Predicting Adverse Drug Events during Methylphenidate Treatment in ADHD : International Collaborative Network Study

Version: 1.0

Date: June 21, 2022

Created by:

Dong Yun Lee, MD, Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea

Chungsoo Kim, PharmD, Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, South Korea

Yunmi Shin, MD, Department of Psychiatry, Ajou University School of Medicine, Suwon, South Korea

Rae Woong Park, MD, PhD, Department of Biomedical Informatics, Ajou University Graduate School of Medicine, Suwon, South Korea

Contact person: Dong Yun Lee – dongyun90@ajou.ac.kr

Principal investigator:

Rae Woong Park, MD, PhD, Professor

Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea

veritas@ajou.ac.kr

Financial supports: This work is supported by the Health Insurance Review & Assessment Service (HIRA). The views expressed are those of the author(s) and not necessarily those of the HIRA. This

research is also funded by the Bio Industrial Strategic Technology Development Program (20003883, 20005021) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea), a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health &Welfare, Republic of Korea (grant number: HR16C0001).

The authors declare the following disclosures: None.

Contents

1 List of abbreviations

AUROC	Area Under the Receiver Operating Characteristic Curve
AAP	American Academy of Pediatrics
ADHD	Attention-Deficit/Hyperactivity Disorder
CDM	Common Data Model
MPH	Methylphenidate
O	Outcome cohort
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcome Partnership
SSRI	Selective Serotonin Reuptake Inhibitor
T	Target cohort
TAR	Time at risk

2 Amendments and Updates

0.1	06 May 2022	C Kim	Initial draft
0.2	20 May 2022	D.Y Lee, C Kim	Finalize draft
1.0	21 June 2022	D.Y Lee, C Kim	Added estimation analysis

3 Executive Summary

The primary objective of this study is to develop and validate patient-level prediction models for patients with Attention-deficit/hyperactivity disorder (ADHD) who were first prescribed methylphenidate (MPH). Thirteen different outcomes will be predicted, including 1) psychosis, 2) mania, 3) tic disorder, 4) sleep disorder, 5) substance abuse disorder, 6) movement disorder, 7) drug induced parkinsonism, 8) tremor, 9) cardiovascular events, 10) hypertension, 11) arrhythmia, 12) traumatic injury, 13) ADHD hospitalization. The time-at-risk period will be defined as from the cohort start date + 7 days to the last date of continuous MPH exposure or the cohort start date + 365 days (maximum). These thirteen prediction models will be developed using three different algorithms: Lasso Logistic Regression, Random Forest, and Extreme Gradient boosting. The secondary objective of this study is to assess clinical outcomes in patients stratified with prediction results.

4 Rationale and Background

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common neurobehavioral disorders[1]. The most used drug for the treatment of ADHD is psychostimulant, which includes MPH,

dextroamphetamine, and lisdexamfetamine, for about 90% of the total anti-ADHD prescription[2, 3]. Although MPH effectively ameliorate the symptoms of ADHD and has the best safety/coverage ratio than other anti-ADHD drugs, adverse events including neuropsychiatric and other medical problems have been reported[4, 5]. Especially, dopaminergic excess or interacting with the dopamine system from MPH treatment can trigger psychotic and tic symptoms.[6, 7] Moreover, the use of MPH has caused concern for increased seizures.[8] These side effects can also affect the medication non-adherence which impact on treatment efficacy.[9] Therefore, early detection and intervention in adverse events associated with MPH is crucial for effective treatments. However, even if associations between MPH and adverse events have been reported, studies predicting individual's probability for adverse events of MPH are still limited.

We aim to develop and validate statistical models for predicting comprehensive adverse events of MPH in patients with ADHD. We will also differentiate models by patient demographics (age groups and sex) to compare the model performance and predictors. We will also assess clinical outcomes in patients stratified with prediction results.

5 Study Objectives

5.1 Objectives

The overall goal of this study is to develop and validate predictive models for various adverse events in MPH to inform the triage and early management of patients with attention deficit/hyperactivity disorder. Also, another goal is assessing clinical outcomes in patients stratified with prediction results.

1) To predict the risk of adverse events due to MPH (psychosis, mania, tic disorder, sleep disorder, substance use disorder, movement disorder, drug induced parkinsonism, tremor, cardiovascular events, hypertension, arrhythmia, traumatic injury, and ADHD hospitalization; all outcomes limited only to a new-onset case) amongst patients with ADHD after prescribing MPH for the first time.

2) To predict the risk of adverse events due to MPH (psychosis, mania, tic disorder, sleep disorder, substance use disorder, movement disorder, drug induced parkinsonism, tremor, cardiovascular events, hypertension, arrhythmia, traumatic injury, and ADHD hospitalization; all outcomes limited only to a new-onset case) amongst subgroup patients (male, female, child & adolescent, and adult groups) with ADHD after prescribing MPH for the first time.

3) To assess the 365-day risk of clinical outcomes (psychiatric hospitalization and suicide) amongst patients stratified with prediction results (psychosis, mania, tic disorder, sleep disorder, substance use disorder, movement disorder, drug induced parkinsonism, tremor, cardiovascular events, hypertension, arrhythmia, traumatic injury, and ADHD hospitalization) with ADHD after prescribing MPH for the first time.

5.1.1 Prediction models

Target Cohorts	Outcome Cohorts
- New MPH users with diagnosis of ADHD AND no exposures to other ADHD medication before the index date (index date: the first prescription date of MPH)	- Psychotic disorders - Mania - Tic disorder - Sleep disorder - Substance use disorder - Movement disorder - Drug induced parkinsonism - Tremor - Cardiovascular events - Hypertension - Arrhythmia - Traumatic injury - Hospitalization with ADHD
- New MPH users in the male population with diagnosis of ADHD AND no exposure to other ADHD medication before the index date (index date: the first prescription date of MPH)	
- New MPH users in the female population with diagnosis of ADHD AND no exposure to other ADHD medication before the index date (index date: the first prescription date of MPH)	
- New MPH users in the adolescent (≤18) population with diagnosis of ADHD AND no exposure to other ADHD medication before the index date (index date: the first prescription date of MPH)	
- New MPH users in the adult (>18) population with a diagnosis of ADHD AND no exposure to other ADHD medication before the index date (index date: the first prescription date of MPH)	

5.1.2 Analysis of clinical outcomes

Study Cohorts	Outcome Cohorts
- New MPH users with diagnosis of ADHD AND no exposures to other ADHD medication before the index date (index date: the first prescription date of MPH) **Target cohort**: patients predicted to have the outcome **Comparator cohort**: patients predicted not to have the outcome	- Psychiatric hospitalization - Suicide
- New MPH users in the male population with diagnosis of ADHD AND no exposure to other ADHD medication before the index date (index date: the first prescription date of MPH) **Target cohort**: patients predicted to have the outcome **Comparator cohort**: patients predicted not to have the outcome	
- New MPH users in the female population with diagnosis of ADHD AND no exposure to other ADHD medication before the index date (index date: the first prescription date of MPH) **Target cohort**: patients predicted to have the outcome **Comparator cohort**: patients predicted not to have the outcome	

- New MPH users in the adolescent (≤18) population with diagnosis of ADHD AND no exposure to other ADHD medication before the index date (index date: the first prescription date of MPH) **Target cohort**: patients predicted to have the outcome **Comparator cohort**: patients predicted not to have the outcome	
- New MPH users in the adult (>18) population with a diagnosis of ADHD AND no exposure to other ADHD medication before the index date (index date: the first prescription date of MPH) **Target cohort**: patients predicted to have the outcome **Comparator cohort**: patients predicted not to have the outcome	

6 Research methods

6.1 Study Design

6.1.1 Overview

This study will be a retrospective, observational, patient-level prediction design. By 'retrospective' we mean the study will use data already collected at the start of the study. By 'observational' we mean no intervention will take place in the course of this study. By 'patient-level' we mean a modelling process wherein an outcome is predicted within a time at risk relative to the target cohort start and/or end date. Prediction is performed using a set of covariates derived using data prior to the start of the target cohort.

Figure 1 illustrates the prediction problem we will address. Among a population at risk, we aim to predict which patients at a defined moment in time (t = 0) will experience some outcome during a time-at-risk (TAR). Prediction is done using only information about the patients in an observation window prior to that moment in time.

We follow the PROGRESS best practice recommendations for model development and the TRIPOD guidance for transparent reporting of the model results.[10, 11]

After developing prediction models, survival analysis will be performed to assess clinical outcomes of patients who have the prediction outcome, as determined by the prediction model.

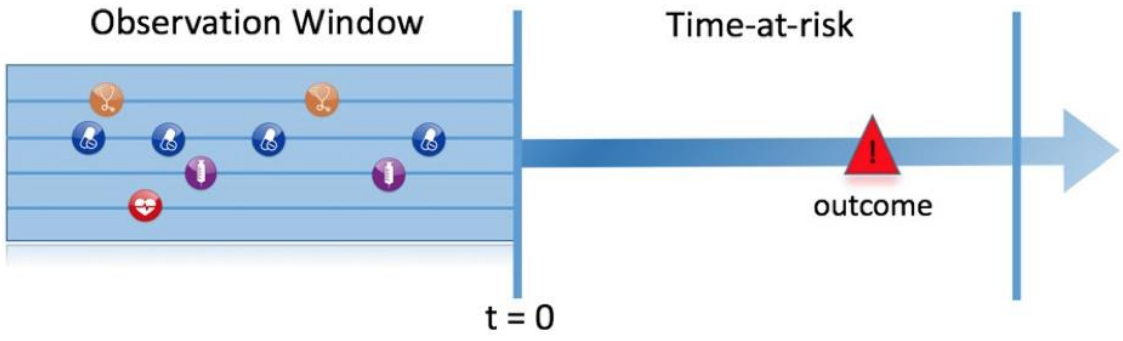

Figure 1: The prediction problem

6.2 Data Source(s)

This study will be conducted using a distributed data network; therefore, data sources may change depending on participating data partners.

Source Full Name	Short Name	Country Code	Data Provenance	Patient Count	History	Patient Type	Data collection
Health Insurance Review and Assessment service database – attention-deficit/hyperactivity disorder subset	HIRA-ADHD	KR	Claims	0.33M	2016.1 – 2021.3	Nationwide health insurance	Anonymized personal identifier, demographics, diagnoses, information on medical procedures and products

6.3 Study population

6.3.1 Target Cohort(s) [T]

The target cohort is the new MPH user group. All subjects in the database will be included who meet the following criteria described below.

Target Cohort (s)	Description
New MPH users	- First MPH prescription in patient's history (index date) - At least 365 days of continuous observation time prior to the index date - ADHD diagnosis for the first time in the patient's history on or before the index date - No other ADHD drugs such as atomoxetine, clonidine, and bupropion before the index date

6.3.2 Outcome Cohort(s) [O]

The outcome cohorts are 13 adverse events of MPH which were already known through the previous research. The description of each outcome is presented in the table below.

Outcome Cohort (s)	Description
Psychotic disorder	Diagnosis of psychotic disorder for the first time
Mania	Diagnosis of mania for the first time
Tic disorder	Diagnosis of tic disorder for the first time
Sleep disorder	Diagnosis of sleep disorder for the first time
Substance use disorder	Diagnosis of substance use disorder for the first time
Movement disorder	Diagnosis of movement disorder for the first time
Drug induced parkinsonism	Diagnosis of drug-induced parkinsonism for the first time
Tremor	Diagnosis of tremor for the first time
Cardiovascular events for the first time	Diagnosis of cardiovascular events including hypertension, arrhythmia, myo-cardiac infarction, cardiomyopathy, or cardiac arrest for the first time
Hypertension	Diagnosis of hypertension for the first time
Arrhythmia	Diagnosis of arrhythmia for the first time
Traumatic injury	Diagnosis of traumatic injury for the first time
Hospitalization with ADHD	Hospitalization with ADHD for the first time

Full descriptions:

The JSON files describing for all the outcome cohorts are available at:

- https://github.com/ABMI/MPH_Safety_Prediction/tree/main/inst/cohort

In order to convert these to a human readable form, import the JSON into a new cohort definition in any instance of ATLAS and reload.

6.3.3 Study population for additional analyses

Since ADHD is with differences according to patient demographics, subgroup analyses will be performed on male, female, child & adolescent, and adult patients. Cohort criteria are the same except additional criteria including sex and age limited to male, female, under 18, or over 18.

A sensitivity analysis using a subset of the target cohort will be conducted.

Target Cohorts	Description
New MPH users in the male	- First MPH prescription in patient's history (index date) - Male - At least 365 days of continuous observation time prior to the index date - ADHD diagnosis for the first time in the patient's history on or before the index date - No other ADHD drugs such as atomoxetine, clonidine, and bupropion before the index date

New MPH users in the female	- First MPH prescription in patient's history (index date) - Female - At least 365 days of continuous observation time prior to the index date - ADHD diagnosis for the first time in the patient's history on or before the index date - No other ADHD drugs such as atomoxetine, clonidine, and bupropion before the index date
New MPH users in the adolescent	- First MPH prescription in patient's history (index date) - Age at index date under 18 (≤ 18 years old) - At least 365 days of continuous observation time prior to the index date - ADHD diagnosis for the first time in the patient's history on or before the index date - No other ADHD drugs such as atomoxetine, clonidine, and bupropion before the index date
New MPH users in the adult	- First MPH prescription in patient's history (index date) - Age at index date over 18 (> 18 years old) - At least 365 days of continuous observation time prior to the index date - ADHD diagnosis for the first time in the patient's history on or before the index date - No other ADHD drugs such as atomoxetine, clonidine, and bupropion before the index date

After the prediction, the target cohort for each predicted outcome will be divided into patients who are predicted to have outcome and patients who are predicted to have none. Survival analysis will be performed using two patient groups.

6.3.4 Time at Risk

The table below describes the Time at Risk (TAR) window start and end for each of the analyses that are executed. The definition of cohort end date will be earlier date among 1) end date of continuous MPH exposure 2) cohort start date + 365 days 3) censoring date.

Time At Risk window	
Start Date	**End Date**
The TAR window starts at +7 days from the index date (first MPH prescription)	The TAR window ends at the cohort end date. The cohort end date is defined as the earliest date among dates below: 1) the last date of continuous MPH exposure 2) + 365 days from the index date 3) the date of censoring (other ADHD medication exposure) 4) the last date of continuous observation

6.4 Statistical Analysis Method(s)

6.4.1 Algorithms

In this study we will apply the Lasso Logistic Regression, Random Forest, and Extreme Gradient Boosting.

Lasso logistic regression belongs to the family of generalized linear models, where a linear combination of the variables is learned and finally a logistic function maps the linear combination to a value between 0 and 1. The lasso regularization adds a cost based on model complexity to the objective function when training the model. This cost is the sum of the absolute values of the linear combination of the coefficients. The model automatically performs feature selection by minimizing this cost. We use the Cyclic coordinate descent for logistic, Poisson and survival analysis (Cyclops) package to perform large-scale regularized logistic regression: https://github.com/OHDSI/Cyclops.

The random forest model uses classification tress as building blocks to construct prediction models. A random forest model is developed by only considering a small subset of the predictors each time it splits. This process results in a reduction of the correlation among the trees, thus making the average of the resulting tress less variable and more reliable

The eXtreme Gradient Boosting (XGBoost) algorithm is a decision tree-based model on the training dataset. XGBoost starts with a simple initial model and its residuals/misclassifications are iteratively improved in subsequent models searching from among all available predictors to try to minimize misclassification. XGBoost was commonly chosen for its interpretability of results and robustness to overfitting

6.4.2 Model Evaluation

The following evaluations will be performed on the model

Evaluation	Description
Box Plots	The prediction distribution boxplots are box plots for the predicted risks of the people in the test set with the outcome (class 1: blue) and without the outcome (class 0: red).
Calibration Plot	The calibration plot shows how close the predicted risk is to the observed risk. The diagonal dashed line thus indicates a perfectly calibrated model. The ten (or fewer) dots represent the mean predicted values for each quantile plotted against the observed fraction of people in that quantile who had the outcome (observed fraction). The straight black line is the linear regression using these 10 plotted quantiles mean predicted vs observed fraction points. The two blue straight lines represented the 95% lower and upper confidence intervals of the slope of the fitted line.

Demographic Summary Plot	This plot shows for females and males the expected and observed risk in different age groups together with a confidence area.
Precision Recall Plot	The precision-recall curve is valuable for dataset with a high imbalance between the size of the positive and negative class. It shows the trade-off between precision and recall for different threshold. High precision relates to a low false positive rate, and high recall relates to a low false negative rate. High scores for both show that the classifier is returning accurate results (high precision), as well as returning a majority of all positive results (high recall). A high area under the curve represents both high recall and high precision.
Prediction Distribution Plots	The preference distribution plots are the preference score distributions corresponding to i) people in the test set with the outcome (red) and ii) people in the test set without the outcome (blue).
ROC Plot	The ROC plot plots the sensitivity against 1-specificity on the test set. The plot shows how well the model is able to discriminate between the people with the outcome and those without. The dashed diagonal line is the performance of a model that randomly assigns predictions. The higher the area under the ROC plot the better the discrimination of the model.
Smooth Calibration Plot	Similar to the traditional calibration shown above the Smooth Calibration plot shows the relationship between predicted and observed risk. the major difference is that the smooth fit allows for a more fine-grained examination of this. Whereas the traditional plot will be heavily influenced by the areas with the highest density of data the smooth plot will provide the same information for this region as well as a more accurate interpretation of areas with lower density. the plot also contains information on the distribution of the outcomes relative to predicted risk. However, the increased information game comes at a computational cost. It is recommended to use the traditional plot for examination and then to produce the smooth plot for final versions.
Test-Train Similarity Plot	The test-train similarity is presented by plotting the mean covariate values in the train set against those in the test set for people with and without the outcome.
Variable Scatter Plot	The variable scatter plot shows the mean covariate value for the people with the outcome against the mean covariate value for the people without the outcome. The size and colour of the dots correspond to the importance of the covariates in the trained model (size of beta) and its direction (sign of beta with green meaning positive and red meaning negative), respectively.

6.4.3 Clinical outcome assessment

Kaplan–Meier and Cox survival analyses of long-term outcomes will be performed. The Kaplan-Meier method and log-rank test will be used to calculate and compare the survival curves stratified by prediction model. Univariate Cox regression models will be used to obtain hazard ratios (HRs) between groups.

6.5 Quality Control

The PatientLevelPrediction package itself, as well as other OHDSI packages on which PatientLevelPrediction depends, use unit tests for validation. More information can be found in the Book of OHDSI at: https://ohdsi.github.io/TheBookOfOhdsi/SoftwareValidity.html

6.6 Tools

To create the study package, ATLAS will be used to specify the cohorts, time-at-risk, covariate and population settings as well as which models will be analysed. Information on this is available in the Book of OHDSI at: https://ohdsi.github.io/TheBookOfOhdsi/OhdsiAnalyticsTools.html#atlas

The package developed in ATLAS will utilise the Patient-Level Prediction R package to run the analysis. More information on this is available at: https://ohdsi.github.io/TheBookOfOhdsi/PatientLevelPrediction.html

This study will be designed using OHDSI tools and run with R. More information about the tools can be found in the Appendix 'Study Generation Version Information'.

7 Data Analysis Plan

7.1 Algorithm Settings

- Model settings #1 LassoLogisticRegressionSettings

Covariates	Settings
seed	
variance	0.01

- Model settings #2 RandomForestSettings

Covariates	Settings
seed	
Max depth	4,10,17
Number of tree features	-1

Number of trees to build	500

- Model settings #3 ExtremeGradientBoostingSettings

Covariates	Settings
seed	
Boosting learning rate	0.01, 0.1
Maximum number of interactions	4,6,17
Maximum number of rows	20
Number of trees to build	10,100

7.2 Covariate Settings

The covariates (constructed using records on or prior to the target cohort start date) are used within this prediction mode include the following. Each covariate needs to contain at least 0.001 subjects to be considered for the model.

Covariates	Settings
VisitCountMediumTerm	FALSE
ObservationShortTerm	TRUE
shortTermStartDays	-30
MeasurementRangeGroupShortTerm	FALSE
ConditionOccurrenceLongTerm	TRUE
DrugEraStartLongTerm	FALSE
VisitCountShortTerm	FALSE
Chads2Vasc	FALSE
ConditionGroupEraStartLongTerm	FALSE
ConditionEraShortTerm	FALSE
Dcsi	FALSE
DrugGroupEraLongTerm	TRUE
DrugGroupEraShortTerm	TRUE
ConditionEraStartLongTerm	FALSE
temporal	FALSE
DemographicsIndexMonth	FALSE
ConditionOccurrencePrimaryInpatientLongTerm	FALSE
ConditionEraAnyTimePrior	FALSE
addDescendantsToInclude	FALSE
ConditionGroupEraStartMediumTerm	FALSE
ProcedureOccurrenceLongTerm	TRUE
DrugExposureLongTerm	TRUE
DrugEraStartShortTerm	FALSE
DistinctIngredientCountMediumTerm	FALSE

DistinctMeasurementCountShortTerm	FALSE
MeasurementRangeGroupLongTerm	FALSE
ConditionGroupEraOverlapping	FALSE
MeasurementRangeGroupMediumTerm	FALSE
DrugGroupEraStartMediumTerm	FALSE
MeasurementAnyTimePrior	FALSE
MeasurementMediumTerm	FALSE
includedCovariateIds	
ConditionOccurrenceAnyTimePrior	FALSE
DistinctConditionCountLongTerm	FALSE
MeasurementValueLongTerm	FALSE
DrugEraShortTerm	FALSE
DrugGroupEraAnyTimePrior	FALSE
DrugEraOverlapping	TRUE
ConditionOccurrencePrimaryInpatientAnyTimePrior	FALSE
ConditionEraMediumTerm	FALSE
ConditionEraOverlapping	FALSE
ConditionEraStartShortTerm	FALSE
ObservationAnyTimePrior	FALSE
VisitConceptCountShortTerm	FALSE
DemographicsEthnicity	FALSE
DistinctIngredientCountLongTerm	FALSE
ConditionOccurrencePrimaryInpatientShortTerm	FALSE
DemographicsAgeGroup	TRUE
DistinctProcedureCountShortTerm	FALSE
DistinctObservationCountMediumTerm	FALSE
includedCovariateConceptIds	
DrugGroupEraStartShortTerm	FALSE
addDescendantsToExclude	FALSE
DrugEraLongTerm	FALSE
DistinctConditionCountShortTerm	FALSE
ConditionGroupEraShortTerm	TRUE
ConditionEraStartMediumTerm	FALSE
VisitCountLongTerm	FALSE
DemographicsRace	FALSE
ProcedureOccurrenceAnyTimePrior	FALSE
DistinctObservationCountLongTerm	FALSE
ProcedureOccurrenceMediumTerm	FALSE
CharlsonIndex	TRUE
DemographicsPriorObservationTime	FALSE
MeasurementShortTerm	FALSE
DistinctProcedureCountMediumTerm	FALSE
ConditionEraLongTerm	FALSE
DrugGroupEraStartLongTerm	FALSE
DemographicsGender	TRUE
DeviceExposureAnyTimePrior	FALSE
ObservationLongTerm	TRUE

DemographicsIndexYearMonth	FALSE
ConditionOccurrenceMediumTerm	FALSE
longTermStartDays	-365
DemographicsAge	FALSE
DrugGroupEraOverlapping	FALSE
DistinctMeasurementCountLongTerm	FALSE
MeasurementRangeGroupAnyTimePrior	FALSE
DistinctConditionCountMediumTerm	FALSE
DrugGroupEraMediumTerm	FALSE
ProcedureOccurrenceShortTerm	TRUE
ObservationMediumTerm	FALSE
ConditionGroupEraAnyTimePrior	FALSE
Chads2	FALSE
DrugExposureAnyTimePrior	FALSE
DeviceExposureLongTerm	FALSE
DemographicsTimeInCohort	FALSE
DistinctMeasurementCountMediumTerm	FALSE
MeasurementValueShortTerm	FALSE
DeviceExposureMediumTerm	FALSE
ConditionGroupEraStartShortTerm	FALSE
ConditionOccurrencePrimaryInpatientMediumTerm	FALSE
MeasurementLongTerm	FALSE
DemographicsIndexYear	FALSE
MeasurementValueMediumTerm	FALSE
DrugEraStartMediumTerm	FALSE
MeasurementValueAnyTimePrior	FALSE
DistinctObservationCountShortTerm	FALSE
DrugEraMediumTerm	FALSE
ConditionGroupEraLongTerm	TRUE
DrugExposureShortTerm	TRUE
DistinctIngredientCountShortTerm	FALSE
DeviceExposureShortTerm	FALSE
mediumTermStartDays	-180
DemographicsPostObservationTime	FALSE
VisitConceptCountLongTerm	FALSE
VisitConceptCountMediumTerm	FALSE
excludedCovariateConceptIds	
ConditionGroupEraMediumTerm	FALSE
DrugExposureMediumTerm	FALSE
DistinctProcedureCountLongTerm	FALSE
DrugEraAnyTimePrior	FALSE
endDays	-1
ConditionOccurrenceShortTerm	TRUE

7.3 Model Development & Evaluation

To build and internally validate the models, we will partition the labelled data into a train set (75%) and a test set (25%).

The hyper-parameters for the models will be assessed using 3-fold cross validation on the train set and a final model will be trained using the full train set and optimal hyper-parameters.

The internal validity of the models will be assessed on the test set. The external validity of the models will be assessed on other databases. We will use the area under the receiver operating characteristic curve (AUROC) to evaluate the discriminative performance of the models and plot the predicted risk against the observed fraction to visualize the calibration. See 'Model Evaluation' section for more detailed information about additional model evaluation metrics

7.4 Analysis Execution Settings

Covariate balance will be summarized in tabular form by showing the mean value (percentage for categorical) for all baseline covariates in the target and comparator cohort, with the associated standardized mean difference computed for each covariate.

For the prediction model there is 5 target cohort evaluated for 13 outcomes over 3 model over 1 covariate setting and over 1 population setting. In total there are 195 analyses performed.

For clinical outcome assessment, there is 5 target-comparator pairs evaluated for 2 outcomes. In total there are 10 analyses performed.

7.5 Strengths and Limitations

Strength

- The analysis can help gain insight into the clinical usefulness of each developed model by identifying whether it is transportable.

Limitations

- Although the CDM standardizes the vocabularies of the datasets, the concept recording distributions are likely to differ between databases and it is unknown how much this will limit model transportability

8 Protection of Human Subjects

The study is using only de-identified data. Confidentiality of patient records will be maintained at all times. All study reports will contain aggregate data only and will not identify individual patients or physicians.

9 Plans for Disseminating and Communicating Study Results

The study protocol will be submitted for publication to an online repository before initiation of the study. Analytic codes will be posted on the online repository after completion of the study. At

least one paper describing the study and its results will be written and submitted for publication to a peer-reviewed scientific journal.

10 References

1. Sayal K, Prasad V, Daley D, Ford T, Coghill D. ADHD in children and young people: prevalence, care pathways, and service provision. *The Lancet Psychiatry*. 2018/02/01/ 2018;5(2):175-186. doi:https://doi.org/10.1016/S2215-0366(17)30167-0
2. Raman SR, Man KKC, Bahmanyar S, et al. Trends in attention-deficit hyperactivity disorder medication use: a retrospective observational study using population-based databases. *The Lancet Psychiatry*. 2018/10/01/ 2018;5(10):824-835. doi:https://doi.org/10.1016/S2215-0366(18)30293-1
3. Bachmann CJ, Wijlaars LP, Kalverdijk LJ, et al. Trends in ADHD medication use in children and adolescents in five western countries, 2005–2012. *European Neuropsychopharmacology*. 2017/05/01/ 2017;27(5):484-493. doi:https://doi.org/10.1016/j.euroneuro.2017.03.002
4. Solmi M, Fornaro M, Ostinelli EG, et al. Safety of 80 antidepressants, antipsychotics, anti-attention-deficit/hyperactivity medications and mood stabilizers in children and adolescents with psychiatric disorders: a large scale systematic meta-review of 78 adverse effects. *World Psychiatry*. 2020;19(2):214-232. doi:https://doi.org/10.1002/wps.20765
5. Schachar RJ, Tannock R, Cunningham C, Corkum PV. Behavioral, Situational, and Temporal Effects of Treatment of ADHD With Methylphenidate. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1997/06/01/ 1997;36(6):754-763. doi:https://doi.org/10.1097/00004583-199706000-00011
6. Kraemer M, Uekermann J, Wiltfang J, Kis B. Methylphenidate-induced psychosis in adult attention-deficit/hyperactivity disorder: report of 3 new cases and review of the literature. *Clinical neuropharmacology*. 2010;33(4):204-206.
7. Park SY, Kim EJ, Cheon K-A. Association between 5-HTTLPR polymorphism and tics after treatment with methylphenidate in Korean children with attention-deficit/hyperactivity disorder. *Journal of child and adolescent psychopharmacology*. 2015;25(8):633-640.
8. Verrotti A, Moavero R, Panzarino G, Di Paolantonio C, Rizzo R, Curatolo P. The challenge of pharmacotherapy in children and adolescents with epilepsy-ADHD comorbidity. *Clinical Drug Investigation*. 2018;38(1):1-8.
9. Kooij J, Rösler M, Philipsen A, et al. Predictors and impact of non-adherence in adults with attention-deficit/hyperactivity disorder receiving OROS methylphenidate: results from a randomized, placebo-controlled trial. *BMC psychiatry*. 2013;13(1):1-10.
10. Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS medicine*. 2013;10(2):e1001381.
11. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine*. 2015;162(1):W1-W73.

11 Appendix: Code Set for Definitions

All codes are available in ATHENA (athena.ohdsi.org)

1. Attention-Deficit/Hyperactivity Disorder

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
438409	Attention deficit hyperactivity disorder	Condition	SNOMED	NO	YES	NO
4047120	Disorders of attention and motor control	Condition	SNOMED	NO	YES	NO

2. Methylphenidate

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
705944	methylphenidate	Drug	RxNorm	NO	YES	NO
21604757	methylphenidate; oral	Drug	ATC	NO	YES	NO

3. Other Anti-ADHD drugs for ADHD

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
742185	Atomoxetine	Drug	RxNorm	NO	YES	NO
221604762	Atomoxetine; oral	Drug	ATC	NO	YES	NO
750982	Bupropion	Drug	RxNorm	NO	YES	NO
21604741	Bupropion; oral	Drug	ATC	NO	YES	NO
21600398	Clonidine; systemic	Drug	ATC	NO	YES	NO

4. Mania

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
4333677	Mania	Condition	SNOMED	NO	YES	NO

5. Psychosis

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
436073	Psychotic disorder	Condition	SNOMED	NO	YES	NO

6. Sleep disorder

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
435524	Sleep disorder	Condition	SNOMED	NO	YES	NO

7. Tic disorder

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
381839	Tic disorder	Condition	SNOMED	NO	YES	NO

8. Substance abuse disorder

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
4279309	Substance abuse	Condition	SNOMED	NO	YES	NO

9. Movement disorder

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
443782	Tremor	Condition	SNOMED	NO	YES	NO
374013	Secondary parkinsonism	Condition	SNOMED	NO	YES	NO
4171569	Parkinsonism due to drug	Condition	SNOMED	NO	YES	NO
375800	Dystonia	Condition	SNOMED	NO	YES	NO

10. Drug-induced parkinsonism

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
374013	Secondary parkinsonism	Condition	SNOMED	NO	YES	NO
4171569	Parkinsonism due to drug	Condition	SNOMED	NO	YES	NO

11. Tremor

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
443782	Tremor	Condition	SNOMED	NO	YES	NO

12. Cardiovascular events

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
316866	Hypertensive disorder	Condition	SNOMED	NO	YES	NO
4185572	Ventricular arrhythmia	Condition	SNOMED	NO	YES	NO
444070	Tachycardia	Condition	SNOMED	NO	YES	NO
315643	Tacharrhythmia	Condition	SNOMED	NO	YES	NO
4248028	Rupraventricular arrhythmia	Condition	SNOMED	NO	YES	NO
4111552	Re-entry ventricular arrhythmia	Condition	SNOMED	NO	YES	NO
44784217	Cardiac arrhythmia	Condition	SNOMED	NO	YES	NO
4068155	Atrial arrhythmia	Condition	SNOMED	NO	YES	NO
4329847	Myocardial infarction	Condition	SNOMED	NO	YES	NO
321319	Cardiomyopathy	Condition	SNOMED	NO	YES	NO
4317150	Sudden cardiac death	Condition	SNOMED	NO	YES	NO
321042	Cardiac arrest	Condition	SNOMED	NO	YES	NO

13. Arrhythmia

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
4185572	Ventricular arrhythmia	Condition	SNOMED	NO	YES	NO
444070	Tachycardia	Condition	SNOMED	NO	YES	NO
315643	Tacharrhythmia	Condition	SNOMED	NO	YES	NO
4248028	Rupraventricular arrhythmia	Condition	SNOMED	NO	YES	NO
4111552	Re-entry ventricular arrhythmia	Condition	SNOMED	NO	YES	NO
44784217	Cardiac arrhythmia	Condition	SNOMED	NO	YES	NO
4068155	Atrial arrhythmia	Condition	SNOMED	NO	YES	NO

14. Hypertension

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
316866	Hypertensive disorder	Condition	SNOMED	NO	YES	NO

15. Traumatic injury

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
440921	Traumatic injury	Condition	SNOMED	NO	YES	NO

16. Hospitalization

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
9201	Inpatient Visit	Visit	Visit	NO	NO	NO
262	Emergency Room and Inpatient Visit	Visit	Visit	NO	NO	NO

17. Suicide

Concept Id	Concept Name	Domain	Vocabulary	Excluded	Descendants	Mapped
4219484	Suicide attempt	Observation	SNOMED	NO	YES	NO
4303690	Intentionally harming self	Observation	SNOMED	NO	YES	NO
4152408	Deliberate self harm	Observation	SNOMED	NO	YES	NO
439235	Self inflicted injury	Condition	SNOMED	NO	YES	NO
435446	Late effect of self inflicted injury	Condition	SNOMED	NO	YES	NO
4152376	Intentional self poisoning	Condition	SNOMED	NO	YES	NO
4075235	Drowning self	Condition	SNOMED	NO	YES	NO