**OHDSI: Prediction of incident major depressive disorder in cardiovascular patients using beta blockers**

**Table of Contents**

[1. List of Abbreviations 3](#_Toc29907679)

[2. Responsible Parties 3](#_Toc29907680)

[3. Executive Summary 3](#_Toc29907681)

[4. Rational & Background 3](#_Toc29907682)

[5. Objective 3](#_Toc29907683)

[6. Methods 4](#_Toc29907684)

[6.1. Study Design 4](#_Toc29907685)

[6.2. Data Source(s) 5](#_Toc29907686)

[6.3. Study Populations 5](#_Toc29907687)

[6.4. Statistical Analysis Method(s) 6](#_Toc29907688)

[6.5. Quality Control 7](#_Toc29907689)

[6.6. Tools 7](#_Toc29907690)

[7. Diagnostics 8](#_Toc29907691)

[8. Data Analysis Plan 8](#_Toc29907692)

[8.1. Algorithm Settings 8](#_Toc29907693)

[8.2. Covariate Settings 8](#_Toc29907694)

[8.3. Model Development & Evaluation 11](#_Toc29907695)

[8.4. Analysis Execution Settings 11](#_Toc29907696)

[9. Strengths & Limitations 11](#_Toc29907697)

[10. Protection of Human Subjects 11](#_Toc29907698)

[11. Plans for Disseminating & Communicating Study Results 12](#_Toc29907699)

[12. Tables & Figures 13](#_Toc29907700)

[12.1. Incidence Rate of Target & Outcome 13](#_Toc29907701)

[12.2. Characterization 13](#_Toc29907702)

[13. Appendices 14](#_Toc29907703)

[13.1. Study Generation Version Information 14](#_Toc29907704)

[13.2. Code List 15](#_Toc29907705)

[13.3. Complete Analysis List 33](#_Toc29907706)

[14. References 34](#_Toc29907707)

# List of Abbreviations

| AUC  BB | Area Under the Receiver Operating Characteristic Curve  Beta Blockers |
| --- | --- |
| CDM  MDD  NHIS-NSC | Common Data Model  Major Depressive Disorder  The National Health Insurance Service-National Sample Cohort |
| O | Outcome Cohort |
| OHDSI | Observational Health Data Sciences & Informatics |
| OMOP | Observational Medical Outcomes Partnership |
| T | Target Cohort |
| TAR | Time at Risk |
|  |  |

# Responsible Parties

Author:

Investigators:

Reviewers:

# Executive Summary

Beta blockers (BB) are known for having depression risk. And also BB is widely prescribed for cardiovascular indications. So We developed a prediction model that predicts who develops incident MDD in BB users for cardiovascular indications. NHIS-NSC, which is a systematically drown sample cohort of the Korean nation-wide claim databases was used for traing. External validation of the developed prediction model will be conducted by patient-level prediction validation package. In this study, we are aiming to verify cross-data compatibility of the model through the OHDSI network.

# Rational & Background

Through prospective studies and meta-analyses, depression is well known adverse prognostic factor for cardiovascular patients such as myocardial infarction and acute coronary syndrome. And for those cardiovascular disease including hypertension and heart failure, beta blockers are widely prescribed drug. As one of adverse effects of beta blockers is depression, it will be meaningful to predict who will develop major depressive disorder in beta blockers for cardiovascular indications.

# Objective

The objective is to validate our simple score patient-level prediction model for the following prediction problems:

| Target Cohorts | Outcome Cohorts | Time at Risk |
| --- | --- | --- |
| Cardiovascular Patients using Beta Blockers | Incident Major Depressive Disorder | Risk Window Start: 1, Add Exposure Days to Start: FALSE, Risk Window End: 365, Add Exposure Days to End: FALSE |

# Methods

## Study Design

This study will follow a retrospective, observational, patient-level prediction design. We define 'retrospective' to mean the study will be conducted using data already collected prior to the start of the study. We define 'observational' to mean there is no intervention or treatment assignment imposed by the study. We define 'patient-level prediction' as a modeling 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. Prediction is done using only information about the patients in an observation window prior to that moment in time.

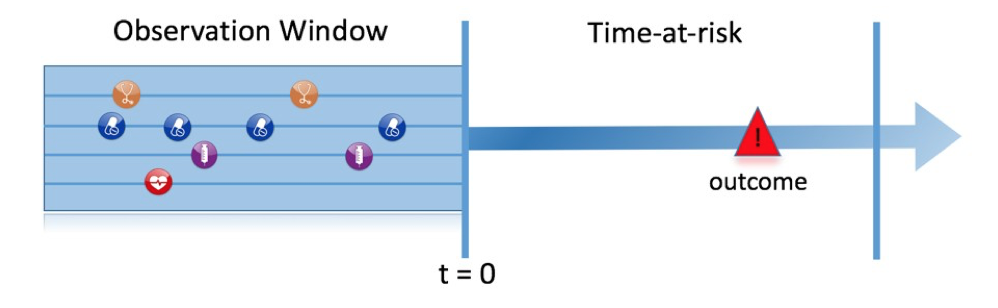


Figure 1: The prediction problem

We follow the PROGRESS best practice recommendations for model development and the TRIPOD guidance for transparent reporting of the model results.

## Data Source(s)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Database Name | Version | Start date | End date | Description |
| Ajou Univ Hosp | V5.3.1 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

## Study Populations

### Target Cohort(s) [T]

Cardiovascular patients using Beta blockers

Initial Event cohort) People having any of the following:

* a drug era of cardiovascular beta-blockers
  + with era length >= 30
  + with age in years at era start >= 20

with continuous observation of at least 365 days prior and 0 days after event index date, and limit initial events to: earliest event per person.

Inclusion Rules) Inclusion Criteria #1: cardiovascular indication

* at least 1 occurrences of a condition occurrence of cardiovascular indication where event starts between 3 days Before and 3 days After index start date

Inclusion Rules) Inclusion Criteria #2: No past history of selected mental disorders

* + exactly 0 occurrences of a condition occurrence of depressive disorder where event starts between all days Before and 0 days After index start date
  + exactly 0 occurrences of a condition occurrence of schizophrenia where event starts between all days Before and 0 days After index start date

Inclusion Rules) Inclusion Criteria #3: No other cancer diagnosis

* exactly 0 occurrences of a drug exposure of ANTIDEPRESSANTS where event starts between all days Before and 0 days After index start date

Limit qualifying cohort to **alle vent per person**

### Outcome Cohorts(s) [O]

Major Depressive Disorder

Initial Event cohort) People having any of the following:

* a condition occurrence of major depressive disorder

for the first time in the person's history

* + - * with continuous observation of at least 0 days prior and 0 days after event index date, and limit initial events to **earliest** **event per person**

.

### Time at Risk

| Time at Risk |
| --- |
| [Time at Risk Settings #1] Risk Window Start: 1, Add Exposure Days to Start: FALSE, Risk Window End: 365, Add Exposure Days to End: FALSE |

### Additional Population Settings

***Population Settings #1***

| Item | Settings |
| --- | --- |
| minTimeAtRisk | 364 |
| requireTimeAtRisk | TRUE |
| addExposureDaysToStart | FALSE |
| riskWindowStart | 1 |
| washoutPeriod | 365 |
| addExposureDaysToEnd | FALSE |
| includeAllOutcomes | TRUE |
| priorOutcomeLookback | 99999 |
| binary | TRUE |
| removeSubjectsWithPriorOutcome | FALSE |
| riskWindowEnd | 365 |
| firstExposureOnly | FALSE |

## Statistical Analysis Method(s)

### Algorithms

In this study we will apply a Lasso Logistic Regression. 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.

### 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 quantile 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 tradeoff 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 color 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. |

## 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

## 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 (17). More information about the tools can be found in the Appendix 'Study Generation Version Information'.

# Diagnostics

Reviewing the incidence rates of the outcomes in the target population prior to performing the analysis will allow us to assess its feasibility. The full table can be found in the 'Table and Figures' section under 'Incidence Rate of Target & Outcome'.

Additionally, reviewing the characteristics of the cohorts provides insight into the cohorts being reviewed. The full table can be found below in the 'Table and Figures' section under 'Characterization'.

# Data Analysis Plan

## Algorithm Settings

LassoLogisticRegressionSettings

|  |  |
| --- | --- |
| Covariates | Settings |
| seed |  |
| variance | 0.01 |

## Model Evaluation

We will use the area under the receiver operating characteristic curve (AUC) 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.

## Analysis Execution Settings

There are 1 target cohorts evaluated for 1 outcomes over 1 models over 1 covariates settings and over 1 population settings. In total there are 1 analysis performed. For a full list refer to appendix 'Complete Analysis List'.

# Strengths & Limitations

* The model is simple to implement
* We are validating the models across a diverse set of patients
* Model validation is possible due to the OHDSI standardizations
* Not all medical events are recorded into the observational datasets and some recordings can be incorrect. This could potentially lead to outcome misclassification.

# Protection of Human Subjects

Confidentiality of patient records will be maintained always. All study reports will contain aggregate data only and will not identify individual patients or physicians. At no time during the study will the sponsor receive patient identifying information except when it is required by regulations in case of reporting adverse events.

# Plans for Disseminating & Communicating Study Results

This work will be submitted to a high impact journal in the field of mental health

# Tables & Figures

## Incidence Rate of Target & Outcome

*<< add incidence here. >>*

## Characterization

*<< add characterization table here. >>*

*<< add results here. >>*

# Appendices

## Study Generation Version Information

Skeleton Version: PatientLevelPredictionStudy - v0.0.1

Identifier / Organization: Janssen Research and Development

## Code List

***Concept Set #1 – Cardiovascular Beta Blockers***

| INVALID\_REASON\_CAPTION | CONCEPT\_ID | CONCEPT\_CODE | CONCEPT\_NAME | DOMAIN\_ID | STANDARD\_CONCEPT\_CAPTION | CONCEPT\_CLASS\_ID | INVALID\_REASON | isExcluded | includeDescendants |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Valid | 1319998 | 149 | Acebutolol | Drug | Standard | V | FALSE | TRUE | FALSE |
| Valid | 1314002 | 1202 | Atenolol | Drug | Standard | V | FALSE | TRUE | FALSE |
| Valid | 40019429 | 371065 | Betaxolol Oral Tablet | Drug | Standard | V | FALSE | TRUE | FALSE |
| Valid | 1338005 | 19484 | Bisoprolol | Drug | Standard | V | FALSE | TRUE | FALSE |
| Valid | 1346823 | 20352 | carvedilol | Drug | Standard | V | FALSE | TRUE | FALSE |
| Valid | 19063575 | 49737 | esmolol | Drug | Standard | V | FALSE | TRUE | FALSE |
| Valid | 1386957 | 6185 | Labetalol | Drug | Standard | V | FALSE | TRUE | FALSE |
| Valid | 1307046 | 6918 | Metoprolol | Drug | Standard | V | FALSE | TRUE | FALSE |
| Valid | 1313200 | 7226 | Nadolol | Drug | Standard | V | FALSE | TRUE | FALSE |
| Valid | 1314577 | 31555 | nebivolol | Drug | Standard | V | FALSE | TRUE | FALSE |
| Valid | 1327978 | 7973 | Penbutolol | Drug | Standard | V | FALSE | TRUE | FALSE |
| Valid | 1345858 | 8332 | Pindolol | Drug | Standard | V | FALSE | TRUE | FALSE |
| Valid | 1353766 | 8787 | Propranolol | Drug | Standard | V | FALSE | TRUE | FALSE |
| Valid | 1370109 | 9947 | Sotalol | Drug | Standard | V | FALSE | TRUE | FALSE |
| Valid | 40087917 | 374135 | Timolol Oral Tablet | Drug | Standard | V | FALSE | TRUE | FALSE |

Cohorts that use this Concept Set:

Cardiovascular Beta Blockers

***Concept Set #2 – Cardiovascular Indications***

| INVALID\_REASON\_CAPTION | CONCEPT\_ID | CONCEPT\_CODE | CONCEPT\_NAME | DOMAIN\_ID | STANDARD\_CONCEPT\_CAPTION | CONCEPT\_CLASS\_ID | INVALID\_REASON | isExcluded | includeDescendants |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Valid | 316866 | 38341003 | Hypertensive disorder | Condition | Standard | V | FALSE | TRUE | FALSE |
| Valid | 321318 | 194828000 | Angina pectoris | Condition | Standard | V | FALSE | TRUE | FALSE |
| Valid | 4329847 | 22298006 | Myocardial infarction | Condition | Standard | V | FALSE | TRUE | FALSE |
| Valid | 316139 | 84114007 | Heart failure | Condition | Standard | V | FALSE | TRUE | FALSE |
| Valid | 317576 | 53741008 | Coronary arteriosclerosis | Condition | Standard | V | FALSE | TRUE | FALSE |

Cohorts that use this Concept Set:

Cardiovascular Beta Blockers

***Concept Set #3 - Depressive disorder***

| INVALID\_REASON\_CAPTION | CONCEPT\_ID | CONCEPT\_CODE | CONCEPT\_NAME | DOMAIN\_ID | STANDARD\_CONCEPT\_CAPTION | CONCEPT\_CLASS\_ID | INVALID\_REASON | isExcluded | includeDescendants |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Valid | 440383 | 35489007 | Depressive disorder | Condition | Standard | V | FALSE | TRUE | FALSE |

Cohorts that use this Concept Set:

Cardiovascular Beta Blockers

***Concept Set #4 - Schizophrenia***

| INVALID\_REASON\_CAPTION | CONCEPT\_ID | CONCEPT\_CODE | CONCEPT\_NAME | DOMAIN\_ID | STANDARD\_CONCEPT\_CAPTION | CONCEPT\_CLASS\_ID | INVALID\_REASON | isExcluded | includeDescendants |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Valid | 435783 | 58214004 | Schizophrenia | Condition | Standard | V | FALSE | TRUE | FALSE |

Cohorts that use this Concept Set:

Cardiovascular Beta Blockers

***Concept Set #5 - Antidepressants***

| INVALID\_REASON\_CAPTION | CONCEPT\_ID | CONCEPT\_CODE | CONCEPT\_NAME | DOMAIN\_ID | STANDARD\_CONCEPT\_CAPTION | CONCEPT\_CLASS\_ID | INVALID\_REASON | isExcluded | includeDescendants |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Valid | 21604686 | N06A | ANTIDEPRESSANTS | Drug | Standard | V | FALSE | TRUE | FALSE |

Cohorts that use this Concept Set:

Cardiovascular Beta Blockers

***Concept Set #6 – Major Depressive Disorder***

| INVALID\_REASON\_CAPTION | CONCEPT\_ID | CONCEPT\_CODE | CONCEPT\_NAME | DOMAIN\_ID | STANDARD\_CONCEPT\_CAPTION | CONCEPT\_CLASS\_ID | INVALID\_REASON | isExcluded | includeDescendants |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Valid | 4152280 | 370143000 | Major depressive disorder | Condition | Standard | V | FALSE | TRUE | FALSE |

Cohorts that use this Concept Set:

Major Depressive Disorder

## Complete Analysis List

Below is a complete list of analysis that will be performed. Definitions for the column 'Covariate Settings ID' can be found above in the 'Covariate Settings' section. Definitions for the 'Population Settings Id' can be found above in the 'Additional Population Settings' section.

| ID | Target Cohort Name | Outcome Cohort Name | Model Settings Id | Model Settings Description | Covariate Settings ID | Population Settings ID |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | Cardiovascular Beta Blockers | Major Depressive Disorder | 1 | Lasso Logistic Regression | 1 | 1 |

*<< add models here >>*

# References