Modeling symptoms of attention deficit hyperactivity disorder (ADHD) and improving its diagnosis using machine learning approaches.

Matthew Bearham

MSc Computer Science

The University of Bath

2021/22

Modeling symptoms of attention deficit hyperactivity disorder (ADHD) and improving its diagnosis using machine learning approaches.

Submitted by: Matthew Bearham

for the degree of MSc in Computer Science

at the University of Bath

Copyright

Declaration

This dissertation is submitted to the University of Bath in accordance with the requirements of the degree of MSc Computer Science in the Department of Computer Science. No portion of the work in this dissertation has been submitted in support of an application for any other degree or qualification of this or any other university or institution of learning. Except where specifically acknowledged, it is the work of the author.

Abstract

This dissertation applies a data-driven approach to address current challenges in the treatment and diagnosis of attention deficit hyperactivity disorder (ADHD). In terms of treatment, the English Prescribing Dataset to observe the trends in ADHD prescriptions over time. One of the interesting aspects is to observe the prevalence of ADHD regionally across England, and correlate this with various socioeconomic metrics (e.g. life expectancy, dispensable income, wellbeing). Then these findings are used to suggest changes to public health policy that have the potential to improve the efficiency of ADHD diagnosis and treatment. During this exploration, a need for more accessible diagnostic tools for ADHD is observed. To address this aspect, behavioural data from the ABCD study (of adolescents) is utilised, and three machine learning models (logistic regression, SVM, and random forest classifier) are developed. These models made remarkable predictions about ADHD diagnosis above the level of chance (AUC = 0.7715). Inability to multitask was identified as the most potent predictor of ADHD, despite not being present on the diagnostic criteria for ADHD. Given further refinement, this model could be used as a tool to automate screening for ADHD.

Acknowledgements

I would like to thank my project supervisor Dr Alok Joshi for providing me with more support than I could have asked for, both academically and logistically. When everything was going wrong (as it did at almost every point in this project), he was always available with words of encouragement and sound advice. Additional thanks to Dr Tom Lancaster for assisting with the data acquisition and helping me find my feet in the dataset, Dr Alessio Guglielmi for his flexibility in supporting this project through uncertain times, and Dr Sarah Bailey for suggesting the NHS EPD dataset while I was in a tough spot. Special thanks to my family for supporting me through my academic career.

Table of Contents

List of Figures 8
List of Tables 9
Introduction 10
1.1. Description of the Problem 10
1.2. Aetiology 10
1.2.1. Genetics 10
1.2.2. Environment 10
1.3. Diagnosis 11
1.3.2. Comorbidity 12
1.4. Treatment 12
1.5. Machine Learning Approach 13
1.6. Aims 13
1.7. Outline 13
Literature and Technology Survey 15
2.1. ADHD-Related Studies 15
2.1.1. Adolescent Brain Cognitive Development (ABCD) Study 15
2.2. Applying ML to ADHD 16
2.2.1. ML Algorithms 16
2.2.2. Examples in Literature 17
2.3. Limitations of Available Literature 18
2.4. Research Questions 18
Epidemiology of ADHD in England 19
3.1. Methods 19
3.1.1. EPD Dataset 19
3.1.2. Geographic Distribution 20
3.1.3. Change over Time 22
3.2. Results 23
3.2.1. Geographic Distribution 23
3.2.2. Cost and Quantity over Time 26
3.3. Discussion 29
3.3.1. Findings 29
3.3.2. Limitations 30

Modelling ADHD in Adolescents 31
4.1. Methods 31
4.1.1. ABCD Dataset 31
4.1.2. Medication Log 32
4.1.3. Temperament 33
4.2. Results 34
4.2.1. Medication Log 34
4.2.2. Temperament 36
4.3. Discussion 40
4.3.1. Findings 40
4.3.2. Limitations 41
Evaluation and Conclusions 42
5.1. Conclusions 42
5.2. Contributions 42
5.3. Reflections 43
5.4. Further Work 43
References 45
Initial Project Plan 49
Data Processing Code 50
Regional Data Analysis 50
Over-time Analysis 52
ABCD Data 56
EPD Dataset Structure 64
Tables from Chapter 3 69
ABCD Data Dictionary 82

List of Figures

Figure 1 - Timeline of events for the ABCD study (ABCD Study, 2022a)..... 15
Figure 2 - The locations of the 21 ABCD study sites across the USA. (ABCD Study, 2022b)..... 16
Figure 3 - Process flow for the processing of EPD data 20
Figure 4 - Map of ADHD/total and ADHD/population by English county 24
Figure 5 - Correlation matrix of ADHD prevalence and a range of socioeconomic measures 25
Figure 6 - Map of ADHD/population against maps of 7 other socioeconomic measures 26
Figure 7 – a) Quantity of ADHD prescriptions between 2014 and 2022 b) Trend in prescriptions for ADHD before and after COVID-19..... 27
Figure 8 – a) Quantity of ADHD drugs prescribed between 2014 and 2022 b) Total cost of ADHD drugs between 2014 and 2022 c) Price of ADHD drugs between 2014 and 2022..... 28
Figure 9 - Process flow of the processing of ABCD data 32
Figure 10 - a) The prevalence of ADHD in ABCD study b) Breakdown of ADHD medications taken in the ABCD study 35
Figure 11 – AUC score of the logistic regression model: a) as the weighting of outcomes was varied during training b) over a narrower range with smaller steps c) trained on the top n features 37
Figure 12 - a) AUC score of the SVM model as the weighting of outcomes was varied during training b) over a narrower range with smaller steps c) trained on the top n features 39
Figure 13 - AUC score of the decision forest model as the weighting of outcomes was varied during training 40

List of Tables

Table 1 - Symptoms of ADHD as listed in the DSM-V (American Psychiatric Association, 2013) 11
Table 2 - Drugs used in the treatment of ADHD and their associated brand names (WebMD, 2021). 12
Table 3 - Socioeconomic metrics and their respective sources 21
Table 4 - Table metadata for the regional analysis of ADHD prevalence 21
Table 5 - Table metadata for the temporal analysis of ADHD prevalence and treatment 22
Table 6 - Top 5 integrated care boards for both ADHD/total and ADHD/population 23
Table 7 - Bottom 5 integrated care boards for both ADHD/total and ADHD/population 23
Table 8 - r value boundaries for correlation categories 25
Table 9 – Terms in the ABCD medication dataset identified as relating to ADHD medication 33
Table 10 - Concurrence of ADHD medications in the ABCD cohort 36
Table 11 - The ten most and ten least indicative features of ADHD 38
Table 12 - Comparison of all metrics for all optimal models 40

Chapter 1

Introduction

1.1. Description of the Problem

Attention deficit hyperactivity disorder (ADHD) is one of the most prevalent childhood neurodevelopmental disorders, usually diagnosed between the ages of 3 and 7 (NHS, 2021). It can be broadly categorised as predominantly inattentive, hyperactive, or a combination of both (Drechsler et al., 2020). At times, diagnosis of ADHD can be particularly challenging as it can manifest differently among individuals, and it is difficult to understand the underlying medical causes for the observed physical symptoms. This is due to the complex interaction of multiple factors which are implicated in the aetiology or pathophysiology of the disorder.

1.2. Aetiology

One of the leading theories suggests that ADHD arises due to both genetic and environmental factors; this could explain the complexity of the disorder, as its mechanism potentially involves several epigenetic alterations and spontaneous mutations at multiple different parts of the genome (Kian et al., 2022). These epigenetic components would be influenced by both genetics and environment.

1.2.1. Genetics

The genetic underpinning of ADHD has been confirmed via different studies. One of the twin studies reported a heritability rate of around 70% (Tarver et al., 2014); another genome-wide association study identified a few nucleotide positions on the human genome that may be involved in the ADHD mechanisms. The two most studied mutations are related to the genes encoding Brain-Derived Neurotrophic Factor (BDNF) and the dopamine transporter (DAT1). The role of these biomarker pathways in ADHD is supported by evidence from neuroimaging studies, where it has been observed that people with ADHD have less prefrontal cortical volume. A mutation in BDNF could explain this observation since it is heavily involved in many stages of neurodevelopment (Mehta et al., 2019). Notably, DAT1 is the target for the most prescribed ADHD medication, methylphenidate, and supports the hypothesis of this pathway's involvement in the pathophysiology of ADHD (Rommelse et al., 2008). While these genes alone cannot fully explain the causes of ADHD, the interaction between genome and environment is a novel research area that could fill these gaps.

1.2.2. Environment

Environmental factors are much more difficult to discern. The two main environmental risk factors proposed are pre-natal smoking (Langley et al., 2005) and low birth weight (Johnson et al., 2010); however, there is limited supporting evidence confirming these as risk factors. A subsequent study found that paternal smoking during pregnancy was just as related as maternal smoking, implying that the correlation may be due to other environmental or genetic factors (Langley et al., 2012).

1.3. Diagnosis

The defining symptoms of ADHD are listed in the diagnostic and statistical manual of mental disorders (DSM-V), and this is commonly used as a reference when diagnosing ADHD. The DSM-V is used by medical professionals worldwide and contains extensive diagnostic criteria for a full range of psychiatric and neurological disorders. As discussed earlier, the symptoms are divided into two categories, the details of which are summarised below:

Inattention	Hyperactivity
a) Makes careless mistakes	a) Often fidgets
b) Difficulty sustaining attention	b) Often leaves seat against expectation
c) Does not listen when spoken to	c) Runs about in inappropriate situations
d) Fails to finish tasks	d) Unable to take part in activities quietly
e) Difficulty organising tasks	e) Is often 'on the go'
f) Avoids tasks requiring sustained mental effort	f) Often talks excessively
g) Often loses things	Impulsivity
h) Often distracted by extraneous stimuli	f) Cannot wait turn in conversation
i) Is often forgetful	g) Has trouble waiting their turn
	h) Often interrupts others

Table 1 - Symptoms of ADHD as listed in the DSM-V (American Psychiatric Association, 2013)

According to DSM-V criteria, a diagnosis is recommended if both symptoms under either category persist for at least six months and to the degree that limits social/academic activities. The diagnosis is then classified under predominantly inattentive, predominantly hyperactive, or combined presentation (American Psychiatric Association, 2013).

In the UK, the National Institute for Health and Care Excellence (NICE) provides diagnostic guidelines to clinicians specifically for children/adolescents. According to the guidelines, if ADHD symptoms persistently affect the quality of life, they should be referred to secondary care (psychiatrist, ADHD specialist). In this setting, the person then carries out a 'full clinical and psychosocial assessment', 'a full developmental and psychiatric history', and obtains 'observer reports and assessment of the person's mental state'. Following this assessment, the person must:

- Meet the DSM-V criteria for ADHD
- Suffer with 'at least moderate psychological, social and/or educational or occupational impairment'
- Have their symptoms 'occurring in 2 or more important settings including social, familial, educational and/or occupational settings'

As the diagnosis process requires very personalised care by an ADHD specialist, there can often be significant waiting times on referrals for ADHD diagnosis, with some waitlists estimated to be as long as six months (ADHD UK, 2022). Additionally, due to the somewhat subjective nature of the symptoms impeding communication and the broad range of ways the condition manifests, there is a need for tools that can provide objective assessments. Such tools would streamline the diagnosis process by providing specialists with accurate information before speaking to the patient (National Institute for Health and Care Excellence [NICE], 2018). Despite this, a specialist is still required due to the challenge of distinguishing ADHD from other similar comorbid conditions.

1.3.2. Comorbidity

It is usually required that a specialist psychiatrist gives ADHD diagnoses due to the complexity of diagnosing a disorder with such subjective behavioural symptoms. It can be particularly challenging to determine whether a patient's symptoms are being caused by ADHD or by some other (or combination of other) similar disorder(s), such as autism spectrum disorder (ASD) or oppositional defiant disorder (ODD). This challenge is exacerbated by the rate of comorbidity between ADHD and other neurodevelopmental and learning disorders such as ASD (70-85%), depression and anxiety disorders (30-50%), reading disorders (15-50%), and ODD (27-55%) (Drechsler et al., 2020). Many of these conditions manifest similar behavioural symptoms to ADHD, making it even harder to discern the actual cause. This challenge is reflected in the DSM-V criteria, updated in 2018 to recognise the substantial comorbidity between ADHD and ASD, which was previously an exclusion criterion in diagnosing ADHD (Drechsler et al., 2020). The condition must be diagnosed accurately so individuals can proceed with the most effective treatment.

1.4. Treatment

ADHD is treated using both pharmacological and non-pharmacological interventions. Pharmacological interventions for ADHD can be categorised into two groups based on their mechanism of action – stimulants and non-stimulants. Cognitive behavioural therapy is often offered on top of these medications to help manage symptoms.

Stimulants (Brand name)	Non-stimulants (Brand name)
Methylphenidate (Ritalin + others)	Atomoxetine (Strattera)
Dexamphetamine (Adderall + others)	Clonidine (Catapres/Kapvay)
Lisdexamfetamine (Vyvanse)	Guanfacine (Intuniv/Tenex)

Table 2 - Drugs used in the treatment of ADHD and their associated brand names (WebMD, 2021)

Stimulant medications are the most conventional treatment for ADHD, having been used in its treatment for decades. Stimulants target and block dopamine reuptake channels, promoting dopaminergic signalling in the brain by increasing dopamine concentrations in the synapses. The therapeutic benefit of stimulants in ADHD has been narrowed down to the increase in catecholaminergic (a subtype of neurotransmitter) signalling in the prefrontal cortex – the area of the brain that controls most human behaviour and attention. The fact that stimulants are effective in treating ADHD backs up the claim that ADHD is caused by irregular catecholaminergic signalling in the prefrontal cortex, as was suggested by the genomic association of dopamine transporters with ADHD (Arnsten, 2006).

Non-stimulant medications are a much more recent development in the treatment of ADHD. These drugs tend to have a longer-acting effect, although maybe not as powerful as stimulants. These are often taken daily in the morning to provide day-long symptom relief. These drugs bind to postsynaptic noradrenaline receptors and activate them, promoting noradrenaline signalling at the synapse. Some patients who do not respond to stimulants may respond to these drugs due to their different mechanisms of action (Dittmann, Hage, and Pedraza, 2018).

Although none are officially licensed for use in treating ADHD, some clinicians may prescribe certain antidepressants as an off-label application. However, antidepressants are not recommended for use in ADHD and are only ever used at the discretion of a medical professional.

In the UK, the recommended procedure for diagnosing and treating ADHD is published by the National Institute for Health and Care Excellence (NICE). These guidelines state that before starting medication, anyone with ADHD needs a full assessment, including a review of their diagnosis, a 'review of (their) physical health', and a 'review of (their) mental health and social circumstances'. If the necessary criteria are met, treatments are offered as follows for children:

- Offer methylphenidate as first-line treatment
- If no response is seen after six weeks, offer lisdexamfetamine
- If responding to lisdexamfetamine but the safety profile is intolerable, offer dexamphetamine
- If not responding/cannot tolerate stimulants, offer atomoxetine or guanfacine
- If none of the above, consult tertiary ADHD service

The only differences in adults are that lisdexamfetamine can be offered as first-line treatment, and guanfacine is never offered. Both guanfacine and atomoxetine are used off-label in ADHD, meaning the drugs are approved for a different indication but may have therapeutic benefits in ADHD (National Institute for Health and Care Excellence [NICE], 2018).

1.5. Machine Learning Approach

Healthcare is a very promising field for applying machine learning (ML) to complex classification problems. ML algorithms are most valuable when identifying complex trends in large datasets; there is an abundance of health data in the world, and the patterns in this data can provide helpful information when considering questions such as disease diagnosis and treatment outcome predictions. Previous studies have shown that ML agents can diagnose disorders with primarily behavioural symptoms, such as depression and anxiety. For example, a random forest model was trained using the results of six cognitive-behavioural tasks taken by participants with depression and anxiety. This model distinguished between the disease and control groups and the anxiety and depression groups (Richter et al., 2021). While many studies have explored this application, especially in major depressive disorder (MDD), the application in ADHD remains relatively unexplored.

1.6. Aims

- Explore the challenges in ADHD diagnosis and treatment through data analysis
    - Observe any regional or over-time trends in ADHD prevalence
    - Observe and predict trends in costs of prescriptions on the NHS
- Conceive tools to address the challenges noted in the initial exploration
    - Use machine learning techniques to model and classify ADHD
    - Improve diagnosis and treatment of ADHD through accessible tools and process automation

1.7. Outline

The first portion of the project will be a preliminary data analysis of data concerning the prescription of medications for ADHD across England, both over time and by region. The primary aim is to identify any patterns or trends in ADHD prevalence and treatment and, based on these, suggest how public health policy might be adjusted to maximise the cost, efficiency, and effectiveness of ADHD management. This investigation will be facilitated through various data analysis and visualisation tools.

The project's second portion will focus on one of the problems identified in the preliminary data exploration phase. While phase one is a surface-level identification of problem areas, this phase will seek ways to solve these problems by using machine learning methods to build up a model of ADHD symptomatology. This phase is more technical and will be heavily informed by the prior findings of the data investigation; many different models will be tested, and their performances will be compared to achieve the highest predictive power possible.

Chapter 2

Literature and Technology Survey

2.1. ADHD-Related Studies

While medical classification is a popular application for machine learning in research, its use in ADHD is limited. This chapter reviews the potential sources of ADHD behavioural data, some potentially useful machine learning techniques, then covers some of the prior applications of machine learning to ADHD in literature.

2.1.1. Adolescent Brain Cognitive Development (ABCD) Study

The ABCD study is an ongoing multi-centre study utilising 21 individual sites across the USA. The study's primary goal is to monitor the cognitive development of a cohort of children as they progress through adolescence, starting at the age of around nine years. Data is collected about all aspects of the participants' lives to measure the impacts physical condition and socioeconomic circumstances have on development; this includes "genetics, mental and physical health, neurocognition, substance use, mobile technology, and culture/environment". Additionally, this study collects various imaging data, including tabulated structural MRI, diffusion MRI, resting fMRI and task-related fMRI images. By collecting this data in visits every six months (2 years for imaging data), the cognitive development of the participants can be tracked, and the influences of external factors assessed. The timeline of events is shown below:

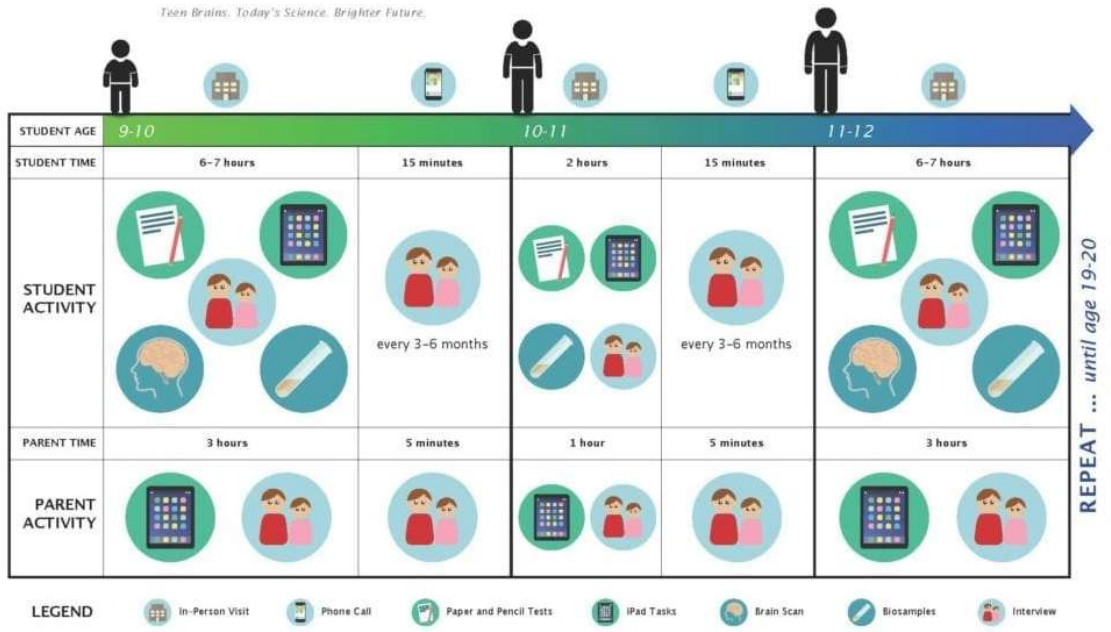


Figure 1 - Timeline of events for the ABCD study (ABCD Study, 2022a).

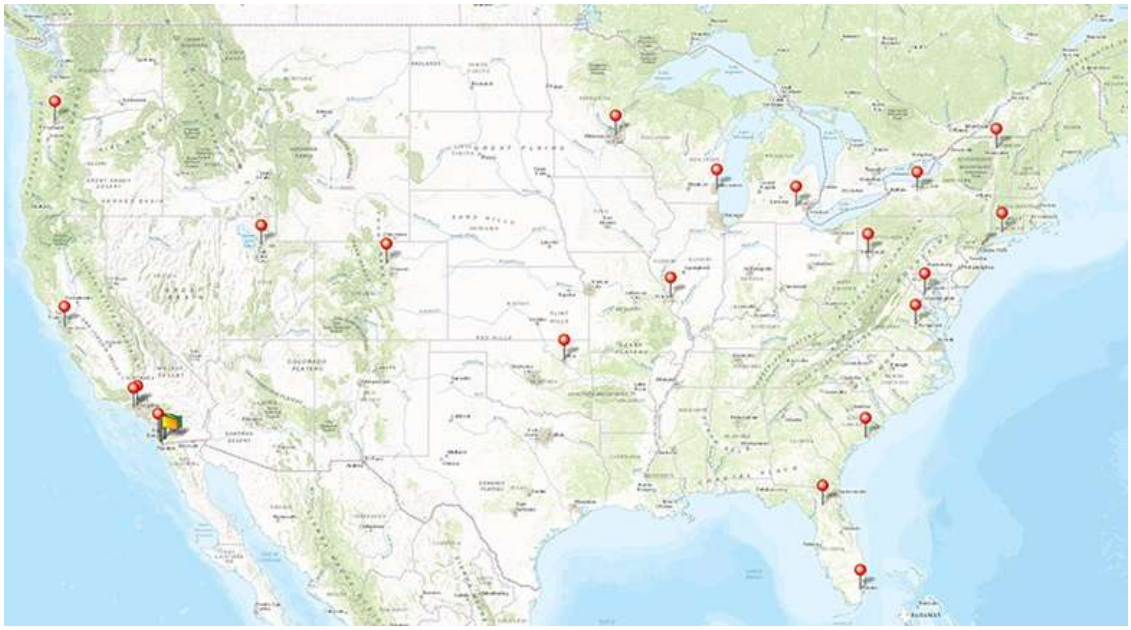


Figure 2 - The locations of the 21 ABCD study sites across the USA. (ABCD Study, 2022b)

The anonymised ABCD dataset is available through NIMH Data Archive, where researchers can submit an access request to NIH. The most current dataset from the study is the 4.0 release.; this contains neuroimaging data from the baseline and 2-year follow-up and complete non-imaging data for the baseline, 6-month, 1-year, and 18-month time points. In addition, interim data is available for the 2-year, 30-month, 3-year and 42-month visits. (Yang, 2021).

2.2. Applying ML to ADHD

2.2.1. ML Algorithms

The technology used in this project primarily consists of the specific machine learning algorithms used to make predictions about treatment outcomes. As a result, the following discusses some algorithms that are currently thought to be applicable in this area.

Deep Neural Networks

Neural networks (NNs) are a very flexible supervised learning model which have risen in prominence recently as computing speed has increased to match their extreme memory and processing requirements. Neural networks consist of a series of layers of 'neurons', all interconnected by a separate transition function. The weightings of these functions can be adjusted through training to correlate the inputs with the expected outputs to make predictions. This model is useful for classification and can produce continuous outputs for regression problems (Kukreja et al., 2016).

Deep neural networks (DNNs) refer to NNs that have more than one hidden layer. The 'deep learning' performed by these agents has the potential to identify trends at a level of complexity that far exceeds human understanding. This complexity is both a blessing and a curse for DNNs. While they are very good at fitting to a training dataset, they risk overfitting to this data if an insufficient quantity is provided. Additionally, while the network can make accurate predictions, it contributes very little to the human understanding of the problem space since it is unclear what the hidden layers actually represent (Liu et al., 2017).

Decision Forests

A decision tree is a supervised classification algorithm that identifies trends in data grouping and uses these to classify future unlabelled inputs. Decision trees are one of the most commonly used classification models due to their ability to spot patterns at a level that would be almost impossible for a human to observe. This model is often applied as a decision forest – a set of *n* different decision trees that are each trained under different restrictions. The final classification is the most frequent output category across all the trees (Criminisi et al., 2012).

One of the limitations of decision forests is that they tend to be greedy. When classifying the data, they prioritise an immediate division over a potentially better one later in the tree. One approach to solving this is combining decision forests with deep learning, as stated by Kontschieder et al. (2015). This model combines the predictive power of decision forests with the backpropagation of the DNNs to allow the trees to sacrifice an immediate reward for a later reward. This paper reports their results as 'on par or superior results when compared to state-of-the-art deep models'. Additionally, this research was partially funded by the pharmaceutical company Novartis, indicating that this algorithm may have been developed to apply it to health data. However, in a study by Yang et al. (2018) comparing neural networks, decision trees, and deep neural decision trees, deep neural decision trees were most often no more effective than neural networks or decision trees alone.

Decision forests have already solved medical classification problems; Pramanik et al. (2021) diagnosed Parkinson's disease using acoustic data with accuracy consistently above 90% across different types of decision forests. Therefore, this algorithm could be an appropriate choice to apply to the problem space.

Graph Neural Networks

A recent development in neural networks is the graph neural network (GNN), which uses deep learning to construct a 'graph'. This complex data type holds lots of relational information between different data elements. This kind of data is not easy to use in traditional DNN structures; however, it can hold much helpful information, particularly in complex classification problems where the relations between features are the crux of the problem (Zhou et al., 2020). GNNs take data in the form of graphs and iteratively update a node with information from its neighbours – eventually, an equilibrium is reached. This process requires much computational power, hence why GNNs were conceived by Gori et al. (2005) but not frequently applied until recently. Nevertheless, this is an exciting new model that could very reasonably be applied to the modelling of ADHD, with some work to convert the data to graph form (Wu et al., 2021).

2.2.2. Examples in Literature

While not as thoroughly investigated as many other psychiatric disorders, some published studies use machine learning to model and classify ADHD.

One study used a machine learning model to distinguish ADHD from ASD, using a dataset of social responsiveness data from individuals with either ASD (n=2775) or ADHD (n=150). The model trained could distinguish ADHD from ASD with an AUC of 0.965 by using only five features from the dataset. Six different models were trained for this study: a categorical lasso, a decision tree, an LDA, a logistic regression, a random forest, and an SVC. Out of all these models, the decision tree and decision forest models performed below the rest of the models, while the other models all performed similarly well. This result is impressive considering the dataset size used, although it may risk overfitting the model.

Only 150 records for participants with ADHD were available, which is fewer than would be desirable for training such models (Duda et al., 2016).

Most endeavours so far have attempted to use imaging data to classify ADHD. Unfortunately, these studies struggle to collect enough data to train reliable models that do not overfit. For example, an SVM model was trained to classify ADHD based on EEG data collected under various conditions. The model classified ADHD vs control correctly around 70% of the time. However, due to the invasive and time-consuming data collection method, only 117 participants could be enrolled (67 with ADHD, 50 without) (Tenev et al., 2014).

Predictions can also be extended to the treatment of ADHD. For example, another study predicted response to methylphenidate using a support vector machine, achieving an accuracy of 84.6%. These predictions were based on genetic data, fMRI scans, and behavioural tests (Kim et al., 2015).

In the literature, the most commonly used models seem to be support vector machines (SVMs) and decision trees. As a result, these will be the first choice of model to use when classifying ADHD in this project. Additionally, logistic regression is frequently used to rank the correlation of each data feature with the classification; this makes it a desirable model to use.

2.3. Limitations of Available Literature

Many of the studies in the available literature use small sample sizes, which limits the reliability of their results due to the risk of overfitting. Consequently, all results need to be taken with a grain of salt as it is not clear whether the models are predictive of the general population. While many studies use imaging data, very few use behavioural data so there are not many reference studies for this aspect.

2.4. Research Questions

This project takes a data-driven approach to addressing some of the current difficulties in diagnosing and treating ADHD. The aims of this project can be divided into two phases:

1. Data-driven identification of current challenges in ADHD diagnosis and treatment
    a. Assess the variance in ADHD prevalence across England, and use socioeconomic factors to rationalise any trends
    b. Use data analysis and visualization techniques to identify trends in ADHD treatment in England, including which drugs are prescribed and the associated costs
    c. Suggest policy changes or areas for further investigation from the above findings
2. Applying machine learning to solve problems in ADHD diagnosis and treatment
    a. Train a machine learning model that can accurately make predictions which can address the previously identified problem area
    b. Optimise the model to maximise predictive capability

Chapter 3

Epidemiology of ADHD in England

3.1. Methods

This chapter covers the exploratory analysis of the treatment and diagnosis of ADHD across England, using prescription data from the NHS.

3.1.1. EPD Dataset

In the first stage of this project, the English Prescribing Data (EPD) dataset was used to develop a general understanding of the impact of ADHD across the country. This impact can be divided into two primary categories: the prevalence of ADHD and how it impacts quality of life, and the prescription of medication to treat ADHD and its associated costs to the NHS. The ideal scenario for assessing these factors would be to analyse diagnosis/prevalence data for ADHD; however, no such data source is accessible. Alternatively, the EPD dataset can be used as a proxy for estimating the number of individuals with ADHD in any given month, assuming that all those diagnosed with ADHD are currently receiving treatment and that all those receiving those specific treatments have ADHD. As a result, this section refers not to the number of ADHD cases but the number of prescriptions issued for ADHD medications in a given month.

The EPD dataset contains information about the location, quantity, and cost of all medications prescribed in England (NHSBSA Data Warehouse, 2022). Location is separated into levels of specificity, ranging from the regional office down to individual practices. Drug names are recorded using British National Formulary (BNF) terminology. The quantity of a drug is given as quantity (the number of drugs in a pack), items (the number of packs prescribed), and the total derived quantity (from multiplying quantity and items). Finally, costs are reported as the net ingredient cost (NIC) and the actual cost to the NHS. Full details of the variables provided in the EPD dataset are given in Appendix C.

A single month of data from the EPD dataset contains vast data. For example, just one month of data from June 2022 is 6.3Gb - assuming consistency across all months, to store the entire dataset would use more than 600Gb of storage. Fortunately, the NHS provides a convenient API for requesting data from these monthly tables. However, one drawback of this method is that the output of each API call is limited to only a few hundred rows, so some of the data pre-processing needs to be done within the API call to meet this requirement.

The python *requests* library allows for integrating a SQL query into the API call, providing an effective way to manipulate the data and create derived data variables before loading the table. Entries relating to ADHD medications were identified using the bnf_chemical_substance field, for which all ADHD medications have codes in the form 0404000 followed by two trailing digits. Functions were written to request specific subsets of data given an input year and month, the code for which can be seen in Appendix B. Full data tables were generated by iterating through all year and month values and merging the tables into a single dataframe using the *pandas* library. This dataframe could then be used

for further processing and analysis.

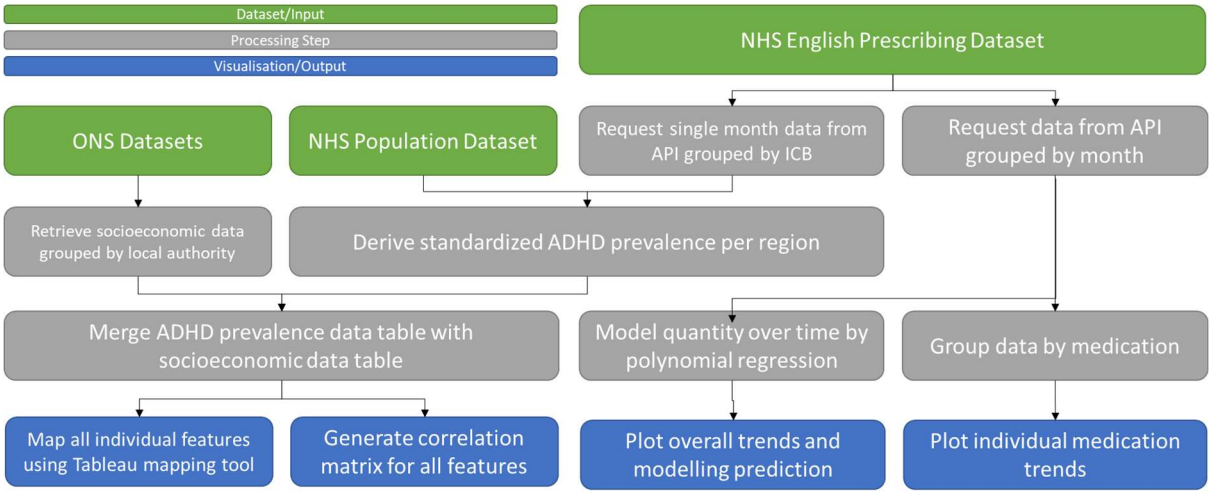


Figure 3 - Process flow for the processing of EPD data

3.1.2. Geographic Distribution

Count of ADHD Medications

To get an idea of how ADHD occurs across England, the prescribing dataset was used to derive the rate of ADHD prevalence in English counties. Data from May 2022 was used as this was the latest data publication at the time of analysis. The closest equivalent to a 'county' field provided in the EPD dataset is the Integrated Care Board (ICB), so a dataframe was generated containing the ICB Name and the sum of total_quantity values, grouped by ICB Name and filtered to ADHD medications. All entries with an ICB Name value of "UNIDENTIFIED" were dropped from the table since these did not convey any interpretable information about the prevalence of ADHD; this resulted in the removal of 13,688 units out of the total 7,170,842 units prescribed. The prescription data at this point must be standardised before any interpretation can take place. Therefore, two different metrics were derived to interpret the prescription quantities differently, with slightly different implications.

ADHD Medications as a % of All Prescriptions

The first derived ADHD metrics looked at the quantity of ADHD prescriptions as a proportion of the total quantity of medications prescribed in a given ICB. A table containing the sum of all medication quantities was generated by repeating the above procedure but removing the ADHD medication filter. From this, the percentage of all medications that were for ADHD was derived. This value emphasises how prevalent ADHD is compared to other health issues in an area and is indicative of the relative healthcare burden of ADHD in that area.

ADHD Medications per unit Population

The second derived ADHD metric is standardised for population. Population count data is freely available from the NHS, but only at the individual clinic level (NHS Digital, 2022). The population values for each ICB were derived by generating a table of all distinct practice_code values and their associated ICB Name. This table was merged with the NHS population data using the practice_code variable, and the sum quantities for each ICB were generated. The number of ADHD medications per unit population was then derived from the count_adhd and number_of_patients values.

Socioeconomic Metrics

The Office of National Statistics offers a range of anonymised open-access datasets providing information about several socioeconomic factors, some of which are divided by region. Variables of interest in exploring ADHD prevalence were economic stability (measured through unemployment and gross disposable household income), access to healthcare (measured through life expectancy), and social well-being (divided into anxiety, happiness, satisfaction, and worthwhileness). Merging these tables with the ADHD medication data required manual assignment of the 'local authority' used in the ONS data to the closest ICB in the NHS data. Some matches were one-to-one, but some were many-to-one or one-to-many. All socioeconomic data was standardised before merging to ensure that the population data was consistent with the geographical boundary used in the data collection. As a result, as long as the regions were approximately consistent between datasets, any correlation should still be visible; this was confirmed by using a table published by NHS England that breaks the ICBs down into the local authorities they contain (NHS England, 2022).

Category	Metric	Source
Economic	Unemployment	(Office for National Statistics, 2022)
	Gross Disposable Household Income	(Office for National Statistics, 2021a)
Health	Life Expectancy	(Office for National Statistics, 2021b)
Social	Anxiety	(Office for National Statistics, 2021c)
	Happiness	(Office for National Statistics, 2021c)
	Satisfaction	(Office for National Statistics, 2021c)
	Worthwhileness	(Office for National Statistics, 2021c)

Table 3 - Socioeconomic metrics and their respective sources

Variable	Description
icb_name	Integrated Care Board Name
count_adhd	The total quantity of ADHD medication prescribed
non_adhd	The total quantity of medication prescribed
percent_total	The percentage of total prescribed medication that was for ADHD
number_of_patients	Number of patients assigned to practices in a given ICB
adhd_per_pop	Quantity of ADHD medication prescribed standardised for population
GDHI/head	Gross Disposable Household Income per person
unemployment	Unemployment Rate
f_life_expectancy	Average Female Life Expectancy
m_life_expectancy	Average Male Life Expectancy
avg_life_expectancy	Average Overall Life Expectancy
anxiety	On a scale where 0 is "not at all anxious" and 10 is "completely anxious", overall, how anxious did you feel yesterday? (0-10)
happiness	Overall, how happy did you feel yesterday? (0-10)
life_satisfaction	Overall, how satisfied are you with your life nowadays? (0-10)
worthwhile	Overall, to what extent do you feel that the things you do in your life are worthwhile? (0-10)

Table 4 - Table metadata for the regional analysis of ADHD prevalence

Visualisation

The resulting data table was loaded into Tableau, and the mapping tool was used to plot the relative values for each metric across the counties of England. Tableau's default county breakdown differs

from the ICGs in the data table, so the counties were manually matched up to the ICGs using the NHS England ICG breakdown document. Since the counties were only used to visualise the data, the maps accurately highlight high and low points nationally for the given metrics. Furthermore, a Pearson correlation matrix was calculated between all the variables using *pandas*; this allowed for a more objective way of identifying any correlated features.

3.1.3. Change over Time

Cost and Quantity by Drug by Month

The EPD dataset contains data from each month from January 2014 to June 2022, which allowed for the modelling of the national prevalence of ADHD over time. As it also describes which medication is prescribed, the relative use of each of the approved ADHD medications can be assessed. Cost and quantity data was retrieved from the NHS EPD API using a SQL query requesting the sum of total_quantity and the sum of actual_cost grouped by each ADHD drug. This request was iteratively made for each month of each year from 2014 to 2022, excluding July to December of 2022. The month and year were then added as columns in the dataframe to produce the data structure described in table 5. Price per unit was derived from the count and cost columns.

Variable	Description
year	Year of the source dataset
month	Month of the source dataset
chemical_substance_bdf_descr	Name of the ADHD drug
count	Total quantity of the drug prescribed that month
cost	Total cost to the NHS of the drug that month
price_per_unit	Cost of buying a single unit of a drug

Table 5 - Table metadata for the temporal analysis of ADHD prevalence and treatment

Visualisation

Firstly, the sum cost and quantity of all ADHD drugs were plotted against time to identify any general trends. Following this, the effect of COVID-19 on the prescription of ADHD drugs was investigated by dividing the data into two subsets – before March 2020 and after March 2020. Next, linear regression was performed on each of these subsets to obtain a growth coefficient pre-COVID-19 and post-COVID-19 describing the rate of increase of ADHD prescriptions; this allowed for modelling future growth in ADHD prescriptions. Furthermore, a more granular metric was calculated by performing the same regression on each year-long period to observe the change in growth over time. Linear regressions were performed using the *NumPy* library's *polyfit()* function, with an n value of 1 specifying a linear regression. For a specified number of dimensions, this function minimises the squared error:

$$E = \sum_{j=0}^{k} |p(x_j) - y_j|^2$$

The cost, quantity, and derived price per unit of each drug were then plotted over time to identify trends in the prescribing and price.

3.2. Results

3.2.1. Geographic Distribution

ADHD Prevalence

Two measures for regional ADHD prevalence were calculated from the NHS EPD dataset: the proportion of prescribed medications for ADHD (referred to as ADHD/total) and the number of units of ADHD medication prescribed per unit of population (referred to as ADHD/pop). These two measures are very closely correlated; this may be expected since both are derived from the count of ADHD medications but standardised using different features. The range of the data shows that the prevalence of ADHD is very different based on location. For example, the highest ADHD/pop county, Kent, had 3.24 times more ADHD cases per person than the lowest, Staffordshire. For the ADHD/total metric, Surrey had 4.62 times more ADHD medication prescribed per medication prescribed than Staffordshire.

The top 5 and bottom 5 ICBs for each measure are shown in tables 6 and 7. Surrey and Kent are in the top 5 for both measures, while the bottom 5 contains North East London, West Yorkshire, North Yorkshire, and Staffordshire. This similarity indicates that a combination of these metrics likely represents the actual ADHD prevalence.

Rank	Integrated Care Board	ADHD/total	Rank	Integrated Care Board	ADHD/Pop
1	NHS Surrey Heartlands	0.1623%	1	NHS Kent and Medway	0.1971
2	NHS Kent and Medway	0.1523%	2	NHS Cheshire and Merseyside	0.1766
3	NHS Mid and South Essex	0.1433%	3	NHS Surrey Heartlands	0.1668
4	NHS South West London	0.1364%	4	NHS Lincolnshire	0.1579
5	NHS Buckinghamshire, Oxfordshire, and Berkshire	0.1323%	5	NHS Norfolk and Waveney	0.1476

Table 6 - Top 5 integrated care boards for both ADHD/total and ADHD/population

Rank	Integrated Care Board	ADHD/total	Rank	Integrated Care Board	ADHD/Pop
38	NHS Black Country	0.05928%	38	NHS Somerset	0.08234
39	NHS North East London	0.05910%	39	NHS West Yorkshire	0.06835
40	NHS West Yorkshire	0.04620%	40	NHS North East London	0.06669
41	NHS Humber and North Yorkshire	0.04343%	41	NHS Humber and North Yorkshire	0.06361
42	NHS Staffordshire and Stoke-on-Trent	0.03515%	42	NHS Staffordshire and Stoke-on-Trent	0.06091

Table 7 - Bottom 5 integrated care boards for both ADHD/total and ADHD/population

The ADHD/pop and ADHD/total are geographically mapped in figure 4. Similarly, this figure demonstrates the close relationship between ADHD/total and ADHD/population. Both maps show particular hotspots in the counties surrounding London, with other high points around Cheshire, Lincolnshire, and Northamptonshire. Conversely, all parts of Yorkshire are consistently low, as are

most counties along the south coast (Somerset, Sussex, Hampshire). Staffordshire is an odd outlier, surrounded by much darker areas (including Cheshire, one of the top 5 for ADHD/pop) yet being the lowest rated in both metrics. One significant outlier is London; many counties surrounding London have very high ADHD prevalence, but London appears relatively light on the map. This finding is likely due to how the data is displayed on the map – 5 separate ICBs represent the 'London' area (as defined by Tableau) in the NHS dataset. These include very high areas (South West London) and very low areas (North East London), which average out to a value representing neither area.

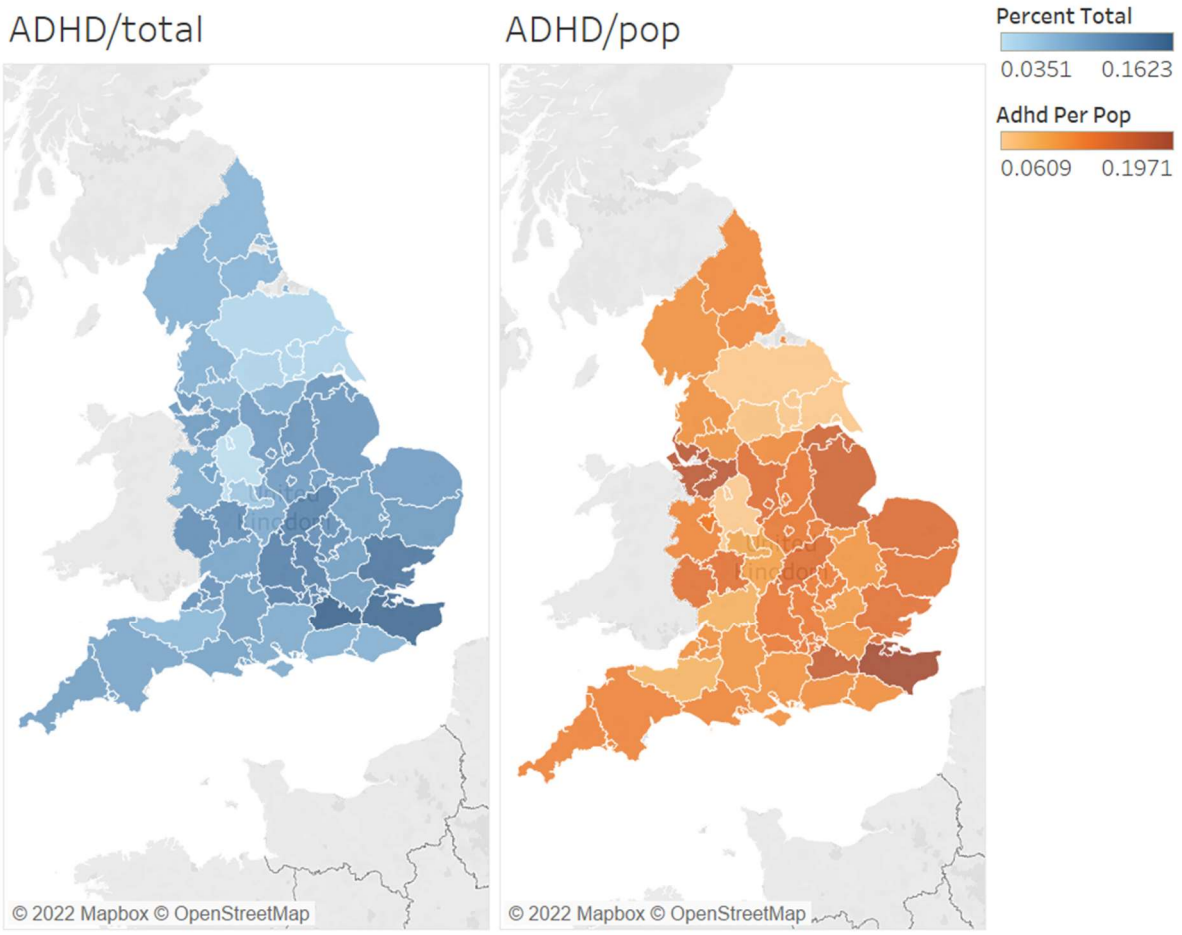


Figure 4 - Map of ADHD/total and ADHD/population by English county

Socioeconomic Factors

The two ADHD prevalence metrics were then compared to a range of other socioeconomic features to identify any correlation with ADHD prevalence. While this correlation does not provide hard evidence of any causality with ADHD, it does provide a springboard to think about how ADHD is diagnosed and its impact on people's lives. The measures used in this study were unemployment rate, gross disposable household income, average life expectancy, and a further four measures of well-being – anxiety, happiness, life satisfaction, and worthwhileness. The correlation matrix for all of these measures is shown in figure 5. This figure shows the correlation between ADHD/total and ADHD/pop with each socioeconomic measure. For this figure, the correlation category boundaries are shown in table 8.

Category	r value	Category	r value
Strong Negative	r <= -0.7	**Strong Positive**	r >= 0.7
Negative	-0.5 <= r < -0.7	**Positive**	0.7 > r >= 0.5
Weak Negative	-0.2 <= r < -0.5	**Weak Positive**	0.5 >= r > 0.2

Table 8 - r value boundaries for correlation categories

The most prominent observation is the 0.77 correlation coefficient between ADHD/total and ADHD/pop, confirming that these two variables are strongly correlated. However, despite these being so correlated with each other, they display differing correlations with the other metrics. This fact is evident for GDHI/head; ADHD/total has an r value of 0.44 (weak correlation), while ADHD/pop has an r value of 0.015 (no correlation). This suggests that places with more economic wealth experience ADHD as a more substantial healthcare burden than other health conditions. Similarly, life expectancy correlates much more with ADHD/total (r=0.47) than ADHD/pop (r=0.22). Another interesting observation is that unemployment is correlated weakly with ADHD/total (r=-0.32) and ADHD/pop (r=-0.29) to almost the same degree. All of the well-being measures showed no significant correlation with either ADHD metric. All socioeconomic measures employed are mapped alongside ADHD/pop in figure 6. This figure reiterates the lack of a strong correlation between ADHD prevalence and socioeconomic factors.

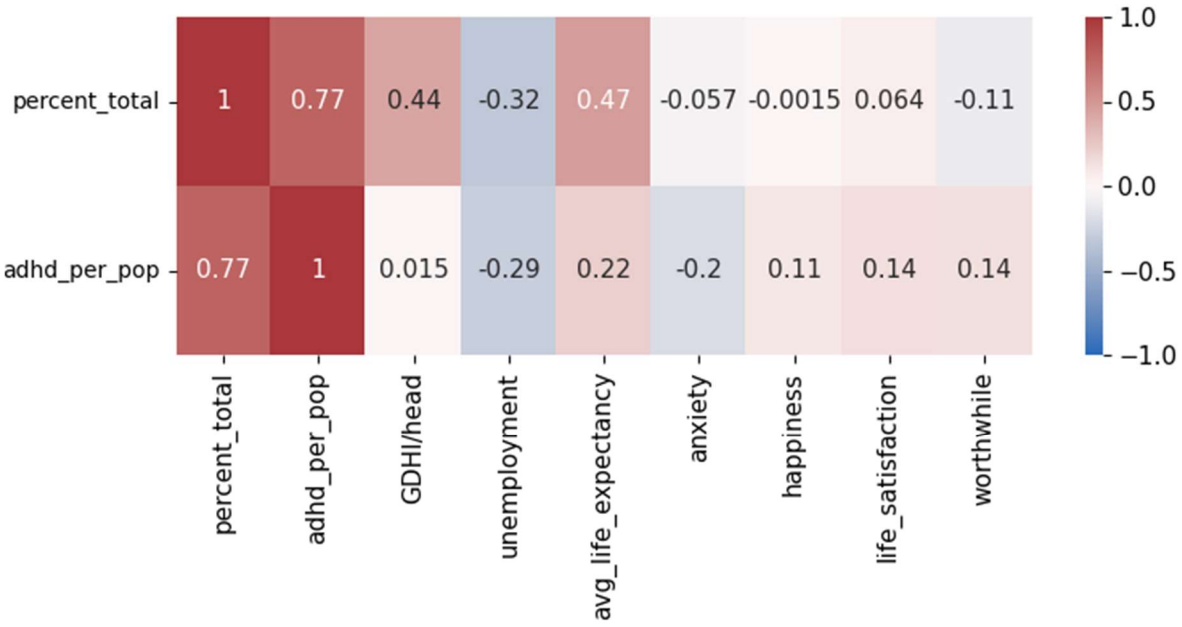


Figure 5 - Correlation matrix of ADHD prevalence and a range of socioeconomic measures

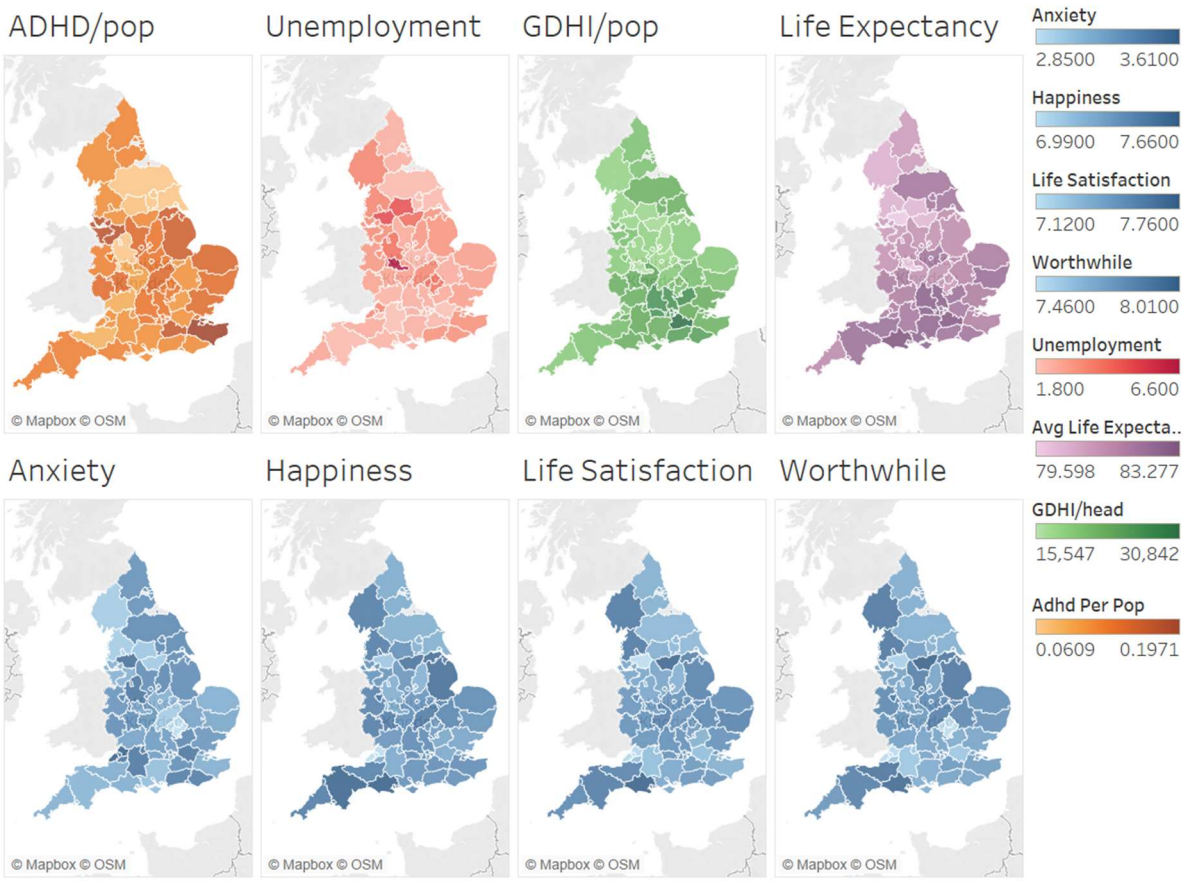


Figure 6 - Map of ADHD/population against maps of 7 other socioeconomic measures

3.2.2. Cost and Quantity over Time

The time-series data from the EPD dataset shows an increase in the prescription of ADHD medications. Figure 7a shows the total quantity of all medications prescribed for ADHD since 2014; the quantity increased linearly until 2020 when the increase became exponential. This observation is supported by the cubic trendline, indicating an exponential factor. The plot seemed to be stable until early 2020 when the number of prescriptions suddenly dropped before beginning an exponential increase. This time point aligned almost perfectly with the beginning of the COVID-19 pandemic, so this relation warranted further investigation.

Figure 7b shows the same plot as figure 7a but is divided into pre-COVID-19 and post-COVID-19, with March 2020 as the cut-off. The red lines represent a linear regression performed on each part of the graph; before March 2020, the increase was 703.7 units/day; after March 2020, the increase was 1991.0 units/day. Figure 7b also shows the growth coefficient calculated for each year as bars. This plot demonstrates the massive spike in prescriptions for ADHD since March 2020, with 2021 increasing at 2443.91 units/day and 2022 increasing at 4996.39 units/day. This change strongly suggests that the COVID-19 pandemic significantly impacted ADHD diagnosis and treatment.

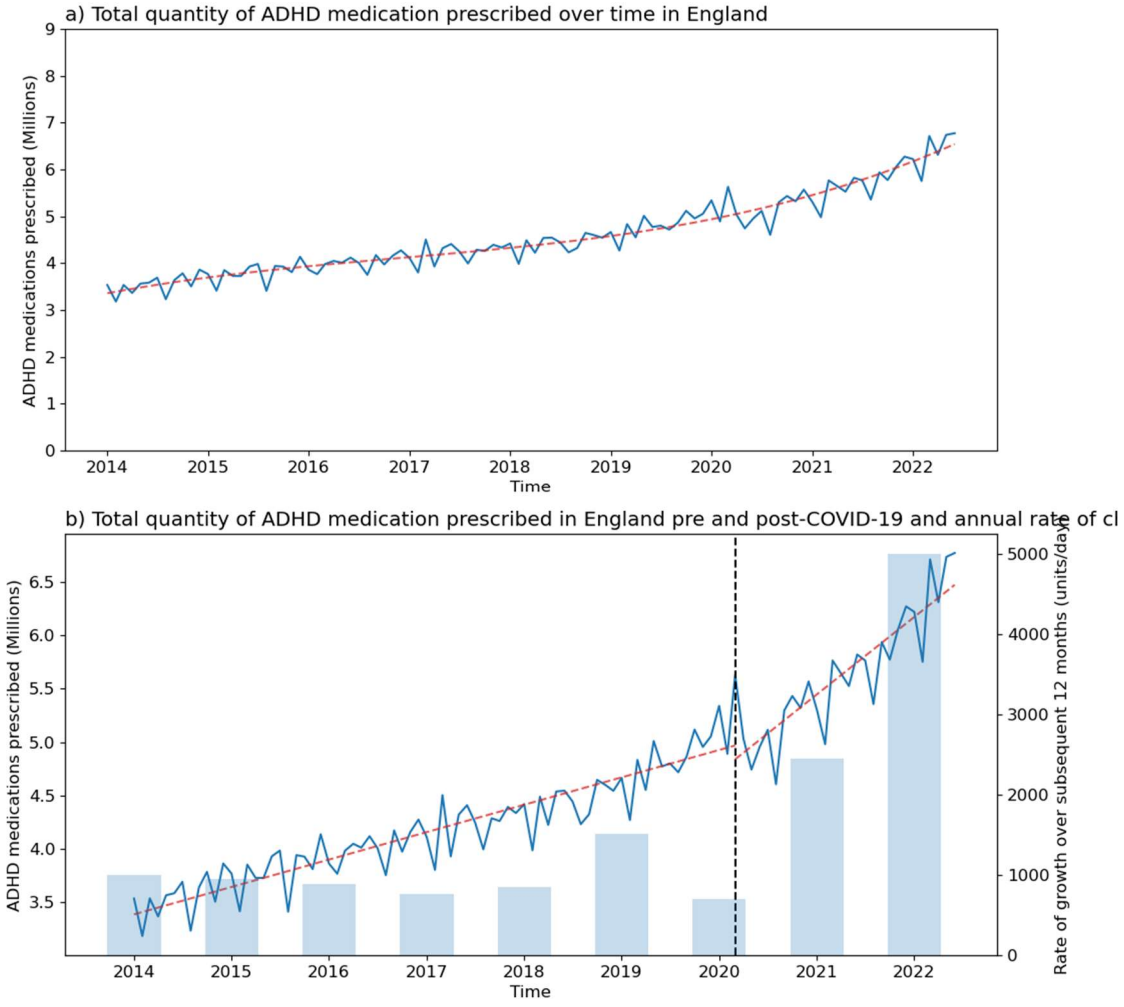


Figure 7 – a) Quantity of ADHD prescriptions between 2014 and 2022 b) Trend in prescriptions for ADHD before and after COVID-19

The total cost and quantity of ADHD prescriptions were divided into five individual medications; figure 8a shows the total number of each medication prescribed. Methylphenidate is the most prescribed medication for ADHD; this is expected, as it is the first-line drug for treating ADHD in the UK, according to the NICE guidelines. All four other drugs (Atomoxetine, Guanfacine, Dexamfetamine, and lisdexamfetamine) are prescribed at very similar rates - the most notable trend amongst these is the exponential increase in Lisdexamfetamine prescriptions. At the start of 2014, Lisdexamfetamine was prescribed over 20,000 times compared to the next lowest drug, atomoxetine, at just over 270000 times. In late 2019, just over five years later, lisdexamfetamine overtook atomoxetine and dexamfetamine to become the second most prescribed medication for ADHD. As of June 2022, almost 1,000,000 units of lisdexamfetamine are prescribed monthly, compared to just over 500,000 units of atomoxetine and dexamfetamine. Guanfacine, atomoxetine and dexamfetamine seem to be increasing much slower, implying that not many new patients are being prescribed these drugs.

The cost plot shown in figure 8b looks similar to the count plot in figure 8a but with a few notable differences. The most obvious difference is that the exponential increase in lisdexamfetamine dose is far more pronounced. The price/unit plot in figure 8c shows that lisdexamfetamine has been consistently priced at around £2.20/unit since 2014. While this is comparable to guanfacine, it far exceeds the pricing for methylphenidate and dexamfetamine, which generally sit at around £0.80/unit. As a result, the increase in prescriptions for lisdexamfetamine has corresponded to an even greater increase in cost to the NHS.

The other noteworthy feature of this plot is the cost and pricing of atomoxetine. In early 2014 atomoxetine was the second highest cost to the NHS of all ADHD medications; however, as of March 2022, atomoxetine was the lowest expense to the NHS out of all five drugs. This change is reflected in the pricing plot, where the price per unit of atomoxetine drops from £2.10 to £1.76 between March and April 2016. In 2021, the price of atomoxetine plummeted; as of June 2022, it currently sits at £0.55 per unit, the cheapest of all ADHD drugs. The impact that pricing can have on the total cost to the NHS emphasises how important health policy is when considering a publicly funded health service.

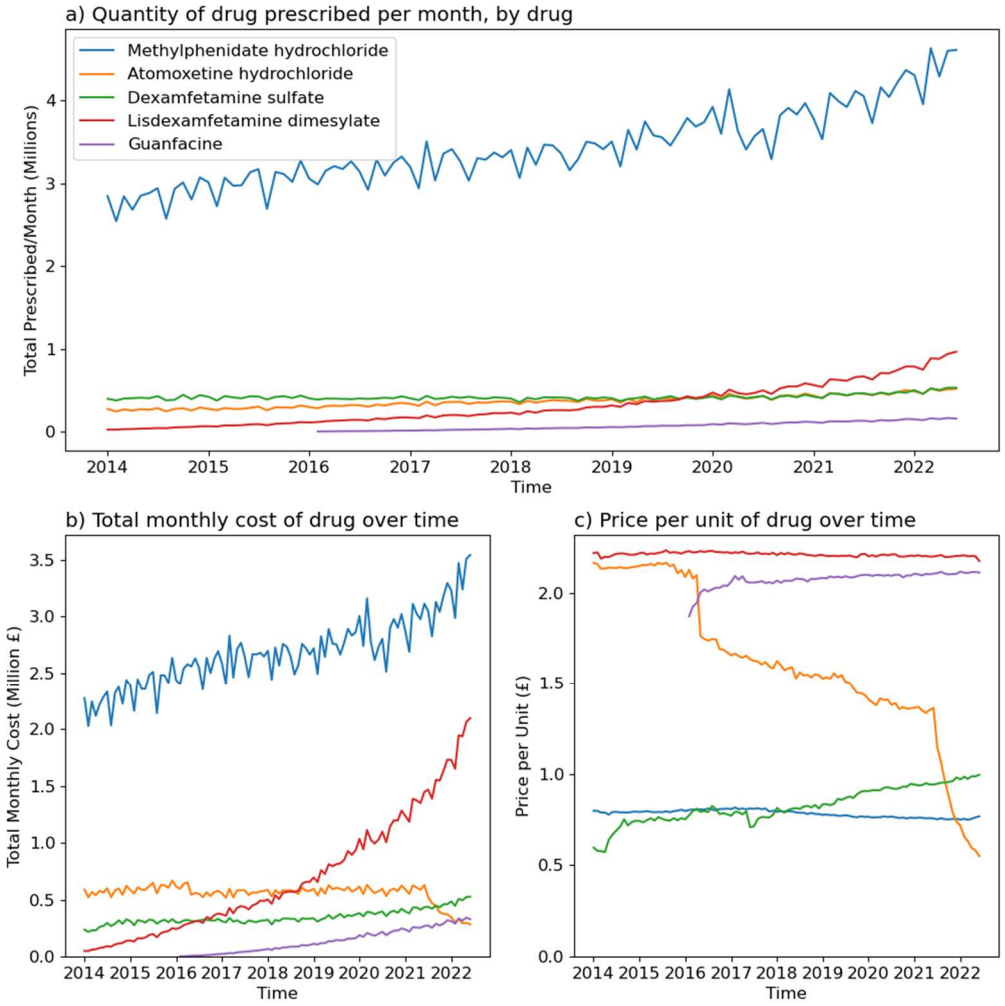


Figure 8 – a) Quantity of ADHD drugs prescribed between 2014 and 2022 b) Total cost of ADHD drugs between 2014 and 2022 c) Price of ADHD drugs between 2014 and 2022

3.3. Discussion

3.3.1. Findings

Impact of Features on Prevalence

As seen from the map, there is a clear difference in the prevalence of ADHD in different counties across England, with the highest county having 3.2 times more cases per person than the lowest. The two most notable correlations in the gathered data were between the proportion of all medications prescribed for ADHD and GDHI/person (0.44) and average life expectancy (0.47). The implication is that areas where people have more disposable income and live longer experience ADHD more often compared to other health conditions. There are two ways to increase this ADHD metric – higher ADHD prevalence or lower prevalence of other conditions; it is essential to consider which is at play here. For GDHI/person, there is almost no correlation with ADHD cases/person (0.015), suggesting that an increased GHDI reduces the frequency of other medical conditions rather than increasing ADHD prevalence.

On the other hand, life expectancy correlates with ADHD cases/person, although slightly weaker (0.22). This slight correlation implies that areas with a longer life expectancy experience ADHD more frequently than those with a shorter life expectancy. Life expectancy is a complicated metric influenced by several socioeconomic conditions in an area but is heavily tied to access to healthcare and behavioural health risks (TheKing'sFund, 2022). Therefore, it is not unreasonable to assume that at least part of the reason for the disparity in ADHD prevalence across England is due to an inability to access the appropriate healthcare resources to receive a diagnosis. One way to address the accessibility problem is to provide tools that are available to everyone – this is where machine learning can help, putting a great deal of diagnostic power in the hands of anyone with access to a computer. This feat is feasible through training a machine learning model that could provide diagnostic classification based on input data that does not require specialist tools to obtain.

Effect of COVID-19

The general trend for ADHD treatments has increased exponentially, especially since the start of the COVID-19 pandemic. When comparing pre-COVID and post-COVID trends, the gradient of the trendline has increased by 2.8 times. The pre-COVID model predicted 6.2 million medications dispensed per month by 2025, compared to the post-COVID model's prediction of 8.3 million. It is unclear from this data why this is the case; however, the diagnosis rate may have increased due to the immense lifestyle changes that accompanied the lockdowns due to the pandemic. For example, the rise in home working for both adults and children has led to parents spending more time observing their children's behaviour, which could increase recognition of ADHD symptoms (Felstead and Reuschke, 2020). Further work should be carried out to investigate the potential causes of this shift in trend.

Prescribing of Medications

While methylphenidate remains the drug of choice for ADHD in England, lisdexamfetamine is becoming a much more common option. This trend could be potentially problematic for the NHS, as lisdexamfetamine is the most expensive option for ADHD, with an associated cost exceeding £2.00 per tablet. Furthermore, the rate of lisdexamfetamine prescription is increasing much faster than that for methylphenidate, so the impact of this cost disparity will become more and more apparent with time. One solution to this problem would be to encourage the prescription of dexamphetamine over lisdexamfetamine since dexamphetamine is under half the price per tablet. Lisdexamfetamine is a pro-

drug, meaning it gets converted into dexamphetamine in the body; as a result, the main difference between the two treatments is the pharmacokinetic profile, with the efficacy of both drugs being very similar. Such a switch from lisdexamfetamine to dexamphetamine could save the NHS a considerable amount – around £1 million per month. Additionally, the price of atomoxetine has decreased substantially over the past eight years; this could be a much cheaper option, provided the efficacy and safety profiles are comparable to first-line treatments.

3.3.2. Limitations

Many assumptions needed to be made to interpret the NHS prescribing data; firstly, it is assumed that the number of people being treated for ADHD is equivalent to the number of people diagnosed with ADHD, which may not be the case. Some people with ADHD may choose not to be on treatment, which is not considered in this analysis. However, to compare regions of England, this assumption is fair as this proportion is likely to be similar regardless of location, meaning that the numbers will all still be proportionate. Similarly, this analysis does not account for someone being prescribed multiple drugs for the same instance of ADHD. A further assumption is that all people receiving these five treatments are being treated for ADHD – while this assumption is reasonable because these medications are only licenced for ADHD, it is possible that they are also being prescribed for off-label indications. Furthermore, some medications need to be taken multiple times a day while some are only taken once; this prevents parallels from being drawn between the number of cases of ADHD and the number of medications prescribed. However, as long as the ratio of these two types of medication is consistent nationally, the comparison should still hold.

One of the most significant limitations in this analysis is data mapping onto the map in Tableau. Values were obtained for specific ICBs from the NHS dataset, which does not match the provided zones for mapping in Tableau. Consequently, some areas had to be mapped one-to-many and others many-to-one. The one-to-many approach is problematic because while the value is valid for the whole area, the individual sub-areas will not be accurately labelled; this is an acceptable compromise, as the map provides a reasonable comparison point visually. However, the more difficult compromise is the many-to-one mapping. The prime example of this is the greater London area, assumed to be one zone in Tableau but made up of 5 distinct ICBs. As a result, the mapped value is an average of all 5 (assuming equal population), which is not representative of any areas due to a lack of population standardisation.

Additionally, a change in the zoning for ICBs in 2020 made it difficult to assess the regional changes over time. Future work could look at which regions have a rapidly growing prevalence or if ADHD is becoming less prevalent in any regions across England.

Chapter 4

Modelling ADHD in Adolescents

4.1. Methods

One of the most significant findings raised in the exploratory analysis was the potential for a tool to facilitate the diagnosis of ADHD. In the next part of this project, behavioural data is used to train a machine learning model to classify individuals as ADHD positive or negative. As this data is accessible, such a tool should be able to improve access to healthcare for individuals with ADHD.

4.1.1. ABCD Dataset

This portion of the project investigated how ADHD symptomatology could be modelled using machine learning methods. Models were trained using data from the Adolescent Brain Cognitive Development (ABCD) Study (Yang 2021). This study collected a range of data from adolescents at regular intervals in order to monitor their cognitive development. Release 4.0 of the dataset was used, containing data from 11,876 participants; the range of data tables in this data release is vast, including questionnaires, medical history, task performance data, personality data, physical exam data, demographic data, and more (complete data dictionary available in Appendix E). As this study used sites across the US, care needs to be taken when considering this data in the context of findings from the previous chapter, ad many of the assumptions for the NHS dataset do not hold for American participants.

This dataset was accessed from the National Institute of Mental Health (NIMH) through the NIMH Data Archive (NDA). This research was carried out under the data use certification granted to Thomas Lancaster at the University of Bath, covering research on the relationships between psychopathology and other metrics. Data was hosted on a cloud-based server and accessed directly in a read-only format by a python script.

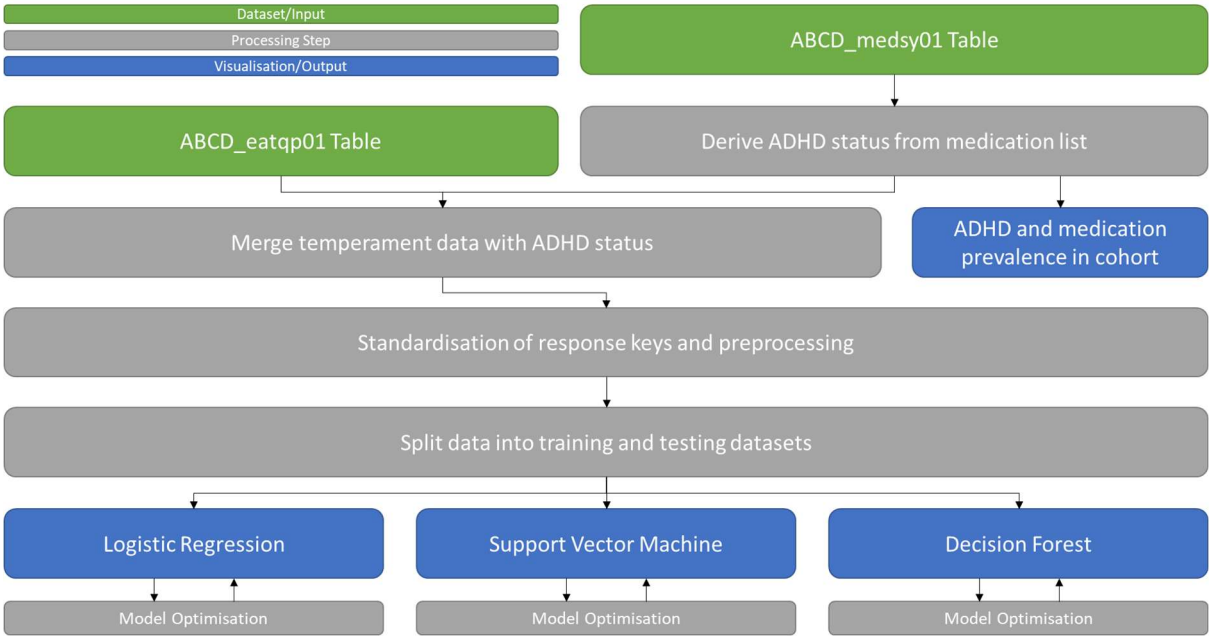


Figure 9 - Process flow of the processing of ABCD data

4.1.2. Medication Log

One of the biggest challenges with this dataset was identifying participants with and without ADHD. While the dataset records medical history, the format of their forms does not include any way to extract information about diagnosed conditions. The workaround was to make some similar assumptions as those made in chapter 3 – that anyone on medication for ADHD had ADHD, and anyone with ADHD was on ADHD medication. The ABCD study dataset contains a table called 'ABCD Parent Medications Survey Inventory' (medsy01) which contains data about any medications the participants were taking. One of the advantages of using this form is its completeness, as it contains data for all 11,867 participants.

Unfortunately, the mesdy01 form recorded this data in a rather unfriendly format, so significant pre-processing was required. First, each participant had 15 groups of columns numbered 1 to 15, referring to the 15 slots in which parents could enter medications their child is on; this made it challenging to filter rows based on the six identified ADHD medications. Furthermore, there was no standardisation of the drug names entered in the free-text field, so the result was a mixture of drug names and brand names, with or without dose. The solution to this problem was to iteratively run a processing step on the data frame, from i = 1 to i = 15 representing each group of columns in the data table. Within each loop, the code iterated through a list of terms, including brands and drug names, and looked for drug names matching each (case insensitive). If a match was found, that subject's key and the matched term were added to a global data frame. The terms used are shown in table 9 (WebMD, 2021).

Dexedrine	Zenzedi	Adderall	Focalin	Methylin
Ritalin	Methylphenidate	Dyanavel	Amphetamine	Dextroamphetamine
Evekeo	Mydayis	Vyvanse	Aptensio	Concerta
Lisdexamfetamine	Quillivant	Quillchew	Azstarys	Metadate
Cotempla	Atomoxetine	Strattera	Guanfacine	Tenex
Intuniv	Clonidine	Kapvay		

Table 9 – Terms in the ABCD medication dataset identified as relating to ADHD medication

Once all the rows had been processed, the resulting data frame was pivoted using the *pandas* pivot_table() function, using the subject key as the index and medication as the column. These columns were summed in groups to produce a table of the identified subjects and the ADHD drugs they were taking. The subject key column was then used in further analysis to identify participants with ADHD.

4.1.3. Temperament

The first table of interest identified was the ABCD Early Adolescent Temperament Questionnaire Parent form (eatpq01), which addresses eight dimensions of behaviour – activation, affiliation, attention, fear, frustration, surgency, inhibitory control, and shyness. These dimensions were of particular interest due to how closely 'attention' and 'inhibitory control' align with the symptoms of ADHD and to see how the other dimensions correlate with these. First, the data table was merged with the table generated from medsy01 to create a new column, 'ADHD', with a value of 0 or 1 indicating the lack or presence of ADHD, respectively. Beyond this, some additional pre-processing was required. Answers in the questionnaire were rated one to five; however, not all scales operated in the same direction (i.e., sometimes five indicated agreement while other times five indicated disagreement with the statement). Therefore, results were standardised so that five always indicated agreement, and one always indicated disagreement. This standardisation facilitated the interpretation of correlation coefficients later on.

The data was separated into a training set (80%) and a testing set (20%) using the train_test_split() function with a random state = 50. The y values (ADHD) were used to stratify the training and testing data, ensuring that the relative class frequencies were preserved. Since the non-ADHD participants far outnumber the ADHD participants by about 9:1, this stratification prevents the testing dataset from only containing non-ADHD participants.

Logistic Regression

The first model explored was a logistic regression, using the data from the eatpq01 table to predict whether the participant had ADHD. The logistic regression model used a limited-memory Broyden-Fletcher-Goldfarb-Shanno (lbfgs) algorithm as the solver for the optimization problem.

The training data was used to train a logistic regression model using the sklearn LogisticRegression object. The class weights were initialised as the inverse of the frequency ratio (0: 1258, 1: 9073) as is standard for imbalanced class frequencies. For these models, performance was measured through several metrics: a confusion matrix of the true positive count, true negative count, false positive count, and false negative count; accuracy, precision, and recall scores; f1 score; and area under the curve (AUC).

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negatuve}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}$$

The weighting ratio for the model was iteratively tested across a range of ratios (0:1 to 2:1), and AUC was plotted. Subsequently, a narrower window of ratios around the previous peak was tested to optimise the model.

Once the optimal class weight had been determined, all of the dataset's features were ranked from most influential to least influential based on the absolute value of their coefficient value in the model. Next, another set of models was trained, using an increasing number of features each time to observe the impact on the model accuracy. This process determined the number of features necessary to predict the presence of ADHD, at which point using more features offered no added predictive capability.

Support Vector Machine

An SVM model was also trained using the svm.SVC object from *sklearn*. Similarly to the logistic regression, the performance was assessed primarily using the AUC score, with additional contribution from the recall and accuracy scores. These were assessed over a range of class weights to determine the optimal weight values for the model.

Decision Forest

Finally, a decision forest model was trained using the RandomForestClassifier from *sklearn.ensemble*. The model was trained using 100 individual decision trees, and models were tested over a range of class weights to observe optimal performance. Again, the measures used for the model's performance were AUC, recall and accuracy.

4.2. Results

4.2.1. Medication Log

The processing of the medication data revealed that 1416 participants were taking medications used for ADHD; this made up about 11.9% of the study population, as represented in figure 10a. Figure 8b shows the breakdown of which ADHD medications participants were taking – methylphenidate was the most common, with 819 participants taking it, while atomoxetine was the least common, being taken by only 80 participants.

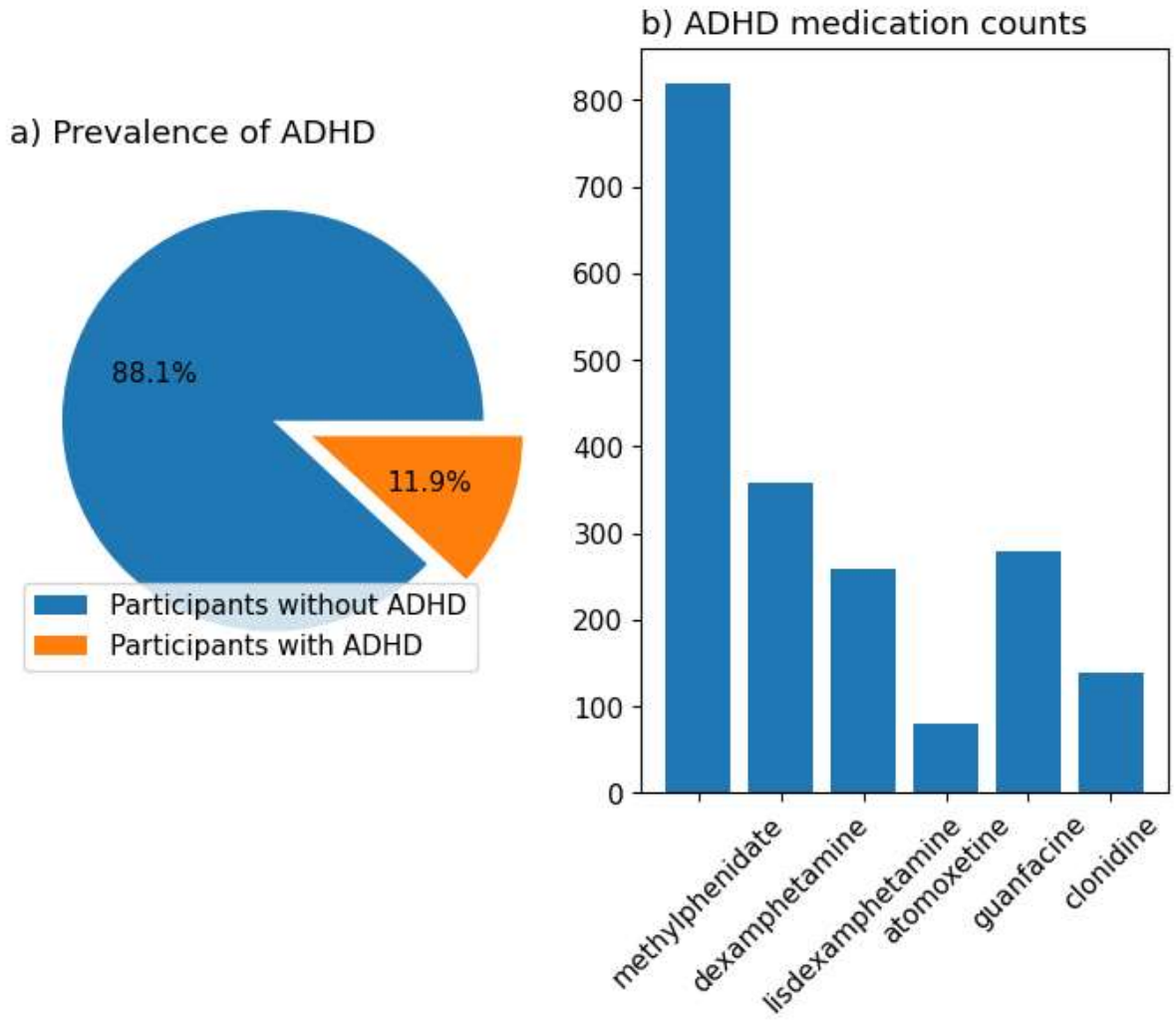


Figure 10 - a) The prevalence of ADHD in ABCD study b) Breakdown of ADHD medications taken in the ABCD study

Some participants were taking more than one different ADHD medication, hence why the sum of the medication counts exceeds the number of cases of ADHD. For example, 321 participants took two different medications, 79 took three, and 14 took four. The most common concurrence was methylphenidate and guanfacine (130 participants) and methylphenidate with dexamphetamine (76 participants). Notably, guanfacine co-prescribed with methylphenidate made up almost half (130/279, 46.6%) of guanfacine occurrences in the medsy01 table. All concurrences are displayed in table 10.

	methylphenidate	dexamphetamine	lisdexamfetamine	atomoxetine	guanfacine	clonidine
methylphenidate	819	76	62	24	130	57
dexamphetamine	76	359	42	12	51	40
lisdexamfetamine	62	42	258	14	46	29
atomoxetine	24	12	14	80	15	19
guanfacine	130	51	46	15	279	25
clonidine	57	40	29	19	25	138

Table 10 - Concurrence of ADHD medications in the ABCD cohort

The CDC estimates that around 9.8% of American children are diagnosed with ADHD, increasing to 10% between ages 6-11 and 13% between ages 12-17 (Centers for Disease Control and Prevention, 2022). This finding generally aligns with the value derived from the ABCD dataset of 11.9% since the participants of this study are split across the 6-11 and 12-17 age groups; hence it is reasonable to conclude that the employed method of recognising participants with ADHD was successful.

4.2.2. Temperament

Of the participants enrolled in the study, 10331 had data available in the eatqp01 table – 9073 without ADHD and 1258 with ADHD.

Logistic Regression

After training, the optimal model achieved a peak AUC value of 0.7451 at a weighting ratio of 0.1110:1; figure 11a/b shows how the AUC changes as the weight of the outcomes were changed during training. The metrics for this model are shown in table 12. This AUC score shows that the model's performance is better than random, indicating that data from these questionnaires is correlated with the presence of ADHD in participants. The high recall and accuracy scores (0.8016 and 0.7025) also suggest that the model is good at identifying ADHD cases. However, the low recall score of 0.2634 suggests that many negative cases are misidentified as ADHD.

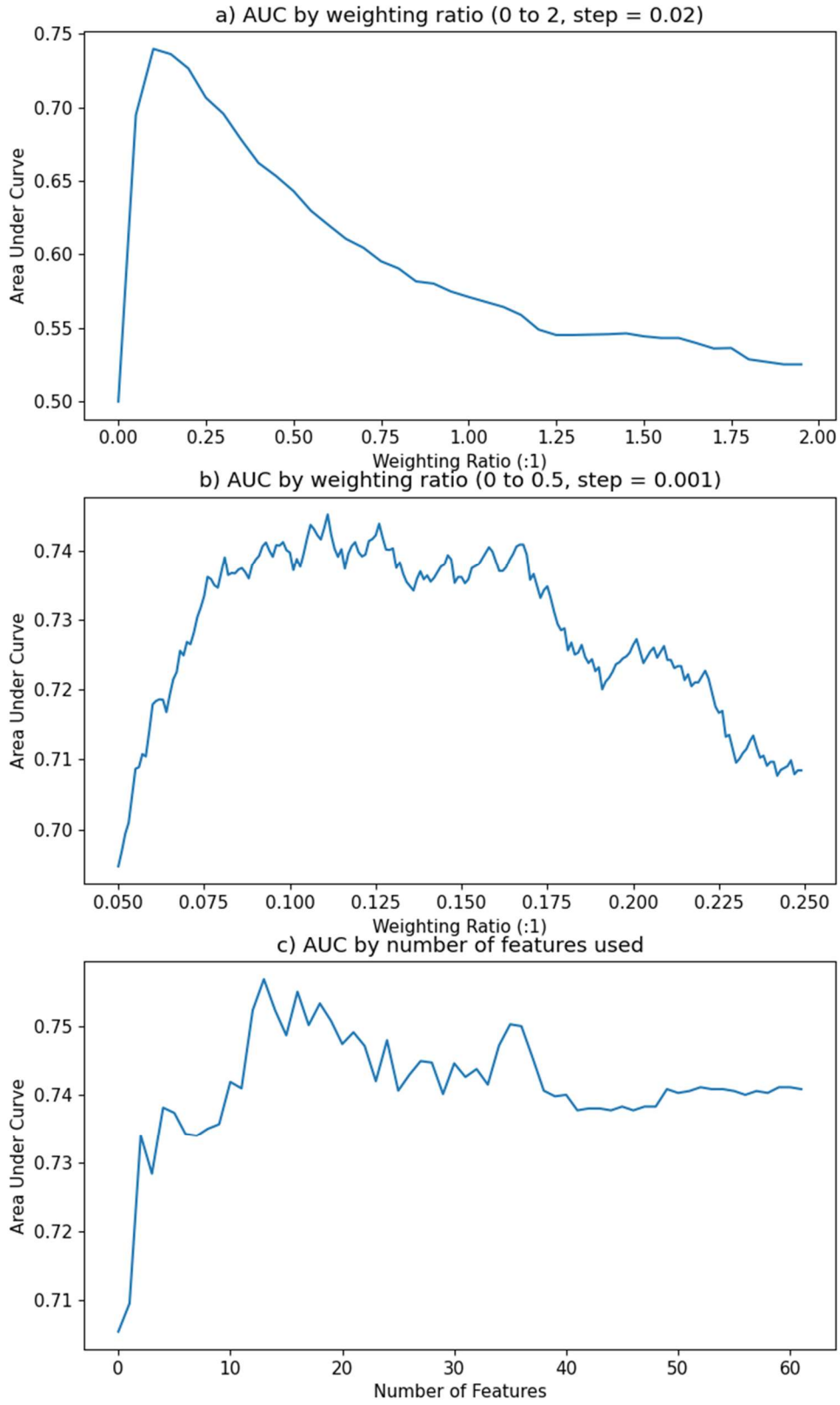


Figure 11 – AUC score of the logistic regression model: a) as the weighting of outcomes was varied during training b) over a narrower range with smaller steps c) trained on the top n features

The features deemed most and least important to the model are summarised in table 11 below. Each statement has a calculated odds ratio – this translates to the change in the likelihood of having ADHD given a one-point increase in that statement. The most noteworthy feature in the model was the ability of a participant to keep track of multiple things happening around them, with an odds ratio of 0.645; this is to say that for every unit increase in the response to this question, the participant was 0.645 times as likely to have ADHD. On the other hand, the least influential feature was participants feeling shy about meeting new people, with an odds ratio of 0.998, implying that this response had almost no impact on the model's classification of a participant.

Rank	Statement	Odds Ratio
1	Is good at keeping track of several different things that are happening around her/him	0.645
2	Finds it easy to really concentrate on a problem	0.770
3	Pays close attention when someone tells her/him how to do something	0.789
4	Usually finishes her/his homework before it's due	0.824
5	Is usually able to stick with his/her plans and goals	0.830
6	Gets irritated when s/he has to stop doing something s/he is enjoying	1.192
7	Has a hard time finishing things on time	0.884
8	Feels like crying over very little things on some days	0.852
9	Usually puts off working on a project until it is due	1.169
10	Likes meeting new people	0.863
…	…	…
53	Feels scared when entering a darkened room at night	0.990
54	Wants to have close relationships with other people	1.010
55	Can generally think of something to say even with strangers	0.990
56	Is not shy	0.991
57	Is energized by being in large crowds of people	0.991
58	If having a problem with someone usually tries to deal with it right away	1.008
59	If very angry might hit someone	1.005
60	Makes fun of how other people look	0.996
61	Likes to be able to share his/her private thoughts with someone else	0.996
62	Feels shy about meeting new people	0.998

Table 11 - The ten most and ten least indicative features of ADHD

Based on these features, the AUC was calculated using the top n features for n = 1 to n = 62, shown in figure 11c. The model reached peak predictive power when trained on the top 14 features, beyond which the features provided no improvement to the prediction.

Support Vector Machine

The optimal SVM model achieved an AUC score of 0.0.7715, with an accuracy score of 0.6918 and a recall score of 0.8770. This model used a class weight of 0.11:1, which matches the optimal ratio for the logistic regression model. The AUC plot recorded during the training process is shown in figure 12a/b. Further analysis, shown in figure 12c, revealed that this model performed best when trained using only the top 16 features (as identified by the logistic regression). The SVM model slightly outperformed the logistic regression model in the AUC score, while both the recall and accuracy scores were comparable for both models. However, similarly to the logistic regression model, this model had a very low precision of 0.2722, meaning it often misdiagnosed control cases as ADHD cases.

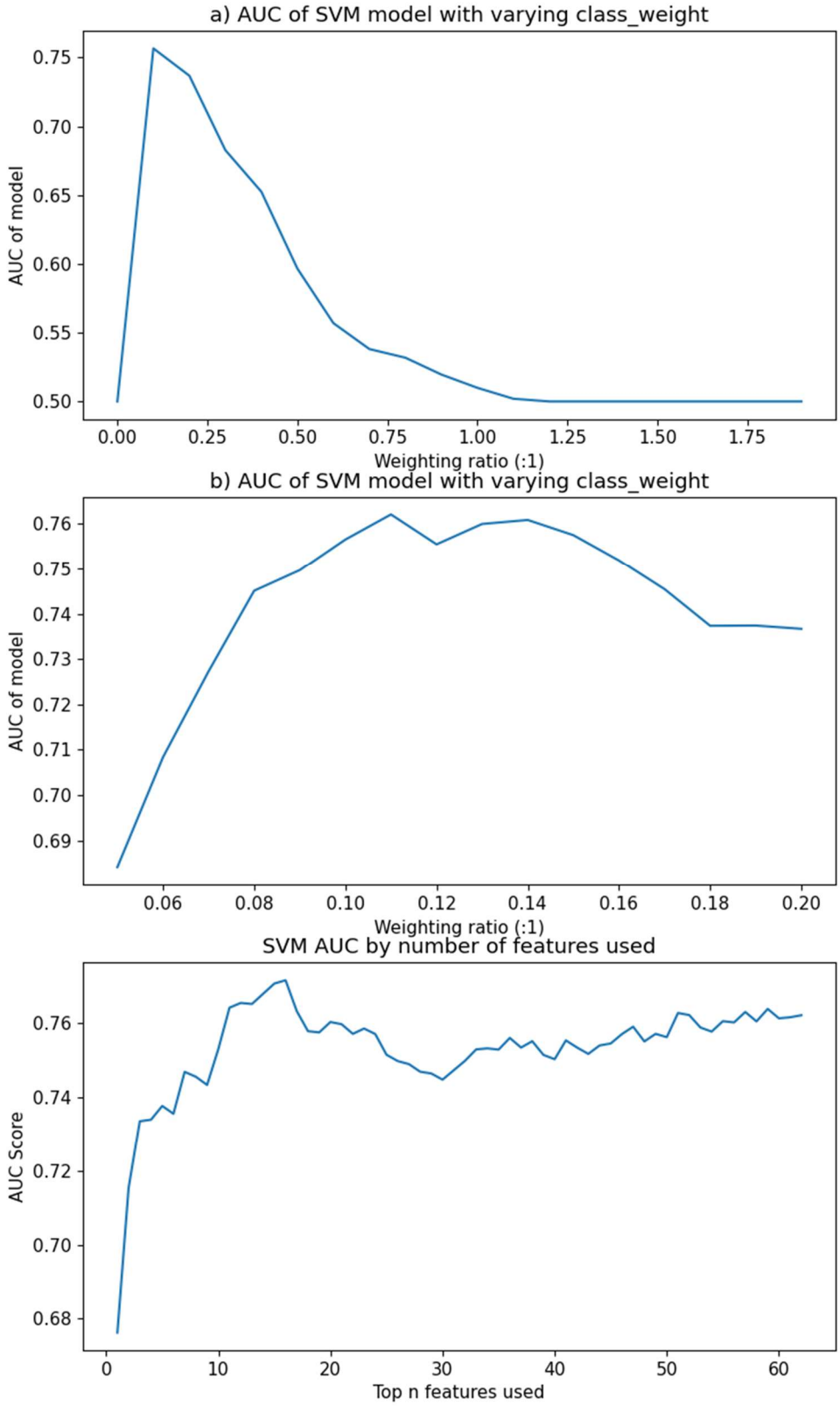


Figure 12 - a) AUC score of the SVM model as the weighting of outcomes was varied during training b) over a narrower range with smaller steps c) trained on the top n features

Decision Forest

The decision forest model was by far the poorest performing of all the models. The highest AUC achieved after iterating across a wide range of weight values was 0.5455. For some reason, no matter what weighting was used for the training, the model would rarely ever predict a case as ADHD – the optimal model only predicted 11 ADHD cases compared to 2056 non-ADHD. This occurrence led to the poor recall and precision scores of 0.0238 and 0.0456. This model performs no better than a random chance model; it could be the case that this model would perform much better with specific optimisations, but to do so is outside of this project's scope due to time constraints.

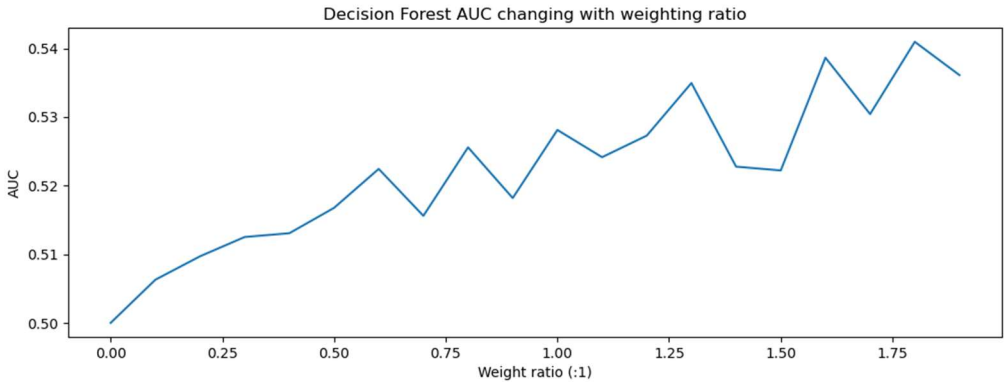


Figure 13 - AUC score of the decision forest model as the weighting of outcomes was varied during training

Model	Accuracy	Precision	Recall	F1 Score	AUC
Logistic Regression	0.7025	0.2634	0.8016	0.3965	0.7451
SVM	0.6918	0.2722	0.8770	0.4017	0.7715
Decision Forest	0.8786	0.0456	0.0238	0.5455	0.5105

Table 12 - Comparison of all metrics for all optimal models

4.3. Discussion

4.3.1. Findings

ADHD Medications in US

The main takeaway from the analysis of the medsy01 data is the differences in drug prescriptions in the US compared to the UK (as seen in chapter 3). While methylphenidate is the majority drug in both populations, American citizens were next most likely to take dexamphetamine, while English citizens were more likely to take lisdexamfetamine. Furthermore, guanfacine is barely prescribed in the UK, being the rarest ADHD medication by far; conversely, it was the 3rd most common just behind dexamphetamine in the US. Likely, such trends are purely due to the differences in the healthcare systems in the two countries; however, it would be interesting to discern through further study the causes and implications of this trend.

Importance of Features

The logistic regression model's coefficients provide valuable insight into the model's understanding of ADHD symptomatology. Out of the top 10 most correlated features, almost all can be related to one of the ADHD diagnosis criteria listed in the DSM-V under the three subcategories 'inattention', 'hyperactivity' and 'impulsivity'. Interestingly, the top-rated feature of 'keeping track of multiple things happening around them' does not align closely with any diagnosis criteria; this feature suggests

an inability to multitask or focus attention on multiple stimuli. It could be argued that this is distinct from the general lack of attention implied by many other diagnostic criteria. Further investigation should be carried out to investigate this as a critical symptom of ADHD, from which the criteria for diagnosis may be adjusted.

One of the interesting features of this questionnaire is that the questions were very clearly targeted towards behaviours in children, whereas the diagnostic criteria in the DSM-V are much more generalised. Since ADHD is so commonly diagnosed in children rather than adults, it could be advantageous to provide diagnostic criteria that are much more relatable to a child's lifestyle. Doing so could make diagnosis smoother by allowing accurate recording of the symptoms rather than running them through the psychiatrist's understanding of the response first.

Comparison of Models

Three different models were trained and refined for this classification problem: a logistic regression, a support vector machine, and a decision forest. The decision forest performed very poorly, while the logistic regression and the SVM demonstrated some ability to distinguish participants with ADHD from those without ADHD. Both successful models tended to over-diagnose ADHD in the population, which led to their low precision scores; however, in the case of medical diagnosis, this is often advantageous, as it is preferable to over-diagnose than under-diagnose. The models had impressive recall scores, both greater than 0.8 – this measures the models' ability to identify the positive cases correctly, which it did well. These results highlight the potential for using these models as a preliminary diagnostic tool to screen out some individuals without ADHD. At the same time, its high recall ensures that few participants with ADHD are given an incorrect result.

4.3.2. Limitations

The primary limitation of this study was the lack of data about the ADHD diagnosis status of the participants, which had to be inferred from the medication data. The result was that all of the cases identified in the data were ADHD that was being treated. Consequently, what the models learned to detect were the symptoms of ADHD under treatment, which may be vastly different from the symptoms of untreated ADHD. As a result, the application of these models to the diagnosis of ADHD in the clinic may be limited. On the other hand, it could be argued that a model that can detect the subdued symptoms of treated ADHD may be even better at identifying untreated ADHD – further study would need to be carried out to test this hypothesis.

A limitation of using the eatqp01 form to train a model is that the questionnaire may have been designed to apply explicitly to ADHD behaviours. A representative model would require questions to target all aspects of the participant's behaviour and personality. Further study could bring together multiple data sources to cover various behavioural aspects. Conversely, it could be argued that as long as the model is predictive and can make accurate classifications, this does not matter – it is more of interest from a research and information perspective.

Chapter 5

Evaluation and Conclusions

5.1. Conclusions

This project has many findings that affect health policy and future research. In the regional analysis of ADHD prevalence, it was demonstrated that there is a disparity in the rate of diagnosis of ADHD across England, the cause of which is unclear. One possibility uncovered by this analysis is limited access to healthcare preventing the diagnosis of ADHD in some areas. This fact is further evidence of the need for more accessible diagnostic tools for ADHD, especially automated tools that can be widely available to the general public. This concept was explored much further in the second part of the project, where it was shown that machine learning models could use parent-reported behavioural data to distinguish adolescents with ADHD from those without ADHD. Support vector classification and logistic regression were the most predictive models after various optimisation steps. While other studies have carried out similar tasks, such as distinguishing ADHD from ASD or diagnosing ADHD from MRI/EEG data, this is the first to do so using behavioural data. This has the apparent advantage of being much more accessible to the public since no specialist tools are required to collect the data; it can all be self-reported or peer-reported.

Additionally, this project had several notable findings regarding the diagnosis and treatment of ADHD in England. An exponential increase in the occurrence of ADHD was identified, which grew more severe after the COVID-19 pandemic. While the specific cause of this phenomenon is not apparent, the existence of the trend shows that the change in lifestyle since the start of the pandemic has significantly impacted the way that ADHD is diagnosed and treated; this certainly warrants further investigation to understand the causes.

Beyond diagnosis, a steep increase in the use of expensive ADHD medications was observed compared to a minimal uptake of cheaper yet comparable drugs. Consequently, the NHS should consider promoting the much cheaper dexamphetamine over the more expensive option lisdexamfetamine, provided that work is carried out to investigate the safety and efficacy implications.

5.2. Contributions

- The novelty of this project is in its use of behavioural data to classify ADHD. While many other studies have made accurate diagnoses using imaging or neurological data, this work shows that it is possible to make predictions using only reported behavioural data. This finding opens the way for tools to be developed that are accessible to everyone without the need for medical devices.
- Noteworthy trends in the treatment of ADHD were found. For example, no previous studies have noted the change in the upward trend in treatments prescribed for ADHD since the onset of COVID-19; this finding warrants further exploration.

- The regional variance of ADHD prevalence has only been investigated using data from 2004 to 2013 and attempted to correlate prevalence with socioeconomic deprivation (Hire et al., 2018). This study is novel in its incorporation of multiple socioeconomic metrics, identifying trends that were not previously seen using the more generic metric.

5.3. Reflections

Regrettably, the full extent of this data could not be explored in this project due to the extreme time constraints placed upon it. Difficulties were encountered in accessing the dataset intended to be used initially. Hence, the project's scope had to be reduced to focus on only one aspect of the ABCD dataset and use more simplistic machine learning models than would have been ideal. If the project had been completed as intended, data from multiple behavioural and personality forms in the ABCD dataset would have been combined, generating a much larger dataset with many more features. Given the time, this could then have been modelled using deep learning to observe more complex trends within the data.

Additionally, data could have been modelled as a graph, using graph neural networks to consider relationships between the data features. Despite this, the good performance of the trained models shows that it is possible to, at least to some extent, make predictions about ADHD status from easily accessible behavioural data. This work will hopefully provide a basis for future works to improve these results, using more sophisticated models to build a tool that can enhance the ADHD diagnosis procedure.

5.4. Further Work

Unfortunately, this project was severely limited in its scope due to time constraints. As previously mentioned, there were findings from the exploratory phase that were not fully investigated that should be more thoroughly examined. Primarily, understanding the marked impact of the COVID-19 pandemic on ADHD treatments prescribed could help provide information on how lifestyle impacts ADHD diagnosis. Furthermore, the financial implications of the trends in prescribing expensive medications should be explored to potentially switch to cheaper medications, freeing up some financial resources for the NHS to spend elsewhere.

In addition, there is much potential for information to be extracted from other data from the ABCD study. For example, the ABCD dataset contains a multitude of questionnaires relating to behaviour, personality, and mental health, which could be combined to expand on the number of features used in the model; this could uncover key features to the diagnosis of ADHD that did not arise in this investigation.

Further study should be carried out once the study has collected more data over the coming years. Eventually, the ABCD data could be used to model how the symptomatology of ADHD changes with age since participants will continue to provide data as they grow up. This model would provide valuable insight into how ADHD diagnosis can be tailored to adolescents of different ages, as ADHD may look different in prepubescents than it does in young adults.

Additionally, more work could be done to optimise the models used in this project, particularly the decision forest model. A review of the current literature suggests that decision forests are a potent model for medical classification problems such as this, so the model used may have had some flaws preventing it from performing well. This idea is reinforced by the fact that the model did not perform poorly but at chance – a sign that something may have gone awry in the training process.

This concept can also be translated to other psychiatric and neurological disorders, many of which also suffer from difficulty objectively conveying and understanding symptoms. Such tools could provide increased access to healthcare, as well as provide medical professionals with additional resources with which to make diagnostic decisions. Both of these applications would reduce waiting times for diagnosis by the NHS, allowing patients to receive treatment and improve their quality of life sooner.

Word Count: 10645

References

NHS, 2021. *Attention deficit hyperactivity disorder* [Online]. Available from: https://www.nhs.uk/conditions/attention-deficit-hyperactivity-disorder-adhd [Accessed 17 August 2022].

Drechsler, R., Brem, S., Brandeis. D., Grunblatt, E., Berger, G. and Walitza, S., 2020. ADHD: Current Concepts and Treatments in Children and Adolescents. *Neuropediatrics*, 51(5), pp.315-335.

Wolraich, M.L., Chan, E., Froehlich, T., Lynch, R.L., Bax, A., Redwine, S.T., Ihyembe, D. and Hagan, J.F., 2019. ADHD diagnosis and treatment guidelines: a historical perspective. *Pediatrics*, 144(4).

Kian, N., Samieefar, N. and Rezaei, N., 2022. Prenatal risk factors and genetic causes of ADHD in children. *World Journal of Pediatrics*, 18, pp.308-319.

Tarver, J., Daley, D. and Sayal, K., 2014. ADHD: an updated review of the essential facts. *Child Care Health Dev*, 40(6), pp.762-774.

Mehta, TR., Monegro, A., Nene, Y., Fayyaz, M. and Bollu, PC., 2019. Neurobiology of ADHD: A Review. *Current Developmental Disorders Reports*, 6, pp.235-240.

Rommelse, NN., Altink, ME., Arias-Vásquez, A., Buschgens, CJ., Fliers, E., Faraone, SV., Buitelaar, JK., Sergeant, JA., Franke, B. and Oosterlaan, J., 2008. A review and analysis of the relationship between neuropsychological measures and DAT1 in ADHD. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 147(8), pp.1536-1546.

Langley, K., Rice, F., Van Den Bree, MB. and Thapar, A., 2005. Maternal smoking during pregnancy as an environmental risk factor for attention deficit hyperactivity disorder behaviour. A review. *Minerva Pediatricia*, 57, pp.359-371.

Johnson, S., Hollis, C., Kochhar, P., Hennessy, E., Wolke, D. and Marlow, N., 2010. Psychiatric disorders in extremely preterm children: longitudinal finding at age 11 years in the EPICure study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 49, pp.453-463.

Langley, K., Heron, J., Smith, GD. and Thapar, A., 2012. Maternal and paternal smoking during pregnancy and risk of ADHD symptoms in offspring: testing for intrauterine effects. *American Journal of Epidemiology*, 176, pp.261-268.

American Psychiatric Association, 2013. Diagnostic and statistical manual of mental disorders: DSM-V, 5th edn, American Psychiatric Association, Washington DC.

WebMD, 2021. *ADHD Medications and Side Effects* [Online]. Available from: https://www.webmd.com/add-adhd/adhd-medication-chart [Accessed 10 September 2022].

Arnsten, AFT., 2006. Stimulants: Therapeutic Actions in ADHD. *Neuropsychopharmacology*, 31, pp.2376-2383.

Dittmann, RW., Hage, A. and Pedraza, J., 2018. Non-stimulants in the treatment of ADHD. *Oxford textbook of ADHD. Oxford University Press, Oxford*, pp.393-401.

Findling, RL., and Dogin, JW., 1998. Psychopharmacology of ADHD: children and adolescents. *Journal of Clinical Psychiatry*, 59, pp.42-49.

ADHD UK, 2022. *Right to Choose (NHS England)* [Online]. Available from: https://adhduk.co.uk/right-to-choose/ [Accessed 05 October 2022].

National Institute for Health and Care Excellence, 2018. *Attention deficit hyperactivity disorder: diagnosis and management* [NICE Guideline No. 87]. https://www.nice.org.uk/guidance/ng87

Richter, T., Fishbain, B., Fruchter, E., Richter-Levin, G., Okon-Singer, H., 2021. Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders. *Journal of Psychiatric Research*, 141, pp. 199-205.

ABCD Study, 2022a. *Protocols* [Online]. Available from: https://abcdstudy.org/scientists/protocols/ [Accessed 20 August 2022].

ABCD Study, 2022b. *Study Sites* [Online]. Available from: https://abcdstudy.org/study-sites/ [Accessed 20 August 2022].

Yang, R., 2021. *Adolescent Brain Cognitive Development Study (ABCD) – Annual Release 4.0* [Online]. Available from: https://dx.doi.org/10.15154/1523041 [Accessed 20 August 2022].

Kukreja, H., Bharath, N., Siddesh, C.S., Kuldeep, S., 2016. An introduction to artificial neural network. *Int J Adv Res Innov Ideas Educ*, 1, pp.27-30.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, FE., 2017. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234, pp. 11-26.

Criminisi, A., Shotton, J., Konukoglu, E., 2012. Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. *Foundations and Trends in Computer Graphics and Vision*, 7(2-3), pp. 81-227.

Kontschieder, P., Fiterau, M., Criminisi, A., Bulo, SR., 2015. Deep neural decision forests. *Proceedings of the IEEE international conference on computer vision*, pp. 1467-1475.

Yang, Y., Morillo, IG., Hospedales, TM., 2018. Deep neural decision trees. *arXiv preprint, arXiv: 1806.06988*.

Pramanik, M., Pradhan, R., Nandy, P., Bhoi, AK., Barsocchi, P., 2021. Machine Learning Methods with Decision Forests for Parkinson's Detection. *Applied Sciences*, 11(2):581.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M., 2020. Graph neural networks: A review of methods and applications. *AI Open*, 1, pp. 57-81.

Gori, M., Monfardini, G., Scarselli, F., 2005. A new model for learning in graph domains. *Proc. IEEE Int. Joint Conf. Neural Netw,* 2, pp. 729-734.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, PS., 2021. A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), pp. 4-24.

Duda, M., Ma, R., Haber, N., Wall, DP., 2016. Use of machine learning for behavioral distinction of autism and ADHD. *Translational Psychiatry*, 5, e732.

Tenev, A., Markovska-Simoska, S., Kocarev, L., Pop-Jordanow, J., Muller, A., Candrian, G., 2014. Machine learning approach for classification of ADHD adults. *International Journal of Psychophysiology*, 93(1), pp. 162-166.

Kim, JW., Sharma, V., Ryan, ND., 2015. Predicting Methylphenidate Response in ADHD Using Machine Learning Approaches. *International Journal of Neuropsychopharmacology*, 18(11), pyv052.

NHSBSA Data Warehouse, 2022. *English Prescribing Dataset* [Online]. Available from: https://opendata.nhsbsa.net/dataset/english-prescribing-data-epd [Accessed 06 August 2022].

NHS Digital, 2022. *Patients Registered at a GP Practice – May 2022* [Online]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/patients-registered-at-a-gp-practice/may-2022 [Accessed 18 September 2022].

NHS England, 2022. *Integrated Care Boards in England* [Online]. Available from: https://www.england.nhs.uk/publication/integrated-care-boards-in-england/ [Accessed 18 September 2022].

Office for National Statistics (ONS), 2022. *CC01 Regional labour market: Claimant Count by unitary and local authority (experimental)* [Online]. Available from: https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment/datasets/claimantcountbyunitaryandlocalauthorityexperimental/current [Accessed 18 September 2022].

Office for National Statistics (ONS), 2021a. *Regional gross disposable household income, UK: 1997 to 2019* [Online]. Available from: https://www.ons.gov.uk/economy/regionalaccounts/grossdisposablehouseholdincome/datasets/regionalgrossdisposablehouseholdincomegdhi [Accessed 18 September 2022].

Office for National Statistics (ONS), 2021b. *Life expectancy estimates, all ages, UK* [Online]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/datasets/lifeexpectancyestimatesallagesuk [Accessed 18 September 2022].

Office for National Statistics (ONS), 2021c. *Annual personal well-being estimates: April 2020 to March 2021* [Online]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/datasets/headlineestimatesofpersonalwellbeing [Accessed 18 September 2022].

TheKing'sFund, 2022. *What is happening to life expectancy in England?* [Online]. Available from: https://www.kingsfund.org.uk/publications/whats-happening-life-expectancy-england#:~:text=Life%20expectancy%20is%20affected%20by,as%20sex%2C%20ethnicity%2C%20disability%20and [Accessed 05 October 2022].

Felstead, A and Reuschke, D., 2020. *Homeworking in the UK: before and during the 2020 lockdown* WISERD Report, Cardiff: Wales Institute of Social and Economic Research. Available for download from: https://wiserd.ac.uk/publications/homeworking-ukand-during-2020-lockdown

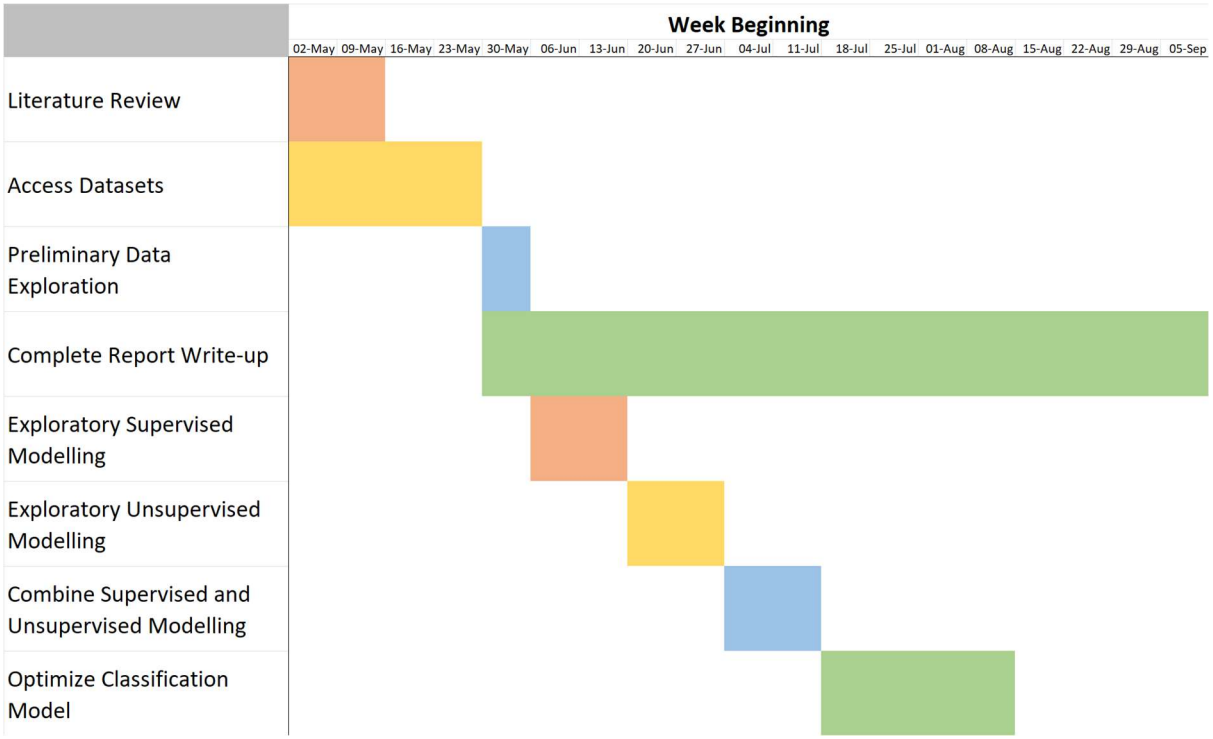
WebMD, 2021. *ADHD Medications and Side Effects* [Online]. Available from: https://www.webmd.com/add-adhd/adhd-medication-chart [Accessed 23 September 2022].

Centers for Disease Control and Prevention (CDC), 2022. *Data And Statistics About ADHD* [Online]. Available from: https://www.cdc.gov/ncbddd/adhd/data.html [Accessed 28 September 2022].

Hire, AJ., Ashcroft, DM., Springate, DA., Steinke, DT., 2018. ADHD in the United Kingdom: Regional and Socioeconomic Variations in Incidence Rates Amongst Children and Adolescents (2004-2013). *Journal of Attention Disorders*, 22(2), pp. 134-142.

Appendix A

Initial Project Plan



Appendix B

Data Processing Code

Regional Data Analysis

```
import pandas as pd
import requests
import urllib.parse
from matplotlib import pyplot as plt
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

font = {'family' : 'normal',
        'size'   : 11}

matplotlib.rc('font', **font)

base_endpoint = 'https://opendata.nhsbsa.net/api/3/action/'
package_list_method = 'package_list'     # List of data-sets in the portal
package_show_method = 'package_show?id=' # List all resources of a data-set
action_method = 'datastore_search_sql?'  # SQL action method
resource_name = "EPD_202205"

## Count of ADHD meds prescribed

base_endpoint = 'https://opendata.nhsbsa.net/api/3/action/'
package_list_method = 'package_list'     # List of data-sets in the portal
package_show_method = 'package_show?id=' # List all resources of a data-set
action_method = 'datastore_search_sql?'  # SQL action method
resource_name = "EPD_202205"
adhd_single_month_query = "SELECT ICB_NAME,sum(TOTAL_QUANTITY) AS count_adhd " \
                  f"FROM `{resource_name}` " \
                  f"WHERE bnf_chemical_substance LIKE '0404000__' " \
                  f"GROUP BY ICB_NAME " \
                  f"ORDER BY icb_name "

adhd_single_month_api_call = f"{base_endpoint}" \
                     f"{action_method}" \
                     "resource_id=" \
                     f"{resource_name}" \
                     "&" \
                     "sql=" \
                     f"{urllib.parse.quote(adhd_single_month_query)}"

adhd_single_month_response = requests.get(adhd_single_month_api_call).json()

adhd_df = pd.json_normalize(adhd_single_month_response['result']['result']['records'])

print(adhd_df)
adhd_df.drop(adhd_df[adhd_df['ICB_NAME'] == 'UNIDENTIFIED'].index, inplace = True)

## Count of all meds prescribed

non_adhd_single_month_query = "SELECT ICB_NAME,sum(TOTAL_QUANTITY) AS count " \
                  f"FROM `{resource_name}` " \
                  f"GROUP BY ICB_NAME " \
```

```
                    f"ORDER BY icb_name "

non_adhd_single_month_api_call = f"{base_endpoint}" \
                        f"{action_method}" \
                        "resource_id=" \
                        f"{resource_name}" \
                        "&" \
                        "sql=" \
                        f"{urllib.parse.quote(non_adhd_single_month_query)}"

non_adhd_single_month_response = requests.get(non_adhd_single_month_api_call).json()

non_adhd_df =
    pd.json_normalize(non_adhd_single_month_response['result']['result']['records'])

## Derive % total

percent_total = 100* adhd_df['count_adhd'] / non_adhd_df['count']
adhd_df['non_adhd'] = non_adhd_df['count']

adhd_df['percent_total'] = percent_total
adhd_df_sorted = adhd_df.sort_values(by=['percent_total'],ascending=False)
adhd_df_sorted['percent_total_index'] = range(len(adhd_df_sorted))

## Adding population data

populations_pre = pd.read_csv("populations.csv")
populations = populations_pre[['CODE','NUMBER_OF_PATIENTS']]

single_month_query = "SELECT DISTINCT icb_name,practice_code " \
                    f"FROM `{resource_name}` "

single_month_api_call = f"{base_endpoint}" \
                        f"{action_method}" \
                        "resource_id=" \
                        f"{resource_name}" \
                        "&" \
                        "sql=" \
                        f"{urllib.parse.quote(single_month_query)}"

single_month_response = requests.get(single_month_api_call).json()

df = pd.json_normalize(single_month_response['result']['result']['records'])

df1 = df.merge(populations,how='left',left_on='practice_code', right_on='CODE')
df1 = df1.groupby('icb_name').sum()

adhd_df_pop =
    adhd_df_sorted.merge(df1,how='left',left_on='ICB_NAME',right_on='icb_name')
    adhd_df_pop = adhd_df_pop.rename(str.lower, axis='columns')

    ### Calculate ADHD meds prescribed per person

    adhd_df_pop['adhd_per_pop'] =
    adhd_df_pop['count_adhd']/adhd_df_pop['number_of_patients']
    adhd_df_pop_sorted = adhd_df_pop.sort_values(by=['adhd_per_pop'],ascending=False)
    adhd_df_pop_sorted['adhd_per_pop_index'] = range(len(adhd_df_pop_sorted))

    ## Adding GDHI data

    gdhi = pd.read_csv('gdhi.csv')
    gdhi = gdhi[['icb_name','GDHI/head']]
    adhd_df_gdhi =
    adhd_df_pop_sorted.merge(gdhi,how='left',left_on='icb_name',right_on='icb_name')
    adhd_df_gdhi_sorted = adhd_df_gdhi.sort_values(by=['GDHI/head'],ascending=False)
    adhd_df_gdhi_sorted['GDHI/head_index'] = range(len(adhd_df_gdhi_sorted))

```

```
## Adding unemployment data

unemployment = pd.read_csv('unemployment.csv')
adhd_df_unemployment = adhd_df_gdhi_sorted.merge(unemployment,how='left',left_on='icb_name',right_on='icb_name')
adhd_df_unemployment_sorted = adhd_df_unemployment.sort_values(by=['unemployment'],ascending=False)
adhd_df_unemployment_sorted['unemployment_index'] = range(len(adhd_df_unemployment_sorted))

## Adding life expectancy data

life = pd.read_csv('life_expectancy.csv')
life['avg_life_expectancy'] = (life['f_life_expectancy'] + life['m_life_expectancy'])/2
adhd_df_life = adhd_df_unemployment_sorted.merge(life,how='left',left_on='icb_name',right_on='icb_name')
adhd_df_life_sorted = adhd_df_life.sort_values(by=['avg_life_expectancy'],ascending=False)
adhd_df_life_sorted['life_expectancy_index'] = range(len(adhd_df_life_sorted))

## Adding lifestyle data

lifestyle = pd.read_csv('lifestyle.csv')
adhd_df_lifestyle = adhd_df_life_sorted.merge(lifestyle,how='left',left_on='icb_name',right_on='icb_name')
variables = adhd_df_lifestyle[['percent_total','adhd_per_pop','GDHI/head','unemployment','avg_life_expectancy','anxiety','happiness','life_satisfaction','worthwhile']]
plt.figure(figsize=(12,6))
sns.heatmap(variables.corr()[:,:5],vmin=-1, vmax=1, annot=True,cmap='vlag')
plt.savefig('Figures/correlation')
plt.show()

adhd_df_sorted = adhd_df_lifestyle.sort_values(by=['icb_name'],ascending=True)
adhd_df_sorted.to_csv('adhd.csv')

plt.figure(figsize=(8.6,6))
plt.subplot(2,1,1)
print(variables.corr())
sns.heatmap(variables.corr()[0:2],vmin=-1, vmax=1, annot=True,cmap='vlag')
plt.yticks(np.arange(2)+0.5,('percent_total','adhd_per_pop'),
           rotation=0, fontsize="10", va="center")
plt.savefig("Figures/correlation_pair")
plt.show()
```

Over-time Analysis

```
import pandas as pd
import re
import requests
import warnings
import urllib.parse
from matplotlib import pyplot as plt
from tqdm.notebook import tqdm
import datetime
import numpy as np
import matplotlib.dates as mdates
import matplotlib

font = {'family' : 'normal',
        'size'   : 11}
```

```
15.
16. matplotlib.rc('font', **font)
17.
18. base_endpoint = 'https://opendata.nhsbsa.net/api/3/action/'
19. package_list_method = 'package_list'     # List of data-sets in the portal
20. package_show_method = 'package_show?id=' # List all resources of a data-set
21. action_method = 'datastore_search_sql?'  # SQL action method
22. adhd = ['0404000L0','0404000M0','0404000U0','0404000S0','0404000V0',]
23.
24. ## Number of each ADHD medication distributed per month
25.
26. # Function requests sum(cost) and sum(quantity) data for a given year/month by drug
27.
28. def get_adhd_data(year,month):
29.     resource_name = "EPD_"+year+month
30.     single_month_query = "SELECT chemical_substance_bnf_descr,sum(TOTAL_QUANTITY) AS
    count,sum(ACTUAL_COST) AS cost " \
31.                          f"FROM `{resource_name}` " \
32.                          f"WHERE bnf_chemical_substance IN
    ('{adhd[0]}','{adhd[1]}','{adhd[2]}','{adhd[3]}','{adhd[4]}') " \
33.                          f"GROUP BY chemical_substance_bnf_descr "
34.
35.     single_month_api_call = f"{base_endpoint}" \
36.                             f"{action_method}" \
37.                             "resource_id=" \
38.                             f"{resource_name}" \
39.                             "&" \
40.                             "sql=" \
41.                             f"{urllib.parse.quote(single_month_query)}"
42.
43.     single_month_response = requests.get(single_month_api_call).json()
44.
45.     df = pd.json_normalize(single_month_response['result']['result']['records'])
46.
47.     df['year'] = year
48.     df['month'] = month
49.
50.     #print(f"{resource_name} done!")
51.
52.     return df
53.
54. # Iterate through all valid year/month combinations, calling the data function for each
    and
55. # concatenating all results
56.
57. results = '.'
58.
59. years = ['2014','2015','2016','2017','2018','2019','2020','2021','2022']
60.
61. months = ['01','02','03','04','05','06','07','08','09','10','11','12']
62.
63. for year in years:
64.     for month in months:
65.         if year == '2022' and month in ['07','08','09','10','11','12']:
66.             pass
67.         else:
68.             if type(results) == str:
69.                 results = get_adhd_data(year,month)
70.             else:
71.                 results =
    pd.concat([results,get_adhd_data(year,month)],ignore_index=True)
72.
73. results["price_per_unit"] = results['cost']/results['count']
74. results.to_csv("over_time_results.csv")
75.
76. # Transform data into matrices for plotting, separating by drug
77.
```

```
labels = []
counts = []
costs = []
datetimes = []

for value in results['chemical_substance_bnf_descr'].unique():
    subset = results.loc[(results['chemical_substance_bnf_descr'] == value)]
    dates = pd.to_datetime(subset[['year', 'month']].assign(DAY=1))
    datetimes.append(dates)
    labels.append(value)
    counts.append(subset['count'].tolist())
    costs.append(subset['cost'].tolist())

# Plot prescriptions over time for each drug

plt.figure(figsize=(12,12))
plt.subplot(2,1,1)
for i in range(len(labels)):
    plt.plot(datetimes[i],[y/1000000 for y in counts[i]],label = labels[i])

#plt.axvline(x=datetime.date(2020,3,1),color='black',ls='--')
plt.xlabel('Time')
plt.ylabel('Total Prescribed/Month (Millions)')
plt.title("a) Quantity of drug prescribed per month, by drug",loc="left")
plt.legend()

# Plot cost over time for each drug

plt.subplot(2,2,3)
for i in range(len(labels)):
    plt.plot(datetimes[i],[j/1000000 for j in costs[i]],label = labels[i])

#plt.axvline(x=datetime.date(2020,3,1),color='black',ls='--')
plt.xlabel('Time')
plt.ylabel('Total Monthly Cost (Million £)')
plt.ylim(bottom=0)
plt.title("b) Total monthly cost of drug over time",loc="left")

# Plot price over time for each drug

plt.subplot(2,2,4)
for i in range(len(labels)):
    plt.plot(datetimes[i],[z/y for y,z in zip(counts[i],costs[i])],label = labels[i])

#plt.axvline(x=datetime.date(2020,3,1),color='black',ls='--')
plt.xlabel('Time')
plt.ylabel('Price per Unit (£)')
plt.ylim(bottom=0)
plt.title("c) Price per unit of drug over time",loc="left")
plt.savefig("Figures/monthly_pricing")
plt.show()

# Add data where is missing

while len(counts[4]) < 102:
    counts[4].insert(0,0)

# Plot prescriptions over time for sum of all drugs

plt.figure(figsize=(12,12))
plt.subplot(2,1,1)
plt.plot(datetimes[0],[sum(x)/1000000 for x in zip(*counts)])

plt.xlabel('Time')
plt.ylabel('ADHD medications prescribed (Millions)')

# Plot regression, n=3 (cubic fit)
```

```
145.
146.  x = mdates.date2num(datetimes[0])
147.  y = [sum(x)/1000000 for x in zip(*counts)]
148.  z = np.polyfit(x, y, 3)
149.  p = np.poly1d(z)
150.
151.  plt.plot(x, p(x), linestyle="--",alpha=0.75,color="tab:red")
152.  plt.ylim(bottom=0,top=9)
153.  plt.title("a) Total quantity of ADHD medication prescribed over time in
    England",loc="left")
154.  #plt.savefig('Figures/prescription_trend')
155.  #plt.show()
156.
157.  # Plot prescriptions over time with vline for start of COVID-19
158.
159.  #plt.figure(figsize=(12,6))
160.  plt.subplot(2,1,2)
161.  plt.plot(datetimes[0],[sum(x)/1000000 for x in zip(*counts)])
162.  plt.axvline(x=datetime.date(2020,3,1),ls='--',color='black')
163.  plt.xlabel('Time')
164.  plt.ylabel('ADHD medications prescribed (Millions)')
165.
166.  # Plot Linear Regressions pre and post COVID-19
167.
168.  nomial = 1
169.
170.  x1 = mdates.date2num(datetimes[0][:75])
171.  y1 = [sum(x) for x in zip(*counts)][:75]
172.  z1 = np.polyfit(x1, y1, nomial)
173.  p1 = np.poly1d(z1)
174.
175.  x2 = mdates.date2num(datetimes[0][74:])
176.  y2 = [sum(x) for x in zip(*counts)][74:]
177.  z2 = np.polyfit(x2, y2, nomial)
178.  p2 = np.poly1d(z2)
179.
180.  years_values = np.zeros(9)
181.
182.  # Linear regression for each year
183.
184.  for i in range(1,10):
185.
186.      temp_values = [sum(x) for x in zip(*counts)]
187.      subset = temp_values[(i-1)*12:(i)*12]
188.
189.      x = mdates.date2num(datetimes[0][(i-1)*12:(i)*12])
190.      z = np.polyfit(x, subset, 1)
191.      p = np.poly1d(z)
192.      print(z[0])
193.      years_values[i-1] = z[0]
194.
195.  print(f"Coefficient pre-COVID-19: {round(z1[0],1)}")
196.  print(f"Coefficient post-COVID-19: {round(z2[0],1)}")
197.
198.  date_2025 = mdates.date2num(datetime.datetime(2025,1,1))
199.  value_2025 = p2(date_2025)
200.  value_2025_old = p1(date_2025)
201.  print(f"Projected quantity/month in 2025: {int(value_2025)}")
202.  print(f"Projected quantity/month in 2025 (old model): {int(value_2025_old)}")
203.  # Plot
204.
205.  plt.plot(x1, p1(x1)/1000000, linestyle="--",alpha=0.75,color="tab:red")
206.  plt.plot(x2, p2(x2)/1000000, linestyle="--",alpha=0.75,color="tab:red")
207.
208.  years_positions = datetimes[0][::12]
209.
210.  ax2 = plt.twinx()
```

```
ax2.bar(years_positions,years_values,width=200,alpha=0.25)
plt.title("b) Total quantity of ADHD medication prescribed in England pre and post-COVID-19",loc="left")
ax2.set_ylabel("Rate of growth over subsequent 12 months (units/day)")
plt.savefig('Figures/prescription_covid')
plt.show()
```

ABCD Data

```
import pandas as pd
from pandasql import sqldf
from tqdm.notebook import tqdm
import numpy as np
from matplotlib import pyplot as plt

from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, precision_score, confusion_matrix,roc_curve, roc_auc_score, precision_score, recall_score, precision_recall_curve
from sklearn.metrics import f1_score
import matplotlib

font = {'family' : 'normal',
        'size'   : 11}

matplotlib.rc('font', **font)


## Deriving the ADHD data from medsy01

df = pd.read_csv(r"X:\CompSci\ResearchProjects\AJoshi\aj2151 - AJoshi\Matt Bearham\Package_1202749\medsy01.txt",sep='\t')

mysql = lambda q: sqldf(q, globals())

adhd_terms = [
    'dexedrine',
    'zenzedi',
    'adderall',
    'focalin',
    'methylin',
    'ritalin',
    'methylphenidate',
    'dyanavel',
    ' amphetamine',
    'dextroamphetamine',
    'evekeo',
    'mydayis',
    'vyvanse',
    'aptensio',
    'concerta',
    'lisdexamfetamine',
    'quillivant',
    'quillichew',
    'azstarys',
    'metadate',
    'cotempla',
    'atomoxetine',
    'strattera',
    'guanfacine',
    'tenex',
    'intuniv',
```

```
    'clonidine',
    'kapvay']

data = []
for i in tqdm(range(15),desc='fields'):
    for term in tqdm(adhd_terms,desc='terms',leave=False):
        temp = df.copy()
        temp['match'] = temp[f'med{i+1}_rxnorm_p'].str.lower().str.contains(term)
        result = temp[temp['match'] == True]['subjectkey']
        result = pd.DataFrame(result)
        result['medication'] = term
        #result = mysql(f"SELECT DISTINCT subjectkey FROM df WHERE med{i+1}_rxnorm_p
LIKE '%{term}%' COLLATE utf8_general_ci")
        data.append(result)

adhd_subjects = pd.concat(data).reset_index().drop(['index'],axis=1)

pivot =
pd.pivot_table(adhd_subjects,index='subjectkey',columns='medication',aggfunc=len,fill_va
lue=0)

medications = pd.DataFrame()
medications['methylphenidate'] = pivot['aptensio'] + pivot['concerta'] +
pivot['cotempla'] + pivot['focalin'] + pivot['metadate'] + pivot['methylin'] +
pivot['methylphenidate'] + pivot['quillichew'] + pivot['quillivant'] + pivot['ritalin']
medications['dexamphetamine'] = pivot['adderall'] + pivot[' amphetamine'] +
pivot['dexedrine'] + pivot['dyanavel'] + pivot['evekeo'] + pivot['zenzedi']
medications['lisdexamphetamine'] = pivot['lisdexamfetamine'] + pivot['vyvanse']
medications['atomoxetine'] = pivot['atomoxetine'] + pivot['strattera']
medications['guanfacine'] = pivot['guanfacine'] + pivot['tenex'] + pivot['intuniv']
medications['clonidine'] = pivot['kapvay'] + pivot['clonidine']
medications = medications.clip(0,1)

adhd_subjects_distinct = pd.concat(data)['subjectkey'].drop_duplicates().reset_index()

num_adhd = len(adhd_subjects_distinct)
num_not = len(df['subjectkey'].unique())-len(adhd_subjects_distinct)
print(f'Number with ADHD: {num_adhd}')
print(f'Number without ADHD: {num_not}')

plt.figure(figsize=(8.5,5.5))
plt.subplot(1,2,1)
plt.pie([num_not,num_adhd],labels=['Participants without ADHD','Participants with
ADHD'],
        autopct='%1.1f%%',labeldistance=None,explode=(0,0.2))
plt.title("a) Prevalence of ADHD",loc="left")
plt.legend(loc=3)
plt.subplot(1,2,2)
plt.bar(medications.columns,medications.sum())
plt.title("b) ADHD medication counts",loc="left")
plt.xticks(rotation=45)

plt.savefig('Figures/rate_adhd')
plt.show()

print(f"{len(medications[medications.sum(axis=1)==2])} participants on 2 different
ADHD medications")
print(f"{len(medications[medications.sum(axis=1)==3])} participants on 3 different
ADHD medications")
print(f"{len(medications[medications.sum(axis=1)==4])} participants on 4 different
ADHD medications")
print(f"{len(medications[medications.sum(axis=1)==5])} participants on 5 different
ADHD medications")
print(f"{len(medications[medications.sum(axis=1)>5])} participants on more than 5
different ADHD medications")

doubles = medications[medications.sum(axis=1)==2]
```

```
109.  triples = medications[medications.sum(axis=1)==3]
110.  quartets = medications[medications.sum(axis=1)==4]
111.
112.  count_doubles =
    doubles.replace(1,'o').groupby(doubles.columns.tolist(),as_index=False).size()\
113.
    .reset_index().rename(columns={0:'records'}).sort_values(by='records',ascending=False)\
114.      .replace(0,'-')
115.  count_triples =
    triples.replace(1,'o').groupby(triples.columns.tolist(),as_index=False).size()\
116.
    .reset_index().rename(columns={0:'records'}).sort_values(by='records',ascending=False)\
117.      .replace(0,'-')
118.  count_quartets =
    quartets.replace(1,'o').groupby(quartets.columns.tolist(),as_index=False).size()\
119.
    .reset_index().rename(columns={0:'records'}).sort_values(by='records',ascending=False)\
120.      .replace(0,'-')
121.
122.  crossfreq = medications.T.dot(medications)
123.  crossfreq.to_csv('crossfreq.csv')
124.
125.  def generate_nice_table(dataframe):
126.      nice_table = []
127.      for index,row in dataframe.iterrows():
128.          drugs = []
129.          for column in row.index:
130.              if row[column] == 'o':
131.                  drugs.append(column)
132.          string =  ', '.join(drugs)
133.          nice_table.append([string,row['records']])
134.      return pd.DataFrame(nice_table).rename(columns={0:'drugs',1:'counts'})
135.
136.  print(generate_nice_table(count_doubles))
137.  print(generate_nice_table(count_triples))
138.  print(generate_nice_table(count_quartets))
139.
140.  ## eatqp01
141.
142.  temprament = pd.read_csv(r"X:\CompSci\ResearchProjects\AJoshi\aj2151 - AJoshi\Matt
    Bearham\Package_1202749\abcd_eatqp01.txt",sep='\t')
143.  temprament['adhd'] = [1 if x in adhd_subjects_distinct['subjectkey'].tolist() else 0
    for x in temprament['subjectkey']]
144.
145.  for i in
    ['eatq_finish_p','eatq_turn_taking_p','eatq_open_present_p','eatq_ski_slope_p','eatq_bef
    ore_hw_p','eatq_distracted_p','eatq_impulse_p','eatq_social_p','eatq_hardly_sad_p','eatq
    _try_focus_p','eatq_no_criticize_p','eatq_puts_off_p','eatq_sidetracked_p','eatq_not_shy
    _p','eatq_meet_p','eatq_rides_scared_p']:
146.      temprament[i].replace({1:5,2:4,4:2,5:1},inplace=True)
147.
148.  temprament = temprament.dropna()
149.  descriptions = temprament.iloc[0]
150.  temprament = temprament.iloc[1:]
151.
152.  X = temprament.iloc[:,10:72]
153.  y = temprament['adhd']
154.  print(y.value_counts())
155.  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, stratify=y,
    random_state=150)
156.
157.  model = LogisticRegression(class_weight={0:1258,1:9073})
158.  model.fit(X_train,y_train)
159.  importance = model.coef_[0]
160.
161.  coefficients = []
162.  for i in range(len(importance)):
```

```
        coefficients.append((X.columns[i],importance[i]))

    to_show = 14

    sorted_coef = sorted(coefficients,key=lambda tup:abs(tup[1]))
    bottom_5 = sorted_coef[:to_show]
    top_5 = sorted_coef[-to_show:]

    print(f'*** The {to_show} least predictive measures were (odds ratio): ***')
    for index,(label,value) in enumerate(bottom_5):
        print(f"{len(coefficients)-index}. {descriptions[label].split(' / ')[0]}:
    {round(np.exp(value),3)}")
    print(f'\n*** The {to_show} most predictive measures were (odds ratio): ***')
    for index,(label,value) in enumerate(top_5):
        print(f"{to_show-index}. {descriptions[label].split('. / ')[0]}:
    {round(np.exp(value),3)}")

    ## Optimising the weighting of the model

    regular = np.arange(0,2,0.05)
    zoomed = np.arange(0.05,0.25,0.001)

    scores_weight = []

    for i in tqdm(regular,desc="varying weights"):

        result = []

        model = LogisticRegression(class_weight={0:i,1:1},max_iter=1000)
        labels = [entry[0] for entry in sorted_coef]

        model.fit(X_train,y_train)

        y_pred = model.predict(X_test)
        matrix = confusion_matrix(y_test, y_pred)

        result.append(i)
        result.append(accuracy_score(y_test,y_pred))
        result.append(roc_auc_score(y_test, y_pred))
        result.append(recall_score(y_test,y_pred))
        for i in range(2):
            for k in range(2):
                result.append(matrix[i][k])
        result.append(precision_score(y_test,y_pred))
        result.append(f1_score(y_test,y_pred))

        scores_weight.append(result)

    scores_weight_df = pd.DataFrame(scores_weight)
    scores_weight_df.columns =
    ['i','accuracy','auc','recall','tn','fp','fn','tp','precision','f1']

    plt.figure(figsize=(8.5,15))
    plt.subplot(3,1,1)
    plt.title("a) AUC by weighting ratio (0 to 2, step = 0.02)")
    plt.plot(regular,scores_weight_df['auc'])
    plt.xlabel('Weighting Ratio (:1)')
    plt.ylabel('Area Under Curve')

    ## Narrower set of weights around previous peak

    scores_weight_zoom = []

    for i in tqdm(zoomed,desc="zoomed weights"):

        result = []

```

```
227.      model = LogisticRegression(class_weight={0:i,1:1},max_iter=1000)
228.      labels = [entry[0] for entry in sorted_coef]
229.
230.      model.fit(X_train,y_train)
231.
232.      y_pred = model.predict(X_test)
233.      matrix = confusion_matrix(y_test, y_pred)
234.
235.      result.append(i)
236.      result.append(accuracy_score(y_test,y_pred))
237.      result.append(roc_auc_score(y_test, y_pred))
238.      result.append(recall_score(y_test,y_pred))
239.      for i in range(2):
240.          for k in range(2):
241.              result.append(matrix[i][k])
242.      result.append(precision_score(y_test,y_pred))
243.      result.append(f1_score(y_test,y_pred))
244.      scores_weight_zoom.append(result)
245.
246.  scores_weight_zoom_df = pd.DataFrame(scores_weight_zoom)
247.  scores_weight_zoom_df.columns =
    ['i','accuracy','auc','recall','tn','fp','fn','tp','precision','f1']
248.  print(scores_weight_zoom_df.iloc[scores_weight_zoom_df.idxmax()[2]])
249.
250.  plt.subplot(3,1,2)
251.  plt.title("b) AUC by weighting ratio (0 to 0.5, step = 0.001)")
252.  plt.plot(zoomed,scores_weight_zoom_df['auc'])
253.  plt.xlabel('Weighting Ratio (:1)')
254.  plt.ylabel('Area Under Curve')
255.
256.  ## Varying the number of features used
257.
258.  scores = []
259.
260.  for j in tqdm(range(1,len(sorted_coef)+1),desc="varying features used"):
261.
262.      result = []
263.
264.      model = LogisticRegression(class_weight={0:1140,1:9275},max_iter=1000)
265.      labels = [entry[0] for entry in sorted_coef]
266.      subset = labels[-j:]
267.
268.      model.fit(X_train[subset],y_train)
269.
270.      y_pred = model.predict(X_test[subset])
271.      matrix = confusion_matrix(y_test, y_pred)
272.
273.      result.append(j)
274.      result.append(accuracy_score(y_test,y_pred))
275.      result.append(roc_auc_score(y_test, y_pred))
276.      result.append(recall_score(y_test,y_pred))
277.      for i in range(2):
278.          for k in range(2):
279.              result.append(matrix[i][k])
280.      result.append(precision_score(y_test,y_pred))
281.      result.append(f1_score(y_test,y_pred))
282.      scores.append(result)
283.
284.  scores_df = pd.DataFrame(scores)
285.  scores_df.columns =
    ['features','accuracy','auc','recall','tn','fp','fn','tp','precision','f1']
286.
287.  plt.subplot(3,1,3)
288.  plt.title("c) AUC by number of features used")
289.  plt.plot(scores_df['auc'])
290.  plt.xlabel('Number of Features')
291.  plt.ylabel('Area Under Curve')
```

```
plt.savefig("Figures/features")
plt.show()

def nicer_confusion_matrix(given_y_test,given_y_pred):
    matrix = confusion_matrix(given_y_test,given_y_pred)
    dictionary = {'actual control':matrix[0],'actual ADHD':matrix[1]}
    df = pd.DataFrame(dictionary)
    df = df.set_index([['predicted control','predicted ADHD']])
    return df

svm_scores = []
svm_labels = []
for i in tqdm(np.arange(0,2,0.1),desc = "SVM weights"):

    svm_model = svm.SVC(class_weight={0:i,1:1})
    svm_model.fit(X_train,y_train)
    svm_pred = svm_model.predict(X_test)
    svm_matrix = nicer_confusion_matrix(y_test, svm_pred)
    svm_acc = accuracy_score(y_test,svm_pred)
    svm_auc = roc_auc_score(y_test,svm_pred)
    svm_recall = recall_score(y_test,svm_pred)
    svm_precision = precision_score(y_test,y_pred)
    svm_f1 = f1_score(y_test,y_pred)
    svm_scores.append(svm_auc)
    svm_labels.append(i)

plt.figure(figsize = (8.5,15))
plt.subplot(3,1,1)
plt.plot(svm_labels,svm_scores)
plt.xlabel("Weighting ratio (:1)")
plt.ylabel("AUC of model")
plt.title("a) AUC of SVM model with varying class_weight")

svm_scores_2 = []
svm_labels_2 = []
for i in tqdm(np.arange(0.05,0.2,0.01),desc = "SVM weights"):

    svm_model = svm.SVC(class_weight={0:i,1:1})
    svm_model.fit(X_train,y_train)
    svm_pred = svm_model.predict(X_test)
    svm_matrix = nicer_confusion_matrix(y_test, svm_pred)
    svm_acc = accuracy_score(y_test,svm_pred)
    svm_auc = roc_auc_score(y_test,svm_pred)
    svm_recall = recall_score(y_test,svm_pred)
    svm_precision = precision_score(y_test,y_pred)
    svm_f1 = f1_score(y_test,y_pred)
    svm_scores_2.append(svm_auc)
    svm_labels_2.append(i)

plt.subplot(3,1,2)
plt.plot(svm_labels_2,svm_scores_2)
plt.xlabel("Weighting ratio (:1)")
plt.ylabel("AUC of model")
plt.title("b) AUC of SVM model with varying class_weight")

svm_scores_3 = []
svm_labels_3 = []
for i in tqdm(range(1,len(sorted_coef)+1),desc="varying features used"):

    svm_model = svm.SVC(class_weight={0:0.11,1:1})

    labels = [entry[0] for entry in sorted_coef]
    subset = labels[-i:]

    svm_model.fit(X_train[subset],y_train)

    svm_pred = svm_model.predict(X_test[subset])
```

```
359.      svm_matrix = nicer_confusion_matrix(y_test, svm_pred)
360.      svm_acc = accuracy_score(y_test,svm_pred)
361.      svm_auc = roc_auc_score(y_test,svm_pred)
362.      svm_recall = recall_score(y_test,svm_pred)
363.      svm_precision = precision_score(y_test,y_pred)
364.      svm_f1 = f1_score(y_test,y_pred)
365.      svm_scores_3.append(svm_auc)
366.      svm_labels_3.append(i)
367.
368.  print(svm_labels_3[svm_scores_3.index(max(svm_scores_3))])
369.
370.  plt.subplot(3,1,3)
371.  plt.title("SVM AUC by number of features used")
372.  plt.xlabel("Top n features used")
373.  plt.ylabel("AUC Score")
374.  plt.plot(svm_labels_3,svm_scores_3)
375.  plt.savefig("Figures/svm_features_vary")
376.  plt.show()
377.
378.  print(np.arange(0.05,0.2,0.01)[svm_scores_2.index(max(svm_scores_2))])
379.
380.  svm_model = svm.SVC(class_weight={0:0.11,1:1})
381.
382.  labels = [entry[0] for entry in sorted_coef]
383.  subset = labels[-16:]
384.  svm_model.fit(X_train[subset],y_train)
385.
386.  svm_pred = svm_model.predict(X_test[subset])
387.  svm_matrix = nicer_confusion_matrix(y_test, svm_pred)
388.  svm_acc = accuracy_score(y_test,svm_pred)
389.  svm_auc = roc_auc_score(y_test,svm_pred)
390.  svm_recall = recall_score(y_test,svm_pred)
391.
392.  svm_matrix = nicer_confusion_matrix(y_test, svm_pred)
393.  svm_acc = accuracy_score(y_test,svm_pred)
394.  svm_auc = roc_auc_score(y_test,svm_pred)
395.  svm_recall = recall_score(y_test,svm_pred)
396.  svm_precision = precision_score(y_test,y_pred)
397.  svm_f1 = f1_score(y_test,y_pred)
398.
399.  print(svm_matrix)
400.  print(f"Accuracy: {svm_acc}")
401.  print(f"AUC: {svm_auc}")
402.  print(f"Recall: {svm_recall}")
403.  print(f"Precision: {svm_precision}")
404.  print(f"F1: {svm_f1}")
405.
406.  forest_scores = []
407.  forest_labels = []
408.  for i in tqdm(np.arange(0,2,0.1),desc = "Forest weights"):
409.      forest = RandomForestClassifier(class_weight={0:i,1:1})
410.      forest.fit(X_train,y_train)
411.      forest_pred = forest.predict(X_test)
412.      forest_auc = roc_auc_score(y_test,forest_pred)
413.      forest_scores.append(forest_auc)
414.      forest_labels.append(i)
415.
416.  plt.figure(figsize=(12,4))
417.  plt.plot(forest_labels,forest_scores)
418.  plt.xlabel("Weight ratio (:1)")
419.  plt.ylabel("AUC")
420.  plt.title("Decision Forest AUC changing with weighting ratio")
421.  plt.savefig("Figures/forest_weights")
422.  plt.show()
423.
424.  forest = RandomForestClassifier(class_weight={0:0.11,1:1})
425.  forest.fit(X_train,y_train)
```

```
426.  forest_pred = forest.predict(X_test)
427.  forest_matrix = nicer_confusion_matrix(y_test, forest_pred)
428.  forest_acc = accuracy_score(y_test,forest_pred)
429.  forest_auc = roc_auc_score(y_test,forest_pred)
430.  forest_recall = recall_score(y_test,forest_pred)
431.  forest_precision = f1_score(y_test,forest_pred)
432.  forest_f1 = precision_score(y_test,forest_pred)
433.
434.  print(forest_matrix)
435.  print(f"Accuracy: {forest_acc}")
436.  print(f"AUC: {forest_auc}")
437.  print(f"Recall: {forest_recall}")
438.  print(f"Precision: {forest_precision}")
439.  print(f"F1 score: {forest_f1}")
```

Appendix C

EPD Dataset Structure

Column	Title	Type	Description
YEAR_MONTH	Year and Month as YYYYMM	number	Example: 201401
REGIONAL_OFFICE_NAME	Regional Office Name	string	The name given to a geographical region by NHS England. Each region supports local systems to provide more joined up and care for patients.
REGIONAL_OFFICE_CODE	Regional Office Code	string	The unique code used to refer to a Regional Office.
ICB_NAME	Integrated Care Board (ICB) Name	string	The name given to a geographical statutory organisation by NHS England that is a smaller division of a Region.
ICB_CODE	Integrated Care Board (ICB) Code	string	The unique code used to refer to an ICB.
PCO_NAME	Primary Care Organisation Name	string	An NHS organisation that commissions or provides care services involving prescriptions that are dispensed in the community. For example: a Sub Integrated Care Board Location (SICBL), an NHS Trust.
PCO_CODE	Primary Care Organisation Code	string	The unique code used to refer to a Primary Care Organisation.
PRACTICE_NAME	Practice Name	string	The name of an organisation that employs one or more prescribers who issue prescriptions that may be dispensed in the community. For example: a GP Practice, an

Column	Title	Type	Description
			Out-of-Hours service, a hospital department within an NHS Trust.
PRACTICE_CODE	Practice Code	string	The unique code used to refer to a Practice.
ADDRESS_1	Address Field 1	string	The Address used by a Practice. This data is supplied by Primary Care Support England (PSCE), NHS England ICBs or the SICBL, whenever a new practice is opened or if a change of details is required.
ADDRESS_2	Address Field 2	string	The Address used by a Practice. This data is supplied by Primary Care Support England (PSCE), NHS England ICBs or the SICBL, whenever a new practice is opened or if a change of details is required.
ADDRESS_3	Address Field 3	string	The Address used by a Practice. This data is supplied by Primary Care Support England (PSCE), NHS England ICBs or the SICBL, whenever a new practice is opened or if a change of details is required.
ADDRESS_4	Address Field 4	string	The Address used by a Practice. This data is supplied by Primary Care Support England (PSCE), NHS England ICBs or the SICBL, whenever a new practice is opened or if a change of details is required.
POSTCODE	Post Code	string	The Address used by a Practice. This data is supplied by Primary Care Support England (PSCE), NHS England ICBs or the SICBL, whenever a new practice is opened or if a change of details is required.

Column	Title	Type	Description
BNF_CHEMICAL_SUBSTANCE	British National Formulary (BNF) Chemical Substance Code	string	A unique code used to refer to a BNF Chemical Substance. For example, 0501013B0
CHEMICAL_SUBSTANCE_BNF_DESCR	British National Formulary (BNF) Chemical Substance Description	string	The name of the main active ingredient in a drug or the type of an appliance. Determined by the British National Formulatory (BNF) for drugs, or the NHS BSA for appliances. For example, Amoxicillin
BNF_CODE	British National Formulary (BNF) Code	string	The unique code used to refer to a BNF Presentation. For example, 0501013B0AAABAB
BNF_DESCRIPTION	British National Formulary (BNF) Description	string	The name given to the specific type, strength, and formulation of a drug; or, the specific type of an appliance. For example, Amoxicillin 500mg capsules
BNF_CHAPTER_PLUS_CODE	British National Formulary (BNF) Description	string	The name given to a British National Formulatory (BNF) Chapter that includes the prescribed product. Includes the numerical code used to refer to the chapter. For example, 05: Infections
QUANTITY	Quantity	number	The quantity of a medicine, dressing or appliance for which an individual item was prescribed and dispensed, for each BNF Presentation. This represents a pseudo pack size, to illustrate the typical range of prescribed quantities of a given presentation. For example, a quantity of 28 for Amoxicillin 500mg capsules means that the

Column	Title	Type	Description
			pack size dispensed was 28 capsules.
ITEMS	Items	number	The number of times a product appears on a prescription form. Prescription forms include both paper prescriptions and electronic messages.
TOTAL_QUANTITY	Total Quantity	number	The total quantity of a drug or appliance that was prescribed. This is calculated by multiplying Quantity by Items. For example, if 2 items of Amoxicillin 500mg capsules with a quantity of 28 were prescribed, total quantity will be 56.
ADQUSAGE	Average Daily Quantity (ADQ)	number	Average Daily Quantity (ADQ) is the typical daily dose of a medication, prescribed to adult patients by GP Practices. This field shows the quantity prescribed multiplied by the strength, which is then divided by the Average Daily Quantity value.
NIC	Net Ingredient Cost (NIC)	number	In GBP. The amount that would be paid using the basic price of the prescribed drug or appliance and the quantity prescribed. Sometimes called the "Net Ingredient Cost" (NIC).The basic price is given either in the Drug Tariff or is determined from prices published by manufacturers, wholesalers or suppliers. Basic price is set out in Parts 8 and 9 of the Drug Tariff. For any drugs or appliances not in Part 8, the price is usually taken from the manufacturer, wholesaler or supplier of the product.

Column	Title	Type	Description
ACTUAL_COST	Actual Cost	number	In GBP. The basic cost after adjustment for the national average discount and some payments to the dispenser. See notes for details of the calculation. The calculation is: Net Ingredient Cost - National Average Discount Percentage + (payment for consumables + out of pocket expenses + payment for containers)
UNIDENTIFIED	Unidentified	string	This field shows data from prescription forms that could not be allocated to a Practice. Please see the guidance notes for a more detailed explanation of this field, and its uses.

Appendix D

Tables from Chapter 3

icb_name	count_adhd	non_adhd	percent_total	number_of_patients	adhd_per_pop	GDHI/head	unemployment_f	life_expectancy_m	life_expectancy_avg	life_expectancy	anxiety	happiness	life_satisfaction	worthwhile
NHS BATH AND NORTH EAST SOMERSET, SWINDO	106603	10573088.5	0.1000	983571	0.1084	23065	1.80	83.86	80.43	82.15	3.48	7.28	7.29	7.57
NHS BEDFORDSHIRE, LUTON AND MILTON KEYNE	139135	11881026.6	0.1171	1083260	0.1284	20538	3.98	82.92	79.22	81.07	2.85	7.35	7.48	7.46
NHS BIRMINGHAM AND SOLIHULL INTEGRATED C	137050	21548178.8	0.0636	1352212	0.1014	19545	5.95	82.13	77.26	79.70	3.54	7.16	7.20	7.70
NHS BLACK COUNTRY INTEGRATED CARE BOARD	142436	24029239 1.8	0.0593	1510729	0.0943	15547	6.60	81.76	77.60	79.68	3.20	7.29	7.47	7.75
NHS BRISTOL, NORTH SOMERSET AND SOUTH GL	120272	10612748 5.7	0.1133	1062335	0.1132	20249	3.40	83.82	80.05	81.93	3.61	6.99	7.15	7.47
NHS BUCKINGHAMSHIRE, OXFORDSHIRE AND BER	256158	19356062 5.4	0.1323	1948711	0.1314	26156	2.33	84.54	81.22	82.88	3.25	7.43	7.52	7.72
NHS CAMBRIDGESHIRE AND PETERBOROUGH INTE	111767	11080775 7	0.1009	1028927	0.1086	20791	2.40	83.45	79.89	81.67	3.21	7.46	7.56	7.84
NHS CHESHIRE AND MERSEYSIDE INTEGRATED C	480573	45331691 5.8	0.1055	2721709	0.1766	20410	3.27	82.87	79.06	80.96	3.12	7.39	7.41	7.73
NHS CORNWALL AND THE ISLES OF SCILLY INT	73913	7288029 2.6	0.1014	598439	0.1235	18869	2.60	83.47	79.52	81.50	3.13	7.52	7.56	7.78
NHS COVENTRY AND WARWICKSHIRE INTEGRATED	105494	11699346 7.7	0.0902	1055677	0.0999	23867	2.80	83.07	79.37	81.22	3.30	7.41	7.59	7.77
NHS DERBY AND DERBYSHIRE INTEGRATED CARE	155829	14869357 8	0.1048	1078868	0.1444	18119	2.50	83.06	79.49	81.27	3.34	7.37	7.42	7.88
NHS DEVON INTEGRATED CARE BOARD	156229	16529947 4.5	0.0945	1272976	0.1227	19890	2.00	84.36	80.46	82.41	3.10	7.64	7.63	7.88
NHS DORSET INTEGRATED CARE BOARD	99987	10230369 3.2	0.0977	821082	0.1218	22613	2.20	85.04	81.52	83.28	3.15	7.66	7.76	7.98
NHS FRIMLEY INTEGRATED CARE BOARD	97538	8016761 1.5	0.1217	814858	0.1197	27632	2.20	84.12	81.29	82.70	3.33	7.42	7.48	7.77
NHS GLOUCESTERSHIRE INTEGRATED CARE BOAR	57905	6025743 4	0.0961	676808	0.0856	22516	2.20	83.68	80.10	81.89	3.21	7.38	7.50	7.74
NHS GREATER MANCHESTER INTEGRATED CARE B	353345	50325181 6.6	0.0702	3202678	0.1103	17767	5.00	81.39	77.80	79.60	3.52	7.12	7.12	7.53
NHS HAMPSHIRE AND ISLE OF WIGHT INTEGRAT	213220	24086779 4.5	0.0885	1924145	0.1108	22535	2.30	84.32	81.14	82.73	3.05	7.40	7.40	7.68
NHS HEREFORDSHIRE AND WORCESTERSHIRE INT	115417	9825850 3	0.1175	817738	0.1411	22490	2.70	83.79	80.16	81.98	3.34	7.51	7.52	7.81
NHS HERTFORDSHIRE AND WEST ESSEX INTEGRA	169022	16742868 6	0.1010	1614786	0.1047	26790	2.50	84.06	80.75	82.40	3.37	7.29	7.36	7.69
NHS HUMBER AND NORTH YORKSHIRE INTEGRATE	112932	26003262 8.1	0.0434	1775517	0.0636	22229	2.00	84.01	80.35	82.18	3.36	7.24	7.32	7.67
NHS KENT AND MEDWAY INTEGRATED CARE BOAR	388445	24855687 5.3	0.1563	1970903	0.1971	22110	3.20	83.57	80.11	81.84	3.14	7.42	7.36	7.75
NHS LANCASHIRE AND SOUTH CUMBRIA INTEGRA	204461	25084528 6.3	0.0815	1819225	0.1124	17827	3.50	82.13	78.49	80.31	2.96	7.51	7.64	7.92
NHS LEICESTER, LEICESTERSHIRE AND RUTLAN	158069	15218218 8	0.1039	1187267	0.1331	16982	2.10	84.02	80.45	82.24	3.13	7.38	7.51	7.83
NHS LINCOLNSHIRE INTEGRATED CARE BOARD	128064	11961562 4.6	0.1071	811190	0.1579	18783	3.10	83.16	79.63	81.39	3.34	7.59	7.59	7.92
NHS MID AND SOUTH ESSEX INTEGRATED CARE	178550	12456862 9.3	0.1433	1255429	0.1422	22741	2.80	83.56	80.33	81.94	3.28	7.30	7.51	7.72
NHS NORFOLK AND WAVENEY INTEGRATED CARE	159801	15485885 5.2	0.1032	1082501	0.1476	19249	2.80	83.81	80.18	81.99	3.14	7.44	7.53	7.81
NHS NORTH CENTRAL LONDON INTEGRATED CARE	204912	18117788 3.9	0.1131	1753725	0.1168	23619	5.70	86.66	81.78	84.22	3.64	6.93	6.78	7.27
NHS NORTH EAST AND NORTH CUMBRIA INTEGRA	374009	47815044 0.5	0.0782	3153940	0.1186	20232	2.30	82.53	79.38	80.95	3.31	7.27	7.36	7.68
NHS NORTH EAST LONDON INTEGRATED CARE BO	158030	26738467 4.5	0.0591	2369569	0.0667	23619	5.50	82.15	77.61	79.88	3.17	7.87	7.35	7.84
NHS NORTH WEST LONDON INTEGRATED CARE BO	237230	23075272 6.9	0.1028	2783728	0.0852	29141	3.80	83.95	79.99	81.97	3.63	7.33	7.55	7.64
NHS NORTHAMPTONSHIRE INTEGRATED CARE BOA	117808	9089103 5.5	0.1296	804839	0.1464	19305	3.40	83.03	79.46	81.24	2.96	7.52	7.49	7.88
NHS NOTTINGHAM AND NOTTINGHAMSHIRE INTEG	146318	13043930 5.8	0.1122	1120897	0.1305	17346	3.00	82.98	79.52	81.25	3.37	7.23	7.37	7.72
NHS SHROPSHIRE, TELFORD AND WREKIN INTEG	63078	7164570 4.8	0.0880	520624	0.1212	18991	3.00	82.96	79.44	81.20	3.29	7.48	7.49	7.84
NHS SOMERSET INTEGRATED CARE BOARD	49106	6631290 7.8	0.0741	596371	0.0823	20827	2.40	84.33	80.62	82.47	3.15	7.50	7.59	7.93
NHS SOUTH EAST LONDON INTEGRATED CARE BO	211767	18412379 8.2	0.1150	2055970	0.1030	28134	5.80	82.53	79.04	80.79	3.69	7.15	7.09	7.52
NHS SOUTH WEST LONDON INTEGRATED CARE BO	206367	15128085 0.7	0.1364	1734036	0.1190	51673	3.40	83.56	80.32	81.94	2.98	7.38	7.51	7.60
NHS SOUTH YORKSHIRE INTEGRATED CARE BOAR	186868	22419194 5.4	0.0834	1604429	0.1165	16653	4.10	81.95	78.36	80.15	2.98	7.56	7.74	8.01
NHS STAFFORDSHIRE AND STOKE-ON-TRENT INT	71443	2032680 14.8	0.0351	1172941	0.0609	17574	4.00	83.16	79.68	81.42	3.48	7.24	7.40	7.71
NHS SUFFOLK AND NORTH EAST ESSEX INTEGRA	146223	13805861 9.1	0.1059	1050121	0.1392	20323	2.80	84.18	80.78	82.48	3.19	7.39	7.59	7.79
NHS SURREY HEARTLANDS INTEGRATED CARE BO	188089	11591652 3.8	0.1623	1127788	0.1668	30842	2.10	84.58	81.69	83.13	3.33	7.42	7.48	7.77
NHS SUSSEX INTEGRATED CARE BOARD	207978	24980321 0.7	0.0833	1823458	0.1141	24580	3.15	84.14	80.49	82.32	3.44	7.40	7.49	7.73
NHS WEST YORKSHIRE INTEGRATED CARE BOARD	179401	38831329 4.9	0.0462	2624688	0.0684	17347	4.80	82.07	78.29	80.18	3.00	7.41	7.37	7.74

year	month	chemical_substance_bnf_descr	count	cost	price_per_unit
2014	1	Methylphenidate hydrochloride	2844406	2273045.43	0.80
2014	1	Atomoxetine hydrochloride	271285	587888.41	2.17
2014	1	Lisdexamfetamine dimesylate	22212	49307.49	2.22
2014	1	Dexamfetamine sulfate	394587	235555.07	0.60
2014	2	Dexamfetamine sulfate	375422	217482.45	0.58
2014	2	Methylphenidate hydrochloride	2541938	2028324.27	0.80
2014	2	Lisdexamfetamine dimesylate	21663	48161.63	2.22
2014	2	Atomoxetine hydrochloride	241539	521259.09	2.16
2014	3	Atomoxetine hydrochloride	265902	567230.95	2.13
2014	3	Methylphenidate hydrochloride	2842642	2243890.74	0.79
2014	3	Dexamfetamine sulfate	397834	229828.06	0.58
2014	3	Lisdexamfetamine dimesylate	27763	60780.63	2.19
2014	4	Methylphenidate hydrochloride	2680664	2117452.88	0.79
2014	4	Dexamfetamine sulfate	402375	230324.72	0.57
2014	4	Lisdexamfetamine dimesylate	29138	64097.86	2.20
2014	4	Atomoxetine hydrochloride	253088	540391.20	2.14
2014	5	Methylphenidate hydrochloride	2851128	2217656.67	0.78
2014	5	Dexamfetamine sulfate	408010	260689.58	0.64
2014	5	Lisdexamfetamine dimesylate	34389	75570.45	2.20
2014	5	Atomoxetine hydrochloride	269542	576734.87	2.14
2014	6	Lisdexamfetamine dimesylate	36349	80246.70	2.21
2014	6	Methylphenidate hydrochloride	2881798	2283945.64	0.79
2014	6	Dexamfetamine sulfate	401919	268746.18	0.67
2014	6	Atomoxetine hydrochloride	262710	561873.56	2.14
2014	7	Dexamfetamine sulfate	427148	294898.51	0.69
2014	7	Methylphenidate hydrochloride	2940702	2331272.96	0.79
2014	7	Lisdexamfetamine dimesylate	40988	90825.59	2.22
2014	7	Atomoxetine hydrochloride	280396	599401.14	2.14
2014	8	Atomoxetine hydrochloride	243840	522447.41	2.14
2014	8	Lisdexamfetamine dimesylate	39083	86676.54	2.22
2014	8	Methylphenidate hydrochloride	2571409	2032449.88	0.79
2014	8	Dexamfetamine sulfate	376435	268070.93	0.71
2014	9	Atomoxetine hydrochloride	272036	581530.59	2.14
2014	9	Methylphenidate hydrochloride	2933336	2318451.49	0.79
2014	9	Dexamfetamine sulfate	383067	288257.27	0.75
2014	9	Lisdexamfetamine dimesylate	48280	107055.56	2.22
2014	10	Dexamfetamine sulfate	442341	317920.57	0.72
2014	10	Methylphenidate hydrochloride	3009476	2373854.13	0.79
2014	10	Lisdexamfetamine dimesylate	51594	114130.26	2.21
2014	10	Atomoxetine hydrochloride	279741	597920.09	2.14
2014	11	Methylphenidate hydrochloride	2804995	2224426.24	0.79
2014	11	Lisdexamfetamine dimesylate	54570	120640.55	2.21
2014	11	Dexamfetamine sulfate	390397	287778.14	0.74

year	month	chemical_substance_bnf_descr	count	cost	price_per_unit
2014	11	Atomoxetine hydrochloride	253473	543053.66	2.14
2014	12	Methylphenidate hydrochloride	3070716	2439146.01	0.79
2014	12	Atomoxetine hydrochloride	289634	621108.91	2.14
2014	12	Dexamfetamine sulfate	439043	326590.12	0.74
2014	12	Lisdexamfetamine dimesylate	62224	137640.49	2.21
2015	1	Atomoxetine hydrochloride	272430	585580.70	2.15
2015	1	Dexamfetamine sulfate	419213	311663.65	0.74
2015	1	Lisdexamfetamine dimesylate	63174	140175.30	2.22
2015	1	Methylphenidate hydrochloride	3012509	2387292.45	0.79
2015	2	Methylphenidate hydrochloride	2722366	2161701.72	0.79
2015	2	Atomoxetine hydrochloride	257511	554838.95	2.15
2015	2	Lisdexamfetamine dimesylate	59668	132584.98	2.22
2015	2	Dexamfetamine sulfate	373421	274708.63	0.74
2015	3	Methylphenidate hydrochloride	3067313	2445003.45	0.80
2015	3	Dexamfetamine sulfate	427958	319868.24	0.75
2015	3	Lisdexamfetamine dimesylate	72824	161083.98	2.21
2015	3	Atomoxetine hydrochloride	282723	608653.17	2.15
2015	4	Atomoxetine hydrochloride	272002	587801.52	2.16
2015	4	Methylphenidate hydrochloride	2971419	2356251.63	0.79
2015	4	Dexamfetamine sulfate	411449	312086.49	0.76
2015	4	Lisdexamfetamine dimesylate	71903	159441.86	2.22
2015	5	Dexamfetamine sulfate	401171	298917.52	0.75
2015	5	Methylphenidate hydrochloride	2976283	2355618.03	0.79
2015	5	Atomoxetine hydrochloride	272917	587181.19	2.15
2015	5	Lisdexamfetamine dimesylate	75462	167233.43	2.22
2015	6	Methylphenidate hydrochloride	3132700	2483681.91	0.79
2015	6	Atomoxetine hydrochloride	285709	618893.25	2.17
2015	6	Dexamfetamine sulfate	425268	325044.27	0.76
2015	6	Lisdexamfetamine dimesylate	85109	189010.97	2.22
2015	7	Dexamfetamine sulfate	422554	319149.00	0.76
2015	7	Methylphenidate hydrochloride	3171527	2510609.39	0.79
2015	7	Atomoxetine hydrochloride	299890	647114.09	2.16
2015	7	Lisdexamfetamine dimesylate	88598	197085.71	2.22
2015	8	Atomoxetine hydrochloride	256309	555404.12	2.17
2015	8	Methylphenidate hydrochloride	2690335	2140173.78	0.80
2015	8	Lisdexamfetamine dimesylate	75795	169440.65	2.24
2015	8	Dexamfetamine sulfate	386742	289133.00	0.75
2015	9	Methylphenidate hydrochloride	3137701	2482949.43	0.79
2015	9	Lisdexamfetamine dimesylate	92320	205074.72	2.22
2015	9	Atomoxetine hydrochloride	291914	627136.35	2.15
2015	9	Dexamfetamine sulfate	417600	311237.50	0.75
2015	10	Methylphenidate hydrochloride	3111852	2479471.00	0.80
2015	10	Lisdexamfetamine dimesylate	97472	216812.53	2.22

year	month	chemical_substance_bnf_descr	count	cost	price_per_unit
2015	10	Atomoxetine hydrochloride	290217	625654.92	2.16
2015	10	Dexamfetamine sulfate	426621	323514.82	0.76
2015	11	Methylphenidate hydrochloride	3016986	2412480.37	0.80
2015	11	Atomoxetine hydrochloride	286783	604502.70	2.11
2015	11	Dexamfetamine sulfate	404138	301982.17	0.75
2015	11	Lisdexamfetamine dimesylate	101118	224715.90	2.22
2015	12	Methylphenidate hydrochloride	3275182	2635346.45	0.80
2015	12	Lisdexamfetamine dimesylate	112382	249162.57	2.22
2015	12	Atomoxetine hydrochloride	314169	668103.46	2.13
2015	12	Dexamfetamine sulfate	433506	328839.16	0.76
2016	1	Methylphenidate hydrochloride	3056548	2437701.74	0.80
2016	1	Atomoxetine hydrochloride	295154	616298.68	2.09
2016	1	Lisdexamfetamine dimesylate	108593	241745.05	2.23
2016	1	Dexamfetamine sulfate	399996	300770.29	0.75
2016	2	Methylphenidate hydrochloride	2986540	2399641.02	0.80
2016	2	Dexamfetamine sulfate	385331	298512.79	0.77
2016	2	Atomoxetine hydrochloride	279988	595718.92	2.13
2016	2	Lisdexamfetamine dimesylate	113614	253458.99	2.23
2016	2	Guanfacine	112	209.71	1.87
2016	3	Atomoxetine hydrochloride	307159	638907.94	2.08
2016	3	Methylphenidate hydrochloride	3152427	2540319.18	0.81
2016	3	Lisdexamfetamine dimesylate	123801	275252.37	2.22
2016	3	Dexamfetamine sulfate	397381	316797.67	0.80
2016	3	Guanfacine	996	1917.44	1.93
2016	4	Atomoxetine hydrochloride	310565	651471.07	2.10
2016	4	Methylphenidate hydrochloride	3207589	2578586.72	0.80
2016	4	Dexamfetamine sulfate	396534	321428.57	0.81
2016	4	Lisdexamfetamine dimesylate	129544	288192.28	2.22
2016	4	Guanfacine	1833	3566.44	1.95
2016	5	Atomoxetine hydrochloride	308092	543254.32	1.76
2016	5	Methylphenidate hydrochloride	3171963	2559855.69	0.81
2016	5	Dexamfetamine sulfate	393547	312078.30	0.79
2016	5	Lisdexamfetamine dimesylate	132986	296658.55	2.23
2016	5	Guanfacine	3351	6712.73	2.00
2016	6	Atomoxetine hydrochloride	319585	558225.20	1.75
2016	6	Lisdexamfetamine dimesylate	139262	310031.81	2.23
2016	6	Methylphenidate hydrochloride	3265126	2630322.19	0.81
2016	6	Dexamfetamine sulfate	389925	311301.63	0.80
2016	6	Guanfacine	3619	7309.63	2.02
2016	7	Methylphenidate hydrochloride	3144487	2551705.56	0.81
2016	7	Atomoxetine hydrochloride	317558	552334.84	1.74
2016	7	Dexamfetamine sulfate	398566	315445.56	0.79
2016	7	Lisdexamfetamine dimesylate	139534	311248.89	2.23

year	month	chemical_substance_bnf_descr	count	cost	price_per_unit
2016	7	Guanfacine	4364	8766.39	2.01
2016	8	Methylphenidate hydrochloride	2921191	2352472.43	0.81
2016	8	Atomoxetine hydrochloride	299004	521864.87	1.75
2016	8	Lisdexamfetamine dimesylate	132878	296504.18	2.23
2016	8	Dexamfetamine sulfate	393707	324653.08	0.82
2016	8	Guanfacine	4683	9460.34	2.02
2016	9	Lisdexamfetamine dimesylate	155893	346999.82	2.23
2016	9	Dexamfetamine sulfate	398592	322712.00	0.81
2016	9	Methylphenidate hydrochloride	3287276	2635358.97	0.80
2016	9	Atomoxetine hydrochloride	323257	564430.23	1.75
2016	9	Guanfacine	6615	13423.95	2.03
2016	10	Methylphenidate hydrochloride	3093981	2501097.14	0.81
2016	10	Atomoxetine hydrochloride	314601	532205.42	1.69
2016	10	Lisdexamfetamine dimesylate	149074	331692.51	2.23
2016	10	Dexamfetamine sulfate	408051	318749.71	0.78
2016	10	Guanfacine	6908	14014.06	2.03
2016	11	Methylphenidate hydrochloride	3249320	2629561.74	0.81
2016	11	Lisdexamfetamine dimesylate	163417	363251.01	2.22
2016	11	Atomoxetine hydrochloride	335986	566788.70	1.69
2016	11	Dexamfetamine sulfate	400949	315082.33	0.79
2016	11	Guanfacine	9056	18487.92	2.04
2016	12	Methylphenidate hydrochloride	3322971	2694817.79	0.81
2016	12	Atomoxetine hydrochloride	344981	577024.25	1.67
2016	12	Lisdexamfetamine dimesylate	170185	378532.63	2.22
2016	12	Dexamfetamine sulfate	424608	331077.72	0.78
2016	12	Guanfacine	10167	20731.50	2.04
2017	1	Methylphenidate hydrochloride	3197142	2584201.05	0.81
2017	1	Lisdexamfetamine dimesylate	167061	370820.67	2.22
2017	1	Atomoxetine hydrochloride	334141	553855.43	1.66
2017	1	Dexamfetamine sulfate	400562	309268.42	0.77
2017	1	Guanfacine	11148	22983.30	2.06
2017	2	Methylphenidate hydrochloride	2939005	2401247.42	0.82
2017	2	Atomoxetine hydrochloride	310132	516656.57	1.67
2017	2	Dexamfetamine sulfate	379334	301127.00	0.79
2017	2	Lisdexamfetamine dimesylate	161393	358908.08	2.22
2017	2	Guanfacine	11643	24379.93	2.09
2017	3	Dexamfetamine sulfate	429420	339952.23	0.79
2017	3	Atomoxetine hydrochloride	360849	596544.14	1.65
2017	3	Methylphenidate hydrochloride	3505301	2832079.07	0.81
2017	3	Lisdexamfetamine dimesylate	192469	427407.70	2.22
2017	3	Guanfacine	14938	30969.35	2.07
2017	4	Lisdexamfetamine dimesylate	170562	380003.75	2.23
2017	4	Atomoxetine hydrochloride	316250	519855.65	1.64

year	month	chemical_substance_bnf_descr	count	cost	price_per_unit
2017	4	Methylphenidate hydrochloride	3034523	2461796.92	0.81
2017	4	Dexamfetamine sulfate	393568	306706.47	0.78
2017	4	Guanfacine	13380	27999.22	2.09
2017	5	Methylphenidate hydrochloride	3357968	2712848.21	0.81
2017	5	Lisdexamfetamine dimesylate	193178	428502.40	2.22
2017	5	Atomoxetine hydrochloride	351483	581707.15	1.66
2017	5	Dexamfetamine sulfate	399612	318062.66	0.80
2017	5	Guanfacine	17834	36764.24	2.06
2017	6	Methylphenidate hydrochloride	3412970	2769315.50	0.81
2017	6	Lisdexamfetamine dimesylate	199313	441720.77	2.22
2017	6	Dexamfetamine sulfate	419846	297686.45	0.71
2017	6	Atomoxetine hydrochloride	356890	583574.47	1.64
2017	6	Guanfacine	19148	39391.30	2.06
2017	7	Atomoxetine hydrochloride	357437	581544.97	1.63
2017	7	Lisdexamfetamine dimesylate	196941	436613.69	2.22
2017	7	Guanfacine	20563	42313.60	2.06
2017	7	Methylphenidate hydrochloride	3266141	2644883.01	0.81
2017	7	Dexamfetamine sulfate	405935	289551.84	0.71
2017	8	Atomoxetine hydrochloride	333377	541079.59	1.62
2017	8	Methylphenidate hydrochloride	3033036	2465022.51	0.81
2017	8	Lisdexamfetamine dimesylate	186446	414487.79	2.22
2017	8	Dexamfetamine sulfate	419572	316419.09	0.75
2017	8	Guanfacine	21269	43809.72	2.06
2017	9	Methylphenidate hydrochloride	3304042	2668659.25	0.81
2017	9	Atomoxetine hydrochloride	351095	563845.01	1.61
2017	9	Lisdexamfetamine dimesylate	203013	449620.31	2.21
2017	9	Guanfacine	24202	49992.08	2.07
2017	9	Dexamfetamine sulfate	403123	308409.41	0.77
2017	10	Methylphenidate hydrochloride	3285916	2665743.09	0.81
2017	10	Guanfacine	24426	50201.05	2.06
2017	10	Lisdexamfetamine dimesylate	208576	462463.72	2.22
2017	10	Atomoxetine hydrochloride	345631	559715.05	1.62
2017	10	Dexamfetamine sulfate	395427	302125.29	0.76
2017	11	Methylphenidate hydrochloride	3371555	2681287.00	0.80
2017	11	Atomoxetine hydrochloride	358848	572623.94	1.60
2017	11	Lisdexamfetamine dimesylate	222067	491574.15	2.21
2017	11	Dexamfetamine sulfate	412255	314173.33	0.76
2017	11	Guanfacine	27183	55982.99	2.06
2017	12	Lisdexamfetamine dimesylate	220203	489138.43	2.22
2017	12	Atomoxetine hydrochloride	363689	576535.72	1.59
2017	12	Methylphenidate hydrochloride	3313867	2648215.41	0.80
2017	12	Guanfacine	29167	59897.52	2.05
2017	12	Dexamfetamine sulfate	407649	316991.65	0.78

year	month	chemical_substance_bnf_descr	count	cost	price_per_unit
2018	1	Methylphenidate hydrochloride	3401830	2698948.76	0.79
2018	1	Lisdexamfetamine dimesylate	227076	501543.42	2.21
2018	1	Atomoxetine hydrochloride	359874	585101.48	1.63
2018	1	Guanfacine	32501	67241.25	2.07
2018	1	Dexamfetamine sulfate	397090	323554.86	0.81
2018	2	Methylphenidate hydrochloride	3065318	2442016.79	0.80
2018	2	Lisdexamfetamine dimesylate	208588	463136.88	2.22
2018	2	Atomoxetine hydrochloride	330127	528809.66	1.60
2018	2	Dexamfetamine sulfate	352915	281587.13	0.80
2018	2	Guanfacine	27996	57564.37	2.06
2018	3	Lisdexamfetamine dimesylate	243806	539678.52	2.21
2018	3	Methylphenidate hydrochloride	3430467	2726625.92	0.79
2018	3	Atomoxetine hydrochloride	370937	584227.80	1.58
2018	3	Dexamfetamine sulfate	406950	321153.51	0.79
2018	3	Guanfacine	36148	74745.35	2.07
2018	4	Methylphenidate hydrochloride	3224794	2567599.07	0.80
2018	4	Lisdexamfetamine dimesylate	227836	505963.60	2.22
2018	4	Atomoxetine hydrochloride	348926	553080.68	1.59
2018	4	Dexamfetamine sulfate	388940	312725.40	0.80
2018	4	Guanfacine	33208	68650.64	2.07
2018	5	Methylphenidate hydrochloride	3466737	2758076.65	0.80
2018	5	Lisdexamfetamine dimesylate	252727	559333.06	2.21
2018	5	Atomoxetine hydrochloride	366521	583412.46	1.59
2018	5	Dexamfetamine sulfate	412443	334310.63	0.81
2018	5	Guanfacine	38518	79643.88	2.07
2018	6	Lisdexamfetamine dimesylate	257166	568627.23	2.21
2018	6	Methylphenidate hydrochloride	3458402	2734402.79	0.79
2018	6	Atomoxetine hydrochloride	378548	578769.94	1.53
2018	6	Guanfacine	39134	81363.61	2.08
2018	6	Dexamfetamine sulfate	410874	334318.32	0.81
2018	7	Methylphenidate hydrochloride	3361432	2648661.51	0.79
2018	7	Atomoxetine hydrochloride	372916	582587.19	1.56
2018	7	Guanfacine	41714	86689.46	2.08
2018	7	Dexamfetamine sulfate	409685	331503.26	0.81
2018	7	Lisdexamfetamine dimesylate	257709	571721.36	2.22
2018	8	Methylphenidate hydrochloride	3158496	2489823.68	0.79
2018	8	Atomoxetine hydrochloride	370122	572343.88	1.55
2018	8	Lisdexamfetamine dimesylate	255952	566800.58	2.21
2018	8	Guanfacine	41543	86197.77	2.07
2018	8	Dexamfetamine sulfate	403844	331690.71	0.82
2018	9	Lisdexamfetamine dimesylate	262299	579332.00	2.21
2018	9	Methylphenidate hydrochloride	3292605	2571440.48	0.78
2018	9	Atomoxetine hydrochloride	355294	550582.66	1.55

year	month	chemical_substance_bnf_descr	count	cost	price_per_unit
2018	9	Dexamfetamine sulfate	370664	306208.52	0.83
2018	9	Guanfacine	42302	87300.48	2.06
2018	10	Guanfacine	48415	100638.74	2.08
2018	10	Methylphenidate hydrochloride	3502352	2759393.53	0.79
2018	10	Dexamfetamine sulfate	410716	335604.83	0.82
2018	10	Atomoxetine hydrochloride	394357	605613.70	1.54
2018	10	Lisdexamfetamine dimesylate	290033	640402.88	2.21
2018	11	Lisdexamfetamine dimesylate	298077	656953.87	2.20
2018	11	Atomoxetine hydrochloride	372570	578454.72	1.55
2018	11	Methylphenidate hydrochloride	3481536	2720303.39	0.78
2018	11	Guanfacine	46914	97434.30	2.08
2018	11	Dexamfetamine sulfate	399150	331599.67	0.83
2018	12	Methylphenidate hydrochloride	3412431	2665120.60	0.78
2018	12	Lisdexamfetamine dimesylate	297141	656398.95	2.21
2018	12	Guanfacine	48853	101647.10	2.08
2018	12	Atomoxetine hydrochloride	374021	578454.74	1.55
2018	12	Dexamfetamine sulfate	410694	333305.14	0.81
2019	1	Methylphenidate hydrochloride	3504419	2720440.60	0.78
2019	1	Atomoxetine hydrochloride	389444	595035.85	1.53
2019	1	Lisdexamfetamine dimesylate	314574	694805.86	2.21
2019	1	Dexamfetamine sulfate	401526	335263.16	0.83
2019	1	Guanfacine	53863	112180.66	2.08
2019	2	Methylphenidate hydrochloride	3203180	2493292.87	0.78
2019	2	Atomoxetine hydrochloride	349420	536811.02	1.54
2019	2	Lisdexamfetamine dimesylate	297386	655356.48	2.20
2019	2	Dexamfetamine sulfate	368357	307642.68	0.84
2019	2	Guanfacine	50809	105737.31	2.08
2019	3	Methylphenidate hydrochloride	3645983	2826335.21	0.78
2019	3	Guanfacine	56867	118943.65	2.09
2019	3	Atomoxetine hydrochloride	393522	601295.86	1.53
2019	3	Lisdexamfetamine dimesylate	344534	759163.81	2.20
2019	3	Dexamfetamine sulfate	392364	325601.34	0.83
2019	4	Guanfacine	56320	117477.56	2.09
2019	4	Lisdexamfetamine dimesylate	329469	726482.42	2.21
2019	4	Atomoxetine hydrochloride	354941	552966.73	1.56
2019	4	Methylphenidate hydrochloride	3408118	2643843.73	0.78
2019	4	Dexamfetamine sulfate	402232	336956.49	0.84
2019	5	Methylphenidate hydrochloride	3746830	2890553.75	0.77
2019	5	Lisdexamfetamine dimesylate	368493	811680.70	2.20
2019	5	Atomoxetine hydrochloride	408449	626960.28	1.53
2019	5	Guanfacine	64758	135362.94	2.09
2019	5	Dexamfetamine sulfate	421415	364288.15	0.86
2019	6	Methylphenidate hydrochloride	3578746	2761407.81	0.77

year	month	chemical_substance_bnf_descr	count	cost	price_per_unit
2019	6	Lisdexamfetamine dimesylate	360393	794655.11	2.20
2019	6	Guanfacine	64121	133685.46	2.08
2019	6	Atomoxetine hydrochloride	381730	590848.19	1.55
2019	6	Dexamfetamine sulfate	387335	332606.08	0.86
2019	7	Atomoxetine hydrochloride	400658	604364.50	1.51
2019	7	Lisdexamfetamine dimesylate	368319	811800.86	2.20
2019	7	Methylphenidate hydrochloride	3556425	2755195.98	0.77
2019	7	Guanfacine	69456	145331.66	2.09
2019	7	Dexamfetamine sulfate	404963	352001.78	0.87
2019	8	Lisdexamfetamine dimesylate	371136	818808.41	2.21
2019	8	Methylphenidate hydrochloride	3455983	2664073.42	0.77
2019	8	Guanfacine	68416	142715.92	2.09
2019	8	Dexamfetamine sulfate	426251	373706.42	0.88
2019	8	Atomoxetine hydrochloride	395686	595518.25	1.51
2019	9	Methylphenidate hydrochloride	3617858	2770276.26	0.77
2019	9	Guanfacine	68792	143515.63	2.09
2019	9	Lisdexamfetamine dimesylate	385865	849250.77	2.20
2019	9	Atomoxetine hydrochloride	395730	582513.98	1.47
2019	9	Dexamfetamine sulfate	395273	343685.95	0.87
2019	10	Lisdexamfetamine dimesylate	422529	928284.35	2.20
2019	10	Methylphenidate hydrochloride	3789056	2892482.16	0.76
2019	10	Atomoxetine hydrochloride	422118	610842.00	1.45
2019	10	Guanfacine	75281	157547.46	2.09
2019	10	Dexamfetamine sulfate	408414	363323.39	0.89
2019	11	Methylphenidate hydrochloride	3685968	2833250.51	0.77
2019	11	Atomoxetine hydrochloride	395878	571697.15	1.44
2019	11	Lisdexamfetamine dimesylate	404032	894272.36	2.21
2019	11	Guanfacine	74696	156748.82	2.10
2019	11	Dexamfetamine sulfate	394301	356247.59	0.90
2019	12	Methylphenidate hydrochloride	3738138	2861220.48	0.77
2019	12	Lisdexamfetamine dimesylate	421188	933149.22	2.22
2019	12	Guanfacine	76995	161685.90	2.10
2019	12	Atomoxetine hydrochloride	409602	587880.49	1.44
2019	12	Dexamfetamine sulfate	407491	369896.19	0.91
2020	1	Methylphenidate hydrochloride	3923798	3005740.15	0.77
2020	1	Lisdexamfetamine dimesylate	469933	1034933.48	2.20
2020	1	Guanfacine	89128	187162.90	2.10
2020	1	Dexamfetamine sulfate	420509	381431.90	0.91
2020	1	Atomoxetine hydrochloride	436920	615723.70	1.41
2020	2	Methylphenidate hydrochloride	3598651	2741339.40	0.76
2020	2	Lisdexamfetamine dimesylate	427116	943065.32	2.21
2020	2	Atomoxetine hydrochloride	393802	548885.50	1.39
2020	2	Guanfacine	82447	172631.31	2.09

year	month	chemical_substance_bnf_descr	count	cost	price_per_unit
2020	2	Dexamfetamine sulfate	387692	353425.93	0.91
2020	3	Methylphenidate hydrochloride	4137391	3160858.93	0.76
2020	3	Lisdexamfetamine dimesylate	505711	1112605.12	2.20
2020	3	Atomoxetine hydrochloride	457668	630856.80	1.38
2020	3	Dexamfetamine sulfate	430953	393018.64	0.91
2020	3	Guanfacine	98591	207015.51	2.10
2020	4	Lisdexamfetamine dimesylate	464144	1025252.10	2.21
2020	4	Methylphenidate hydrochloride	3634810	2775187.42	0.76
2020	4	Atomoxetine hydrochloride	415727	588111.27	1.41
2020	4	Guanfacine	92470	194041.30	2.10
2020	4	Dexamfetamine sulfate	425360	387726.03	0.91
2020	5	Methylphenidate hydrochloride	3407813	2616792.99	0.77
2020	5	Lisdexamfetamine dimesylate	448939	993025.01	2.21
2020	5	Dexamfetamine sulfate	402356	372121.44	0.92
2020	5	Atomoxetine hydrochloride	396344	557799.07	1.41
2020	5	Guanfacine	87262	182844.62	2.10
2020	6	Methylphenidate hydrochloride	3571794	2732515.14	0.77
2020	6	Lisdexamfetamine dimesylate	465917	1028454.82	2.21
2020	6	Dexamfetamine sulfate	415562	385651.13	0.93
2020	6	Atomoxetine hydrochloride	407485	573877.19	1.41
2020	6	Guanfacine	95166	199736.86	2.10
2020	7	Methylphenidate hydrochloride	3654857	2803843.61	0.77
2020	7	Lisdexamfetamine dimesylate	496306	1097961.38	2.21
2020	7	Atomoxetine hydrochloride	431399	595834.07	1.38
2020	7	Guanfacine	103825	217515.75	2.10
2020	7	Dexamfetamine sulfate	429221	396057.31	0.92
2020	8	Methylphenidate hydrochloride	3291959	2513879.87	0.76
2020	8	Lisdexamfetamine dimesylate	454064	1004667.88	2.21
2020	8	Guanfacine	89499	187973.78	2.10
2020	8	Atomoxetine hydrochloride	382748	531899.86	1.39
2020	8	Dexamfetamine sulfate	386667	355215.93	0.92
2020	9	Methylphenidate hydrochloride	3822560	2902855.46	0.76
2020	9	Guanfacine	100552	211227.15	2.10
2020	9	Dexamfetamine sulfate	424215	391176.82	0.92
2020	9	Atomoxetine hydrochloride	430010	583455.82	1.36
2020	9	Lisdexamfetamine dimesylate	520197	1144821.04	2.20
2020	10	Methylphenidate hydrochloride	3911280	2978499.02	0.76
2020	10	Guanfacine	108495	226343.78	2.09
2020	10	Lisdexamfetamine dimesylate	544190	1197377.74	2.20
2020	10	Atomoxetine hydrochloride	436313	596579.74	1.37
2020	10	Dexamfetamine sulfate	432410	402849.69	0.93
2020	11	Methylphenidate hydrochloride	3831900	2901853.46	0.76
2020	11	Dexamfetamine sulfate	413485	384773.84	0.93

year	month	chemical_substance_bnf_descr	count	cost	price_per_unit
2020	11	Lisdexamfetamine dimesylate	543343	1196517.60	2.20
2020	11	Atomoxetine hydrochloride	423462	575630.29	1.36
2020	11	Guanfacine	108778	228118.06	2.10
2020	12	Methylphenidate hydrochloride	3968880	3022357.20	0.76
2020	12	Lisdexamfetamine dimesylate	581507	1282756.58	2.21
2020	12	Guanfacine	117745	246802.15	2.10
2020	12	Atomoxetine hydrochloride	457639	622755.22	1.36
2020	12	Dexamfetamine sulfate	443088	419502.01	0.95
2021	1	Methylphenidate hydrochloride	3779361	2868256.47	0.76
2021	1	Dexamfetamine sulfate	420267	396022.07	0.94
2021	1	Lisdexamfetamine dimesylate	560758	1235891.71	2.20
2021	1	Atomoxetine hydrochloride	431837	589298.60	1.36
2021	1	Guanfacine	112458	236878.60	2.11
2021	2	Methylphenidate hydrochloride	3533032	2689541.19	0.76
2021	2	Atomoxetine hydrochloride	402203	550431.12	1.37
2021	2	Dexamfetamine sulfate	404336	379517.49	0.94
2021	2	Lisdexamfetamine dimesylate	536878	1185198.12	2.21
2021	2	Guanfacine	104205	219116.53	2.10
2021	3	Methylphenidate hydrochloride	4091112	3112596.73	0.76
2021	3	Atomoxetine hydrochloride	463470	625745.23	1.35
2021	3	Lisdexamfetamine dimesylate	629855	1387607.46	2.20
2021	3	Dexamfetamine sulfate	458189	430342.87	0.94
2021	3	Guanfacine	122992	257947.68	2.10
2021	4	Methylphenidate hydrochloride	3992303	3024959.52	0.76
2021	4	Atomoxetine hydrochloride	454422	606766.28	1.34
2021	4	Dexamfetamine sulfate	454561	431958.73	0.95
2021	4	Lisdexamfetamine dimesylate	622918	1372633.33	2.20
2021	4	Guanfacine	121167	255001.53	2.10
2021	5	Dexamfetamine sulfate	436500	411618.94	0.94
2021	5	Methylphenidate hydrochloride	3922953	2977197.98	0.76
2021	5	Atomoxetine hydrochloride	433395	586011.73	1.35
2021	5	Guanfacine	120025	252440.93	2.10
2021	5	Lisdexamfetamine dimesylate	613185	1349447.42	2.20
2021	6	Atomoxetine hydrochloride	461066	628276.07	1.36
2021	6	Dexamfetamine sulfate	459600	433337.84	0.94
2021	6	Methylphenidate hydrochloride	4116056	3114566.29	0.76
2021	6	Lisdexamfetamine dimesylate	657442	1446988.22	2.20
2021	6	Guanfacine	127737	269105.23	2.11
2021	7	Methylphenidate hydrochloride	4054349	3049737.30	0.75
2021	7	Dexamfetamine sulfate	461748	437386.89	0.95
2021	7	Lisdexamfetamine dimesylate	665533	1467233.95	2.20
2021	7	Atomoxetine hydrochloride	454677	522050.26	1.15
2021	7	Guanfacine	129412	273029.46	2.11

year	month	chemical_substance_bnf_descr	count	cost	price_per_unit
2021	8	Methylphenidate hydrochloride	3727427	2826343.41	0.76
2021	8	Atomoxetine hydrochloride	440389	467305.57	1.06
2021	8	Lisdexamfetamine dimesylate	628633	1388590.79	2.21
2021	8	Guanfacine	119492	252901.39	2.12
2021	8	Dexamfetamine sulfate	441025	420156.39	0.95
2021	9	Lisdexamfetamine dimesylate	706972	1554280.15	2.20
2021	9	Methylphenidate hydrochloride	4160160	3129750.39	0.75
2021	9	Atomoxetine hydrochloride	465453	445591.90	0.96
2021	9	Guanfacine	136896	288339.45	2.11
2021	9	Dexamfetamine sulfate	469025	447499.22	0.95
2021	10	Methylphenidate hydrochloride	4043840	3042230.84	0.75
2021	10	Lisdexamfetamine dimesylate	703500	1546681.97	2.20
2021	10	Guanfacine	130240	273875.47	2.10
2021	10	Atomoxetine hydrochloride	443832	388727.02	0.88
2021	10	Dexamfetamine sulfate	452551	433129.14	0.96
2021	11	Methylphenidate hydrochloride	4224263	3185355.04	0.75
2021	11	Dexamfetamine sulfate	478144	459849.45	0.96
2021	11	Atomoxetine hydrochloride	476208	375961.42	0.79
2021	11	Lisdexamfetamine dimesylate	742749	1635155.54	2.20
2021	11	Guanfacine	137249	288793.67	2.10
2021	12	Methylphenidate hydrochloride	4368819	3296011.77	0.75
2021	12	Lisdexamfetamine dimesylate	785264	1730954.48	2.20
2021	12	Atomoxetine hydrochloride	499435	369213.22	0.74
2021	12	Dexamfetamine sulfate	468452	459747.30	0.98
2021	12	Guanfacine	151233	318491.58	2.11
2022	1	Methylphenidate hydrochloride	4307197	3232386.54	0.75
2022	1	Lisdexamfetamine dimesylate	784154	1729064.92	2.21
2022	1	Atomoxetine hydrochloride	483618	346997.72	0.72
2022	1	Guanfacine	146971	311228.78	2.12
2022	1	Dexamfetamine sulfate	499282	484911.52	0.97
2022	2	Methylphenidate hydrochloride	3957308	2988337.44	0.76
2022	2	Dexamfetamine sulfate	453040	446509.60	0.99
2022	2	Lisdexamfetamine dimesylate	747822	1651069.33	2.21
2022	2	Guanfacine	136810	288960.56	2.11
2022	2	Atomoxetine hydrochloride	458062	301013.94	0.66
2022	3	Methylphenidate hydrochloride	4633627	3472689.36	0.75
2022	3	Atomoxetine hydrochloride	518136	328217.45	0.63
2022	3	Dexamfetamine sulfate	520460	506729.33	0.97
2022	3	Guanfacine	158366	334162.98	2.11
2022	3	Lisdexamfetamine dimesylate	883416	1945231.32	2.20
2022	4	Methylphenidate hydrochloride	4291677	3240604.74	0.76
2022	4	Lisdexamfetamine dimesylate	878693	1936813.87	2.20
2022	4	Dexamfetamine sulfate	500528	494693.25	0.99

year	month	chemical_substance_bnf_descr	count	cost	price_per_unit
2022	4	Atomoxetine hydrochloride	492468	293185.20	0.60
2022	4	Guanfacine	149392	316078.55	2.12
2022	5	Methylphenidate hydrochloride	4601512	3509326.77	0.76
2022	5	Lisdexamfetamine dimesylate	936616	2062845.29	2.20
2022	5	Atomoxetine hydrochloride	508692	295895.19	0.58
2022	5	Guanfacine	161879	342402.13	2.12
2022	5	Dexamfetamine sulfate	529051	522433.21	0.99
2022	6	Methylphenidate hydrochloride	4612394	3543984.50	0.77
2022	6	Lisdexamfetamine dimesylate	963443	2096657.61	2.18
2022	6	Guanfacine	155258	328064.61	2.11
2022	6	Atomoxetine hydrochloride	514849	283395.98	0.55
2022	6	Dexamfetamine sulfate	528205	526214.20	1.00

Appendix E

ABCD Data Dictionary

Title	Short Name	Categories	Shared Subjects
ABCD dMRI Post Processing QC	abcd_dmriqc01	Evaluated Data	11,610
ABCD dMRI RSI Part 1	abcd_drsip101	Evaluated Data	11,811
ABCD dMRI RSI Part 2	abcd_drsip201	Evaluated Data	11,811
ABCD dMRI RSI Part 3	abcd_drsip301	Evaluated Data	11,811
ABCD dMRI RSI Part 4	abcd_drsip401	Evaluated Data	11,811
ABCD dMRI RSI Part 7	abcd_drsip701	Evaluated Data	11,811
ABCD sMRI Destrieux Parcellation Part 1	abcd_mrisdp10201	Evaluated Data	11,811
ABCD sMRI Destrieux Parcellation Part 2	abcd_mrisdp20201	Evaluated Data	11,811
ABCD sMRI Part 1	abcd_smrip10201	Evaluated Data	11,811
ABCD sMRI Part 2	abcd_smrip20201	Evaluated Data	11,811
ABCD ABCL Scores	abcd_abcls01	Parenting	10,522
ABCD ACS Post Stratification Weights	acspsw03	Socioeconomic	11,876
ABCD Adult Behavior Checklist	abcd_adbc01	Behavior	10,414
ABCD BIRD task Trial Level Behavior Data	abcd_bird_tlbd01	Trial-level Behavior	0
ABCD Barkley Deficits in Executive Functioning Scale	barkley_exec_func01	Cognitive	6,251
ABCD Breastfeeding Questionnaire	breast_feeding01	Parenting	6,251
ABCD Brief Problem Monitor-Teacher Form For Ages 6-18 (BPMT)	abcd_bpmt01	Behavior	11,876
ABCD Cash Choice Task	cct01	Cognitive	11,876
ABCD Child Nutrition Assessment	abcd_cna01	Food	11,225
ABCD Children's Report of Parental Behavioral Inventory	crpbi01	Parenting	11,876

Title	Short Name	Categories	Shared Subjects
ABCD Covid 19 Fitbit Survey	covid19_fitbit_survey01	Activity	0
ABCD Covid 19 Geocoded Data	covid19_geocoded_data01	Linked Data	0
ABCD Cyber Bully	abcd_cb01	Aggression	10,522
ABCD Delay Discounting Trial Level Behavior	abcd_ddtlb01	Task Based	10,406
ABCD Developmental History Questionnaire	dhx01	Med History	11,876
ABCD Difficulty in Emotion Regulation	diff_emotion_reg_p01	Emotions	6,251
ABCD Early Adolescent Temperament Questionnaire Parent	abcd_eatqp01	Personality	10,414
ABCD Emotion Regulation Questionnaire	emotion_reg_erq_feelings01	Emotions	6,251
ABCD Emotional Stroop Task Trial Level Behavior	abcd_esttlb01	Task Based	10,399
ABCD Family History Assessment Part 1	fhxp102	Med History	11,876
ABCD Family History Assessment Part 2	fhxp201	Med History	11,876
ABCD Flanker Millisecond task Trial Level Behavior Data	abcd_flankerms_tlbd01	Trial-level Behavior	0
ABCD Follow-Up Scheduling Screener	abcd_screen02	Questionnaire	10,414
ABCD FreeSurfer QC	abcd_fsurfqc01	Evaluated Data	11,816
ABCD Game of Dice Summary Scores	abcd_gdss01	Task Based	10,414
ABCD Game of Dice Trial Level Behavior	abcd_gdtlb01	Task Based	9,593
ABCD Hormone Saliva Salimetric Scores	abcd_hsss01	Phys Exam	11,876
ABCD International Physical Activity Questionnaire	internat_physical_activ01	Activity	6,251
ABCD Irma Substudy Child	abcd_isc01	Questionnaire	408
ABCD Irma Substudy Parent	abcd_ip01	Questionnaire	466
ABCD Kiddie Schedule for Affective Disorders and Schizophrenia 5 Parent	ksads2daic_use_only_p01	Diagnostic	6,251
ABCD Kiddie Schedule for Affective Disorders and Schizophrenia 5 Youth	ksads2daic_use_only01	Mental Health	6,251
ABCD Little Man Task Summary Scores	lmtp201	Task Based	11,876
ABCD Little Man Task Trial Level Behavior	abcd_lmtlb01	Task Based	11,826
ABCD Longitudinal Parent Demographics Survey	abcd_lpds01	Demographics	11,876
ABCD Longitudinal Parent Diagnostic Interview for DSM-V Background Items Full (KSAD)	abcd_lpksad01	Diagnostic	11,488

Title	Short Name	Categories	Shared Subjects
ABCD Longitudinal Parent Medical History Questionnaire	abcd_lpmh01	Med History	11,488
ABCD Longitudinal Parent Ohio State Traumatic Brain Injury Screen-Short Modified (OTBI)	abcd_lpohstbi01	Trauma	11,488
ABCD Longitudinal Parent Sports and Activities Involvement Questionnaire (SAIQ)	abcd_lpsaiq01	Activity	11,488
ABCD Longitudinal Summary Scores Medical History	abcd_lssmh01	Med History	11,488
ABCD Longitudinal Summary Scores Sports Activity	abcd_lsssa01	Activity	11,488
ABCD Longitudinal Summary Scores Traumatic Brain Injury	abcd_lsstbi01	Trauma	11,488
ABCD Longitudinal Tracking	abcd_lt01	Summary	11,876
ABCD MR Findings	abcd_mrfindings02	Summary	11,876
ABCD MRI Info	abcd_mri01	Summary	11,832
ABCD Mobil Tech from EARS Company	abcd_mte01	Questionnaire	67
ABCD Mobil Tech from EARS Raw Data	abcd_earsraw01	Task Based	68
ABCD Mobil Tech from Vibrent Company	abcd_mtv01	Questionnaire	59
ABCD Multi-Group Ethnic Identity-Revised (MEIM-R)	multigrp_ethnic_id_meim01	Cultural Identity	6,251
ABCD Multidimensional Neglectful Behavior Scale	neglectful_behavior01	Parenting	6,251
ABCD NIH Toolbox Trial Level Behavior	abcd_tb_tlb01	Cognitive	11,876
ABCD Occupation Survey Parent	abcd_occsp01	Questionnaire	10,522
ABCD Other Resilience	abcd_ysr01	Social Adjustment	11,876
ABCD Pain Questionnaire	abcd_pq01	Pain	10,522
ABCD Parent Acculturation Survey Modified from PhenX (ACC)	pacc01	Questionnaire	11,876
ABCD Parent Adult Self Report Raw Scores Aseba (ASR)	pasr01	Questionnaire	11,876
ABCD Parent Adult Self Report Scores Aseba (ASR)	abcd_asrs01	Questionnaire	11,876
ABCD Parent Child Behavior Checklist Raw Scores Aseba (CBCL)	abcd_cbcl01	Behavior	12,120
ABCD Parent Child Behavior Checklist Scores Aseba (CBCL)	abcd_cbcls01	Behavior	11,876
ABCD Parent Community Risk and Protective Factors (CRPF)	abcd_crpf01	Questionnaire	11,876
ABCD Parent Demographics Survey	pdem02	Demographics	11,876

Title	Short Name	Categories	Shared Subjects
ABCD Parent Diagnostic Interview for DSM-V (KSADS) ADHD Individual Questions	attn_deficit_hyperactiv_p01	ADHD	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Agoraphobia Individual Questions	agoraphobia_p01	Mental Health	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Alcohol Use Disorder Individual Questions	alcohol_use_disorder_p01	Substance Use	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Autism Spectrum Disorders Individual Questions	autism_spectrum_dis_p01	Autism Spectrum Disorders	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Bipolar Disorders Individual Questions	bipolar_disorders_p01	Mental Health	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) DMDD Individual Questions	disruptive_mood_dysreg_p01	Emotions	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Depressive Disorders Individual Questions	depressive_disorders_p01	Mental Health	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Drug Use Disorder Individual Questions	drug_use_disorders_p01	Substance Use	3,189
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Eating Disorders Individual Questions	eating_disorders_p01	Eating Disorder	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Generalized Anxiety Disorder Individual Questions	generaled_anx_disorder_p01	Mental Health	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Homicidality Individual Questions	homicidality_p01	Questionnaire	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) ODD Individual Questions	opp_defiant_disorder_p01	Mental Health	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Panic Disorder Individual Questions	panic_disorder_p01	Mental Health	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Psychotic Disorders Individual Questions	psychosis_p01	Mental Health	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Separation Anxiety Disorder Individual Questions	separation_anxiety_p01	Mental Health	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Sleep Problems Individual Questions	sleep_problems_p01	Sleep	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Specific Anxiety Disorder Individual Questions	social_anxiety_disorder_p01	Mental Health	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Specific Phobia Individual Questions	specific_phobia_p01	Mental Health	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Suicidality Individual Questions	suicidality_p01	Suicidality	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Tic Disorder Individual Questions	tic_disorders_p01	Mental Health	11,876
ABCD Parent Diagnostic Interview for DSM-V (KSADS) Traumatic Events	abcd_ptsd01	PTSD	11,876
ABCD Parent Diagnostic Interview for DSM-V Background Items Full (KSADS-5)	dibf01	Social Adjustment	11,876

Title	Short Name	Categories	Shared Subjects
ABCD Parent Diagnostic Interview for DSM-V Full (KSADS-5)	abcd_ksad01	Diagnostic	11,876
ABCD Parent Diagnostic Interview for DSM-V(KSADS) Obsessive Compulsive Disorder Individual Questions	obs_compulsive_disorder_p01	Mental Health	11,876
ABCD Parent Family Environment Scale-Family Conflict Subscale Modified from PhenX (FES)	fes02	Parenting	11,876
ABCD Parent Family History Summary Scores	abcd_fhxssp01	Med History	11,876
ABCD Parent Fitbit Baseline	abcd_pfb01	Questionnaire	164
ABCD Parent Fitbit Followup	abcd_pff01	Questionnaire	11,876
ABCD Parent Gender Identity	abcd_pgi01	Questionnaire	11,488
ABCD Parent KSADS Conduct Disorder	abcd_pksadscd01	Diagnostic	11,876
ABCD Parent Life Events	abcd_ple01	Life Events	11,488
ABCD Parent Medical History Questionnaire (MHX)	abcd_mx01	Med History	11,876
ABCD Parent Medications Survey Inventory Modified from PhenX (PMP)	medsy01	Treatment	11,876
ABCD Parent Mexican American Cultural Values Scale Modified (MACV)	macv01	Social Adjustment	11,876
ABCD Parent Mobil Tech Postassessment	abcd_mtpa01	Questionnaire	10,414
ABCD Parent Mobil Tech Preassessment	abcd_mtpap01	Questionnaire	10,414
ABCD Parent Multi-Group Ethnic Identity-Revised Survey (MEIM)	abcd_meim01	Social Adjustment	11,876
ABCD Parent Neighborhood Safety/Crime Survey Modified from PhenX (NSC)	abcd_pnsc01	Questionnaire	11,876
ABCD Parent Ohio State Traumatic Brain Injury Screen-Short Modified (OTBI)	abcd_otbi01	Trauma	11,876
ABCD Parent Parent General Behavior Inventory-Mania (PGBI)	abcd_pgbi01	Questionnaire	11,876
ABCD Parent Participant Last Use Survey Day 2 3 4 (PLUS)	plus01	Questionnaire	11,876
ABCD Parent PhenX Community Cohesion	abcd_pxccp01	Social Adjustment	10,414
ABCD Parent Pubertal Development Scale and Menstrual Cycle Survey History (PDMS)	abcd_ppdms01	Phys Characteristics	11,876
ABCD Parent School Attendance and Grades	abcd_saag01	Questionnaire	10,522

Title	Short Name	Categories	Shared Subjects
ABCD Parent Screen Time Survey (STQ)	stq01	Questionnaire	11,876
ABCD Parent Screentime Questionnaire	screentime_psq_p01	Activity	6,251
ABCD Parent Short Social Responsiveness Scale	abcd_pssrs01	Social Responsiveness	11,225
ABCD Parent Sleep Disturbance Scale for Children (SDS)	abcd_sds01	Sleep	11,876
ABCD Parent Sports and Activities Involvement Questionnaire (SAIQ)	abcd_saiq02	Activity	11,876
ABCD Parent Survey of Substance Use Density, Storage, and Exposure	abcd_pssudse01	Substance Use	10,522
ABCD Parent Vancouver Index of Acculturation-Short Survey (VIA)	abcd_via01	Social Adjustment	11,876
ABCD Parental Monitoring Survey	pmq01	Parenting	11,876
ABCD Parental Rules on Substance Use	prq01	Substance Use	11,876
ABCD Pearson Scores	abcd_ps01	Cognitive	11,876
ABCD Peer Experiences Questionnaire	abcd_peq01	Social Adjustment	10,522
ABCD Perceived Stress Scale	perceived_stress_p01	Stress	6,251
ABCD Pet Ownership	pet_ownership01	Questionnaire	6,251
ABCD Post-assessment Parent Survey for Fitbit Protocol	abcd_fbpap01	Questionnaire	10,414
ABCD Post-assessment Youth Survey for Fitbit Protocol	abcd_fbpay01	Questionnaire	10,414
ABCD Pre-assessment Parent Survey for Fitbit Protocol	abcd_fbprp01	Questionnaire	10,414
ABCD Pre-assessment Youth Survey for Fitbit Protocol	abcd_fbpry01	Questionnaire	10,414
ABCD Prodromal Psychosis Scale	pps01	Psychosis	11,876
ABCD Pubertal Hormone Saliva	sph01	Phys Exam	11,876
ABCD RA Scanning Checklist and Notes	abcd_ra01	Questionnaire	11,876
ABCD RECMEM Task Trial Level Behavior	recog_memory_trial_level01	Cognitive	10,858
ABCD Raw Data for the Timeline Followback (TLFB) Calendar	abcd_tlfb_tlbd01	Substance Use	0
ABCD Recommended Imaging Inclusion	abcd_imgincl01	Evaluated Data	11,832
ABCD SMARTE Task Trial Level Behavior	smarte_trial_level_behav01	Cognitive	1,478

Title	Short Name	Categories	Shared Subjects
ABCD School Risk and Protective Factors Survey	srpf01	Social Adjustment	11,876
ABCD Screener	abcd_screen01	Questionnaire	11,876
ABCD Social Development Follow Up Alabama Parenting	soc_dev_fu_alabama01	Parenting	1,606
ABCD Social Development Follow Up Child Difficulties in Emotion Regulation	soc_dev_fu_diff_emo_reg01	Emotions	1,606
ABCD Social Development Follow Up Child Feedback	soc_dev_fu_feedback01	Questionnaire	1,606
ABCD Social Development Follow Up Child Firearms	soc_dev_fu_firearms01	Questionnaire	1,606
ABCD Social Development Follow Up Child Peer Behavior	soc_dev_fu_peer_behav01	Relationships	1,606
ABCD Social Development Follow Up Child Personality Disposition	soc_dev_fu_personality01	Personality	1,606
ABCD Social Development Follow Up Child Victimization	soc_dev_fu_victimize01	Trauma	1,606
ABCD Social Development Follow Up Parent Alabama Parenting	soc_dev_fu_alabama_p01	Parenting	1,606
ABCD Social Development Follow Up Parent Emotion Regulation	soc_dev_fu_diff_emo_reg_p01	Emotions	1,606
ABCD Social Development Follow Up Parent Feedback	soc_dev_fu_feedback_p01	Questionnaire	1,606
ABCD Social Development Follow Up Parent Firearms	soc_dev_fu_firearms_p01	Questionnaire	1,606
ABCD Social Development Follow Up Parent Neighborhood	soc_dev_fu_neighbor_p01	Social Adjustment	1,606
ABCD Social Development Follow Up Parent Personality Disposition	soc_dev_fu_personality_p01	Personality	1,606
ABCD Social Development Follow Up Parent Reported Delinquency	soc_dev_fu_rep_delinq_p01	Questionnaire	1,606
ABCD Social Development Follow Up Parent Victimization	soc_dev_fu_victimize_p01	Trauma	1,606
ABCD Social Development Follow Up Reported Delinquency	soc_dev_fu_rep_delinq01	Questionnaire	1,606
ABCD Social Development Follow Up Visit Type	soc_dev_fu_visit_type01	Relationships	1,606
ABCD Social Influence Summary Scores	abcd_siss01	Social Adjustment	10,414
ABCD Social Influence Task Trial Level Behavior	abcd_sitlb01	Task Based	9,511
ABCD Specialty Summary Score	abcd_sss01	Well-Being	11,876
ABCD Sports Activities Read/Music – Parent	sports_activ_read_music_p01	Activity	6,251

Title	Short Name	Categories	Shared Subjects
ABCD Sports Activities Read/Music – Youth	sports_activ_read_music01	Activity	6,251
ABCD Stanford Mental Arithmetic Response Time Evaluation	smarte_sumscores01	Cognitive	6,251
ABCD Sum Scores Culture & Environment Parent	abcd_sscep01	Social Adjustment	11,876
ABCD Sum Scores Culture & Environment Youth	abcd_sscey01	Social Adjustment	11,876
ABCD Sum Scores Mobil Tech Youth	abcd_ssmty01	Med History	11,876
ABCD Sum Scores Physical Health Parent	abcd_ssphp01	Med History	11,876
ABCD Sum Scores Physical Health Youth	abcd_ssphy01	Med History	11,876
ABCD Sum Scores Traumatic Brain Injury	abcd_tbi01	Trauma	11,876
ABCD Summary Scores Brief Problem Monitor-Teacher Form for Ages 6-18	abcd_ssbpmtf01	Behavior	11,876
ABCD Summary Scores Developmental History	abcd_devhxss01	Med History	11,876
ABCD Summary Scores Medical History	abcd_medhxss01	Med History	11,876
ABCD Summary Scores Sports Activity	abcd_spacss01	Activity	11,876
ABCD Summary Scores Substance Use	abcd_suss01	Substance Use	11,876
ABCD Task fMRI MID Average Beta Weights Destrieux Parcellations Part 1	abcd_midabwdp01	Evaluated Data	10,982
ABCD Task fMRI MID Average Beta Weights Destrieux Parcellations Part 2	abcd_midabwdp202	Evaluated Data	10,982
ABCD Task fMRI MID Average Beta Weights Part 1	midaparc03	Evaluated Data	10,982
ABCD Task fMRI MID Average Beta Weights Part 2	midaparcp203	Evaluated Data	10,982
ABCD Task fMRI MID Average SEM Destrieux Parcellations Part 1	abcd_midasemdp101	Evaluated Data	10,982
ABCD Task fMRI MID Average SEM Destrieux Parcellations Part 2	abcd_midasemdp202	Evaluated Data	10,982
ABCD Task fMRI MID Average Standard Error of the Mean Part 1	abcd_midasemp102	Evaluated Data	10,982
ABCD Task fMRI MID Average Standard Error of the Mean Part 2	abcd_midasemp202	Evaluated Data	10,982
ABCD Task fMRI MID Behavior	abcd_mid02	Task Based	11,876
ABCD Task fMRI MID Run 1 Beta Weights Destrieux Parcellations Part 1	abcd_midr1bwdp101	Evaluated Data	10,979
ABCD Task fMRI MID Run 1 Beta Weights Destrieux Parcellations Part 2	abcd_midr1bwdp202	Evaluated Data	10,979

Title	Short Name	Categories	Shared Subjects
ABCD Task fMRI MID Run 1 Beta Weights Part 1	abcd_midr1bwp102	Evaluated Data	10,979
ABCD Task fMRI MID Run 1 Beta Weights Part 2	abcd_midr1bwp202	Evaluated Data	10,979
ABCD Task fMRI MID Run 1 SEM Destrieux Parcellations Part 1	abcd_tmidr1semdp101	Evaluated Data	10,979
ABCD Task fMRI MID Run 1 SEM Destrieux Parcellations Part 2	abcd_tmidr1semdp202	Evaluated Data	10,979
ABCD Task fMRI MID Run 1 Standard Error of the Mean Part 1	abcd_midsemp102	Evaluated Data	10,979
ABCD Task fMRI MID Run 1 Standard Error of the Mean Part 2	abcd_midsemp202	Evaluated Data	10,979
ABCD Task fMRI MID Run 2 Beta Weights Destrieux Parcellations Part 1	abcd_tr2bwdp01	Evaluated Data	10,736
ABCD Task fMRI MID Run 2 Beta Weights Destrieux Parcellations Part 2	abcd_tr2bwdp202	Evaluated Data	10,736
ABCD Task fMRI MID Run 2 Beta Weights Part 1	midr2bwp102	Evaluated Data	10,736
ABCD Task fMRI MID Run 2 Beta Weights Part 2	midr2bwp202	Evaluated Data	10,736
ABCD Task fMRI MID Run 2 SEM Destrieux Parcellations Part 1	abcd_tr2semdp101	Evaluated Data	10,736
ABCD Task fMRI MID Run 2 SEM Destrieux Parcellations Part 2	abcd_tr2semdp202	Evaluated Data	10,736
ABCD Task fMRI MID Run 2 Standard Error of the Mean Part 1	abcd_midr2semp102	Evaluated Data	10,736
ABCD Task fMRI MID Run 2 Standard Error of the Mean Part 2	abcd_midr2semp202	Evaluated Data	10,736
ABCD Task fMRI MID Trial Level Behavior	abcd_midtlb01	Task Based	11,101
ABCD Task fMRI REC Behavior	mribrec02	Task Based	11,876
ABCD Task fMRI SST Average Beta Weights	mrisst02	Evaluated Data	10,898
ABCD Task fMRI SST Average Beta Weights Destrieux Parcellations Part 1	abcd_tfsstabwdp101	Evaluated Data	10,898
ABCD Task fMRI SST Average Beta Weights Destrieux Parcellations Part 2	abcd_tfsstabwdp201	Evaluated Data	10,898
ABCD Task fMRI SST Average SEM Destrieux Parcellations Part 1	abcd_tfsstasemdp101	Evaluated Data	10,898
ABCD Task fMRI SST Average SEM Destrieux Parcellations Part 2	abcd_tfsstasemdp201	Evaluated Data	10,898
ABCD Task fMRI SST Average Standard Error of the Mean	mrisstsem01	Evaluated Data	10,898
ABCD Task fMRI SST Behavior	abcd_sst02	Task Based	11,876
ABCD Task fMRI SST Run 1 Beta Weights	mrisstr1bw01	Evaluated Data	10,897

Title	Short Name	Categories	Shared Subjects
ABCD Task fMRI SST Run 1 Beta Weights Destrieux Parcellations Part 1	abcd_tfsstr1bwdp101	Evaluated Data	10,897
ABCD Task fMRI SST Run 1 Beta Weights Destrieux Parcellations Part 2	abcd_tfsstr1bwdp201	Evaluated Data	10,897
ABCD Task fMRI SST Run 1 SEM Destrieux Parcellations Part 1	abcd_tfsstr1semdp101	Evaluated Data	10,897
ABCD Task fMRI SST Run 1 SEM Destrieux Parcellations Part 2	abcd_tfsstr1semdp201	Evaluated Data	10,897
ABCD Task fMRI SST Run 1 Standard Error of the Mean	mrisstr1sem01	Evaluated Data	10,897
ABCD Task fMRI SST Run 2 Beta Weights	mrisstr2bw01	Evaluated Data	10,646
ABCD Task fMRI SST Run 2 Beta Weights Destrieux Parcellations Part 1	abcd_tfsstr2bwdp101	Evaluated Data	10,646
ABCD Task fMRI SST Run 2 Beta Weights Destrieux Parcellations Part 2	abcd_tfsstr2bwdp201	Evaluated Data	10,646
ABCD Task fMRI SST Run 2 SEM Destrieux Parcellations Part 1	abcd_tfsstr2semdp101	Evaluated Data	10,646
ABCD Task fMRI SST Run 2 SEM Destrieux Parcellations Part 2	abcd_tfsstr2semdp201	Evaluated Data	10,646
ABCD Task fMRI SST Run 2 Standard Error of the Mean	mrisstr2bwsem01	Evaluated Data	10,646
ABCD Task fMRI SST Trial Level Behavior	abcd_sst_tlb01	Task Based	11,028
ABCD Task fMRI nBack Average Beta Weights	nback_bwroi02	Evaluated Data	10,830
ABCD Task fMRI nBack Average Beta Weights Destrieux Parcellations Part 1	abcd_tfabwdp101	Evaluated Data	10,830
ABCD Task fMRI nBack Average Beta Weights Destrieux Parcellations Part 2	abcd_tfabwdp201	Evaluated Data	10,830
ABCD Task fMRI nBack Average SEM Destrieux Parcellations Part 1	abcd_tnbasemdp101	Evaluated Data	10,830
ABCD Task fMRI nBack Average SEM Destrieux Parcellations Part 2	abcd_tnbasemdp201	Evaluated Data	10,830
ABCD Task fMRI nBack Average Standard Error of the Mean	nbackallsem01	Evaluated Data	10,830
ABCD Task fMRI nBack Behavior	abcd_mrinback02	Task Based	11,693
ABCD Task fMRI nBack Run 1 Beta Weights	nbackr101	Evaluated Data	10,826
ABCD Task fMRI nBack Run 1 Beta Weights Destrieux Parcellations Part 1	abcd_tfncr1bwdp101	Evaluated Data	10,826
ABCD Task fMRI nBack Run 1 Beta Weights Destrieux Parcellations Part 2	abcd_tfncr1bwdp201	Evaluated Data	10,826
ABCD Task fMRI nBack Run 1 SEM Destrieux Parcellations Part 1	abcd_tfnbr1semdp101	Evaluated Data	10,826

Title	Short Name	Categories	Shared Subjects
ABCD Task fMRI nBack Run 1 SEM Destrieux Parcellations Part 2	abcd_tfnbr1semdp201	Evaluated Data	10,826
ABCD Task fMRI nBack Run 1 Standard Error of the Mean	nbackr1sem01	Evaluated Data	10,826
ABCD Task fMRI nBack Run 2 Beta Weights	nbackr201	Evaluated Data	10,629
ABCD Task fMRI nBack Run 2 Beta Weights Destrieux Parcellations Part 1	abcd_tfnbr2bwdp101	Evaluated Data	10,629
ABCD Task fMRI nBack Run 2 Beta Weights Destrieux Parcellations Part 2	abcd_tfnbr2bwdp201	Evaluated Data	10,629
ABCD Task fMRI nBack Run 2 SEM Destrieux Parcellations Part 1	abcd_tfnbr2dp101	Evaluated Data	10,629
ABCD Task fMRI nBack Run 2 SEM Destrieux Parcellations Part 2	abcd_tfnbr2dp201	Evaluated Data	10,629
ABCD Task fMRI nBack Run 2 Standard Error of the Mean	nbackr2sem01	Evaluated Data	10,629
ABCD Task fMRI nBack Trial Level Behavior	abcd_nbacktlb01	Task Based	9,891
ABCD The Stanford Education Data Archive (SEDA) part 2	led_school_part_201	Linked Data	10,153
ABCD The Stanford Education Data Archive (SEDA) part 3	led_school_part_301	Linked Data	10,527
ABCD The Stanford Education Data Archive (SEDA) part 4	led_school_part_401	Linked Data	10,527
ABCD The Stanford Education Data Archive (SEDA) part 5	led_school_part_501	Linked Data	10,416
ABCD The Stanford Education Data Archive(SEDA) Part 1	led_school_part_101	Linked Data	10,561
ABCD Timeline Follow-back Survey Calendar Scores (TLFB)	abcd_tlfb01	Substance Use	11,876
ABCD Twin Zygosity Rating	abcd_tztab01	Phys Characteristics	11,876
ABCD Youth 7-Up Mania Items	abcd_y7mi01	Mania	11,368
ABCD Youth Acculturation Survey Modified from PhenX (ACC)	yacc01	Questionnaire	11,876
ABCD Youth Alcohol Measures	abcd_yam01	Substance Use	11,488
ABCD Youth Alcohol Screen	yalcs01	Phys Exam	11,876
ABCD Youth Anthropometrics Modified From PhenX (ANT)	abcd_ant01	Phys Characteristics	11,876
ABCD Youth Behavioral Inhibition/Behavioral Approach System Scales Modified from PhenX (BIS/BAS)	abcd_bisbas01	Questionnaire	11,876
ABCD Youth Block Food Screen	abcd_bkfs01	Food	10,522

Title	Short Name	Categories	Shared Subjects
ABCD Youth Blood Analysis	abcd_ybd01	Phys Exam	10,414
ABCD Youth Blood Pressure	abcd_bp01	Phys Characteristics	10,414
ABCD Youth Brief Problem Monitor	abcd_bpm01	Behavior	11,733
ABCD Youth Community Risk and Protective Factors	abcd_ycrpf01	Substance Use	10,522
ABCD Youth Delay Discounting Sum Scores	abcd_yddss01	Task Based	11,368
ABCD Youth Diagnostic Interview for DSM-V (KSADS) Alcohol Use Disorder Individual Questions	alcohol_use_disorder01	Substance Use	11,876
ABCD Youth Diagnostic Interview for DSM-V (KSADS) Bipolar Disorders Individual Questions	bipolar_disorders01	Mental Health	11,876
ABCD Youth Diagnostic Interview for DSM-V (KSADS) Conduct Disorder Individual Questions	conduct_disorder01	Conduct Disorder	11,876
ABCD Youth Diagnostic Interview for DSM-V (KSADS) DMDD Individual Questions	disruptive_mood_dysreg01	Emotions	11,876
ABCD Youth Diagnostic Interview for DSM-V (KSADS) Depressive Disorders Individual Questions	depressive_disorders01	Mental Health	11,876
ABCD Youth Diagnostic Interview for DSM-V (KSADS) Drug Use Disorder Individual Questions	drug_use_disorders01	Substance Use	7,808
ABCD Youth Diagnostic Interview for DSM-V (KSADS) Eating Disorders Individual Questions	eating_disorders01	Eating Disorder	11,876
ABCD Youth Diagnostic Interview for DSM-V (KSADS) Generalized Anxiety Disorder Individual Questions	generaled_anx_disorder01	Mental Health	11,876
ABCD Youth Diagnostic Interview for DSM-V (KSADS) Sleep Problems Individual Questions	sleep_problems01	Sleep	11,876
ABCD Youth Diagnostic Interview for DSM-V (KSADS) Specific Anxiety Disorder Individual Questions	social_anxiety_disorder01	Mental Health	11,876
ABCD Youth Diagnostic Interview for DSM-V (KSADS) Suicidality Individual Questions	suicidality01	Suicidality	11,876
ABCD Youth Diagnostic Interview for DSM-V 5 (KSADS-5)	abcd_ksad501	Diagnostic	11,876
ABCD Youth Diagnostic Interview for DSM-V Background Items 5 (KSADS-5)	abcd_yksad01	Social Adjustment	11,876
ABCD Youth Discrimination Measure	abcd_ydmes01	Social Adjustment	11,453
ABCD Youth Edinburgh Handedness Inventory Short Form (EHIS)	abcd_ehis01	Cognitive	11,891
ABCD Youth Emotional Stroop Task	abcd_yest01	Task Based	11,368
ABCD Youth Family Environment Scale-Family Conflict Subscale Modified from PhenX (FES)	abcd_fes01	Parenting	11,876
ABCD Youth Fitbit Baseline	abcd_yfb01	Questionnaire	162
ABCD Youth Fitbit Daily Physical Activity Summaries	abcd_fbdpas01	Activity	7,439
ABCD Youth Fitbit Daily Sleep Summaries	abcd_fbdss01	Sleep	5,950

Title	Short Name	Categories	Shared Subjects
ABCD Youth Fitbit Followup	abcd_yff01	Questionnaire	150
ABCD Youth Fitbit Weekly Physical Activity Summaries	abcd_fbwpas01	Activity	7,076
ABCD Youth Fitbit Weekly Sleep Summaries	abcd_fbwss01	Sleep	4,709
ABCD Youth Gender Identity	abcd_ygi01	Questionnaire	11,488
ABCD Youth Genetic Blood (RUCDR)	biocf01	Phys Exam	11,876
ABCD Youth Genetic Saliva (RUCDR)	abcd_ygs01	Phys Exam	11,876
ABCD Youth Gish2	abcd_gish2y01	Relationships	10,522
ABCD Youth Hair Results	abcd_yhr01	Substance Use	1,262
ABCD Youth Hair Sample	abcd_hers01	Substance Use	11,876
ABCD Youth Life Events	abcd_yle01	Life Events	11,488
ABCD Youth Marijuana Illicit Drug Measures	abcd_ymidm01	Substance Use	11,488
ABCD Youth Mexican American Cultural Values Scale	abcd_macvsy01	Questionnaire	10,522
ABCD Youth Mid Year Phone Interview Substance Use	abcd_ymypisu01	Substance Use	11,672
ABCD Youth Mobil Tech Postassessment	abcd_mtpay01	Questionnaire	10,414
ABCD Youth Mobil Tech Preassessment	abcd_mtpry01	Questionnaire	10,414
ABCD Youth Monetary Incentive Delay Task Survey Post Scan Questionnaire	abcd_monet01	Task Based	11,876
ABCD Youth Munich Chronotype Questionnaire	abcd_mcqc01	Questionnaire	10,522
ABCD Youth NIH TB Summary Scores	abcd_tbss01	Cognitive	11,876
ABCD Youth NIH Toolbox Positive Affect Items	abcd_ytbpai01	Cognitive	11,716
ABCD Youth Neighborhood Safety/Crime Survey Modified from PhenX (NSC)	abcd_nsc01	Questionnaire	11,876
ABCD Youth Neurognition Survey Session	neurocog_youth_session01	Cognitive	11,876
ABCD Youth Nicalert	abcd_yn01	Phys Exam	11,488
ABCD Youth Nicotine Measures	abcd_ynm01	Substance Use	11,488
ABCD Youth Participant Last Use Survey Day 1 2 3 4 (PLUS)	abcd_plus01	Questionnaire	11,876
ABCD Youth Peer Behavior Profile	abcd_pbp01	Personality	10,522
ABCD Youth Peer Network Health Protective Scaler	abcd_pnhps01	Activity	10,522
ABCD Youth Post Scan Questionnaire 2	abcd_ypsq201	Questionnaire	11,876

Title	Short Name	Categories	Shared Subjects
ABCD Youth Post Scan Questionnaire 1	abcd_ypsq101	Questionnaire	11,876
ABCD Youth Pre Scan Questionnaire 1	abcd_ypre101	Questionnaire	11,876
ABCD Youth Pre Scan Questionnaire 2	abcd_ypre201	Questionnaire	11,876
ABCD Youth Pubertal Development Scale and Menstrual Cycle Survey History (PDMS)	abcd_ypdms01	Phys Characteristics	11,876
ABCD Youth Rescan Monetary Incentive Delay Task Survey Post Scan Questionnaire	abcd_prepost01	Task Based	11,876
ABCD Youth School Attendance and Grades	abcd_ysaag01	Questionnaire	10,522
ABCD Youth Screen Time Survey (STQ)	abcd_stq01	Questionnaire	11,876
ABCD Youth Snellen Vision Screener (SVS)	abcd_svs01	Phys Exam	11,876
ABCD Youth Substance Use Attitudes	abcd_ysua01	Substance Use	11,488
ABCD Youth Substance Use Interview	abcd_ysu02	Substance Use	11,876
ABCD Youth Substance Use Introduction and Patterns	abcd_ysuip01	Substance Use	11,488
ABCD Youth Summary Scores BPM and POA	abcd_yssbpm01	Behavior	11,733
ABCD Youth Teeth Collection	bteeth01	Phys Characteristics	11,876
ABCD Youth Toxicology Test	abcd_ytt01	Phys Exam	11,876
ABCD Youth Vancouver Index of Acculturation	vancouver_identity_accult01	Social Adjustment	6,251
ABCD Youth Wills Problem Solving Scale	abcd_ywpss01	Behavior	11,368
ABCD Youth Youth Risk Behavior Survey Exercise Physical Activity (YRB)	abcd_yrb01	Activity	11,876
ABCD dMRI DTI Destrieux Parcellations Part 1	abcd_ddtidp101	Evaluated Data	11,811
ABCD dMRI DTI Destrieux Parcellations Part 2	abcd_ddtidp201	Evaluated Data	11,811
ABCD dMRI DTI Full Destrieux Parcellation Part 1	abcd_ddtifp101	Evaluated Data	11,811
ABCD dMRI DTI Full Destrieux Parcellation Part 2	abcd_ddtifp201	Evaluated Data	11,811
ABCD dMRI DTI Full Part 1	abcd_dmdtifp101	Evaluated Data	11,811
ABCD dMRI DTI Full Part 2	abcd_dmdtifp201	Evaluated Data	11,811

Title	Short Name	Categories	Shared Subjects
ABCD dMRI DTI Part 1	abcd_dti_p101	Evaluated Data	11,811
ABCD dMRI DTI Part 2	abcd_dti_p201	Evaluated Data	11,811
ABCD dMRI RSI Part 5	abcd_drsip501	Evaluated Data	11,811
ABCD dMRI RSI Part 6	abcd_drsip601	Evaluated Data	11,811
ABCD rsfMRI Destrieux	abcd_mrirsfd01	Evaluated Data	11,614
ABCD rsfMRI Gordon Network Correlations	abcd_betnet02	Evaluated Data	11,614
ABCD rsfMRI Network to Subcortical ROI Correlations	mrirscor02	Evaluated Data	11,614
ABCD rsfMRI Temporal Variance	abcd_mrirstv02	Evaluated Data	11,614
ABCD sMRI Destrieux Parcellation Part 3	abcd_mrisdp30201	Evaluated Data	11,811
ABCD sMRI Part 3	abcd_smrip30201	Evaluated Data	11,811
ABCD sMRI T2w Post Processing QC	abcd_t2wqc01	Evaluated Data	11,488
Automated Post-Processing QC Metrics	abcd_auto_postqc01	Evaluated Data	11,811
Genomics Sample	genomics_sample03	Omics	59,961
MRI QC Raw Part 1	mriqcrp10301	Evaluated Data	11,841
MRI QC Raw Part 2	mriqcrp20301	Evaluated Data	11,841
MRI QC Raw Part 3	mriqcrp302	Evaluated Data	11,841
Manual fMRI Post-Processing QC	abcd_fmriqc01	Evaluated Data	11,738
Mobile Data	aurora01	Activity	7,032
Parent ABCD COVID-19 Questionnaire	pabcdcovid19questionnaire01	Questionnaire	6,010
Parent Prosocial Behavior Survey	psb01	Behavior	11,876
Processed MRI Data	fmriresults01	Evaluated Data	14,671
Residential History Derived Scores	abcd_rhds01	Socioeconomic	11,876
Social Development Child Alabama Parenting Questionnaire	abcd_socdev_child_alabam01	Parenting	2,300
Social Development Child Difficulties in Emotion Regulation	abcd_socdev_child_emr01	Emotions	2,300

Title	Short Name	Categories	Shared Subjects
Social Development Child Feedback	abcd_socdev_child_fb01	Questionnaire	2,300
Social Development Child Firearms	abcd_socdev_child_fa01	Questionnaire	2,300
Social Development Child Peer Behavior	abcd_socdev_child_pb01	Relationships	2,300
Social Development Child Personality Disposition	abcd_socdev_child_pdis01	Personality	2,300
Social Development Child Reported Delinquency	abcd_socdev_child_rde01	Questionnaire	2,300
Social Development Child Victimization	abcd_socdev_child_vic01	Trauma	2,300
Social Development Contact Track	abcd_socdev_ctr01	Questionnaire	11,876
Social Development Parent Alabama Parenting Questionnaire	abcd_socdev_p_alabama01	Parenting	2,300
Social Development Parent Difficulties in Emotion Regulation	abcd_socdev_p_emr01	Emotions	2,300
Social Development Parent Feedback	abcd_socdev_p_fb01	Questionnaire	2,300
Social Development Parent Firearms	abcd_socdev_p_fa01	Questionnaire	2,300
Social Development Parent Neighborhood	abcd_socdev_p_nbh01	Social Adjustment	2,300
Social Development Parent Personality Disposition	abcd_socdev_p_pdis01	Personality	2,300
Social Development Parent Reported Delinquency	abcd_socdev_p_rde01	Questionnaire	2,300
Social Development Parent Victimization	abcd_socdev_p_vic01	Trauma	2,300
Social Development Visit Type	abcd_socdev_vt01	Questionnaire	2,300
Sum Scores Mental Health Parent	abcd_mhp02	Summary	11,876
Sum Scores Mental Health Youth	abcd_mhy02	Summary	11,876
UPPS-P for Children Short Form (ABCD-version)	abcd_upps01	Personality	11,876
Youth ABCD COVID-19 Questionnaire	yabcdcovid19questionnaire01	Questionnaire	6,010
Youth Prosocial Behavior Survey	abcd_psb01	Behavior	11,876

Metadata from each individual table is available from https://nda.nih.gov/data_dictionary.html?submission=ALL&source=ABCD%2BRelease%2B4.0 .