



Introducing the CostUtilization Package: Standardizing Health Economic Analysis in OHDSI

Alpha Release Overview

Audience: OHDSI HEVA Workgroup

Date: October 2025



The Need for Standardized Economic Analysis

Challenge:

- Methodological Opacity: HEOR studies often rely on ad hoc, custom SQL implementations, hindering reproducibility and validation.
- Semantic Ambiguity: Inconsistent definitions of "Cost," "Charge," and "Payment" lead to incomparable results across the network.
- Integration Gap: Difficulty in integrating economic outcomes seamlessly into standard OHDSI analytical pipelines (e.g., CohortMethod, PLP).

Vision:

- To enable "actuarial-quality" economic analysis at scale across the OHDSI network.
- Provide a HADES-conformant tool that standardizes the feature engineering of cost and utilization metrics.
- Transform health economics from a siloed activity into a standard component of OHDSI evidence generation.



What is the CostUtilization Package?

Key Features of the R Package

1. Complies with HADES standards for cost metrics
2. Connects to OMOP CDM databases
3. Executes optimized SQL queries for cost calculation
4. Outputs might be standardized OHDSI covariateData objects
5. Uses validated SQL logic for CDM version 6.0
6. Applies visit-centric analytical framework
7. Supports diverse cost analysis applications



Alpha Release Status and The Path Forward

Current Status: Alpha Release

- Core architecture and foundational logic are implemented.
- Testing is currently restricted to simulated data.
- Not ready for general use due to external dependencies.

Critical Dependencies (The Blockers):

- OMOP CDM v5.5 (Pending Approval):
 - The package relies exclusively on the normalized COST table structure proposed in CDM v5.5 (reintegrated from v6.0).
 - This structure uses `cost_concept_id` (meaning) and `cost_type_concept_id` (provenance) for essential flexibility.

Themis Conventions (Pending Finalization):

- Standardized semantic meaning of cost data is essential for "actuarial-quality" results.
- The package assumes conformance to these conventions (e.g., standardized currency).



Architecture and Design Philosophy

HADES Conformant Structure:

- R/: User-facing functions (e.g., `calculateCostOfCare()`).
 - Handles connection, parameterization, and output formatting.
- inst/sql/: The core analytical engine.
 - A large, parameterized SQL script.

Why a Standalone Package?

- FeatureExtraction uses "hyper-parameterized" SQL for many simple clinical features.
- CostUtilization requires a more monolithic, "opinionated" SQL optimized for the complexities of the normalized COST table.
- Direct integration was challenging due to these fundamental differences in SQL architecture.

The Integration Strategy:

- The output might be structurally identical to *FeatureExtraction's* output (the *covariateData* object).



Seamless Integration with the HADES Workflow

Workflow Integration:

1. Define Cohorts (Atlas)
2. Generate Features
 - FeatureExtraction (Clinical Covariates) - covariateData
 - CostUtilization (Economic Covariates) – covariateData or other generic output
3. Combine (rbind/union) (in progress)
4. Execute Analysis
 - CohortMethod
 - PatientLevelPrediction
 - Characterization



The Analytical Pattern: A Modular, Visit-Centric Approach

Design Principle:

- Costs must be anchored to clinical context.
- A cost without a visit_occurrence_id is considered uninterpretable and is excluded.

The Standardized Query Pattern (Based on Technical Tutorial):

Step 0: Calculate Denominator (Person-Time at Risk)

- Accurately censor observation time using OBSERVATION_PERIOD to avoid bias.

Step 1: Identify Relevant Encounters (The Anchor)

- Select VISIT_OCCURRENCE records within the analysis window.
- Apply filters (e.g., by visit type, or by clinical events occurring during the visit).

Step 2: Link Encounters to Costs (The Numerator)

- Join to the COST table using visit_occurrence_id.
- Filter by cost_concept_id (e.g., Allowed Amount) and currency_concept_id.

Step 3: Calculate Standardized Rates

- Combine Numerator and Denominator (e.g., Cost PPPM).



Analytical Flexibility: From Broad to Granular Costing (Visit-Level)

Visit Cost Scenarios and Adjustments

1. Base Case: Overall care cost for all settings
2. Encounter Type: Costs for specific visit types
3. Event-Based: Expenses for visits with designated events
4. Concurrent Events: Costs for visits with multiple events
5. Note: Cost includes entire visit for all scenarios



Advanced Use Case: Scenario 5 - Micro-Costing

Objective:

- Isolate the specific line-item cost of an intervention itself, rather than the cost of the entire visit.

Use Case:

- Essential for Health Technology Assessments (HTA) or detailed economic evaluations (e.g., the cost of the chemotherapy drug itself).

Prerequisites:

- VISIT_DETAIL table must be populated.
- Costs must be attributed at the line level using visit_detail_id in domain tables AND the COST table.

Logic Shift:

- The analysis shifts focus from VISIT_OCCURRENCE to VISIT_DETAIL.
- The COST table is joined via visit_detail_id.
- Requires filtering the COST table to ensure the cost is linked to the VISIT_DETAIL domain (using cost_event_field_concept_id).



Ensuring Reliability: The Testing Framework

The Challenge:

- The standard OHDSI test dataset (Eunomia) does not contain COST data, nor does it conform to CDM v5.5.

The Innovation: On-the-Fly Data Generation

The package's test suite dynamically generates a compliant test environment:

- Establish standard Eunomia environment.
- Programmatically create the CDM v5.5 COST table structure.
- Synthesize and inject cost data, linked relationally to existing Eunomia visits.
- Execute unit tests and validate outputs.

Path to HADES Conformance:

- In Progress: Finalizing unit tests to achieve high code coverage.
- Automated: Continuous Integration (CI) via GitHub Actions ensures stability.
- Future: Real-world validation against production CDM v5.5 datasets (essential for Beta release).



Challenges and Community Adoption

Key Challenges:

- Standards Adoption (CDM v5.5/Themis):
 - The primary blocker. Adoption is required before the package can be widely used or validated.
- Data Quality and Conformance:
 - Real-world cost data is complex.
 - The package will eventually include built-in checks to enforce Themis conformance and may halt execution if data is non-conformant.
- International Applicability:
 - The current logic, while proven, needs validation against diverse international reimbursement models.

Adoption Strategy:

- Positioning the package as the reference implementation of the HEVA workgroup's proposed standards.
- Actively seeking early adopter partners to test and validate the package on real-world data.



Conclusion and Call to Action

Summary:

- The **CostUtilization** package offers a standardized, HADES-compliant framework to elevate the quality and reproducibility of health economic analysis in OHDSI.
- A Call to Action for the HEVA Workgroup:
 - Engage in Standards Development:
 - Your active participation in finalizing CDM v5.5 and Themis conventions is critical.
 - Review the Code:
 - The alpha_release branch is public. We welcome your technical feedback on GitHub.
 - github.com/OHDSI/CostUtilization (alpha_release branch)
 - Become an Early Tester:
 - Do you have cost data? Are you planning to migrate to CDM v5.5?
 - Let's partner to test and validate this tool.

Q&A