Optum Integrated Claims ETL

Authors: Anthony Sena, Ajit Londhe, Chris Knoll, Erica Voss

# Introduction

The purpose of this document is to describe the Extract, Transform, Load (ETL) mapping of the licensed data from Optum Integrated (United Health Group) into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) V5. This data set represents patients in terms of their claims and EHR records. In this ETL specification, we will model the healthcare interactions for patients that have both medical and prescription insurance coverage. When EHR activity is present for patients during these observation periods, these will also be recorded. This document will reference the following Optum ETL specifications:

***Optum Extended SES:***

<https://github.com/OHDSI/ETL-CDMBuilder/blob/master/man/OPTUM_EXTENDED/OptumExtended_ETL_CDM_V5.docx>

***Optum Oncology:***

<https://github.com/OHDSI/ETL-CDMBuilder/blob/master/man/OPTUM_ONCOLOGY/OptumOncologyEHR_CDM_V5.docx>

The Optum Integrated ETL specification will be based on logic from one or both of the referenced Optum ETL specifications. The rationale is that the Integrated data set will have both claims and can be processed similarly to the data found in Optum SES. The same holds true for the EHR data as that follows the same format as the Optum Oncology data set.

This document will mainly focus on those CDM database tables that have processing unique to Integrated and reference the supporting ETL specifications where applicable.

# Optum data management practices

Optum has communicated that the Optum Extended SES product will receive enhancements ahead of Optum Integrated. For example, the SES data product may contain additional columns and/or tables in the data supplied in Q1 2017 and these changes to the “claims” tables will not be carried forward to Optum Integrated until sometime later. As a result, there may be scenarios when the structure of the claims data contained in SES will differ from Integrated which may result in differences in the way these claims need to be processed. We wanted to note this here since this document references the SES ETL specification and at this time, this document does not attempt to account for those differences since the data structure is identical.

# Data Mapping

This section describes how the source files are mapped into the CDM. Refer to each ETL reference for notes on conventions used.

## Person

This table will follow a unique approach for Integrated. This is due to the fact that the SES logic requires fields in the MEMBER\_DETAIL table that are not present in the Integrated tables. Furthermore, we want to only ETL those patients that have both medical & prescription coverage. In order to do this, we will JOIN the following source tables together as specified below:

select \*

from patient p

inner join member\_detail md on p.ptid = md.ptid

Once this is set, follow the logic laid out in the Oncology ETL.

## Observation\_Period

Note: only create observation\_period entries for the subset of person entries that were obtained in the Person section above.

This table will follow a unique approach for Integrated. We want to construct observation periods that reflect the time when a person had active claims and clinical coverage to ensure we have a complete picture of a patient’s interactions during that period. The following procedure is used to create these periods:

* Construct candidate eligibility observation periods for claims coverage from MEMBER\_DETAIL (follow the Optum SES approach). These candidate observation periods will be referred to as **#elig\_periods** with fields for ptid, elig\_start\_date and elig\_end\_date respectively.
* Construct candidate clinical observation periods from PATIENT (follow the Optum Oncology approach). These candidate observation periods will be referred to as **#clin\_periods** with fields for ptid, clin\_start\_date and clin\_end\_date respectively.
* **NOTE:** A patient should only have 1 clinical observation period but may have > 1 eligibility observation period.
* Join #elig\_periods and #clin\_periods on ptid to obtain a list of all claims eligibility periods and the corresponding clinical observation period. Call this table #obs\_periods
* Next apply the following logic to construct the observation periods:

SELECT

person\_id

, CASE

WHEN DATEDIFF(d, clin\_start\_date, elig\_start\_date) < 0 THEN clin\_start\_date

ELSE elig\_start\_date

END observation\_period\_start\_date

, CASE

WHEN DATEDIFF(d, clin\_end\_date, elig\_end\_date) > 0 THEN clin\_end\_date

ELSE elig\_end\_date

END observation\_period\_end\_date

, 38000280 as period\_type\_concept\_id

from #obs\_periods

where elig\_start\_date <= clin\_end\_date

AND clin\_start\_date <= elig\_end\_date -- Any overlap in periods

;

The query above attempts to find the latest of the start\_dates and the earliest of the end\_dates to ensure that the observation period is restricted to the time where we are confident that both clinical and claims data are present.

## Visit\_Occurrence

Logic to use: **Both**

**Notes:** Both ETL specifications use a system-generated id’s for visit\_occurrence\_id and so we will need to account for the “union” of these two sets of visits when generating IDs.

***In a future version, we’ll review this approach to potentially change this to collapse to a single visit\_occurrence record where applicable. This decision will need to be noted in the metadata.***

## Payer\_Plan\_Period

Logic to use: **SES**

## Care\_Site

Logic to use: **SES**

## Location

Logic to use: **Oncology**

Note: SES ETL requires fields in the member\_detail table that are not present in Integrated.

## Provider

Logic to use: **Oncology**

Note: SES logic will NOT work with integrated as it does not contain the DEA column on the rx\_claims table

## Death

Logic to use: **Oncology**

***Notes:***

***Per Optum’s documentation, the native Patient.Deceased\_indicator column:*** *Flag indicating if patient is known to be deceased as of analytic end date (sourced from EMRs and the Social Security Administration Death Master File)*

## All Other CDM Tables

Logic to use: **Both**