LEGEND data model

Contents

[1 Fields with minimum values 3](#_Toc5285762)

[2 Study specification 3](#_Toc5285763)

[2.1 Indications 3](#_Toc5285764)

[2.1.1 Table: indication 3](#_Toc5285765)

[2.2 Analyses 3](#_Toc5285766)

[2.2.1 Table: cohort\_method\_analysis 3](#_Toc5285767)

[2.2.2 Table: covariate\_analysis 3](#_Toc5285768)

[2.2.3 Table: incidence\_analysis 4](#_Toc5285769)

[2.3 Exposures 4](#_Toc5285770)

[2.3.1 Table: single\_exposure\_of\_interest 4](#_Toc5285771)

[2.3.2 Table: combi\_exposure\_of\_interest 4](#_Toc5285772)

[2.3.3 Table: exposure\_group 5](#_Toc5285773)

[2.4 Outcomes 5](#_Toc5285774)

[2.4.1 Table: outcome\_of\_interest 5](#_Toc5285775)

[2.4.2 Table: negative\_control\_outcomes 5](#_Toc5285776)

[2.4.3 Table: positive\_control\_outcomes 5](#_Toc5285777)

[3 Generated results 6](#_Toc5285778)

[3.1 Metadata 6](#_Toc5285779)

[3.1.1 Table: database 6](#_Toc5285780)

[3.1.2 Table: exposure\_summary 6](#_Toc5285781)

[3.1.3 Table: comparison\_summary 6](#_Toc5285782)

[3.1.4 Table: attrition 7](#_Toc5285783)

[3.1.5 Table: covariate 7](#_Toc5285784)

[3.1.6 Table: cm\_follow\_up\_dist 8](#_Toc5285785)

[3.2 Main results 9](#_Toc5285786)

[3.2.1 Table: cohort\_method\_results 9](#_Toc5285787)

[3.2.2 Table: incidence 10](#_Toc5285788)

[3.3 Diagnostics 10](#_Toc5285789)

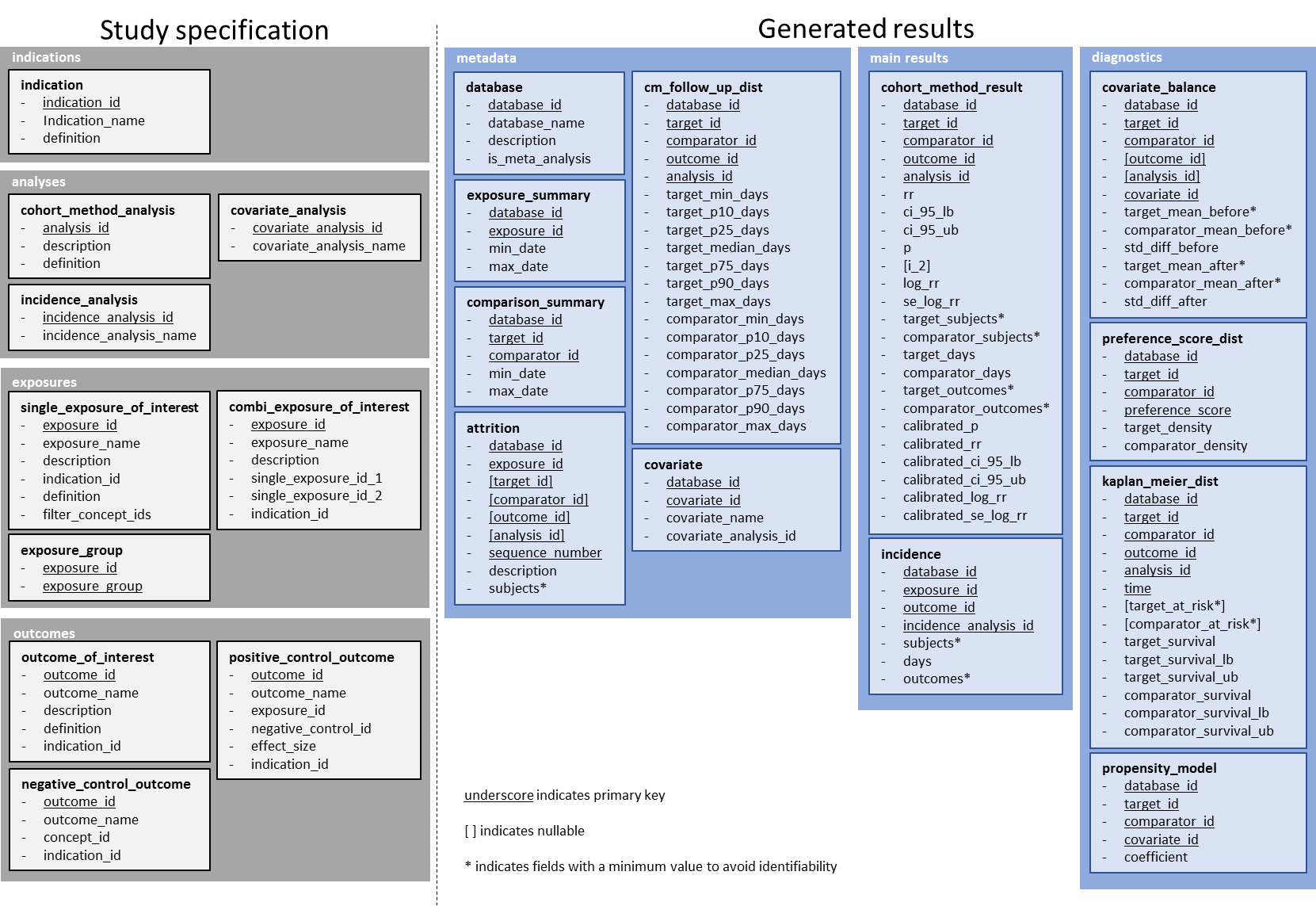
[3.3.1 Table: covariate\_balance 10](#_Toc5285790)

[3.3.2 Table: preference\_score\_dist 11](#_Toc5285791)

[3.3.3 Table: kaplan\_meier\_dist 11](#_Toc5285792)

[3.3.4 Table: propensity\_model 12](#_Toc5285793)

This document describes the data model for storing the output of the LEGEND analyses. This model will be used both to communicate results from a data site to the LEGEND coordinating center, as well as for disseminating the LEGEND evidence.



# Fields with minimum values

Some fields contain patient counts or fractions that can easily be converted to patient counts. To prevent identifiability, these fields are subject to a minimum value. When the value falls below this minimum, it is replaced with the negative value of the minimum. For example, if the minimum subject count is 5, and the actual count is 2, the value stored in the data model will be -5, which could be represented as ‘<5’ to the user. Note that the value 0 is permissible, as it identifies no persons.

These fields have been marked with \* in the preceding diagram, and are noted ‘with min value’ in the Type column in the table definitions below.

# Study specification

The first set of tables are not specific to a database, but rather provide a reference for linking results generated in databases. These can be thought of as the study specifications.

## Indications

LEGEND focuses on indications: for each indication a set of drugs and outcomes is selected, and evidence is generated for these.

### Table: indication

Lists the indications for which LEGEND is executed.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| indication\_id | varchar | A unique identifier for an indication, e.g. ‘Depression’. |
| indication\_name | varchar | The full name for the indication, e.g. ‘Pharmaceutically treated major depressive disorder’. |
| definition | varchar | OHDSI SQL or JSON object defining the nesting cohort. |

## Analyses

### Table: cohort\_method\_analysis

Lists the analyses that will be executed by the CohortMethod package.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| analysis\_id | integer | A unique identifier for an analysis. |
| description | varchar | A description for an analysis, e.g. ‘On-treatment’. |
| definition | varchar | A CohortMethod JSON object specifying the analysis. |

### Table: covariate\_analysis

Lists the covariate analyses that will be executed by the FeatureExtraction package, as well as several LEGEND-specific custom covariate builders. Each analysis can generate one or more covariates. For example, the age group analysis creates binary covariates for each 5-year age group.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| covariate\_analysis\_id | integer | A unique identifier for a covariate analysis. |
| covariate\_analysis\_name | varchar | A name for a covariate analysis, e.g. ‘Demographics: age group’. |

### Table: incidence\_analysis

Lists the incidence rate and fraction analyses that will be performed. For example, one analysis may focus on the incidence during the on-treatment window, and another may focus on the intent-to-treat window.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| incidence\_analysis\_id | varchar | A unique identifier for an incidence analysis. E.g. ‘On-treatment’. |
| description | varchar | A description for an incidence analysis |

## Exposures

Exposures can be exposures to drugs, procedures, or combinations of these. The exposure IDs used in the two exposure-of-interest tables do not overlap.

### Table: single\_exposure\_of\_interest

Lists all single exposures considered for a specific indication. Can include exposures to drugs or to procedures.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| exposure\_id | integer | A unique identifier for an exposure. |
| exposure\_name | varchar | A name for the exposure, e.g. ‘Sertraline’. |
| description | varchar | A longer description, e.g. “First use of sertraline” |
| indication\_id | varchar | Foreign key referencing the indication. |
| definition | varchar | OHDSI SQL or ATLAS cohort definition JSON for constructing the exposure. |
| filter\_concept\_ids | varchar | A list of concept IDs for which all descendants and ancestors must be excluded when fitting a propensity model. |

### Table: combi\_exposure\_of\_interest

Lists all combined exposures considered for a specific indication. These can include combinations of 2 or 3 treatments. For 2 treatments, single\_exposure\_id\_3 will be set to NULL.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| exposure\_id | integer | A unique identifier for an exposure. |
| exposure\_name | varchar | A name for the exposure, e.g. ‘Sertraline & trazadone’. |
| description | varchar | A longer description, e.g. “First use of sertraline in combination with duloxetine” |
| single\_exposure\_id\_1 | integer | Foreign key referencing the single\_exposure\_of\_interest table |
| single\_exposure\_id\_2 | integer | Foreign key referencing the single\_exposure\_of\_interest table |
| indication\_id | varchar | Foreign key referencing the indication. |

### Table: exposure\_group

Links exposures to exposure groups. One exposure can belong to multiple exposure groups.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| exposure\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| exposure\_group | varchar | A name of an exposure group, e.g. ‘Drug’, ‘Drug class’, or ‘Drug major class’. |

## Outcomes

Outcomes can be distinguished into outcomes of interest, where the true effect size is unknown and of interest, negative control outcomes where the true effect size is known to be 1, and positive control outcomes where the true effect size is of a known magnitude greater than 1. The outcome IDs used in the three outcome tables do not overlap.

### Table: outcome\_of\_interest

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| outcome\_id | integer | A unique identifier for an outcome. |
| outcome\_name | varchar | A name for the outcome, e.g. ‘Stroke. |
| description | varchar | A longer description, e.g. “First occurrence of a stroke diagnosis code in an inpatient setting.” |
| definition | varchar | OHDSI SQL or JSON object defining the outcome. |
| indication\_id | varchar | Foreign key referencing the indication. |

### Table: negative\_control\_outcomes

Negative control outcomes are derived from a single concept ID.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| outcome\_id | integer | A unique identifier for an outcome. |
| outcome\_name | varchar | A name for the outcome, e.g. ‘Ingrown nail’. |
| concept\_id | integer | The concept ID defining the negative control. |
| indication\_id | varchar | Foreign key referencing the indication. |

### Table: positive\_control\_outcomes

Positive controls are synthesized by injecting simulated outcomes into negative controls.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| outcome\_id | integer | A unique identifier for an outcome. |
| outcome\_name | varchar | A name for the outcome, e.g. ‘Ingrown nail RR=2’. |
| exposure\_id | integer | The exposure for which the signal is injected. A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| negative\_control\_id | integer | The negative control used to create the positive control. A foreign key referencing outcome\_id field in the negative\_control\_outcomes table. |
| effect\_size | float | The simulated effect size for the positive control. |
| indication\_id | varchar | Foreign key referencing the indication. |

# Generated results

The second set of tables contain the results generated on each database.

## Metadata

For each database, some meta data is captured.

### Table: database

Lists the databases that have contributed data. To identify meta-analyses estimates across databases, a dummy database record is created where the is\_meta\_analysis flag is set to 1.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| database\_id | varchar | A unique identifier for a database, e.g. ‘MDCD’. |
| database\_name | varchar | The full name for the database, e.g. ‘Truven MarketScan Multi-state Medicaid (MDCD)’. |
| description | varchar | A longer description, e.g. ‘Truven Health MarketScan® Multi-State Medicaid Database (MDCD) adjudicated US health insurance claims for Medicaid enrollees from multiple states …’ |
| is\_meta\_analysis | integer | Does the record pertain a meta-analysis across databases? (0=no, 1=yes) |

### Table: exposure\_summary

Provides summary statistics for the exposure cohorts, independent of other exposure cohorts.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| database\_id | varchar | Foreign key referencing the database. |
| exposure\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| min\_date | date | The earliest date when the exposure was observed in the database. |
| max\_date | date | The earliest date when the exposure was observed in the database. |

### Table: comparison\_summary

Provides summary statistics for the comparison between two exposure cohorts.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| database\_id | varchar | Foreign key referencing the database. |
| target\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| comparator\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| min\_date | date | The earliest date when both target and comparator were observed in the database. |
| max\_date | date | The latest date when both target and comparator were observed in the database. |

### Table: attrition

Provides the number of people in the exposure cohorts after each step of the analyses. Because some steps are related to a specific comparison or even analysis, the target, comparator, and analysis ID can optionally also be specified.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| database\_id | varchar | Foreign key referencing the database. |
| exposure\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| target\_id | integer nullable | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| comparator\_id | integer  nullable | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| outcome\_id | integer  nullable | A foreign key referencing the outcome\_of\_interest table. |
| analysis\_id | integer nullable | A foreign key referencing the cohort\_method\_analysis table. |
| sequence\_number | integer | The place in the sequence of steps defining the final analysis cohort. 1 indicates the original exposed population without any inclusion criteria. |
| description | varchar | A description of the last restriction, e.g. “Removing persons with the outcome prior”. |
| subjects | integer with min value | The number of subjects in the cohort. |

### Table: covariate

Lists the covariates constructed in a database.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| database\_id | varchar | Foreign key referencing the database. |
| covariate\_id | integer | A unique identified for a covariate. |
| comparator\_name | varchar | A name for a covariate, e.g. ‘Age group: 20-25 years’. |
| covariate\_analysis\_id | integer | A foreign key referencing the covariate\_analysis table. |

### Table: cm\_follow\_up\_dist

Contains the distribution of follow up time in the target and comparator groups for a specific cohort method analysis. Only outcomes of interest are included.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| database\_id | Varchar | Foreign key referencing the database. |
| target\_id | Integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| comparator\_id | Integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| outcome\_id | Integer | A foreign key referencing the outcomes\_of\_interest, negative\_control\_outcome, or positive\_control\_outcome table. |
| analysis\_id | Integer | A foreign key referencing the cohort\_method\_analysis table. |
| target\_min\_days | integer | The minimum number of observation days for a person. |
| target\_p10\_days | integer | The 10th percentile of number of observation days for a person in the target group. |
| target\_p25\_days | integer | The 25th percentile of number of observation days for a person in the target group. |
| target\_median\_days | integer | The median number of observation days for a person in the target group. |
| target\_p75\_days | integer | The 75th percentile of number of observation days for a person in the target group. |
| target\_p90\_days | integer | The 90th percentile of number of observation days for a person in the target group. |
| target\_max\_days | integer | The maximum number of observation days for a person in the target group. |
| comparator\_min\_days | integer | The minimum number of observation days for a person in the comparator group. |
| comparator\_p10\_days | integer | The 10th percentile of number of observation days for a person in the comparator group. |
| comparator\_p25\_days | integer | The 25th percentile of number of observation days for a person in the comparator group. |
| comparator\_median\_days | integer | The median number of observation days for a person in the comparator group. |
| comparator\_p75\_days | integer | The 75th percentile of number of observation days for a person in the comparator group. |
| comparator\_p90\_days | integer | The 90th percentile of number of observation days for a person in the comparator group. |
| comparator\_max\_days | integer | The maximum number of observation days for a person in the comparator group. |

## Main results

These tables contain the main results of LEGEND.

### Table: cohort\_method\_results

Contains the results produced by the CohortMethod package for the main effects. Also contains calibrated p-values and confidence intervals. Meta-analysis estimates are also stored in this table.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| database\_id | varchar | Foreign key referencing the database. |
| target\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| comparator\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| outcome\_id | integer | A foreign key referencing the outcomes\_of\_interest, negative\_control\_outcome, or positive\_control\_outcome table. |
| analysis\_id | integer | A foreign key referencing the cohort\_method\_analysis table. |
| rr | float | The estimated relative risk (hazard ratio). |
| ci\_95\_lb | float | The lower bound of the 95% confidence interval of the relative risk. |
| ci\_95\_ub | float | The upper bound of the 95% confidence interval of the relative risk. |
| p | float | The two-sided p-value considering the null hypothesis of no effect. |
| i\_2 | float nullable | The I2 measure of between-database heterogeneity (for meta-analyses estimates only). |
| log\_rr | float | The log of the relative risk. |
| se\_log\_rr | float | The standard error of the log of the relative risk. |
| target\_subjects | integer with min value | The number of subject in the target cohort. |
| comparator\_subjects | integer with min value | The number of subject in the comparator cohort. |
| target\_days | Integer | The number of days observed in the target cohort. |
| comparator\_days | Integer | The number of days observed in the comparator cohort. |
| target\_outcomes | integer with min value | The number of outcomes observed in the target cohort. |
| comparator\_outcomes | integer with min value | The number of outcomes observed in the comparator cohort. |
| calibrated\_p | float | The calibrated p-value. |
| calibrated\_rr | float | The calibrated relative risk (hazard ratio). |
| calibrated\_ci\_95\_lb | float | The lower bound of the calibrated 95% confidence interval of the relative risk. |
| calibrated\_ci\_95\_ub | float | The upper bound of the calibrated 95% confidence interval of the relative risk. |
| calibrated\_log\_rr | float | The log of the calibrated relative risk. |
| calibrated\_se\_log\_rr | float | The standard error of the log of the calibrated relative risk. |

### Table: incidence

Contains the data for computing incidence rates and fractions for the outcomes of interest in the exposure cohorts.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| database\_id | varchar | Foreign key referencing the database. |
| exposure\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| outcome\_id | integer | A foreign key referencing the outcomes\_of\_interest table. |
| incidence\_analysis\_id | varchar | A foreign key referencing the incidence\_analysis table. |
| subjects | integer with min value | The number of subjects in the exposure cohort. |
| days | integer | The number of days observed in the exposure cohort. |
| outcomes | integer with min value | The number of outcomes observed in the exposure cohort. |

## Diagnostics

### Table: covariate\_balance

Contains the covariate balance statistics for each comparison. To save space, balance for all covariates is only computed once for each target-comparator pair, using propensity score matching and stratification. Only a subset of covariates is reported for each outcome-analysis combination.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| database\_id | varchar | Foreign key referencing the database. |
| target\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| comparator\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| outcome\_id | integer nullable | A foreign key referencing the outcomes\_of\_interest table. |
| analysis\_id | integer nullable | A foreign key referencing the cohort\_method\_analysis table. |
| covariate\_id | integer | A foreign key referencing the covariate table. |
| target\_mean\_before | float with min value | The mean value of the covariate in the target cohort before propensity score adjustment. |
| comparator\_mean\_before | float with min value | The mean value of the covariate in the comparator cohort before propensity score adjustment. |
| std\_diff\_before | float | The standardized difference of the means between the target and comparator cohort before propensity score adjustment. |
| target\_mean\_after | float with min value | The mean value of the covariate in the target cohort after propensity score adjustment. |
| comparator\_mean\_after | float with min value | The mean value of the covariate in the comparator cohort after propensity score adjustment. |
| std\_diff\_after | float | The standardized difference of the means between the target and comparator cohort after propensity score adjustment. |

### Table: preference\_score\_dist

Provides the preference score distribution for each comparison.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| database\_id | varchar | Foreign key referencing the database. |
| target\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| comparator\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| preference\_score | float | A preference score value. |
| target\_density | float | The distribution density for the target cohort at the given preference score. |
| comparator\_density | float | The distribution density for the comparator cohort at the given preference score. |

### Table: kaplan\_meier\_dist

Contains data to display as a Kaplan-Meier plot for each comparison. The number of subjects at risk in the target and comparator cohorts will only be provided for specific pre-defined time points.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| database\_id | varchar | Foreign key referencing the database. |
| target\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| comparator\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| outcome\_id | integer | A foreign key referencing the outcomes\_of\_interest table. |
| analysis\_id | integer | A foreign key referencing the cohort\_method\_analysis table. |
| time | integer | Time in days since cohort start. |
| target\_at\_risk | integer with min value nullable | The number of subjects still at risk in the target cohort. |
| comparator\_at\_risk | integer with min value nullable | The number of subjects still at risk in the comparator cohort. |
| target\_survival | float | The estimated survival fraction in the target cohort. |
| target\_survival\_lb | float | The lower bound of the 95% confidence interval of the survival fraction in the target cohort. |
| target\_survival\_ub | float | The upper bound of the 95% confidence interval of the survival fraction in the target cohort. |
| comparator\_survival | float | The estimated survival fraction in the comparator cohort. |
| comparator\_survival\_lb | float | The lower bound of the 95% confidence interval of the survival fraction in the comparator cohort. |
| comparator\_survival\_ub | float | The upper bound of the 95% confidence interval of the survival fraction in the comparator cohort. |

### Table: propensity\_model

Contains the propensity models.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| database\_id | varchar | Foreign key referencing the database. |
| target\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| comparator\_id | integer | A foreign key referencing the single\_exposure\_of\_interest or combi\_exposure\_of\_interest table. |
| covariate\_id | integer | A foreign key referencing the covariate table. |
| coefficient | float | The coefficient (beta) for the covariate in the propensity model. |