

# Package ‘CohortPathways’

March 24, 2025

**Type** Package

**Title** Create Pathways from Target to Event Cohorts

**Version** 0.0.1

**Date** 2024-10-20

**Maintainer** Gowtham Rao <rao@ohdsi.org>

**Description** Software tool designed to compute the temporal relationship defined as pathways between any two instantiated cohorts. The cohorts are input as Target and event cohorts.

**Depends** DatabaseConnector (>= 5.0.0),  
R (>= 4.1.0)

**Imports** checkmate,  
dplyr,  
lifecycle,  
rlang,  
SqlRender,  
tidyr

**Suggests** remotes,  
testthat,  
withr

**License** Apache License

**RoxygenNote** 7.3.2

**Roxygen** list(markdown = TRUE)

**Encoding** UTF-8

**Language** en-US

**URL** <https://github.com/OHDSI/CohortPathways>

**BugReports** <https://github.com/OHDSI/CohortPathways/issues>

## Contents

executeCohortPathways . . . . .	2
Index	5

---

executeCohortPathways *Execute cohort pathway analysis.*

---

## Description

This function reproduces the logic behind "Cohort Pathways" (sometimes referred to as "Treatment Pathways") as defined in ATLAS. A *pathway* is the sequence of events that occur after a cohort entry, often used to compare or visualize patients' treatment patterns. The function takes a set of target cohorts (the index cohorts) and event cohorts (the "steps" in the pathway) and computes the frequency and order in which these events occur.

## Usage

```
executeCohortPathways(
  connectionDetails = NULL,
  connection = NULL,
  cohortDatabaseSchema,
  cohortTableName = "cohort",
  tempEmulationSchema = getOption("sqlRenderTempEmulationSchema"),
  targetCohortIds,
  eventCohortIds,
  minCellCount = 5,
  allowRepeats = FALSE,
  maxDepth = 5,
  collapseWindow = 30
)
```

## Arguments

- |                      |  |
|----------------------|--|
| connectionDetails    | An object of type <code>createConnectionDetails</code> (from the <b>DatabaseConnector</b> package) providing the details to connect to the database. If NULL, then connection must be provided.        |
| connection           | An existing database connection (as created by <code>connect</code> ). If NULL, a new connection will be established using connectionDetails at the start of the function, and closed upon completion. |
| cohortDatabaseSchema | Schema where the target and event cohorts are stored. For SQL Server, this should include both the database and schema name (e.g., "mydatabase.dbo").  |
| cohortTableName      | Name of the table containing your cohorts (default = "cohort").  |
| tempEmulationSchema  | For databases without true temp table support (e.g., Oracle), specify a schema where you have write privileges to emulate temp tables.   |
| targetCohortIds      | Integer vector of cohort definition IDs representing one or more <i>target</i> cohorts (the index populations whose pathways you want to explore).   |
| eventCohortIds       | Integer vector of cohort definition IDs representing one or more <i>event</i> cohorts (the steps in the pathways, e.g., treatments or clinical events).  |

minCellCount	The minimum cell count to display (default = 5). Any group or count below this threshold should be redacted or masked if you're following a strict privacy policy.
allowRepeats	Boolean; if TRUE, an event cohort can reappear multiple times in the same patient's pathway sequence (default = FALSE).
maxDepth	Maximum number of steps in any pathway (default = 5). Pathways longer than this threshold are truncated.
collapseWindow	An integer indicating the number of days within which events are considered the same "step" (default = 30 days). This window effectively groups closely occurring events to reduce spurious variability.

## Details

**Overview** Cohort Pathways are commonly used to visualize how patients move through different treatments or events over time. Within ATLAS, these pathways are depicted as “sunburst” diagrams and flow charts. This function provides a programmatic way to replicate that logic outside of ATLAS, especially useful when you need to run a pathway analysis in a fully scripted R environment.

### Key Steps

1. **Identify Cohorts:** You must already have *target* and *event* cohorts instantiated in your results schema (i.e., the cohorts exist in the cohortTableName table).
2. **Specify Pathway Parameters:** Options such as allowRepeats, collapseWindow, and maxDepth control how the event sequences are aggregated:
  - allowRepeats: If TRUE, an event cohort can appear more than once at different time points in the pathway sequence for the same person.
  - collapseWindow: Events that occur within this many days of each other (default = 30 days) will be considered part of the same step, reducing the “noise” of very frequent transitions.
  - maxDepth: The maximum length of a pathway (default = 5 steps). Longer sequences are truncated to keep the analysis and visualization more interpretable.
3. **Analyze:** The function creates internal temporary tables (or uses your specified tempEmulationSchema if necessary) to calculate how many people experience each distinct pathway. It then returns several data frames capturing:
  - pathwayAnalysisStatsData: Summary statistics by target cohort, including total counts and the distribution of event occurrences.
  - pathwaysAnalysisPathsData: The distinct sequences (paths) observed, with counts of how many persons followed each sequence.
  - pathwaysAnalysisEventsData: A more granular breakdown of how events contribute to different pathway steps.
  - pathwayAnalysisCodesData and related tables: Helper mappings and indices that translate event cohorts into pathway “codes.”

**Cell Count Restrictions** The argument minCellCount allows you to enforce a minimum cell count, consistent with privacy rules. Any group or count below minCellCount can be masked or handled downstream according to your organization's policies.

### Reference

- *The Book of OHDSI* (Chapter: Characterization, Section: Cohort Pathways in ATLAS)
- OHDSI Forums: <https://forums.ohdsi.org/t/cohort-pathways-in-atlas-faq/9511/2> <https://forums.ohdsi.org/t/reproducing-a-treatment-pathway-study-atlas-json-files/9450>

**Value**

A named list of data frames capturing key pathway results:

`pathwayAnalysisStatsData` High-level summary statistics about the number of people in each target cohort and the distribution of events.

`pathwaysAnalysisPathsData` Rows describing each distinct pathway (sequence of events) observed, along with counts of persons following each path.

`pathwaysAnalysisEventsData` Detailed event-level information, allowing you to see how each event fits into a given path.

`pathwayAnalysisCodesData` Mapping of event cohort IDs to pathway "codes" for easy interpretation of combinatorial steps.

`pathwaycomboIds`, `pathwayAnalysisCodesLong`, `isCombo` Additional helper tables describing combos of events.

**Examples**

```
## Not run:
# Example: Execute a pathway analysis for two target cohorts and three event cohorts.

# 1) Create connection details or use an existing connection.
connDetails <- DatabaseConnector::createConnectionDetails(
  dbms = "postgresql",
  server = "myserver/mydb",
  user = "someuser",
  password = "DontSharePasswords"
)

# 2) Run the function
result <- executeCohortPathways(
  connectionDetails = connDetails,
  cohortDatabaseSchema = "my_results_schema",
  cohortTableName = "my_cohort",
  targetCohortIds = c(123, 456), # e.g., 'diabetes' and 'hypertension' cohorts
  eventCohortIds = c(789, 1011, 1213), # e.g., 'metformin', 'insulin', 'ace_inhibitors'
  allowRepeats = TRUE,
  maxDepth = 4,
  collapseWindow = 14
)

# 3) Inspect the returned data frames for pathway summaries and details
names(result)
head(result$pathwaysAnalysisPathsData)

## End(Not run)
```

# Index

connect, [2](#)  
createConnectionDetails, [2](#)  
executeCohortPathways, [2](#)