

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

OMOP Common Data Model (CDM)

ETL Mapping Specification Template

September 23, 2009

Indiana University – Regenstrief Institute

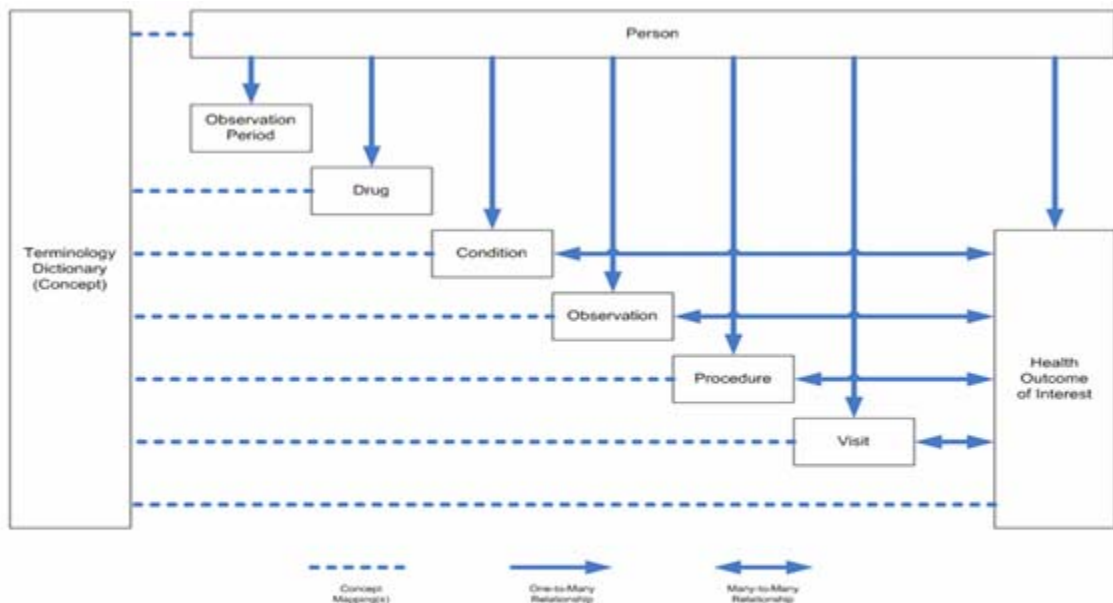


Table of Contents

1.0 Introduction	4
2.0 Source Data Mapping Approach	4
3.0 Source Data Mapping	6
3.1 Data Mapping	6
3.1.1 Table Name: PERSON	6
3.1.2 Table Name: DRUG_EXPOSURE	7
3.1.3 Table Name: CONDITION_OCCURRENCE	9
3.1.4 Table Name: VISIT_OCCURRENCE	13
3.1.5 Table Name: PROCEDURE_OCCURRENCE	13
3.1.6 Table Name: OBSERVATION	15
3.1.7 Table Name: OBSERVATION_PERIOD	17
3.2 Source Independent Data Mapping	18
3.2.1 Table Name: DRUG_ERA	18
3.2.2 Table Name: CONDITION_ERA	19
3.3 Reference Tables	20
3.3.1 Table Name: DRUG_EXPOSURE_REF	20
3.3.2 Table Name: CONDITION_OCCURRENCE_REF	20
3.3.3 Table Name: PROC_OCCURRENCE_REF	23
3.3.4 Table Name: OBSERVATION_TYPE_REF	24
3.3.5 Table Name: VOCABULARY_REF	25
3.3.6 Table Name: RELATIONSHIP_TYPE	26

Document Control

Change Record

Date	Author	Version	Change Reference
21-Sep-09	OMOP	1.0	New document, describes OMOP ETL for Distributed Partners

Contributors

Name	Organization	Title

Reviewers

Name	Role	Title	Date Reviewed
Mark Khayter	OMOP	Technical Consultant	23-Sep-2009

Document References

Document Title	Type of Reference	Document Location
OMOP ETL Mapping Specification		OMOP Basecamp

1.0 Introduction

This document reflects the requirements, assumptions, business rules and transformations for the implementation of the Common Data Model (CDM) as implemented by [Regenstrief Institute](#). The initial ETL process was built using data and transformations as applicable to GE and Thomson.

The purpose of this document is to describe the ETL mapping of the ~~proprietary or licensed~~ data (for which [Regenstrief Institute](#) serves as the custodian) from the [INPC](#) into the OMOP Common Data Model. **Note that the INPC is a database which receives data from many entities, and Regenstrief Institute put into the CDM only those entities' data which it had permission to use in the OMOP project. The Regenstrief Institute made an additional decision to use, for the CDM, data only for those patients who live in the nine-county Central Indiana area (Indianapolis and the surrounding eight counties).**

It is based on the OMOP ETL Specifications. General information that is covered by the OMOP ETL Specification will not be covered in this document, but a detailed discussion of the [Regenstrief Institute](#)-specific aspects of mapping and converting data to the standard CDM is provided.

The document is composed of three main sections:

- Source Data Mapping. Describes major tables of the CDM schema and special data handling required for each table.
- Source Independent Data Mapping. Describes mapping process of the Drug and Condition Era's.
- Data Mapping Reference tables.

In each section, the tables and their mapping are individually reviewed along with any source specific rules and exceptions.

The intended audience for this document will include both researchers that want to use the experience and learning in order to incorporate them into their own CDM construction.

2.0 Source Data Mapping Approach

In the OMOP ETL Specifications, this section covers the high-level assumptions and approach to extraction, transformation and loading (ETL) of raw source data into the Common Data Model (CDM). This high-level approach should be equivalent between the data sources obtained by OMOP and [Regenstrief Institute](#). However, if a significant divergence becomes necessary and meaningful, it should be discussed here.

Regenstrief Oracle Database: In the OMOP project *per se*, the Regenstrief Institute did not transform data from its numerous supplier institutions' raw data feeds into the

OMOP CDM. Rather, in the OMOP project, the Regenstrief Institute transformed data from *the Regenstrief Oracle database* into the OMOP CDM. The raw data feeds undergo a fair bit of processing before making it into the Regenstrief Oracle database; our OMOP team is therefore downstream.

The raw data feeds undergo pre-processing, which includes data validation, syntactic and semantic standardization, etc., before being “persisted” in the Regenstrief Oracle database.

As a downstream consumer of the data, the Regenstrief OMOP team benefits from these pre-processing steps.

INPC Institutions: Validating the data loaded into the CDM was challenging in part because the CDM does not provide a field for storing the identity of the institution from which the data were derived. There are three specific uses that we envision for such an institution “label”:

- 1) *Comparisons and validation.* It would be easier to compare the CDM data against the INPC Oracle data from whence it came, for validation, troubleshooting, etc.
- 2) *Purges.* It would be much easier to purge data from the CDM when necessary—e.g. there were times during the initial CDM ETL when it would have been desirable to purge all data for a particular institutional source (and to re-start that query thread).
- 3) *Late additions* of institutions’ data (institutions previously not loaded) would be easier.

We do, in fact, prefer that the working version of the CDM does not contain these institutional source “label” data, but having these data available as an intermediate step would facilitate troubleshooting. The data use agreements under which Regenstrief is the custodian of the data restrict leeway for reporting aggregate results containing any comparisons between the data sources. That restriction is why it would be best for the final version of the CDM not to retain the institutional source as a data element. While this may be more of a problem for a multi-faceted regional health information organization like Regenstrief, it will also be important for clinical data sources that are drawing data from multiple ancillary systems.

Dynamic, real-world data: The INPC system is dynamic with new data constantly being added. In some cases this is new, real time data such as laboratory results, but it can also be a medication history which contains data from the previous year or more. The CDM represents a “snap-shot” of these data at a single point in time.

3.0 Source Data Mapping

This section will describe mapping process and ETL conversions of data received from your data into Common Data Model.

3.1 Data Mapping

Describe here how your data are provided, and in what technology (relational database system, SAS files etc.) the CDM will be represented.

3.1.1 TABLE NAME: PERSON

Describe how the Person mapping and transformations are designed.

The field mapping is performed as follows:

We weighed the options for which patients to include carefully. While including all of the more than 10 million patients in the INPC would create a interesting dataset to characterize, it would also pose added complexity.

We also considered including any patients for whom the INPC contained longitudinal claims data but we thought there would be many other large claims datasets that could be characterized in OMOP.

We ultimately elected to include those patients for whom we have the richest longitudinal data (the 9 counties, 2004-2009).

We believe this population will be unique among OMOP's core and distributed partner data sets and will offer a rich environment for examining the methods that rely on clinical data.

Regenstrief's underlying probabilistic global patient ID matching links data obtained from querying the Regenstrief Oracle database on a patient-institution pair basis.

The "global patient ID" service is maintained as a core function of the INPC. The global patient ID is determined using a sophisticated, unsupervised matching algorithm designed for the idiosyncrasies of real-world healthcare data.

Destination Field	Source Field	Applied Rule	Comment
PERSON_ID	Not applicable	This is a generated field in the OMOP database.	
YEAR_OF_BIRTH	RI has a date of birth field.	Extracts the year from date of birth	RI has a table at the person level for this information. The year of birth was rounded off to 1920 if <1920.

Destination Field	Source Field	Applied Rule	Comment
GENDER_CONCEPT_ID	RI has two relevant fields: gender, and an ID which describes what type of gender code is being used by the gender field.	The two RI source fields were mapped to the OMOP gender_concept_id.	RI has a table at the person level for this information.
RACE_CONCEPT_ID	RI has two relevant fields: race, and an ID which describes what type of race code is being used by the race field.	The two RI source fields were mapped to the OMOP race_concept_id.	Quite a few different entities send data to RI. The types of race data that they send, and hence that RI stores, vary. Some have a few letters as values. Others have a wider array of letters as values. Others have numbers as values.
LOCATION_CONCEPT_ID	RI has a field for the patient's (home address) zip code.	Extracts the first three digits	RI has a table at the person level for this information.
SOURCE_PERSON_KEY	RI has a global person ID.		
SOURCE_GENDER_CODE	Please see the above cell re gender_concept_id.	RI put the two RI fields together into one, connected by two carets.	RI has a table at the person level for this information.
SOURCE_LOCATION_CODE	Please see the above cell re location_concept_id.		RI has a table at the person level for this information.
SOURCE_RACE_CODE	Please see the above cell re race_concept_id.	RI put the two RI fields together into one, connected by two carets.	RI has a table at the person level for this information.

3.1.2 TABLE NAME: DRUG_EXPOSURE

Describe how the Drug_Exposure mapping and transformation are designed.

The field mapping is performed as follows:

For the December 2009 iteration (iteration 2) of the RI OMOP CDM, we are mapping HCPSCS procedure codes that refer to the use of medications to the Drug_Exposure table.

Destination Field	Source Field	Applied Rule	Comment
-------------------	--------------	--------------	---------

Destination Field	Source Field	Applied Rule	Comment
DRUG_EXP OSURE_ID	Not applicable	This is a generated field in the OMOP database.	
DRUG_EXP OSURE_STA RT_DATE	There are fill dates in the RI medication data table		RI medication data table
DRUG_EXP OSURE_EN D_DATE	Fill dates and days supply	RI used both the date of last refill and the number of days supplied to generate the end_date.	RI medication data table
PERSON_ID	Not applicable	This is a generated field in the OMOP database.	
DRUG_CON CEPT_ID	RI has two relevant fields: the drug code, and an ID which describes what type of drug code is being used by the drug code field.	<p>For this CDM, all the drug codes available from Regenstrief's source (Oracle) data were NDC codes.</p> <p>RI used OMOP's NDC to RxNorm mapping, followed by OMOP's RxNorm to OMOP Concept mapping.</p> <p>In not necessarily the next iteration of the CDM, but in some future iteration of the CDM, once RI (a group separate from this OMOP project team) migrates additional medication and medication orders data from the Vax to the RI Oracle database, RI will also use an additional mapping sequence to take RI dictionary terms into OMOP concepts (via some combination of GPI, NDC, and RxNorm).</p>	RI medication data table
DRUG_EXP OSURE_TYP E		Based on a similar type field that RI uses in its Oracle database, all of which have the "prescription written" value, RI thus far in the CDM has used OMOP type "2" across the board here, to signify "prescription written." However, please see the comment at right.	Based on the better understanding gained during the OMOP project mapping, RI is going to correct this error in its own database (and in the next iteration of the CDM), to revise the value from "prescription written" to "prescription filled." However, once additional data are migrated from Vax to Oracle, some of the new data (medication orders) will be assigned the "prescription written" value.
STOP_REAS ON	None	None	RI does not have stop reasons.

Destination Field	Source Field	Applied Rule	Comment
REFILLS	RI stores number of refills	Directly translates	RI medication data table
DRUG_QUANTITY	There is an amount given (minimum amount) field.	Directly translates	RI medication data table The data in this field appear to be problematic.
DAYS_SUPPLY	There is field that shows days supply.	Directly translates	RI medication data table
SOURCE_DRUG_CODE	Please see the above cell re drug_concept_id.	RI put the two RI fields together into one, connected by two carets.	

3.1.3 TABLE NAME: CONDITION_OCCURRENCE

Describe how the Condition_Occurrence mapping and transformation are designed.

RI has mapped two different source tables to the CONDITION_OCCURRENCE table. One is a large, multifaceted table that contains diagnoses and many other clinical variables. The second is a table that exclusively stores diagnoses. The latter table for a period of years was the sole “destination” table for incoming healthcare payer diagnoses. To help improve the completeness with which RI Oracle diagnosis data is transferred into the CDM, it was necessary to use both of the source tables described. RI therefore divided the ETL table below into two.

The field mapping (using the diverse table with many clinical variables) is performed as follows:

Destination Field	Source Field	Applied Rule	Comment
CONDITION_OCCURRENCE_ID	Not applicable	This is a generated field in the OMOP database.	
CONDITION_START_DATE	Time of admit	A separate encounter table was tapped into to get this timestamp.	An encounter table separate from the table with many clinical variables was needed here.

Destination Field	Source Field	Applied Rule	Comment
PERSON_ID	Not applicable	This is a generated field in the OMOP database.	
CONDITION_END_DATE	Time of discharge	A separate encounter table was tapped into to get this timestamp.	An encounter table separate from the table with many clinical variables was needed here.
CONDITION_OCCURRENCE_TYPE	RI has two relevant fields: the care setting code, and an ID which describes what type of care setting code is being used by the care setting code field.	<p>The two RI source fields were manually mapped to the OMOP condition occurrence type.</p> <p>e.g., RI mapped the RI locations as follows:</p> <p>Observation as observation Ambulatory as ambulatory Series as outpatient Series (unlimited) as outpatient Doctor referral as doctor referral Unknown as unknown Preadmit as inpatient</p>	<p>An encounter table separate from the table with many clinical variables was needed.</p> <p>RI used OMOP definitions: 18 Inpatient Header 1 48 Outpatient Header 1 and then 200 Observation, null, 1 201 Ambulatory, null, 1 202 Doctor Referral, null, 1 203 Unknown, null, 1 204 Emergency, null, 1</p>
CONDITION_CONCEPT_ID	RI has three relevant fields: 1) the field for the “Question”; 2) the field for the “answer” (or the “value” stored under the “Question”); and 3) a field indicating whether the above are coded (versus text or other).	RI manually reviewed a set of 16,000 “Question” fields to select those to use for diagnoses to transfer into the CDM and into the OMOP vocabulary..	This process tapped into the table with the many clinical variables.
STOP_REASON	None	None	RI does not have stop reasons.

Destination Field	Source Field	Applied Rule	Comment
DX_QUALIFIER	Please see the above cell for condition_concept_id.	RI manually defined dx_qualifier as needed (to distinguish the different "Question" fields referred to two rows above.	
SOURCE_CONDITION_CODE	Please see the above cell for condition_concept_id.		As noted above, this procedure tapped into the table with the many clinical variables.

The field mapping (using the table that exclusively stores diagnoses) is performed as follows:

Destination Field	Source Field	Applied Rule	Comment
CONDITION_OCCURRENCE_ID	Not applicable	This is a generated field in the OMOP database.	
CONDITION_START_DATE	Time of admit	A separate encounter table was tapped into to get this timestamp.	An encounter table separate from the table with of diagnoses was needed here.
PERSON_ID	Not applicable	This is a generated field in the OMOP database.	
CONDITION_END_DATE	Time of discharge	A separate encounter table was tapped into to get this timestamp.	An encounter table separate from the table with of diagnoses was needed here.

Destination Field	Source Field	Applied Rule	Comment
CONDITION_OCCURRENCE_TYPE	RI has two relevant fields: the care setting code, and an ID which describes what type of care setting code is being used by the care setting code field.	The two RI source fields were manually mapped to the OMOP condition occurrence type. e.g., RI mapped the RI locations as follows: Observation as observation Ambulatory as ambulatory Series as outpatient Series (unlimited) as outpatient Doctor referral as doctor referral Unknown as unknown Preadmit as inpatient	An encounter table separate from the table of diagnoses was needed. RI used OMOP definitions: 18 Inpatient Header 1 48 Outpatient Header 1 and then 200 Observation, null, 1 201 Ambulatory, null, 1 202 Doctor Referral, null, 1 203 Unknown, null, 1 204 Emergency, null, 1
CONDITION_CONCEPT_ID	RI has two relevant fields: the diagnosis code, and an ID which describes what type of diagnosis code is being used by the diagnosis code field.	These are ICD9 codes and are mapped to OMOP's ICD9 codes.	This step uses the table that exclusively stores diagnoses.
STOP_REASON	None	None	RI does not have stop reasons.
DX_QUALIFIER	Please see the above cell for condition_concept_id.	For all diagnoses extracted from the table that exclusively stores diagnoses, RI entered "major diagnosis" here.	
SOURCE_CONDITION_CODE	Please see the above cell for condition_concept_id.	Please see the above row re condition_concept_id.	This step uses the table that exclusively stores diagnoses.

3.1.4 TABLE NAME: VISIT_OCCURRENCE

Describe how the Visit_Occurrence mapping and transformation are designed.

The field mapping is performed as follows:

Destination Field	Source Field	Applied Rule	Comment
VISIT_OCCURRENCE_ID	Not applicable	This is a generated field in the OMOP database.	
VISIT_START_DATE	Time of admit	An encounter table was tapped into to get this timestamp.	An encounter table was tapped into here.
VISIT_END_DATE	Time of discharge	An encounter table was tapped into to get this timestamp.	An encounter table was tapped into here.
PERSON_ID	Not applicable	This is a generated field in the OMOP database.	
VISIT_CONCEPT_ID	RI has two relevant fields: the care setting code, and an ID which describes what type of care setting code is being used by the care setting code field.	The two RI source fields were manually mapped to the OMOP condition occurrence type. e.g., RI mapped the RI locations as follows: Observation as observation Ambulatory as ambulatory Series as outpatient Series (unlimited) as outpatient Doctor referral as doctor referral Unknown as unknown Preadmit as inpatient	An encounter table was tapped into here. RI used OMOP definitions: 18 Inpatient Header 1 48 Outpatient Header 1 and then 200 Observation, null, 1 201 Ambulatory, null, 1 202 Doctor Referral, null, 1 203 Unknown, null, 1 204 Emergency, null, 1
SOURCE_VISIT_CODE	Please see the above cell re visit_concept_id.	RI put the two RI fields together into one, connected by two carets. (the care setting code, and the ID which describes what type of care setting code is being used by the care setting code field)	An encounter table was tapped into here.

3.1.5 TABLE NAME: PROCEDURE_OCCURRENCE

Describe how the Procedure_Occurrence mapping and transformation are designed.

The field mapping is performed as follows:

An example of the type of informal validation we have done is to examine the frequencies with which procedures are stored in the Regenstrief Oracle database, by INPC institution. These explorations of descriptive data help inform our understanding of what “should” appear in the CDM.

Destination Field	Source Field	Applied Rule	Comment
PROCEDURE_OCCURRENCE_ID	Not applicable	This is a generated field in the OMOP database.	
PROCEDURE_DATE	There is a time stamp in the RI Oracle procedure data table.		There is an RI Oracle procedure data table.
PERSON_ID	Not applicable	This is a generated field in the OMOP database.	
PROCEDURE_CONCEPT_ID	RI has two relevant fields: the procedure code, and an ID which describes what type of procedure code is being used by the procedure code field.	RI mapped its field pairs (please see the cell at left) into OMOP's ICD9 procedure codes or OMOP's CPT codes.	RI Oracle procedure data table.
SOURCE_PROCEDURE_CODE	Please see the above cell re procedure_concept_id.	RI put the two RI fields together into one, connected by two carets.	RI Oracle procedure data table.
PROCEDURE_OCCURRENCE_TYPE	Procedure priority level, and please see also 3.1.3. Condition	RI used a combination of the procedure priority level and the care setting information described in 3.1.3 above, in order to query the OMOP procedure_occurrence_reference table.	RI Oracle procedure data table for the procedure priority, and the separate RI encounter table for the care setting.

Destination Field	Source Field	Applied Rule	Comment
	Occurrence Type		

3.1.6 TABLE NAME: OBSERVATION

Describe how the Observation mapping and transformation are designed.

The field mapping is performed as follows:

We encountered, particularly in the Regenstrief Oracle table from which much of the RI OMOP Observation table data are derived, a variety of data which could not be accommodated in OMOP’s default length of 68 characters, so we extended the maximum length to 1000 characters.

If the field length is too short, clinical findings may be truncated. OMOP might consider including a distribution of lengths found in the observation table to help choose an appropriate maximum length empirically.

For the December 2009 iteration (iteration 2) of the RI OMOP CDM, we have added family history elements to the Observation table.

We have also examined the set of unmapped local dictionary terms (after applying as many of the mappings as we could find in the Regenstrief concept mapping table and in other available standards) to verify that few of the unmapped observation data terms seem relevant to the HOIs. Those few that do appear to have possible relevance we are continuing to take a look at for possible mapping in the iteration 2 CDM. We also are going to include the unmapped observation data codes in the CDM, for possible future mapping if the CDM is used in the future (beyond its use in the next 0-4 months of the OMOP project).

Destination Field	Source Field	Applied Rule	Comment
OBS_OCCURRENCE_ID	Not applicable	This is a generated field in the OMOP database.	
PERSON_ID	Not applicable	This is a generated field in the OMOP database.	

Destination Field	Source Field	Applied Rule	Comment
SOURCE_OBS_CODE	RI used two relevant fields: 1) the RI source field for the Observation "Question"; and 2) a field indicating what type of "Question" it is (local dictionary, LOINC code, etc.).	The two items at left, connected by two carets.	This process tapped into the diverse table with the many clinical variables.
OBS_CONCEPT_ID	Please see the above row for the two fields.	RI used its own concept mapping table in Oracle to map as many of the "Question" codes as possible to other standard code "languages" that exist among OMOP's "languages." RI then uses OMOP's concept table to map to the OMOP concept ID. Any RI concepts for which there was not a readily-on-hand map to a "language" that RI and OMOP already have in common have not been mapped to the CDM, at least for this iteration.	The diverse table with the many clinical variables.
OBS_VALUE_AS_NUMBER	Numeric value	Directly translates	The diverse table with the many clinical variables.
OBS_DATE	There is a well-populated timestamp in this RI table.	Directly translates	The diverse table with the many clinical variables.
OBS_RANGE_LOW	RI has a single string normal range field which contains both the low and the high.	RI parsed the RI single string into high and low.	The diverse table with the many clinical variables.
OBS_RANGE_HIGH	RI has a single string normal range field which contains both the low and the high.	RI parsed the RI single string into high and low.	The diverse table with the many clinical variables.

Destination Field	Source Field	Applied Rule	Comment
OBS_TYPE	Please see the cell a few rows above regarding source_obs_code	RI examined its “Question” code types and mapped those into OMOP’s obs_type.	The diverse table with the many clinical variables.
OBS_VALUE_AS_STRING	Text value that gets “displayed” in RI systems		The diverse table with the many clinical variables.
OBS_VALUE_AS_CONCEPT_ID	Value_if_type_is_coded_code, value_if_type_is_coded_sys_id	As a few rows above for OBS_CONCEPT_ID, RI did the same for OBS_VALUE_AS_CONCEPT_ID	The diverse table with the many clinical variables.
OBS_UNITS_CONCEPT_ID	For each “Question” code in the RI concept table, there is (unless it’s missing) a unit code and a code indicating what type of code the unit code is.	RI examined the units and units type, and checked whether they mapped (in RI’s own database’s concept mapping) to unit types that exist in OMOP. If yes, RI mapped those units to the OMOP units concept ID.	The diverse table with the many clinical variables.

3.1.7 TABLE NAME: OBSERVATION_PERIOD

Describe how the Observation_Period mapping and transformation are designed.

The field mapping is as follows:

Destination Field	Source Field	Applied Rule	Comment
OBSERVATION_PERIOD_ID	Not applicable	This is a generated field in the OMOP database.	
OBSERVATION_PERIOD_START_DATE	Time of admit	Earliest {time of admit} ever, across INPC institutions. For this reason, this table and the era tables were constructed only after all the other OMOP CDM tables were constructed.	Encounter table and the RI medication data table
OBSERVATION_PERIOD_END_DATE	Time of discharge	Latest {time of discharge} ever, across INPC institutions. For this reason, this table and the era tables were constructed only after all the other OMOP CDM tables were constructed.	Encounter table and the RI medication data table

Destination Field	Source Field	Applied Rule	Comment
PERSON_ID	Not applicable	This is a generated field in the OMOP database.	
PERSON_STAT US_CONCEPT_ID	Timestamp of death (see cell at right)	For a particular patient, RI looked for a timestamp of death. No timestamp of death implied that the patient was alive.	The table with the many clinical variables.
RX_DATA_AVAILABILITY		RI generated a second record, for this table, for each patient who ever had any medication data stored in the INPC institutions (that is, those that were used in this project). This second record used the period from earliest to latest prescription fill, and set rx_data_availability to “Y.”	RI medication data table

3.2 Source Independent Data Mapping

The following mapping processes ought to work independent of the source feed. Describe here if significant changes have to be made.

Unless otherwise specified in the sections below, Source Independent Data Mapping will follow specifications as defined in ETL Mapping Specification document.

3.2.1 TABLE NAME: DRUG_ERA

All Drug Eras are recorded in the DRUG_ERA table based on the following field mapping:

We did some informal validation here, comparing the number of unique drug eras produced when the persistence window was varied (e.g. from 30 to 31 days). We also clarified with the OMOP team an era-definition question with regard to the timing of the subsequent prescription in any series of two, and we are making sure to generate the eras by ingredient rather than by more “granular” medication data records.

Destination Field	Source Field	Applied Rule	Comment
DRUG_ERA_ID	Not applicable	This is a generated field in the OMOP database.	
DRUG_ERA_START_DATE	Drug_exposure_start_date from 3.1.2	For the same person-id and same drug_concept_id, aggregate all the drug_exposure_start_dates together, find the earliest one.	Drug_exposure table 3.1.2

Destination Field	Source Field	Applied Rule	Comment
DRUG_ERA_END_DATE	Drug_exposure_end_date from 3.1.2	For the same person-id and same drug_concept_id, aggregate all the drug_exposure_end_dates together, find the latest one.	Drug_exposure table 3.1.2
PERSON_ID	Not applicable	This is a generated field in the OMOP database.	
DRUG_EXPOSURE_TYPE		RI is using a persistence window of 30 days, which has drug_exposure_type as 7. After we run the persistence window 30 day table, we are also running this table a second time, to generate an additional data table, with persistence window 0 days.	
DRUG_CONCEPT_ID	Drug_concept_id	The drug_concept_id from the drug_exposure table 3.1.2	Drug_exposure table 3.1.2
DRUG_EXPOSURE_COUNT		Count of the number of records aggregated from the drug_exposure table	Drug_exposure table 3.1.2

3.2.2 TABLE NAME: CONDITION_ERA

Condition Era table is constructed through an aggregation of individual Condition Occurrences recorded in the CONDITION_OCCURRENCE table.

All Condition Eras are recorded in the CONDITION_ERA table based on the following field mapping:

Destination Field	Source Field	Applied Rule	Comment
CONDITION_ERA_ID	Not applicable	This is a generated field in the OMOP database.	
CONDITION_ERA_START_DATE	Condition_start_date	For the same person-id and same condition_concept_id, aggregate all the records together and find the earliest condition_start_date	Condition_occurrence table 3.1.3
PERSON_ID	Not applicable	This is a generated field in the OMOP database.	
CONFIDENCE		RI is not using this optional field.	RI is not using this optional field.
CONDITION_ERA_END_DATE	Condition_end_date	For the same person-id and same condition_concept_id, aggregate all the records together and find the latest	Condition_occurrence table 3.1.3

Destination Field	Source Field	Applied Rule	Comment
		condition_end_date	
CONDITION_CONCEPT_ID	Condition_concept_id	Condition_occurrence table 3.1.3	Condition_occurrence table 3.1.3
CONDITION_OCCURRENCE_TYPE		RI is using a persistence window of 30, which has condition_occurrence_type "65".	
CONDITION_OCCURRENCE_COUNT		Count of records aggregated by same person, same condition_concept_id	Condition_occurrence table 3.1.3

3.3 Reference Tables

The following contain reference tables that were derived from the OMOP Thomson and GE source data. They reflect the content of those databases. It is assumed that you will update these tables to describe your data more adequately.

3.3.1 TABLE NAME: DRUG_EXPOSURE_REF

Drug Exposure Types are used to define the indicators from which exposures have been extracted. They also define the characteristics of the exposure and the level of aggregation. The following Drug Exposure Types are allowed.

Drug Exposure Type	Drug Exposure Type Description	Persistence Window (In Days)
1	Prescription Dispensed	
2	Prescription Written	
3	Medication List	
4	Physician Administered Drug (Identified as Procedure)	
5	Inpatient Administration	
6	Drug Era – 0 day window	0
7	Drug Era – 30 days window	30

3.3.2 TABLE NAME: CONDITION_OCCURRENCE_REF

Condition Occurrence Reference table serves as the reference listing of various types of Condition Occurrences recorded for analysis. The Condition Occurrence Type conveys

the indicator(s) from which the Condition Occurrence was captured and defines the characteristic of the occurrence and the level of aggregation.

This table is loaded based on a reference list of Occurrence types, descriptions and persistence window settings. The current listing is as follows:

Condition Occurrence Type	Condition Occurrence Type Description	Condition Occurrence Position	Persistence Window (in days)
1	Inpatient Detail	Primary	
2	Inpatient Detail	1	
3	Inpatient Detail	2	
4	Inpatient Detail	3	
5	Inpatient Detail	4	
6	Inpatient Detail	5	
7	Inpatient Detail	6	
8	Inpatient Detail	7	
9	Inpatient Detail	8	
10	Inpatient Detail	9	
11	Inpatient Detail	10	
12	Inpatient Detail	11	
13	Inpatient Detail	12	
14	Inpatient Detail	13	
15	Inpatient Detail	14	
16	Inpatient Detail	15	
17	Inpatient Header	Primary	
18	Inpatient Header	1	
19	Inpatient Header	2	
20	Inpatient Header	3	
21	Inpatient Header	4	
22	Inpatient Header	5	
23	Inpatient Header	6	
24	Inpatient Header	7	
25	Inpatient Header	8	
26	Inpatient Header	9	

Condition Occurrence Type	Condition Occurrence Type Description	Condition Occurrence Position	Persistence Window (in days)
27	Inpatient Header	10	
28	Inpatient Header	11	
29	Inpatient Header	12	
30	Inpatient Header	13	
31	Inpatient Header	14	
32	Inpatient Header	15	
33	Outpatient Detail	1	
34	Outpatient Detail	2	
35	Outpatient Detail	3	
36	Outpatient Detail	4	
37	Outpatient Detail	5	
38	Outpatient Detail	6	
39	Outpatient Detail	7	
40	Outpatient Detail	8	
41	Outpatient Detail	9	
42	Outpatient Detail	10	
43	Outpatient Detail	11	
44	Outpatient Detail	12	
45	Outpatient Detail	13	
46	Outpatient Detail	14	
47	Outpatient Detail	15	
48	Outpatient Header	1	
49	Outpatient Header	2	
50	Outpatient Header	3	
51	Outpatient Header	4	
52	Outpatient Header	5	
53	Outpatient Header	6	
54	Outpatient Header	7	
55	Outpatient Header	8	

Condition Occurrence Type	Condition Occurrence Type Description	Condition Occurrence Position	Persistence Window (in days)
56	Outpatient Header	9	
57	Outpatient Header	10	
58	Outpatient Header	11	
59	Outpatient Header	12	
60	Outpatient Header	13	
61	Outpatient Header	14	
62	Outpatient Header	15	
63	Problem List		
64	Condition Era		0
65	Condition Era		30
66	Death at Discharge		
Please see above in 3.1.3. (page 8) re the 200 series			

3.3.3 TABLE NAME: PROC_OCCURRENCE_REF

Procedure Occurrence Reference table serves as the reference listing of various types of Procedure Occurrences recorded for analysis. The Procedure Occurrence Type conveys the indicator(s) from which the Procedure Occurrence was captured, and defines the characteristic of the occurrence.

This table is loaded based on a reference list of occurrence types, position and descriptions. The current listing is as follows:

Procedure Occurrence Type	Procedure Occurrence Type Description	Procedure Occurrence Position
1	Inpatient Detail	Primary
2	Inpatient Detail	1
3	Inpatient Header	Primary
4	Inpatient Header	1
5	Inpatient Header	2
6	Inpatient Header	3

Procedure Occurrence Type	Procedure Occurrence Type Description	Procedure Occurrence Position
7	Inpatient Header	4
8	Inpatient Header	5
9	Inpatient Header	6
10	Inpatient Header	7
11	Inpatient Header	8
12	Inpatient Header	9
13	Inpatient Header	10
14	Inpatient Header	11
15	Inpatient Header	12
16	Inpatient Header	13
17	Inpatient Header	14
18	Inpatient Header	15
19	Outpatient Detail	Primary
20	Outpatient Detail	1
21	Outpatient Header	Primary
21	Outpatient Header	1
22	Outpatient Header	2
23	Outpatient Header	3
24	Outpatient Header	4
25	Outpatient Header	5
26	Outpatient Header	6
27	EHR Order	

3.3.4 TABLE NAME: OBSERVATION_TYPE_REF

Assignment of an Observation type is essential to determine the type of source data, level of standardization, and coding, as well as the type of result recorded for the Observation. The Observation Types include the following.

- Lab Observation Numeric Result
- Lab Observation Text
- Lab Observation Concept Code Result

- Numeric Observations from EHRs (e.g., blood pressure). These are tracked separately and not rolled into other Lab Observation categories
- EHR observations with text results (e.g., reason for visit)
- Chief Complaint

Data in the OBSERVATION_TYPE_REF table is as follows:

Observation_Type	Observation_Type_Desc
LON	Lab Observation Numeric Result
LOT	Lab Observation Text
LOC	Lab Observation Concept Code Result
HER	Observation recorded from Electronic Health Records
TEM	Observation recorded from Electronic Health Records with text results
CHC	Chief Complaint

We have added an “FMH” (Family Medical History) in the above table.

3.3.5 TABLE NAME: VOCABULARY_REF

The Vocabulary Reference entity includes a list of all standard terminologies from which Concepts have been extracted for observational analysis using the Common Data Model. The reference table is populated with a single record for each Vocabulary source and includes a descriptive name for the Vocabulary source.

Data in the VOCABULARY_REF table is as follows:

VOCABULARY_CODE	VOCABULARY_NAME
01	SNOMED
02	ICD9 CM
03	ICD9 Procedure
04	CPT
05	HCPCS
06	LOINC
07	NDFRT

VOCABULARY_CODE	VOCABULARY_NAME
08	RxNorm
09	NDC
52	THOMSON
51	GE
15	MedDRA
10	GPI
54	OMOP Intermediate Concept – Drug
55	OMOP Generic
53	OMOP Intermediate Concept-Condition
11	UCUM
12	HL7 ADMINISTRATIVE SEX
13	CDC RACE/ETHNICITY
14	CMS PLACE of SERVICE

3.3.6 TABLE NAME: RELATIONSHIP_TYPE

A Concept Relationship is standardized via the Relationship Type entity. The Relationship Type codes are adopted from SNOMED-CT. Where the relationships are hierarchical, the Relationship Type captures the “IS A” string that identifies it as a Subtype relationship. Where the relationship is an Object Attribute Value relationship, the Relationship Type holds the Concept that defines the Attribute.

Data in the RELATIONSHIP_TYPE table is as follows:

RELATIONSHIP_TYPE	RELATIONSHIP_DESCRIPTION
001	LOINC Map To
002	RXNORM Has precise ingredient
003	RXNORM Has tradename
004	RXNORM Has dose form
005	RXNORM Has form
006	RXNORM Has ingredient
007	RXNORM Constitutes

RELATIONSHIP_TYPE	RELATIONSHIP_DESCRIPTION
008	RXNORM Contains
009	RXNORM Reformulation of
010	Subsumes
011	NDFRT Has DoseForm
012	NDFRT Induces
013	NDFRT May Diagnose
014	NDFRT Has PE
015	NDFRT CI PE
016	NDFRT Has Ingredient
017	NDFRT CI ChemClass
018	NDFRT Has MoA
019	NDFRT CI MoA
020	NDFRT Has PK
021	NDFRT May Treat
022	NDFRT CI With
023	NDFRT May Prevent
024	NDFRT Has Active Metabolites
025	NDFRT Site of Metabolism
026	NDFRT Effect May Be Inhibited By
027	NDFRT Has Chemical Structure
028	NDFRT RXN RELA
120	MedDRA Has Hierarchy Level
121	MedDRA Has System Organ Class
122	MedDRA Has High Level Group Term
123	MedDRA Has High Level Term
124	MedDRA Has Preferred Term
101	OMOP Intermediate Condition Concept To SNOMED
102	OMOP Intermediate Drug Concept To RxNorm
041	Indirect morphology

RELATIONSHIP_TYPE	RELATIONSHIP_DESCRIPTION
072	Procedure site – Direct
060	Scale type
093	CPT – SNOMED
074	Procedure device
032	Pathological process
045	Has intent
050	Episodicity
037	Occurrence
056	Associated morphology
042	Indirect device
030	Procedure site
091	Hli ICD9CM Procedure to SNOMED Category
083	Using substance
080	Surgical approach
043	Has specimen
089	Hli ICD-9-CM to SNOMED Category
049	Finding site
054	Component
040	Interprets
039	Laterality
038	Method
068	Subject relationship context
062	Specimen procedure
036	Access
065	Specimen source topography
085	Clinical course
087	Finding method
064	Specimen source morphology
048	Has active ingredient

RELATIONSHIP_TYPE	RELATIONSHIP_DESCRIPTION
058	Measurement Method
044	Has interpretation
070	After
076	Finding context
053	Direct device
071	Associated procedure
073	Procedure site - Indirect
063	Specimen source identity
092	Hli ICD9CM Procedure to SNOMED Specific
035	Revision status
079	Associated with
046	Has focus
031	Priority
033	Part of
075	Procedure morphology
029	Recipient category
061	Time aspect
082	Using energy
057	Associated finding
069	Has dose form
067	Due to
047	Has definitional manifestation
088	Finding informer
094	CPT EQUAL SNOMED
059	Property
055	Causative agent
077	Procedure context
051	Direct substance
090	Hli ICD-9-CM to SNOMED Specific

RELATIONSHIP_TYPE	RELATIONSHIP_DESCRIPTION
078	Temporal context
034	Severity
084	Using access device
086	Route of administration
052	Direct morphology
081	Using device
066	Specimen substance

Additional notes re “informal” validation activities (beyond those described above with some of the specific tables):

We performed various spot-checks and cross-checks of source data tables vs. CDM database tables.

In some cases (e.g., for procedure codes), this process revealed that the source data dictionary needed an update.

In another case (for the observation table), this process showed that the way that Oracle itself handles {null} in the SQL statement that deals with “if not {1, 2, or 3}” is illogical: Oracle was not treating {null} as being not {1, 2, 3}. (This issue arose after 1) we divided our {large, multifaceted table that contains diagnoses and many other clinical variables} into three “chunks” (diagnoses, drugs, and observations -- the latter being everything that is not a diagnosis or a drug), and 2) one of the source data institutions turned out to have a great many null values as observation values -- a great many such that the discrepancy was clearly manifested.)

We also did some initial runs of OSCAR and found some records with start date > end date (though not an inordinate number of such records -- as you know such anomalies are to be expected in a small percentage of cases in large databases).