basis of the relationship between a set of characteristics (in our case, the Current View information) and a dependent variable (number of appointments). The aim is to find subsets of a sample that can be identified by their presenting information, so that the members within each subset attend a similar number of appointments. All presenting information from the Current View Tool was used, as well as the following summary indicators: the maximum problem rating, the number of problems rated moderate or above, and the number of problems rated severe.

Regression trees can lead to overfitting, i.e. the optimisation of a model on the data set at hand, which may not generalise to the population. To avoid overfitting, we employed the cross-validation strategy suggested by Kuhn & Johnson (2013). We took ten random "development" samples (stratified by age and gender), each comprising 50% of cases. Each time, the remaining 50% of cases were designated the "test sample". A set of nested trees of varying sizes was grown on each development sample, using a range of stopping criteria. Two candidates for "best"

tree were then chosen for each development sample by testing the trees' predictive power on the associated test sample, using (1) a traditional R-square, and (2) an outlier-resistant R-square based on medians rather than means (respectively, R_1^2 and R_9^2 as defined in Kvalseth (1985)).

Conceptual classification

We defined a conceptual classification of children coming to CAMHS based on our review of best practice guidelines from the UK National Institute for Health and Care Excellence (NICE). We reviewed 19 existing NICE guidelines, which were either written specifically for, or made reference to, children (Vostanis, et al., 2015). Since the guidelines judged treatment intensity to be influenced by symptom severity and impairment, but not by contextual factors, we designed our categories to account only for presenting problems. Each relevant NICE guideline (e.g. NICE guideline CG31 "Obsessive-Compulsive Disorder") is represented by one cluster ("Getting Help: OCD"). All categories are mutually exclusive, with the exception of the grouping "Neurodevelopmental Assessment", which implies a recommendation of specialist assessment and may co-exist with other groupings.[1] The groups that make up the conceptual classification are displayed in Table 1.

A period of contact is assigned to a particular cluster based on presenting problem ratings. Clinical judgement was used to define assignment rules. A period of contact was assigned to a NICE group if the problem rating corresponding to the associated

---

[1] The NICE guideline CG78 ("Borderline Personality Disorder") is represented in our conceptualisation, but was not included in the algorithmic allocation, as we judged that presenting information was not sufficient to identify personality disorder. In practice, allocation to this grouping would be matter of clinical judgement using additional information from case histories not captured by the Current View form.

diagnosis was either moderate or severe (e.g., in the case of "Getting Help: OCD", this applies to the presenting problem "Compelled to do or think things (OCD)"). Some level of comorbidity was deemed acceptable within a NICE guidance category (e.g. mild depression within OCD), as long as there was a clear primary problem. Where comorbidity suggested the inapplicability of a purely delivered treatment according to NICE guidance (e.g. when severe OCD was comorbid with severe depression), the case was allocated to one of several mixed presentation groups. These groups also collect cases whose Current View ratings suggested that none of the presenting problems associated with any particular NICE guideline was present.

--- Table 1 about here ---

Model comparison

Using the ten test samples, we compared the classifications from regression trees, k-medoids cluster analysis, and conceptual grouping with respect to their ability to account for variation in number of appointments between CAMHS users. We used mixed-effects negative binomial regression with a random effect for provider. The dependent variable was "Number of appointments after the first session", i.e. the minimum possible value was zero (indicating that the child had not returned after the first appointment). We used Akaike's Information Criterion (AIC) and Schwartz' Bayesian Information Criterion (BIC) to compare models. All analyses were carried out using the R software (R Core Team, 2014). Mixed effects models were estimated using the glmmADMB package (Fournier et al., 2012).

Results

Description of the sample

The demographic compositions of the Full and Analysis Samples are given in Table 2. As we have shown elsewhere (Wolpert et al., 2015), these distributions were broadly similar to those of other large data sets of English outpatient CAMHS, although older teenagers may have been slightly overrepresented in our sample.

--- Table 2 about here ---

The distributions of presenting problems, context factors and complexity and EET problems in the Analysis Sample are shown in Figures 1b – 1d. Full tables are available in the online supplement, Tables S1 – S3. The distributions in the Full Sample (not shown here) are very similar to those in the Analysis Sample.

--- Figure 1 about here ---

The number of appointments in the Analysis Sample ranged from 0 to 101; their distribution is displayed in Figure 1a. The mean was 4.96, and the three quartiles were: Q1: 1; Median: 3; Q3: 6. Because of the relatively short observation period (22 months), and the need to consider only cases that were open and closed within this period, the sample is likely to be biased towards shorter periods of contact. We make no claim to have calculated unbiased estimates of parameters of the

11

distribution of "number of appointments". What we are interested in is a comparison of different groups within a given classification.

One feature of interest in these data is the considerable variation in measured resource provision between providers. Table 3 shows the quartiles and means of the appointments distribution in each of the eleven providers. We see that this average varies between 2.9 in Provider E and 12.9 in Provider C.

--- Table 3 about here ---

K-medoids cluster analysis

We now turn to an account of the results of each of our three methods of classification. K-medoids cluster analysis resulted in a poor fit. Average silhouette values above 0.51 indicate reasonable fit; we found no average silhouette above 0.10. Inspection of the clusters revealed that many included an indeterminate mix of periods of contact with varying presenting problems, so that clinical meaningfulness was doubtful. Nonetheless, we selected the 6- and 26-cluster solutions for further testing, since they represented local maxima of average silhouettes values. The resulting clusters are illustrated in the online supplement, Figures 3 and 4a-e.

Regression Trees

Our analysis yielded 20 regression trees (ten pairs of nested regression trees from each development sample). Inspection showed that different development samples

led to qualitatively different classifications, suggesting that the regression tree method did not yield a reliable classification of our data. This is also illustrated in the online supplement, Figures 1 and 2.

Model comparison

For each of the ten test samples, we compared five classifications: the 6- and 26-cluster solutions from the k-medoids cluster analysis, two regression trees (selected by traditional and outlier-resistant R-square), and the conceptual classification. Using the AIC, in nine out of ten samples the conceptual classification performed best; once the "traditional R-square" tree performed best. Using the BIC, a "traditional R-square" model performed best six times (but note that this is a different model each time), while the 6-cluster model from k-medoids cluster analysis performed best four times. Full results are presented in the online supplement, Table S4. In general, the AIC tends to select larger models, whereas BIC tends to select more parsimonious models. Overall, we concluded that there was little evidence that statistical methods outperformed our conceptual classification. Regression trees and cluster analysis maximize the statistical fit in a particular sample, and therefore ought to outperform a conceptual classification that has a tenuous relation to reality. This is not what we find here. In addition, the k-medoids clusters had doubtful clinical meaningfulness, while the regression tree method did not yield a reliable classification. We therefore choose to describe and investigate the conceptual classification more closely.

Conceptual classification: description

We present the 18-group conceptual classification, focusing on estimates of group membership and the relationship to number of appointments. Table 1 presents our estimates of group membership proportions. The best estimates are derived from the Full Sample, taking into account presenting information from both open and closed cases. The three largest groups are not symptom-specific: "Signposting and Self-Management Advice" (ADV), "Difficulties Not Covered by Other Groupings" (DNC) and "Difficulties of Severe Impact" (DSI). Between them, these three groups are estimated to account for about 52% of children and young people coming to CAMHS. Around 9% of children are estimated to be in one of the two groups defined by specific co-occurring difficulties ("Co-Occurring Behavioural and Emotional Difficulties" [BEM] and "Co-Occurring Emotional Difficulties" [EMO]). Around 39% of children are estimated to belong to a grouping referring to a specific NICE guideline for treatment. The counts and percentages from the Analysis Sample are given here to enable comparison.

Figure 2 shows the distributions of number of appointments by group membership. Children assigned to the "Signposting and Self-management advice" group attended on average a relatively small number of appointments (the median for this group was two appointments, the third quartile was four appointments). The three groups that belong to the "Getting More Help" supergrouping, "Eating Disorders" (EAT), "Psychosis" (PSY) and "Difficulties of Severe Impact" (DSI), attended on average more appointments than the members of most other groups. Most of the groups within the "Getting Help" supergrouping are somewhere in between the others.

There is large within-group variation, which indicates that most of the variation in the number of appointments is not explained by the groupings.

--- Figure 2 about here ----

Complexity factors, contextual problems, and education/employment/training issues

Next we investigated whether the prediction of number of appointments can be further improved by taking into account complexity factors, contextual problems and education/employment/training (EET) issues. We used mixed negative binomial regression to compare two models: Model 1 predicts the number of appointments using the 18-group conceptual classification and a random effect to account for provider-level variation. Model 2 is the same as Model 1, except that complexity factors, contextual problems and EET issues have been added as covariates.

We observed the following model fit indices: AIC: 21704.6 for Model 1, compared to 21690.4 for Model 2. BIC: 21833.2 for Model 1, compared to 21941.1 for Model 2. So Model 2 slightly outperforms Model 1 according to the AIC, but Model 1 is clearly preferred according to the BIC. Overall, this suggests that there is at best weak evidence for an additional effect of complexity factors, contextual problems or EET issues, above the effect captured in the conceptual grouping. We present estimates from Model 2 to further illustrate this point.

Table 4 shows rate ratios estimated from Model 2. Rate ratios represent the ratio of

the predicted number of appointments for a given group over the predicted number

of appointments for a reference group. Here, the reference group are children in

"ADV" without any complicating factors. For example, the rate ratio of 2.01 for DEP

indicates that a child in this group is predicted to attend twice as many sessions as a

child in the ADV group. For contextual problems and EET issues, the coefficients

represent the effect of having a given problem or issue rated as severe compared to

a rating of none or not known. A rate ratio of 1 represents absence of effect.

Table 4 shows that the estimated effects of all additional factors are small relative to

the effects of the groupings. Of the 19 additional factors, only "EET attendance

issues" has a 99% confidence interval that does not include 1.

--- Table 4 about here ---

Overall model fit and variation between providers

We computed two types of adjusted R^2 statistics: the adjusted traditional R^2 and the

adjusted outlier-resistant R^2 (respectively, R_{a1}^2 and R_{a9}^2 as defined in Kvalseth

(1985). These R^2 values suggest that the models account for about 12 % of the

variation in number of appointments, or 33 % when using the outlier-resistant

statistic. The large difference between these two values suggests that a large

amount of unpredictability in the number of appointments is caused by outliers with

very large numbers of sessions.

If the conceptual classification was used for payment purposes or benchmarking,

taking the provider-level variation into account is not meaningful. When we re-estimated Model 2 without the random effect for provider, the adjusted traditional R^2 statistic was 5 %, and the outlier-resistant R^2 statistic was 13 %. This means that unexplained provider-level variation is large relative to the estimated effects of the groupings: around 7 % according to the traditional R^2, and 20 % according to the outlier-resistant R^2 The number of appointments a child is predicted to attend depends at least as much on which provider the child goes to, as on the presenting difficulties of the child.

Discussion

This paper is the first to model the relationship between presenting information and number of appointments attended in users of English child and adolescent mental health services. We compared three methods of classifying periods of contact in CAMHS to define clusters, or groups, of children and young people that have similar needs, and whose treatment requires similar amounts of resource on average. Two statistical methods of classification, k-medoids cluster analysis and regression trees, did not yield reliable or clinically meaningful systems of categories. A third approach, the conceptual grouping based on current NICE guidelines for the treatment of children and young people with mental health difficulties, is clinically meaningful and did at least as well as the two statistical approaches in predicting number of appointments.

There was at best weak evidence for an influence of complexity factors, contextual

problems, and education/employment/training (EET) issues, after controlling for presenting problem information encoded in our conceptual classification. Our analysis suggests that only EET attendance issues may additionally increase average number of appointments. School attendance was also found to be a predictor of treatment cost in child and adolescent mental health services in Australia (Buckingham, et al., 1998) and New Zealand (Gaines, et al., 2003).

In its current form, our classification makes only a modest contribution to the attempt to account for differences in resource provision between CAMHS providers. The question remains why presenting problems have such a weak relationship with number of appointments, and why provider-level variation is so large relative to variation between children with different presentations. We consider three sets of reasons: [a] data quality; [b] omission of important individual-level factors; [c] lack of standardization of practice.

Data quality: Problems with data quality may have compromised our ability to accurately estimate the effect of our groupings on number of appointments (Vostanis, et al., 2015). Different providers may have been more or less thorough in recording appointments, so that some of the provider-level variation may be due to different recording practices, rather than actual differences in treatment provision. Also, we were able to look at number of appointments only at the level of an individual period of contact within a provider, and could not 'follow' cases that had been closed because they had been referred to another provider for further treatment or assessment. Thus, a child that was 'referred on' after a single session looks like a case of 'low resource use' in our data, but may in fact have continued

prolonged treatment elsewhere.

Omission of important individual-level factors: We may not have measured all relevant variables. First, we had no indicators of case history. Second, we were unable to measure treatment goals or preferences held by the children or their parents (Independent Hospital Pricing Authority, 2015a). Third, important information about a child or young person may often not be discovered in the first few sessions, but emerge only in the course of treatment. Models such as the Behavioral Model of Health Services Use would usefully inform future research into important factors for which there is currently a lack of data (Andersen, 1995). Finally, we were unable to collect indirect treatment activity. This may have limited our ability to distinguish the effect of complexity factors on resource use, as one might surmise that some of the additional work required for complex cases may involve activity other than face-to-face interaction with the child or family. Among the direct activities, our main analysis uses the number of appointments only. However, results of sensitivity analyses suggest that our conclusions do not change if we take into account the duration of appointments, and the number and professions of staff present (Wolpert, et al., 2015).

Lack of standardisation of practice: Our data may point to a lack of standardisation of practice: the same child, with the same presenting problems, may receive different treatment depending on which provider they happen to present at. Further research should aim to verify this observation, and, if confirmed, illuminate it. Current evidence is insufficient to decide which of these reasons apply, or their relative importance. Efforts to encourage the collection of complete, reliable and

high quality data about the children seen and the treatment provided in CAMHS should be made an issue of high importance.

Notwithstanding these issues, our classification has much to recommend it. It is based on best practice guidelines by NICE. Assigning a CAMHS user to a particular group implies the applicability of guidelines for treatment, making the classification clinically meaningful. The classification also has an inbuilt flexibility. If best practice guidelines change, additional groups could be defined. The effect of changes to best practice guidelines (say, on resource use) in a particular category could in principle be investigated, if data on category allocation and other relevant variables were rigorously collected in CAMHS.

There is no automatic read- across of the case-mix groupings suggested here to the clusters developed for adult mental health in the UK as part of currency development (Self, Rigby, Leggett, & Paxton, 2008). Children and young people are different to adult service users, in terms of the types of difficulty, relevant contextual factors and forms of service provision. This necessitated a case mix categorisation that did not map easily onto existing adult mental health models. If more integrated all age services were to be developed both classifications would need to considered to create an integrated all age case mix categorisation. For the moment, however, we consider our conceptual grouping to be a useful starting point for the development of a casemix classification to inform payment and quality monitoring in outpatient CAMHS.

A follow-up study should attempt to validate our classification by comparing our

assignment algorithm to assignment based on clinical judgement (Self & Painter, 2009; Self, et al., 2008). Such a study may well lead to a refinement of our method of assigning children to groupings. Further research should consider how to incorporate inpatient episodes into a casemix classification (Beecham, Green, Jacobs, & Dunn, 2009).

References

Andersen, R. M. (1995). Revisiting the behavioural model and access to medical care: does it matter? *Journal of Health and Social Behavior, 36*(1), 1-10.

Beecham, J., Green, J., Jacobs, B., & Dunn, G. (2009). Cost variation in child and adolescent psychiatric inpatient treatment. *European child & adolescent psychiatry, 18*(9), 535-542.

Buckingham, B., Burgess, P., Solomon, S., Pirkis, J., & Eagar, K. (1998). Developing a Casemix Classification for Mental Health Services. Volume 1: Main Report. Canberra.

Care Pathways and Packages Project. (2015). The legacy. Bradford, Leeds, Rotherham & Sheffield: Care Pathways and Packages Project: Developing Currencies for Mental Health.

Costa, A., Poss, J. W., & McKillop, I. (2015). Contemplating case mix. A primer on case mix classification and management. *Healthcare Management Forum, 28*(1), 12-15.

CYP IAPT. (2013). Children and Young People's Improving Access to Psychological Therapies. Dataset Specification V. 3.1.2, 15.07.2013.

Denham-Vaughan, S., & Clark, M. (2012). Transformation of payment by results in mental health: ensuring a recovery focused and socially inclusive system. *Mental Health and Social Inclusion, 16*(3), 155-158.

Fetter, R. B., Shin, Y., Freeman, J. L., Averill, R. F., & Thompson, J. D. (1980). Case Mix Definition by Diagnosis-Related Groups. *Medical Care, 18*(2, Supplement), 1-53.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Maunder, M. N., Nielsen, A., . . .

Magnusson, A. (2012). AD Model Builder : using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software, 27*(2), 233-249.

Gaines, P., Bower, A., Buckingham, B., Eagar, K., Burgess, P., & Green, J. (2003). New Zealand Mental Health Classification Outcomes Study: Final Report. Auckland: Health Research Council of New Zealand.

Gower, J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics, 27*(4), 857-871.

Halsteinli, V., Kittelsen, S. a., & Magnussen, J. (2010). Productivity growth in outpatient child and adolescent mental health services: the impact of case-mix adjustment. *Social science & medicine (1982), 70*(3), 439-446.

Harris, M., Legge, N., Diminic, S., Carstensen, G., McKeon, G., Siskind, D., . . . Whiteford, H. (2013). Mental health service cost drivers - an international literature review. Final report for Stage B of the Definition and Cost Drivers for Mental Health Services project: The University of Queensland.

Health and Social Care Information Centre. (2006). End Stage Report - Mental Health Casemix Classification Development. Leeds.

Hornbrook, M. C. (1980). Hospital case mix: its definition, measurement and use. Part i: the conceptual framework. *Medical Care Review, 39*(1), 1-43.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics, 15*(3), 651-674.

Independent Hospital Pricing Authority. (2015a). Development of the Australian mental health care classification - public consultation paper 2. Sydney: Independent Hospital Pricing Authority.

Independent Hospital Pricing Authority. (2015b). Development of the Australian Mental Health Care Classification. Public Consultation Paper 1: Independent Hospital Pricing Authority.

Jones, M., Hopkins, K., Kyrke-Smith, R., Davies, R., Vostanis, P., & Wolpert, M. (2013). Current View Tool Completion Guide. London.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken: Wiley.

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. New York: Springer.

Kvalseth, T. (1985). Cautionary Note About R2. *The American Statistician, 39*(4), 279-285.

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.

NHS England. (2015). Developing a new approach to palliative care funding. Leeds: Palliative Care Team.

Ogles, B. M., Carlson, B., Hatfield, D., & Karpenko, V. (2008). Models of case mix adjustment for Ohio mental health consumer outcomes among children and adolescents. *Administration and Policy in Mental Health and Mental Health Services Research, 35*, 295-304.

Phillips, S. D., Kramer, T. L., Compton, S. N., Burns, B. J., & Robbins, J. M. (2003). Case-mix adjustment of adolescent mental health treatment outcomes. *The Journal of Behavioural Health Services and Research, 30*(1), 125-136.

R Core Team. (2014). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Rousseeuw, P. J. (1987). Silhouettes : a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53-65.

Royal College of Psychiatrists. (2014). Statement on Mental Health Payment Systems (formerly Payment by Results). Position Statement PS01/2014. London: Royal College of Psychiatrists.

Self, R., & Painter, J. (2009). Study: To improve and demonstrate the structural properties of the care clusters that form the basis of the PbR currency development programme. Wakefield.

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.

Vostanis, P., Martin, P., Davies, R., De Francesco, D., Jones, M., Sweeting, R., . . . Wolpert, M. (2015). Development of a framework for prospective payment for child mental health services. *Journal of Health Services Research & Policy, 20*(4), 202-209.

Wolpert, M., Vostanis, P., Young, S., Clark, B., Davies, R., Fleming, I., . . . Whale, A. (2015). Child and adolescent mental health services payment system project: final report. London: CAMHS Press

Yeomans, D. (2014). Clustering in mental health payment by results: a critical summary for the clinician. *Advances in Psychiatric Treatment, 20*(4), 227-234.