

# Assignment: Strategus Study Specification Build

**Designing a Reproducible OHDSI Study Package (No Execution Required)**

## Overview

In this assignment, you will use **Strategus** to design an observational study package.

Because not everyone will have access to the data required, you are **not required to run the study**. Instead, you will build a complete study specification and package infrastructure that could be executed later.

This is a three-part assignment. In total, this is worth 40% of your final grade.

## Learning Aims

By completing this assignment, you should be able to:

- Translate a study question into an OHDSI-compatible study specification.
- Build Strategus study specification scripts for multiple analysis sections.
- Organize cohort definitions and JSON artifacts required for a study package.
- Document and justify key design decisions in a concise final write-up.

## Assignment Structure and Due Dates

### Part 1: Diagnostics / Characterization / Incidence Plan [10% of final grade]

- **Due:** Week 7, March 6th, midnight
- Build the specification components for diagnostics, characterization, and incidence.
- Focus on code mechanics and package structure.

## **Part 2: Population Estimation Plan (CohortMethod/SCCS/other) [10% of final grade]**

- **Due:** Week 8, March 13th, midnight
- Build the specification components for population-level estimation (CohortMethod).
- Focus on code mechanics and package structure.

**Note:** Should your research question not lend itself to these approaches, reach out to us.

## **Part 3: Final Integrated Product [20% of final grade]**

- **Due:** March 20th, midnight
- Integrate Parts 1 and 2 into a coherent final project.
- Include the required write-up (see below).

### **Core Technical Requirements**

The final project should include:

- A Strategus study specification build script.
- Study package infrastructure for a realistic OHDSI study project.
- Cohort definitions and supporting JSON artifacts.
- Organized files/folders so the project is reviewable and reproducible.

The study specification must successfully compile to JSON. At minimum, include the assets needed to demonstrate study design readiness (even though execution is not required).

For this assignment, **negative controls are optional and not required**.

### **Understanding Module Defaults**

Every module's `createModuleSpecifications()` has default values for its parameters. Before accepting any defaults, you must inspect them:

```
# Example: inspect all parameters and their defaults
formals(cdModule$createModuleSpecifications)
```

You should do this for **every module you use**. In your code, explicitly list all parameters (not just the ones you are overriding) with inline comments showing the default value. For example:

```

cohortDiagnosticsModuleSpecifications <- cdModule$createModuleSpecifications(
  runInclusionStatistics = TRUE,                      # default: TRUE
  runIncludedSourceConcepts = TRUE,                   # default: TRUE
  runOrphanConcepts = TRUE,                          # default: TRUE
  runTimeSeries = FALSE,                            # default: FALSE
  ...
)

```

This forces you to think about what each module does by default and whether those defaults are appropriate for your study. In your write-up, briefly justify any defaults you changed and explain why the ones you kept are appropriate.

## Module Reference Documentation

Consult the online documentation for each module's full parameter list and descriptions:

**Strategus module classes:**

- [CohortGeneratorModule](#)
- [CohortDiagnosticsModule](#)
- [CohortIncidenceModule](#)
- [CharacterizationModule](#)
- [CohortMethodModule](#)
- [SelfControlledCaseSeriesModule](#)

**Underlying HADES package docs (for sub-object functions like `createOutcomeDef`, `createTimeAtRiskDef`, etc.):**

- [CohortGenerator](#)
- [CohortDiagnostics](#)
- [CohortIncidence](#) — see especially the [reference index](#)
- [Characterization](#)
- [CohortMethod](#)
- [SelfControlledCaseSeries](#)

## Data and Cohort Choices

You have two main paths for choosing your research question and cohorts:

**Option 1: Bring Your Own Question** You can use cohort definitions you or someone else has created to answer a novel scientific question of your interest, regardless of whether you have the data to run it.

- **Trade-off:** This path is best if you want to realistically explore how to build a study package for a specific scientific question, even if you won't be able to execute it during the class.

**Option 2: Use a Eunomia Dataset** You are allowed to use one of the available [Eunomia datasets](#) (e.g., GiBleed, MIMIC, or Synthea). For this path, you should use cohort definitions that make sense for that specific dataset (you can use the bundled cohort definitions or create simple ones).

**Important:** Please do **not** use the GiBleed dataset for your assignment submission, since we are using GiBleed in class as the live demo. Choose a different Eunomia dataset (for example, MIMIC or Synthea) or bring your own question/cohort assets.

- **Trade-off:** This path is best if you want to be able to actually run the study package end-to-end and see the results, as the data is readily available.

It is up to you which path you choose.

## Sourcing Cohorts

If your cohort definitions are available on an accessible ATLAS instance, you are **encouraged** to pull definitions via WebAPI. You can create a script named something like `DownloadDefinitions.R`

You can also manually download and organize cohorts without WebAPI pulls using Atlas exporter or CAPR.

## Deliverables

### Part 1

- A brief outline of your study question and rationale.
- Code and configuration for diagnostics/characterization/incidence planning, with all `createModuleSpecifications()` parameters listed explicitly and defaults annotated.
- Study specification JSON file.
- Relevant cohort and JSON assets for this section.

## **Part 2**

- Code and configuration for Population Estimation (e.g., CohortMethod) population estimation planning or other approved method, with all `createModuleSpecifications()` parameters listed explicitly and defaults annotated.
- Study specification JSON file.
- Relevant cohort and JSON assets for this section.

## **Part 3**

- Integrated final Strategus study build project.
- Complete, organized cohort/JSON/specification infrastructure.
- Final write-up (required only with Part 3).

### **Required Write-Up (Part 3 Only)**

Submit one concise write-up that includes:

1. **Study question and rationale:** this need not be detailed to the level of a “production quality” study protocol, but detailed enough to explain the basic study design and rationale behind it.
2. **Cohort definitions** (target/comparator/outcome and related choices)
3. **Decision rationale by section**
  - Why diagnostics/characterization/incidence choices were made
  - Why population estimation choices were made

For **Parts 1 and 2**, with the exception of a very brief write-up of the question/rationale, you may focus on building the code and infrastructure. A full, polished write-up is not required until Part 3.

### **Submission Format**

Preferred:

- GitHub repository URL

Accepted alternative:

- ZIP archive of the full project directory

Your submission should make it easy to review folder structure, scripts, and study assets. Please email us GitHub submissions; otherwise, upload your ZIP archive to Canvas.

## **Evaluation Focus**

Work will be evaluated based on:

- Completeness of Strategus study specification components
- Quality and organization of package infrastructure
- Cohort/JSON asset readiness and consistency
- Clarity and quality of final study rationale (Part 3 write-up)