Predicting methylphenidate response in attention deficit hyperactivity disorder: A preliminary study

Blair A Johnston, David Coghill, Keith Matthews and J Douglas Steele

Abstract
Methylphenidate (MPH) is established as the main pharmacological treatment for patients with attention deficit hyperactivity disorder (ADHD). Whilst MPH is generally a highly effective treatment, not all patients respond, and some experience adverse reactions. Currently, there is no reliable method to predict how patients will respond, other than by exposure to a trial of medication. In this preliminary study, we sought to investigate whether an accurate predictor of clinical response to methylphenidate could be developed for individual patients, using sociodemographic, clinical and neuropsychological measures. Of the 43 boys with ADHD included in this proof-of-concept study, 30 were classed as responders and 13 as non-responders to MPH, with no significant differences in age nor verbal intelligence quotient (IQ) between the groups. Here we report the application of a multivariate analysis approach to the prediction of clinical response to MPH, which achieved an accuracy of 77% (p = 0.005). The most important variables to the classifier were performance on a ‘go/no go’ task and comorbid conduct disorder. This preliminary study suggested that further investigation is merited. Achieving a highly significant accuracy of 77% for the prediction of MPH response is an encouraging step towards finding a reliable and clinically useful method that could minimise the number of children needlessly being exposed to MPH.

Keywords
ADHD, attention deficit hyperactivity disorder, children, conduct disorder, drug activity, drug sensitivity, methylphenidate, pattern recognition, predictive methods, therapeutic drugs

Introduction
Whilst around 65% of children diagnosed with attention deficit activity disorder (ADHD) will respond to and tolerate the stimulant drug methylphenidate (MPH), which is the most commonly used pharmacological therapy for ADHD, 35% do not (Hodgkins et al., 2012). Denney and Rapport (1999) evaluated several models designed to predict the response to MPH. None of the proposed models were robust, leading them to conclude that a comprehensive model of the MPH response would need to include both biological and behavioural components. Coghill et al. (2007) investigated the relationship between the MPH response and several neuropsychological measures, in a randomised placebo-controlled trial design: They found that poor performance on a ‘delayed matching to sample’ (DMtS) task at baseline was the only pre-treatment correlate of clinical response. Their data support the conclusion of Denney and Rapport (1999) that a comprehensive model of the MPH response would likely rely on a wide range of measures, rather than a single measure (Coghill et al., 2007).

A discriminant analysis by Buitelaar et al. (1995) attempting to distinguish responders and non-responders achieved 81% accuracy using a combination of demographic and behavioural measures. Unfortunately, cross-validation, a model validation approach for assessing how the analysis generalises to novel data, was not used in this study and it is, therefore, unclear whether the accuracy achieved could be replicated in any independent data set; however, this study suggests that it may be possible to predict treatment response in individual subjects by using demographic and behavioural measures (Buitelaar et al., 1995). The key variables highlighted in this discriminant analysis include intelligence quotient (IQ), inattention, age, symptom severity and anxiety level (Buitelaar et al., 1995).

The response to MPH in ADHD has been linked with a number of ADHD subtypes and comorbidities. Children and adolescents with predominantly hyperactive symptoms of ADHD have long been suggested to be more responsive to MPH than those with ADHD without hyperactivity (Barkley et al., 1991). In addition, comorbidities, such as internalising disorders, were associated with reduced likelihood of response to MPH in one study (DuPaul et al., 1994), but no association was found for anxiety, in another (Diamond et al., 1999). Furthermore, MPH is shown to have positive effects on children and adolescents with conduct disorder (CD), with and without ADHD (Klein et al., 1997). Therefore, there are several strands of evidence suggesting that the MPH treatment response could, in principle, be predictable using data that is often collected within routine clinical practice.

Division of Neuroscience, Medical Research Institute, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK.

Corresponding author:
Blair A Johnston, Division of Neuroscience, Medical Research Institute, Ninewells Hospital and Medical School, University of Dundee, Mailbox 6, Dundee DD1 9SY, UK.
Email: blair.a.johnston@gmail.com
The application of machine learning-based techniques, which use cross-validation for replication, to predict treatment response in psychiatry is a rapidly developing field. Machine learning techniques tend to outperform more traditional statistical methods, such as discriminant analysis, as they are more suited to be able to classify imperfect data and complex non-linear data (Cleophas and Zwinderman, 2013; Moss et al., 2012). To date, the majority of treatment response prediction studies in psychiatry have used neuroimaging data. A notable focus of the literature is on major depressive disorder (MDD) treatment response prediction, perhaps because there are a large number of potential treatments that can take many months (or years) to evaluate for each patient, although effective treatment of psychotic symptoms in schizophrenia can be similarly slow to evaluate. Identifying accurate predictors of response to one therapy versus another, for a variety of syndromes, could have clear benefits for clinical practice and patients.

The performance of a classification study is typically assessed through three variables: the accuracy (the ratio of correctly-classified subjects to total subjects), the sensitivity (in this study, this relates to the percentage of non-responders whom are correctly identified as such) and the specificity (in this study, this relates to the percentage of responders whom are correctly identified).

Using a cross-validation approach, Costafreda et al. (2009a) attempted to predict the response to either antidepressant medication (fluoxetine: accuracy = 89%, sensitivity = 89% and specificity = 89%) or to cognitive behavioural therapy (CBT), which was not significant, in MDD patients using neuroimaging data. Using the same data as Costafreda et al. (2009a), but a different machine learning technique, Nouretdinov et al. (2011) report results comparable to the Costafreda results.

Gong et al. (2011) report an accuracy of 70% (specificity = 70% and sensitivity = 70%) when using images of grey matter and 65% using white matter (specificity = 74%, sensitivity = 57%), when predicting the response to antidepressant treatment in medication-naive MDD participants. Fu et al. (2008) also attempted to predict which depressed patients would respond to the antidepressant medication fluoxetine, correctly identifying 75% of the partial or non-responders, and 62% of the full responders. Costafreda et al. (2009b) were able to classify responders and non-responders to CBT with 71% sensitivity and 86% specificity.

After training a classifier to predict between a medication-naive MDD group and a healthy control group, Gao et al. (2012) classified previously-depressed patients whom achieved clinical remission through treatment. When using no feature selection (an approach often used in prediction studies to identify the most relevant variables, for use in the creation of a model), the group in remission were predicted to be healthy controls in 88% of participants, but all 16 participants were predicted to be controls when feature selection was used (Gao et al., 2012). When attempting to predict treatment responders versus treatment-resistant patients, Liu et al. (2012) achieved 83% accuracy (using either grey or white matter images).

Machine learning techniques have also been applied to predict treatment responses in other psychiatric disorders, such as social anxiety disorder. Using pre-treatment scores on the Liebowitz Social Anxiety Scale (LSAS) and functional magnetic resonance imaging data, Doehrmann et al. (2013) successfully predicted post-CBT treatment LSAS scores. Using pre-treatment electroencephalography data, Khodayari-Rostamabad et al. (2010) were able to predict the response to clozapine therapy in schizophrenic patients with 85% accuracy.

As the classification techniques discussed above and implemented in this study are data driven, it is possible that the variables (or brain regions) identified as predictive of treatment response may not be identified in future datasets. There are many possible reasons for this, such as differences in the subject recruitment criteria, the definition of response or correlated variables. These issues are discussed in more detail in the results and discussion sections.

To our knowledge, currently there are no published studies, for any psychiatric disorder, that successfully predict medication response in individual subjects, using demographic variables, clinical variables and neuropsychological task scores. Certainly, no study successfully predicts the response to MPH in patients with ADHD; therefore, the aim of the present study was to predict the MPH response using a multivariate approach, using a subset of demographic, clinical and neuropsychological measures identified by Coghill et al. (2007) as being most likely linked to medication response.

Methods

Subjects

The participant data used in this analysis were from a 12-week, placebo-controlled, double blind, randomised, crossover trial of MPH therapy in boys diagnosed with hyperkinetic disorder / ADHD by Coghill et al. (2007). Of the 75 boys with ADHD included in the original study, we excluded 32, either due to incomplete data or to ensure there were no significant differences in age or the British Picture Vocabulary Scale (BVPS), a proxy measure of verbal IQ, between the identified groups of responders and non-responders. Of the 43 boys with ADHD and with complete data sets who were included in the present study, we classed 30 as responders and 13 as non-responders to MPH.

Clinical, demographic and neuropsychological variables

The clinical response to MPH was determined using the method of Jacobson and Truax (1991). Using this method, both ‘clinically significant change’ and ‘reliable change’ between placebo and MPH dose are required as criteria to determine full response. Clinically significant change was defined as a transition from the ‘dysfunctional’ to the ‘functional’ range (i.e. < 65) on the Conners’ Global Index total t-score. Reliable change is a measure of how much each subjects’ scores change during treatment and is calculated by dividing the difference between the pre-treatment and post-treatment scores by the standard error (SE) of the difference between the two scores (defined as the reliable change index (RCI) < – 1.96). Subjects who experienced significant adverse side effects, irrespective of symptom changes that resulted in MPH treatment being stopped, were also classed as non-responders (see Jacobson and Truax (1991) for a full consideration of the relevant theory and application of this approach and Coghill et al. (2007) for details of the original study).

Coghill et al. (2007) identify 13 variables, using principal components analysis, as being the most likely to distinguish responders to MPH from non-responders: three demographic variables, three clinical variables and various measures from four neuropsychological tasks. The demographic variables were the
scores on the British Picture Vocabulary Scale (BPVS), 2nd edition (Dunn et al., 1997), a measure of verbal IQ; decimal age; and standardised socioeconomic deprivation (SIMD) scores. The clinical variables were the presence of comorbid oppositional defiant disorder (ODD) or CD and the $t$-score baseline for the Parents’ ADHD Conners’ questionnaire. We took four additional neuropsychological task measures from a ‘go/no go’ task and three from the ‘CANTAB’ Visual Memory Battery (pattern recognition, spatial recognition and DmTS total percentage of correct $z$ scores, adjusted for age and BPVS).

The ‘go/no go’ task involved the subjects being presented with a sequence of letters and numbers on a screen. The ‘Type 1’ block corresponded to a ‘switch’ trial, where subjects were required to withhold response when the stimulus changed from letters to numbers, or vice-versa. The ‘Type 2’ block corresponded to ‘non-switch’ trials, where the subjects were required to withhold response when a letter was presented, if the previous stimulus was a letter, and similarly for numbers. For both ‘Type 1’ and ‘Type 2’, the output variables were the errors for distractors (ERD), a measure of the average number of times the subject responded when they were required to inhibit their response, and the reaction time to target stimuli (RTT), a measure of the reaction time when a correct response (key press) was required. Four variables were extracted from the ‘go/no go’ task: go/no go-Type 1 RTT and ERD, and go/no go-Type 2 RTT and ERD. The tasks in the visual memory battery of the CANTAB tested for the ability to recognise a previously-presented abstract pattern in a forced choice procedure for the pattern recognition task; the ability to recognise the spatial locations of target stimuli for the spatial recognition task; and the ability to remember the visual features of a complex, abstract, target stimulus, and to select from a choice of four patterns after a variable delay in the DmTS task (Coghill et al., 2007). All variables above are described in more detail by Coghil et al. (2007).

**Variable preparation**

As a first step, the variables were normalised to reduce errors due to scaling. This involved all scores being brought within the range of 0–1 by subtracting the minimum value and dividing by the range (maximum – minimum). Normalisation aims to ensure that the automated selection of variables is based on their predictive value, rather than on their relative variability or magnitude. There are many different approaches to normalisation. Whilst this approach is one of the simplest methods of rescaling the data, it must be acknowledged that the approach is more easily influenced by outliers than other techniques, as it depends on individual values (the maximum and minimum) rather than more stable statistics, such as the mean and standard deviation.

In order to compare the data contained in this study with the study by Buitelaar et al. (1995), a discriminant analysis was performed without cross-validation. These results are in the supplementary material.

**Individual scan classification**

In order to investigate which variables predict a clinical response, we applied a linear Support Vector Machine (SVM) (Vapnik, 1995, 1998) pattern recognition method to the data, with leave-one-out cross-validation (LOOCV). We used a second (inner) LOOCV procedure for parameter selection, to ensure that predictions were made only on novel data. We performed all analyses in Matlab 2012a (The Mathworks Inc., Natick, MA, USA) and the Matlab-based calculations used the SVM toolbox (Schwaighofer, 2001) and custom Matlab scripts.

In other words, for each boy we first constructed a training set that consisted of the remaining 42 boys. For each of these sets, we began by identifying the combination of variables that best predicted the classification of responder/non-responder, within the training set. The optimal combination of variables is then used to predict the MPH response of the novel boy.

The feature selection technique (used to identify the variables that contribute most to prediction) involved ranking the variables from largest to smallest differences between groups, within each training set (excluding the left-out subject, to ensure it is novel to the classifier). The lowest-ranked variable was iteratively removed from the classification process, until there was only one variable remaining. This approach optimised the number of variables required to classify the data. We calculated the classification accuracy, sensitivity and specificity at each stage.

As there were far fewer non-responders than responders, the model optimisation method was modified to avoid the ‘class imbalance problem’ (Mourão-Miranda et al., 2011; Theodoridis and Koutroumbas, 2006). The class imbalance problem is a common issue in binary classification studies that occurs when one group has more subjects than the other group. The larger group tends to ‘overwhelm’ the classifier, making the larger group considerably more likely to be predicted (Theodoridis and Koutroumbas, 2006). The class imbalance problem tends not to be an issue when using large datasets and classifying well-separated groups; however, as this preliminary study contains a large imbalance, a relatively small number of subjects and no significant differences between groups in any individual variable, class imbalance required consideration.

Typically, binary classification studies optimise SVM and feature selection parameters using the training stage accuracy (the accuracy obtained in the inner LOOCV procedure, not including the novel subject); however, due to the imbalance between the groups in this study, if in a poorly-optimised classifier, all subjects were classed as responders and the accuracy would be 70% (30/43 subjects would be classified correctly, as there are 30 responders and 13 non-responders). Therefore, in order to minimise the class imbalance problem, we used the training stage sensitivity (more accurately classifying those in the group with the fewest subjects, the non-responders) to optimise the SVM and feature selection, instead of the usual training stage accuracy.

The combination of variables that achieved the highest sensitivity for prediction were then selected for training and testing the SVM, to predict the response of the novel, ‘left out’ subject. If two or more combinations of variables obtained the maximum sensitivity value during the training stage, then the accuracy was used as a secondary selection parameter. Notably, as feature selection took place for each training set, it was possible for a different combination of variables to be selected for the prediction of each subject.

**Results**

**Participant characteristics**

Age and verbal IQ did not differ significantly ($t$-test, $p > 0.1$) between responders and non-responders to MPH. The mean age
Table 1. Clinical descriptors for responders and non-responders to MPH. Variables are shown as mean (SD).

<table>
<thead>
<tr>
<th></th>
<th>Responders (<em>n</em> = 30)</th>
<th>Non-responders (<em>n</em> = 13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPVS percentile rank</td>
<td>40.27 (30.58)</td>
<td>32.15 (28.84)</td>
</tr>
<tr>
<td>decimal age</td>
<td>11.19 (2.39)</td>
<td>11.26 (2.99)</td>
</tr>
<tr>
<td>diagnosis of oppositional defiant disorder*</td>
<td>21/30</td>
<td>9/13</td>
</tr>
<tr>
<td>diagnosis of conduct disorder*</td>
<td>14/30</td>
<td>2/13</td>
</tr>
<tr>
<td>deprivation score</td>
<td>4.27 (1.72)</td>
<td>4.08 (1.32)</td>
</tr>
<tr>
<td>t-score baseline for Parents’ ADHD Conners’</td>
<td>78.07 (4.25)</td>
<td>80.08 (4.03)</td>
</tr>
<tr>
<td>deprivation score</td>
<td>4.27 (1.72)</td>
<td>4.08 (1.32)</td>
</tr>
<tr>
<td>t-score baseline for Parents’ ADHD Conners’</td>
<td>78.07 (4.25)</td>
<td>80.08 (4.03)</td>
</tr>
</tbody>
</table>

*Chi-square tests (the others were t-tests).

ADHD: Attention deficit hyperactivity disorder; BPVS: British Picture Vocabulary Scale; DMtS: delayed matching to sample task; ERD: errors for distractors; MPH: methylphenidate; NS: not significant; RTT: reaction time to target stimuli; z score: a measure of how many standard deviations an individual score is from the mean.

The SVM approach achieved a classification accuracy for MPH response of 76.7% (sensitivity = 0.54, specificity = 0.87, $\chi^2 = 7.8$ and $p = 0.005$ (Figure 1)). Our preliminary data suggested a chi-square effect size (w) of 0.42, indicating a w in the medium-large range.

Presence of CD and the ERD in one of the go/no go blocks (Coghill et al., 2007) were the only variables that were selected in all predictions, although RTT in the same go/no go block was also selected in a high proportion of predictions (38/43 subjects). On average, four variables were used for each prediction. The variables which were never selected for predictions were decimal age, presence of ODD and SIMD score; and notably, DMtS total percent correct z score, which was the variable previously highlighted by Coghill et al. (2007) as the only baseline neuropsychological univariate correlate of clinical response. The number of times that each variable was used in each ‘leave-one-out’ prediction is summarised in Table 2.

Variables that were not selected were not necessarily without predictive value. For example, if several variables are strongly correlated, then some may be removed automatically during feature selection, as multiple variables containing similar information do not improve classification (Guyon and Elisseeff, 2003). An uncorrelated set of variables is best for prediction. For example, the ‘Type 1’ Go/No Go block, reflecting a ‘switch’ condition, was selected more often in the prediction of MPH response than the ‘Type 2, non-switch’ Go/No Go block. As 9 out of 13 of the variables failed the Shapiro-Wilk test for normality, which is required for parametric statistical testing (e.g. Pearson’s coefficient of correlation), we performed a non-parametric Spearman’s rank-order (Spearman’s rho) correlation analysis. This revealed that the RTT and the ERD scores from each of the blocks were significantly correlated with the corresponding variable in the other type (RTT: $\rho = 0.76; p < 0.001$ and ERD: $\rho = 0.78; p < 0.001$). Therefore, the selective inclusion of Type 1 tasks is most likely due to a strong correlation between these variables. Similarly, the unexpected omission of the DMtS variable may be explained by the significant correlations with the Parents’ ADHD Conners’ Questionnaire ($\rho = 0.44; p = 0.004$) and pattern recognition ($\rho = 0.35; p = 0.023$) scores.
### Table 2. Frequency of variable selection for prediction.

<table>
<thead>
<tr>
<th>Variable selection (Frequency)</th>
<th>Leave-one-out loops per variable selected (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPVS percentile rank</td>
<td>2</td>
</tr>
<tr>
<td>decimal age</td>
<td>0</td>
</tr>
<tr>
<td>presence of ODD</td>
<td>0</td>
</tr>
<tr>
<td>presence of CD</td>
<td>43</td>
</tr>
<tr>
<td>deprivation score</td>
<td>0</td>
</tr>
<tr>
<td>t-score baseline</td>
<td>24</td>
</tr>
<tr>
<td>Parents’ ADHD Conner’s score</td>
<td></td>
</tr>
<tr>
<td>Go/No Go Type 1 RTT</td>
<td>38</td>
</tr>
<tr>
<td>Go/No Go Type 2 RTT</td>
<td>2</td>
</tr>
<tr>
<td>Go/No Go Type 1 ERD</td>
<td>43</td>
</tr>
<tr>
<td>Go/No Go Type 2 ERD</td>
<td>9</td>
</tr>
<tr>
<td>Pattern recognition z score</td>
<td>12</td>
</tr>
<tr>
<td>Spatial recognition z score</td>
<td>1</td>
</tr>
<tr>
<td>DMtS total % correct z score</td>
<td>0</td>
</tr>
</tbody>
</table>

ADHD: Attention deficit hyperactivity disorder; BPVS: British Picture Vocabulary Scale; CD: conduct disorder; DMtS: delayed matching to sample task; ERD: errors for distractors; ODD: oppositional defiant disorder; RTT: reaction time to target stimuli; z score: a measure of how many standard deviations an individual score is from the mean.

### Discussion

By using sociodemographic, clinical and neuropsychological data with a machine learning approach, it was possible to predict the response to MPH in a population of boys with ADHD, with 77% accuracy. Despite the relatively small number of subjects and variables included in this preliminary study, this accuracy is comparable with prediction studies for other psychiatric disorders that have used neuroimaging data. Whilst the 77% accuracy achieved in this study is statistically significant and comparable with larger studies of treatment response, it would probably not be high enough to be considered useful for clinical work. Therefore, improving on the 77% accuracy achieved in this preliminary study appears to require a larger dataset (more participants and a wider range of variables) and, ideally, equally-sized groups. Also, further improvement may be obtained by combining the best sociodemographic, clinical and neuropsychological measures with genetic and/or neuroimaging data, as suggested by Denney and Rapport (1999) and Coghill et al. (2007). As cross-validation was used, each prediction was made on completely novel data; therefore, if an entirely independent set of samples were used, we would expect a similar accuracy of prediction as reported in this study, assuming more (or similar numbers of) subjects were included, with identical recruitment criteria to Coghill et al. (2007).

Whilst Buitelaar et al. (1995) identified (among other variables not included in this study) that age, IQ and symptom severity were possible predictors of MPH response, only the baseline Parents’ Conner’s symptom severity score was used in more than 5% of the predictions in the present study, with 56% of predictions using this score.

Despite the attempts to maximise sensitivity that are detailed in the methods, this approach still suffers from a low accuracy (54%) in predicting non-responders to treatment. This is likely due to the large class imbalance between groups and emphasises the need for balanced groups in future studies, however, it also raises an important clinical question: If 100% accuracy cannot be achieved, is it better to sacrifice poor sensitivity for maximal specificity? In other words, is it better to maximise the number of correctly-predicted treatment responders at the cost of incorrectly predicting that a higher proportion of non-responders would respond to treatment? It could be argued that this is a better scenario than incorrectly predicting that responders to treatment would not respond, given that, in comparison with the current procedures, the former reduces but does not eliminate the number of children and adolescents whom are needlessly taking MPH, whilst maximising the number of responders being given an effective treatment.

Considering that this preliminary study is based on a small number of subjects and a small number of variables, it may be possible to increase accuracy further, by including more subjects and more variables. It is important to emphasise that despite the fact this study used cross-validation for internal replication, a larger study, with high-quality data obtained from multiple research groups, is recommended. Assuming the study will be replicated with an improved accuracy in a larger study, a possible approach towards clinical application is to split the data into two balanced groups and perform 2-fold cross-validation (train a classifier in one group, then test using the second group (and vice-versa), creating two classifiers. Provided that these two classifiers produce accurate predictions that are similar to each other, they can be combined to create a ‘function’, to predict the MPH response. This function could then be tested in a clinical setting, whereby the MPH response is predicted prior to treatment, patients receive treatment-as-usual irrespective of the prediction, and the accuracy of prediction is assessed through follow-up. There are too few subjects included in this preliminary study to generate a function of MPH response, using this 2-fold cross-validation approach, but it may be possible in larger samples.

Another approach could be to incrementally ‘update’ the predictor by training on all subjects with a known response, predicting the new subject(s) prior to treatment, and then after the treatment-as-usual, adding the new subject(s) in, and with their known responses, to update the classifier. In principle, this approach would continually improve its own robustness,
as the classifier would be given increasing information over time.

An alternative approach to binary classification is probabilistic prediction. Popular machine learning approaches that could provide probabilistic predictions include: Gaussian processes and Relevance Vector Machines (Rasmussen and Williams, 2006; Tipping, 2001). These have the potential to provide an estimate of the likelihood of response to treatment, or various treatments, or could predict the ‘magnitude of response or non-response’, thus removing the arbitrary cut-off between responders and non-responders.

A number of study limitations must be acknowledged. Some have been highlighted already, such as the class imbalance problem between responders and non-responders, and the small number of subjects and variables included in the study. It should be noted that the current study uses cross-validation and so it estimates the accuracy of prediction for completely independent, future samples; however, it is not an estimate for predictive accuracy if different participant inclusion and exclusion criteria are used.

Treatment adherence, an important factor that could influence response to treatment (Charach and Fernandez, 2013), was assessed by pill count and clinical enquiry. Using these measures, adherence was consistently high in the vast majority of participants; however, this is not a completely reliable and valid measure of treatment adherence, and so it is a limitation of this study. Furthermore, the Jacobson-Truax criteria, although conceptually robust and valid, may be considered too stringent when describing clinical response.

Further work is required to identify which set of criteria best distinguishes responders and non-responders to treatment. Additionally, it is well known that accuracy of a predictor is dependent on the prevalence of the feature that is being predicted; so if a sub-group of children and adolescents with ADHD were recruited whom were less likely responders, then the predictive accuracy may be lower. Our preliminary study was aimed at medication-naive patients recruited from out-patient referrals; therefore, further work would be required to determine whether this technique could be adapted to ensure that the technique remained robust when applied to patients with multiple comorbidities.

Our finding of a prediction accuracy of 77% for MPH response in ADHD is an encouraging step towards identifying a reliable technique could be adapted to ensure that the technique remained robust when applied to patients with multiple comorbidities.

Further work is required to identify which set of criteria best distinguishes responders and non-responders to treatment. Additionally, it is well known that accuracy of a predictor is dependent on the prevalence of the feature that is being predicted; so if a sub-group of children and adolescents with ADHD were recruited whom were less likely responders, then the predictive accuracy may be lower. Our preliminary study was aimed at medication-naive patients recruited from out-patient referrals; therefore, further work would be required to determine whether this technique could be adapted to ensure that the technique remained robust when applied to patients with multiple comorbidities.

Acknowledgements
We thank Dr. Benson Mwangi, who specialises in machine learning, for advice on the method and all volunteers who agreed to participate in this study.

Declaration of Conflicting Interests
The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: BAJ was funded by a SINAPSE-SPIRIT industry partnership with Siemens Medical. DC received research funding from Lilly, Shire, Janssen and Vifor; honoraria for consultancy, advisory boards and speaker fees from Lilly, Shire, Janssen, Medice, Flynn, Novartis and Vifor.

KM chaired advisory boards for studies of deep brain stimulation for obsessive-compulsive disorder, sponsored by Medtronic. He has received educational grants from Cyberonics and from Schering Plough; has received research project funding from Lundbeck, Merck Serono, Reckitt Benckiser; and funding from Saint Jude Medical for a multi-centre, clinical trial of deep brain stimulation for depression. He also received travel and accommodation support to attend meetings, from Medtronic and Saint Jude Medical.

JDS has received research funding via an honorarium associated with a lecture from Wyeth.

Funding
The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a Tenovus-Scotland initiative (a local trust) and by SINAPSE (www.sinapse.ac.uk), which included a SINAPSE-SPIRIT industry partnership with Siemens Medical (a SINAPSE studentship for author BAJ). Our funding sources played no part in the design of the study, analysis or interpretation of the data, nor manuscript preparation.

References

Johnston et al.


