FISABIO-HSRP OMOP ETL design v3.1

AUTHOR
Francisco Sanchez-Saez

PUBLISHED
June 25, 2024

DOCUMENT HISTORY

Table 1: version changelog

Version Number	Date	Changes
1.0	2023-02-23	
1.1	2023-04-06	1. The pipeline of FISABIO-HSRP unit for the ETL VID-OMOP has been updated. The curation of the source tables before the transformation process has been enhanced. Additionally, a source data quality check has been implemented to distinguish between potential issues in the source data and potential issues in the transformation to the OMOP CDM process. The following changes have been made: ◦ The **GAIA** table, which is a processed table, has been redesigned, and the source tables that make it up (**pres**, **fact**, **rele**, and **tx**) have been added. ◦ The **CONG** table, which contains information about congenital anomalies, has been added. ◦ A new column, *mandatory*, indicating if it is mandatory to extract a variable in the source table, has been added. 2. Some variable names of the source tables have been updated in the data dictionary and the data mapping section: ◦ In **CEX**: ▪ *fecha* has been updated to *fecha_consulta*. ◦ In **AED**: ▪ *momento_reg* has been updated to *fecha_registro*. ▪ *momento_alta* has been updated to *fecha_alta*. ◦ In **EOS**: ▪ *id_embarazo* has been updated to *embarazo_id*. ▪ *fecha_fin* has been updated to *fecha_fin_emb*. ▪ *tipo_fin_rn1* has been updated to *resultado_rn1*. ▪ *tipo_fin_rn2* has been updated to *resultado_rn2*. ▪ *tipo_fin_rn3* has been updated to *resultado_rn3*.

An asterisk (*) next to the version number indicates that this version has an associated release, with the version date as the identifier.

Version Number	Date	Changes
		○ In **MDR**: ■ *sip* has been updated to *sip_madre*. 3. The figure in Appendix 1 has been modified.
1.2	2023-05-08	1. The diagnoses collected in **CEX** and **AED** can be mixed in the same observation, i.e., some ICD10ES and some ICD9CM codes. As a result, *tipo_codigo* has been split into *tipo_codigo1*, *tipo_codigo2*, *tipo_codigo3*, and *tipo_codigo4* for the **CEX** table, and *tipo_codigo1* and *tipo_codigo2* for the **AED** table. 2. The graphs of the mapping between the source and target tables have been updated according to the changes collected in v.1.1 and v.1.2.
1.3	2023-07-08	1. Fixed an issue in the description of **CONG** and **SIV** tables in Appendix 2. 2. The data mapping has been updated with some changes in the origin tables: ○ In the **SIP** (curated) table, the *sexo* variable now has the categories 'Female', 'Male', and 'Unknown' instead of the previous 'M', 'H', and 'U'. 3. The format of the document has been improved, following specific conventions: ○ Table names should be in **bold**. ○ Column/variable names should be in *italics*. ○ Category names should be enclosed in single quotation marks ' '. 4. The version changelog table has been named and enumerated. 5. The information in the '**From sip to person**' specifications has been updated (Table 2 and Figure 3) to include the variable *dpto_salud* from the origin table. 6. The information in the Target table: **care_site** has been updated with new specifications regarding *care_site_name* and *care_site_source_value*. 7. Changed the specification of 'NULL' to 'NA_character_', 'NA_Date_', or 'NA_integer_' as these are the values used in R. 8. The intermediate tables **pcv_to_visit_occurrence**, **cex_to_visit_occurrence**, **mbds_to_visit_occurence**, and **aed_to_visit_occurrence** has been renamed to **from_pcv_to_visit_occurrence**, **from_cex_to_visit_occurrence**, **from_mbds_to_visit_occurence**, and **from_aed_to_visit_occurrence** respectively. 9. In the '**From sip to observation_period**' specifications, the *period_type_concept_id* has been changed from '32827: EHR encounter record' to '32828: EHR episode record'.

An asterisk (*) next to the version number indicates that this version has an associated release, with the version date as the identifier.

Version Number	Date	Changes
		10. In the '**From mbds to visit_occurrence**' and '**From aed to visit_occurrence**' specifications, the comments for *admitted_from_concept_id*, and *discharged_to_concept_id* have been enhanced. 11. The ICD9 and ICD10 code systems have been updated to the more precise names ICD9CM and ICD10ES, respectively. 12. A new mapping has been added to create pregnancy eras in the '**From MDR to condition_occurrence**' subsection. 13. The specifications of '**From eos to condition_occurrence**' have been enhanced. 14. In the '**From mbds to procedure_occurrence**' specifications, the *modifier_source_value* has been changed from 'NA_character_' to 'MBDS'. 15. In the '**From PMR to death**' subsection, a note has been added to indicate that the ETL process for this table is not currently possible. 16. The specifications of '**From gaia to drug_exposure**' have been enhanced. 17. The **fact_relationship** table cannot be populated because we don't have information about the sex of the children, which is required to create their person observations. A note has been added in that section to highlight this limitation. 18. After performing the DQD analysis, the values of certain variables have been modified from NA_integer_ or NA_real to 0 in order to comply with the completeness tests. 19. After performing the DQD analysis, the values of certain 'datetime' variables have been modified from NA_Date_ to the Date value. 20. The expected timeline for milestone 3 has been revised from 'month 10' to 'month 11' in the **NEXT STEPS** section.
1.4	2023-08-18	1. The fields of the **measurement** table have been reordered according to the order of the Specification Document for the OMOP Common Data Model, v5.4.
2.0	2023-09-18	Major update incorporating recommendations from the SME (IOMED) to enhance the ETL process. The following significant changes have been implemented: 1. The content of the **condition_occurrence**, **procedure_occurrence**, and **drug_exposure** tables, whose standardized concepts domain was different from their respective table, has been relocated. To achieve this, the target tables **device_exposure** and **observation** have been created. 2. The target table **source_to_concept_map** has been created to store the source-to-concept mapping.

An asterisk (*) next to the version number indicates that this version has an associated release, with the version date as the identifier.

Version Number	Date	Changes
		3. The derived tables **condition_era** and **drug_era** have been created.
2.1 *	2023-09-25	1. An error in the calculation of *drug_exposure_end_datetime*, and *drug_exposure_end_datetime* has been corrected. 2. A typo in the definition of the type of vocabulary has been corrected, resulting in ICD10 codes now being mapped from the AED and CEX tables.
2.2	2023-10-18	corrected **MBDS** Fields in Appendix 2.
3.0 *	2024-05-12	1. The observation periods have been improved, merging periods for the same person where the next period starts on the same day as the previous one ended. 2. Vocabulary update to version August 2023 (preferred for DARWIN studies). 3. Update of some terms due to VID catalogue update to version 2.0: ○ Renamed variable *fecha_nac_hijo* to *fecha_fin_emb* in table **MDR**. 4. Major update incorporating recommendations from EHDEN: i. Drugs mapped to Clinical Drug Boxes or Quantified Clinical Drugs. ii. CPT4 vocabulary downloaded.
3.1 *	2024-06-14	1. The **cdm_source** table information is appended to this document at the end of Part I of the DATA MAPPING section. Note that the github repository has been renamed from 'OHDSI-VID_to_OMOP_ETL' to 'VID2OMOP' and the has been changed from 756265 (v.5.4.0) to 798878 (v.5.4.1). 2. The MILESTONES section has been removed as the EHDEN project has been completed. 3. Blood pressure measurements are mapped. 4. In the version column, if there is a corresponding release, it is marked with '*'. 5. The TECHNICAL INFRASTRUCTURE section has been updated to reflect the current process for storing and archiving CDM releases. 6. A description of the **source_to_concept_map** table has been added to Part II of the DATA MAPPING section. 7. The DATA MAPPING section has been restructured to include Part III, which presents the derived tables (**condition_era** and **drug_era**). 8. The MILESTONES section has been removed due to the completion of the EHDEN project. 9. The '**From gaia to drug_exposure**' section has been updated in order to reflect the changes in the drugs mapping. Since version 3.0 the mapping is done using the *pres_farma_desc* and *pres_farma_cod*, which

An asterisk (*) next to the version number indicates that this version has an associated release, with the version date as the identifier.

Version Number	Date	Changes
		is a description of the *Clinical Drug Boxes* or the *Quantified Clinical Drugs*. 10. A note indicating that this table is not yet mapped in the OMOP CDM is added to the **CONG** table in APPENDIX 2. 11. APPENDIX 1 is updated to include the **TESTS** table. 12. The VOCABULARY MAPPING section has been rewritten to include a subsection called *Vocabularies not available in ATHENA*.
3.2	2024-06-25	1. Calculate *days_supply* as *drug_exposure_end_date* - *drug_exposure_start_date* + 1 since *days supply* is not the same as *env_durac*. Practically, *days supply* is *env_durac* + 1. (THIS IS NOT YET IMPLEMENTED. IT IS COMMENTED IN THE 3.01 SCRIPT)

An asterisk (*) next to the version number indicates that this version has an associated release, with the version date as the identifier.

1 INTRODUCTION

This document describes how VID database is converted to the OMOP Common Data Model (CDM) version 5.4. This is a collaborative effort by the European Health Data and Evidence Network (EHDEN) project and FISABIO-HSRP unit. It describes the definition of the ETL that will be used in the implementation.

The document is elaborated and maintained into a *.qmd (quarto)* file using RStudio. Quarto files can be knitted, through markdown format and pandoc converter into several output formats, as it is represented in the Figure 1. For each relevant version of the document, from the same *.qmd* file, two outputs are generated: a *.docx document* and an *.html document*.

Figure 1: quarto workflow. Source: https://quarto.org/docs/faq/rmarkdown.html

The Valencia Health System Integrated Database (VID) is a set of multiple, public, population-wide electronic databases for the Valencia Region, the fourth most populated Spanish region, with about 5 million inhabitants and an annual birth cohort of 48 000 newborns, representing 10.7% of the Spanish population and around 1% of the European population. The VID provides exhaustive longitudinal information including sociodemographic and administrative data (sex, age, nationality, date of death,

etc.), clinical (diagnoses, procedures, diagnostic tests, imaging, etc.), pharmaceutical (prescription, dispensing) and healthcare utilization data from hospital care, emergency departments, specialized care (including mental and obstetrics care), primary care and other public health services. It also includes a set of associated population databases and registers of significant care areas such as vaccines, cancer, rare diseases, congenital anomalies, metabolic diseases, perinatal mortality, microbiology (including COVID-19 test results register) and others, and also public health databases from the population screening programmes. All the information in the VID databases can be linked at the individual level through a single personal identification code. The databases were initiated at different moments in time, but all in all the VID provides comprehensive individual-level data fed by all the databases from 2008 to date.

In the VID database, each study leads to a different extraction. However, there are a set of bases that are usually used in the projects by FISABIO-HSRP unit. These are: **SIP**, **PCV**, **CEX**, **MBDS**, **AED**, **DIAGNOSES**, **GAIA**, **SIV**, **MDR**, **PMR**, **EOS**, and **REDMIVA**.

2 TECHNICAL INFRASTRUCTURE

2.1 Overview

The ETL is performed in R 4.4.0 on a machine with 64 GB of RAM. Data transformation is performed using *Tidyverse* (*dplyr*) and *data.table* packages. The target tables are stored in *.csv* files during the process. Once the quality of the ETL has been assessed, the tables are stored in a *PostgreSQL* database.

The VID extractions are obtained specifically for each study. Therefore, we will develop the ETL pipeline in a specific study. In this document, the selected study is the Consign study. This is a large study whose extraction contains all the source tables presented in this document for 1 964 588 women from 2018 to 2021.

2.2 Active and archived releases

Regarding to the active and archived releases the architecture is as follows:

There are two databases hosted on the Postgres server. The *n* release and the *n-1* release (or the archived release of interest). Older releases are archived.

The archiving process is triggered when a new CDM is released. A backup of the PostgreSQL database is made using the *pgdump* utility, resulting in a backup of both the database schemas and data. The backup is stored together with the results of the various tools run in connection with this release:

- output_achilles (with the corresponding *condition_treemap.json*).

- output dqd.
- output_catalogue_export.
- output_cdm_onboarding.
- output_dashboard_export.

To access an archived release in the future, the process is as follows: - Remove the n-1 release from the PostgreSQL database to free up space. - Restore the desired archived release using the pgrestore utility. With this architecture we maintain two live releases: the latest and another release of interest.

The releases of the CDM instances and their status in the PostgreSQL server are presented in Table 2.

Table 2: CDM releases

Release	status	Database ID
ETL_VID_CONSIGN_2023_09_25	Archived	(-)
ETL_VID_CONSIGN_2024_05_12	Active	postgres2
ETL_VID_CONSIGN_2024_06_14	Active	postgres

3 DATA MAPPING

The data mapping process consists of three parts. Part I was designed using Rabbit-in-a-Hat and involves mapping from source tables to target tables. Conversely, Part II refines the contents of the target tables created in Part I and add the **source_to_concept_map** table. Finally, Part III presents the derived tables (**condition_era** and **drug_era**).

3.1 Part I Overview

In the Figure 2 is depicted the relationship among the VID source tables and OMOP CDM v 5.4 target tables.

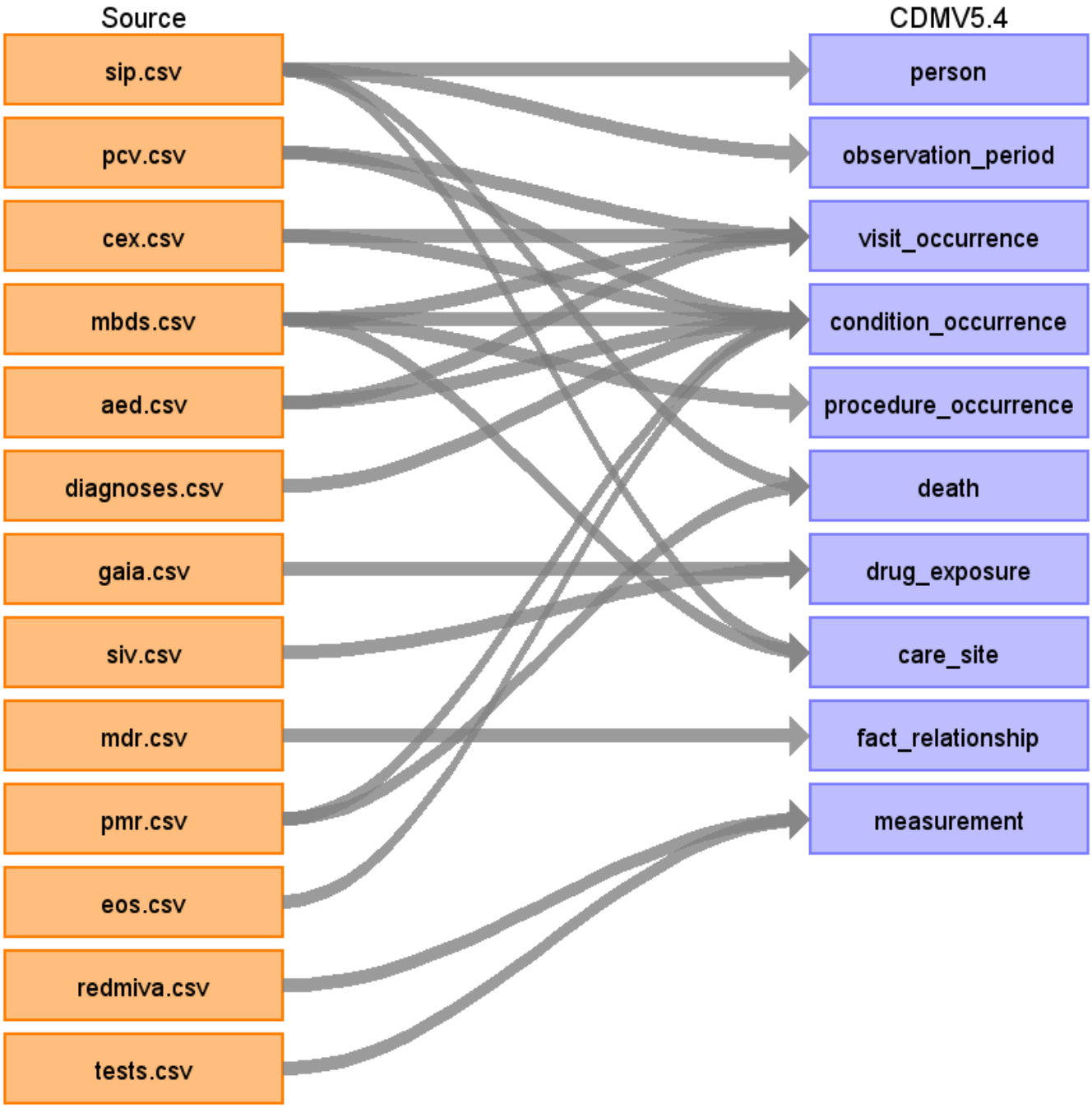


Figure 2: Overview of VID-OMOP CDM v5.4 ETL diagram

3.1.1 Target table: person

The target table **person** is populated with the information from the source table **SIP**.

3.1.1.1 From sip to person

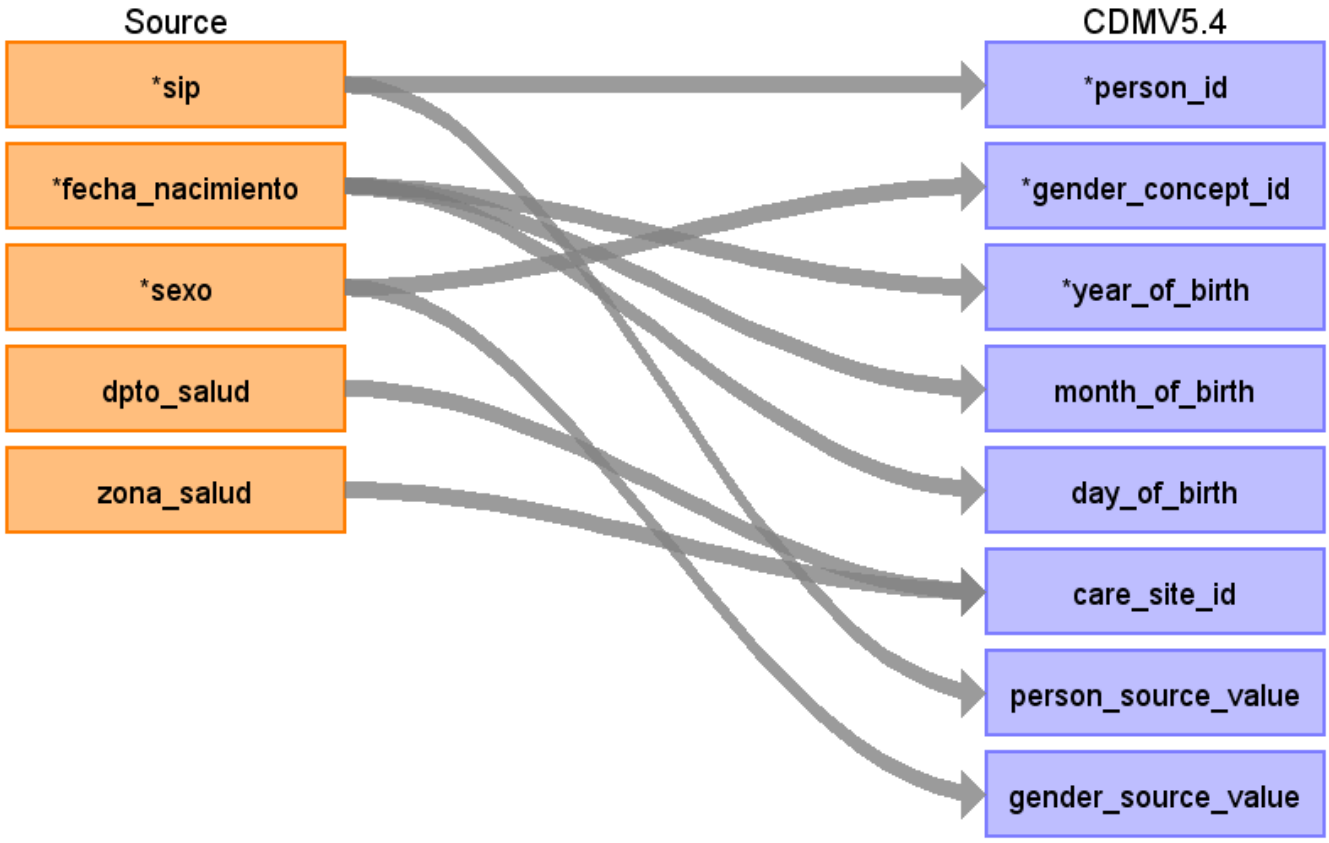


Figure 3: sip to person mapping diagram

Table 3: sip to person mapping description

Destination Field	Source Field	Logic	Comment
person_id	*sip*	Each different pseudonimyzed ID is converted to an integer starting with 1.	Autogenerate.
gender_concept_id	*sexo*	case_when(*sexo* == 'Male' ~ 8507, *sexo* == 'Female' ~ 8532, T ~ NA_integer_)	When *sexo* is 'Male' *gender_concept_id* i Male'. When *sexo* is the *gender_concept_* '8532: Female'. Oth drop person: after t mapping of the *gender_concept_id*, table with: person | filter(!is.na(*gender_c*
year_of_birth	*fecha_nacimiento*	str_sub(*fecha_nacimiento*, 1, 4) |> as.integer()	*fecha_nacimiento* is format variable (YY dd).

Destination Field	Source Field	Logic	Comment
month_of_birth	*fecha_nacimiento*	str_sub(*fecha_nacimiento*, 6, 7) |> as.integer()	*fecha_nacimiento* is format variable (YY dd).
day_of_birth	*fecha_nacimiento*	str_sub(*fecha_nacimiento*, 9, 10) |> as.integer()	*fecha_nacimiento* is format variable (YY dd).
birth_datetime	*fecha_nacimiento*	*fecha_nacimiento*	*fecha_nacimiento* in format (YYYY-mm-d
race_concept_id			0
ethnicity_concept_id			0
location_id			NA_integer_
provider_id			NA_integer_
care_site_id	*dpto_salud*, *zona_salud*	Variables *dpto_salud* and *zona_salud* are converted to a *care_site_id* (see the **care_site** target table).	
person_source_value	*sip*		
gender_source_value	*sexo*		
gender_source_concept_id			0
race_source_value			NA_character_
race_source_concept_id			0
ethnicity_source_value			NA_character_
ethnicity_source_concept_id			0

3.1.2 Target table: observation_period

The target table **observation_period** is populated with the information from the source table **SIP**.

3.1.2.1 From sip to observation_period

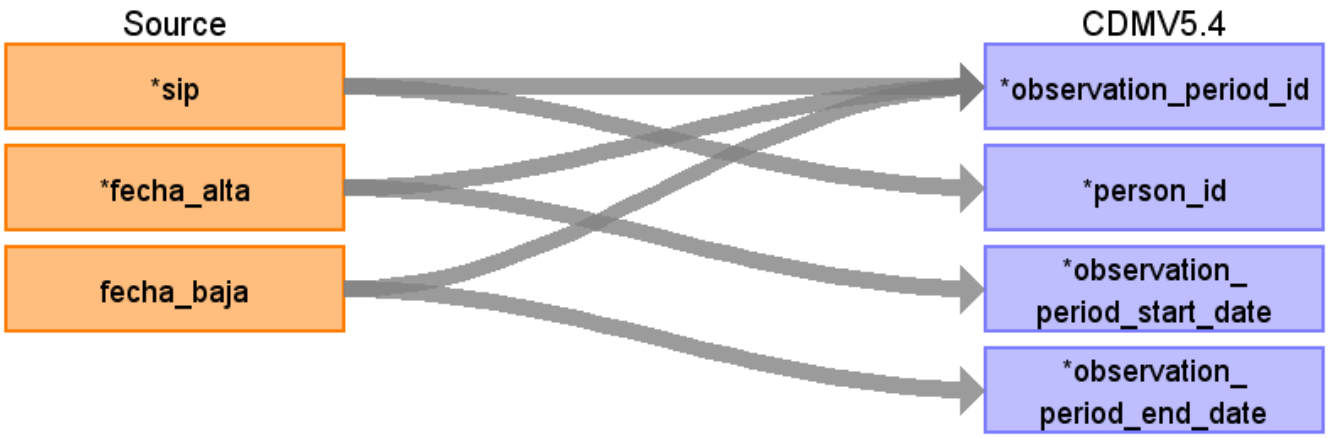


Figure 4: sip to observation_period mapping diagram

Table 4: sip to observation_period mapping description

Destination Field	Source Field	Logic	Comment
observation_period_id	*sip, fecha_alta, and fecha_baja*	Autogenerate. Create a new observation period (integer) for each person and different observation periods (*fecha_alta* to *fecha_baja*).	
person_id	*sip*		
observation_period_start_date	*fecha_alta*		
observation_period_end_date	*fecha_baja*		
period_type_concept_id			All observation periods are obtained from the Population Information System (SIP). The records in **SIP** were created when an encounter is produced (without specifying the type). The category

Destination Field	Source Field	Logic	Comment
			that best fits is '32828: EHR episode record'.

3.1.3 Target table: visit_occurrence

The target table **visit_occurrence** is populated with the information from the source tables **PCV**, **CEX**, **MBDS** and **AED**.

In the mapping process of the visit occurrence, we will create the following intermediate tables with the *visit_occurrence_id* and the *diag_cod* together:

- **from_pcv_to_visit_occurrence**.
- **from_cex_to_visit_occurrence**.
- **from_mbds_to_visit_occurrence**.
- **from_aed_to_visit_occurrence**.

3.1.3.1 From pcv to visit_occurrence

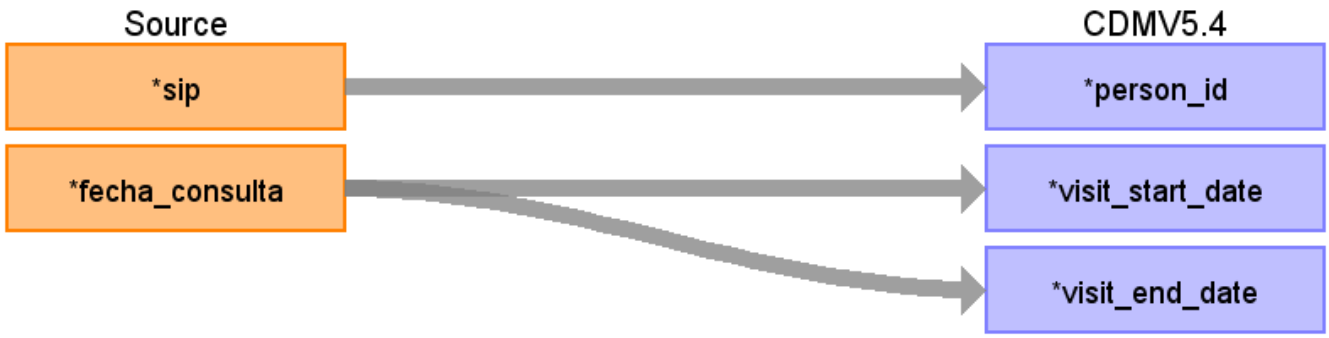


Figure 5: pcv to visit_occurrence mapping diagram

Table 5: pcv to visit_occurrence mapping description

Destination Field	Source Field	Logic	Comment
visit_occurrence_id			Autogenerate: from 1 to n_pcv when source table is **PCV**.
person_id	*sip*		
visit_concept_id			In **PCV** there are primary care visits. The Concept ID is '9202: Outpatient Visit'.
visit_start_date	*fecha_consulta*		
visit_start_datetime	*fecha_consulta*		
visit_end_date	*fecha_consulta*		
visit_end_datetime	*fecha_consulta*		
visit_type_concept_id			In **PCV** there are primary care visits. The Concept ID is '32834: EHR outpatient note'.
provider_id			NA_integer_
care_site_id			NA_integer_
visit_source_value			PCV
visit_source_concept_id			0
admitted_from_concept_id			0
admitted_from_source_value			NA_character_
discharged_to_concept_id			0
discharged_to_source_value			NA_character_
preceding_visit_occurrence_id			Once all the source tables that contribute to visit_occurence are populated, we will order the *visit_occurrence_id* for each *person_id*

Destination Field	Source Field	Logic	Comment
			and complete this field.

3.1.3.2 From cex to visit_occurrence

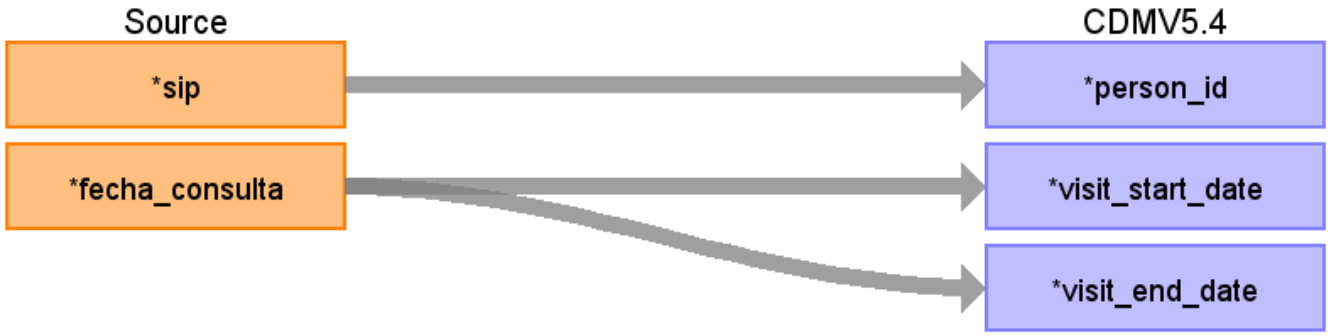


Figure 6: cex to visit_occurrence mapping diagram

Table 6: cex to visit_occurrence mapping description

Destination Field	Source Field	Logic	Comment
visit_occurrence_id			Autogenerate: from n_pcv + 1 to n_pcv + n_cex when source table is **CEX**.
person_id	*sip*		
visit_concept_id			In **CEX** there are specialist care visits. The Concept ID is '9202: Outpatient Visit'.
visit_start_date	*fecha_consulta*		
visit_start_datetime	*fecha_consulta*		
visit_end_date	*fecha_consulta*		
visit_end_datetime	*fecha_consulta*		

Destination Field	Source Field	Logic	Comment
visit_type_concept_id			In **CEX** there are specialist care visits. The Concept ID is '32834: EHR outpatient note'.
provider_id			NA_integer_
care_site_id			NA_integer_
visit_source_value			CEX
visit_source_concept_id			0
admitted_from_concept_id			0
admitted_from_source_value			NA_character_
discharged_to_concept_id			0
discharged_to_source_value			NA_character_
preceding_visit_occurrence_id			Once all the source tables that contribute to visit_occurence are populated, we will order the *visit_occurrence_id* for each *person_id* and complete this field.

3.1.3.3 From mbds to visit_occurrence

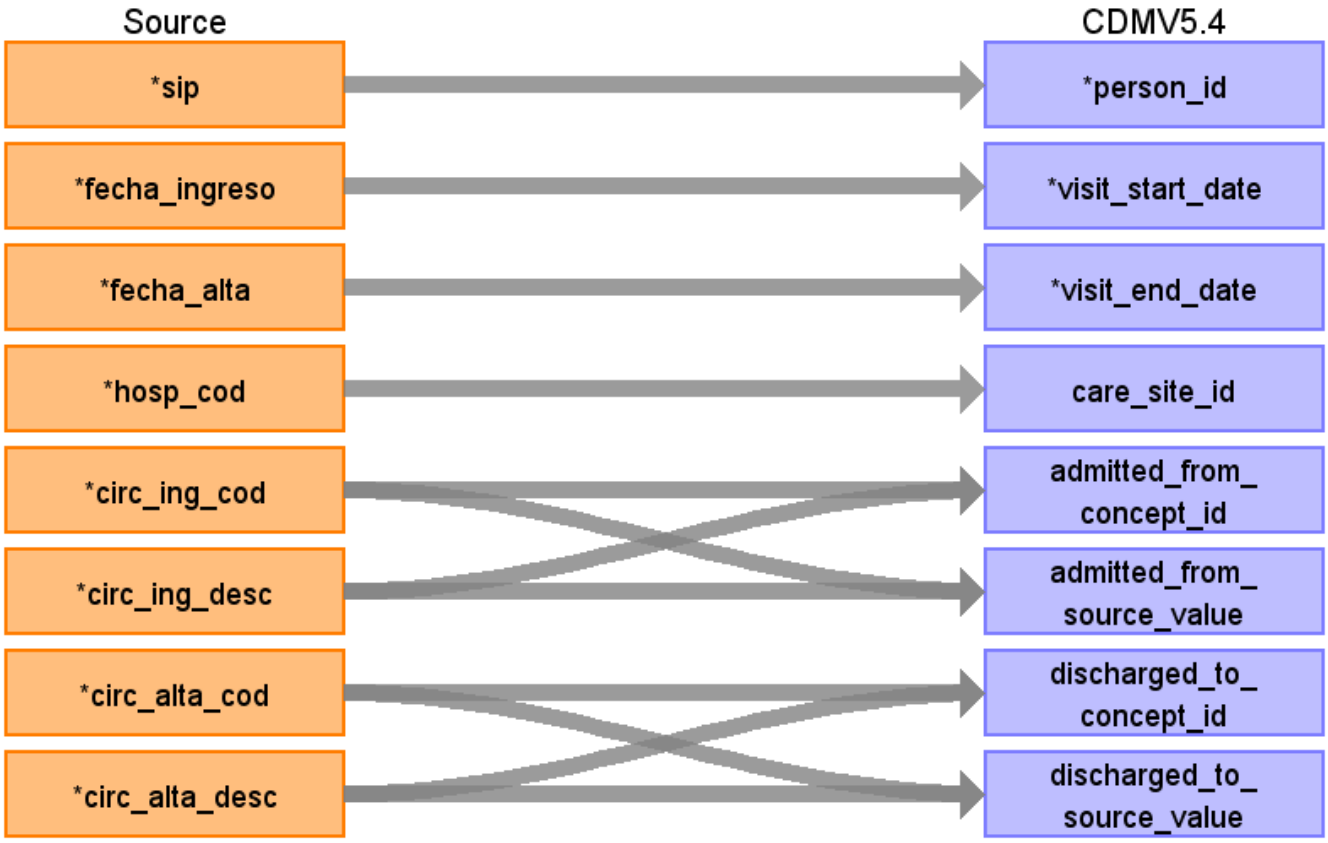


Figure 7: mbds to visit_occurrence mapping diagram

Table 7: mbds to visit_occurrence mapping description

Destination Field	Source Field	Logic	Comment
visit_occurrence_id			Autogenerate: from n_pcv + n_cex + 1 to n_pcv + n_cex + n_mbds when source table is **MBDS**.
person_id	*sip*		
visit_concept_id			In **MBDS** there are hospital admissions. The Concept ID is '8717: Inpatient Hospital'.
visit_start_date	*fecha_ingreso*		
visit_start_datetime	*fecha_ingreso*		
visit_end_date	*fecha_alta*		
visit_end_datetime	*fecha_alta*		

Destination Field	Source Field	Logic	Comment
visit_type_concept_id			In **MBDS** there are hospital discharge summaries. The Concept ID is '32824: EHR discharge summary'.
provider_id			NA_integer_
care_site_id	*hosp_cod*		
visit_source_value			MBDS
visit_source_concept_id			0
admitted_from_concept_id	*circ_ing_cod*, *circ_ing_desc*		Admission Standardized CONCEPT ID. When the *circ_ing_desc* is 'Programado' or 'CMA' the *admitted_from_concept_id* is '0'. When the circ_ing_desc is 'Urgente' the *admitted_from_concept_id* is '9203: Emergency Room Visit'.
admitted_from_source_value	*circ_ing_cod*, *circ_ing_desc*		Source admission code + description. The structure is: {*circ_ing_cod*}: {*circ_ing_desc*}.
discharged_to_concept_id	*circ_alta_cod*, *circ_alta_desc*		Standardized discharge concept ID following the conventions outlined in the *circ_alta* to *discharged_to_concept_id* mapping.
discharged_to_source_value	*circ_alta_cod*, *circ_alta_desc*		Source discharge code + description. The structure is: {*circ_alta_cod*}: {*circ_alta_desc*}.
preceding_visit_occurrence_id			Once all the source tables that contribute to visit_occurence are populated, we will order the *visit_occurrence_id* for

Destination Field	Source Field	Logic	Comment
			each *person_id* and complete this field.

Table 8: circ_alta to discharged_to_concept_id mapping

circ_alta_cod	circ_alta_desc	discharged_to_concept_id
-2	[Sin referencia]	0
-1	[Vacio]	0
0	[No aplica]	0
1	Domicilio	0
2	Equipo atención primaria	0
3	Consultas externas	0
4	Hospital de Día	0
5	Unidad de Hospitalización a domicilio	0
6	Alta voluntaria	0
7	Traslado Hospital de agudos	38004279: General Acute Care Hospital
8	Traslado Hospital Media larga estancia	38004277: Long Term Care Hospital
9	Traslado Residencia o Centro Socio-Sanitario asistido	8676: Nursing Facility
10	Éxitus	0
11	Fuga	0
13	Alta disciplinaria	0
14	Unidad de Salud Mental	38004284: Psychiatric Hospital
15	Hospitalización	8717: Inpatient Hospital
16	Desconocido	0
17	Urgencias	9203:Emergency Room Visit

circ_alta_cod	circ_alta_desc	discharged_to_concept_id
26	COVID-seguimiento	0
99	Otros	0

3.1.3.4 From aed to visit_occurrence

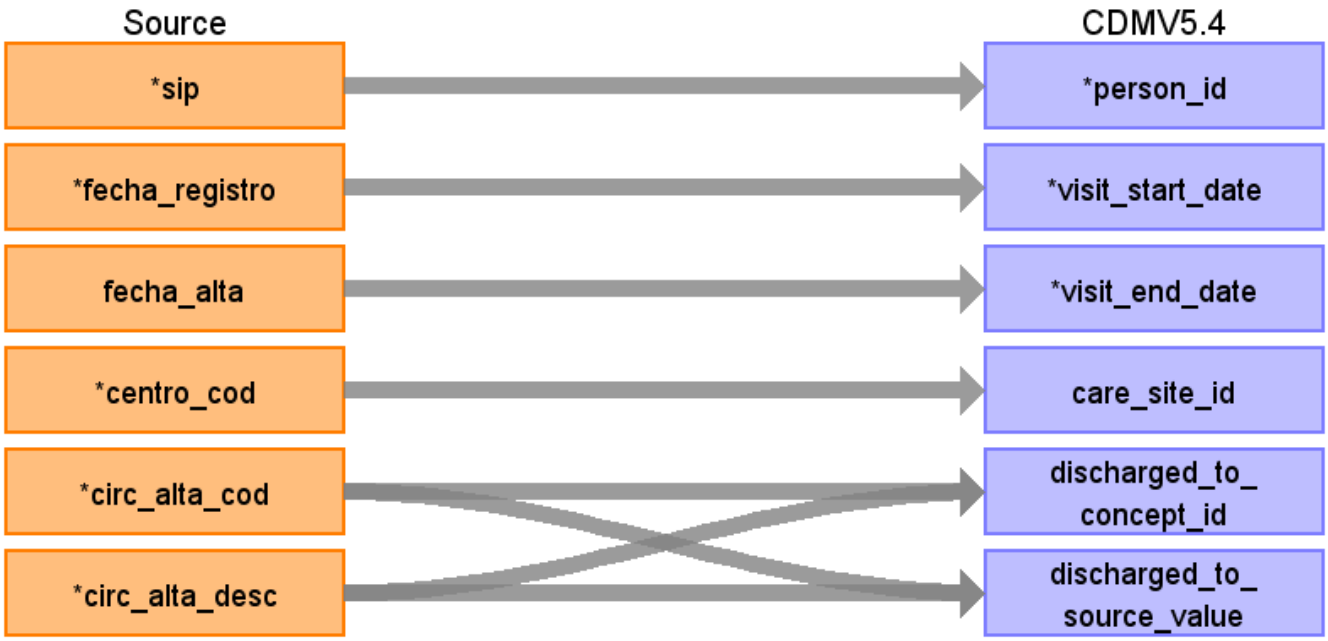


Figure 8: aed to visit_occurrence mapping diagram

Table 9: aed to visit_occurrence mapping description

Destination Field	Source Field	Logic	Comment
visit_occurrence_id			Autogenerate: from n_pcv + n_cex + n_mbds + 1 to n_pcv + n_cex + n_mbds + n_aed when source table is **AED**.
person_id	*sip*		
visit_concept_id			In **AED** there are emergency visits. The Concept ID is

Destination Field	Source Field	Logic	Comment
			'9203: Emergency Room Visit'.
visit_start_date	*fecha_registro*		
visit_start_datetime	*fecha_registro*		
visit_end_date	*fecha_alta*		
visit_end_datetime	*fecha_alta*		
visit_type_concept_id			In **AED** there are emergency visits. The Concept ID is '32826: EHR emergency room note'.
provider_id			NA_integer_
care_site_id			NA_integer_
visit_source_value			AED
visit_source_concept_id			0
admitted_from_concept_id			0
admitted_from_source_value			NA_character_
discharged_to_concept_id	*circ_alta_cod*, *circ_alta_desc*		Standardized discharge concept ID following the conventions outlined in the *circ_alta* to *discharged_to_concept_id* mapping.
discharged_to_source_value	*circ_alta_cod*, *circ_alta_desc*		Source discharge code + description. The structure is: {*circ_alta_cod*}: {*circ_alta_desc*}.
preceding_visit_occurrence_id			Once all the source tables that contribute to visit_occurence are populated, we will order the

Destination Field	Source Field	Logic	Comment
			visit_occurrence_id for each *person_id* and complete this field.

3.1.4 Target table: condition_occurrence

The target table **condition_occurrence** is populated with the information from the source tables **PCV**, **CEX**, **MBDS**, **AED**, **DIAGNOSES**, and **EOS**.

3.1.4.1 From pcv to condition_occurrence

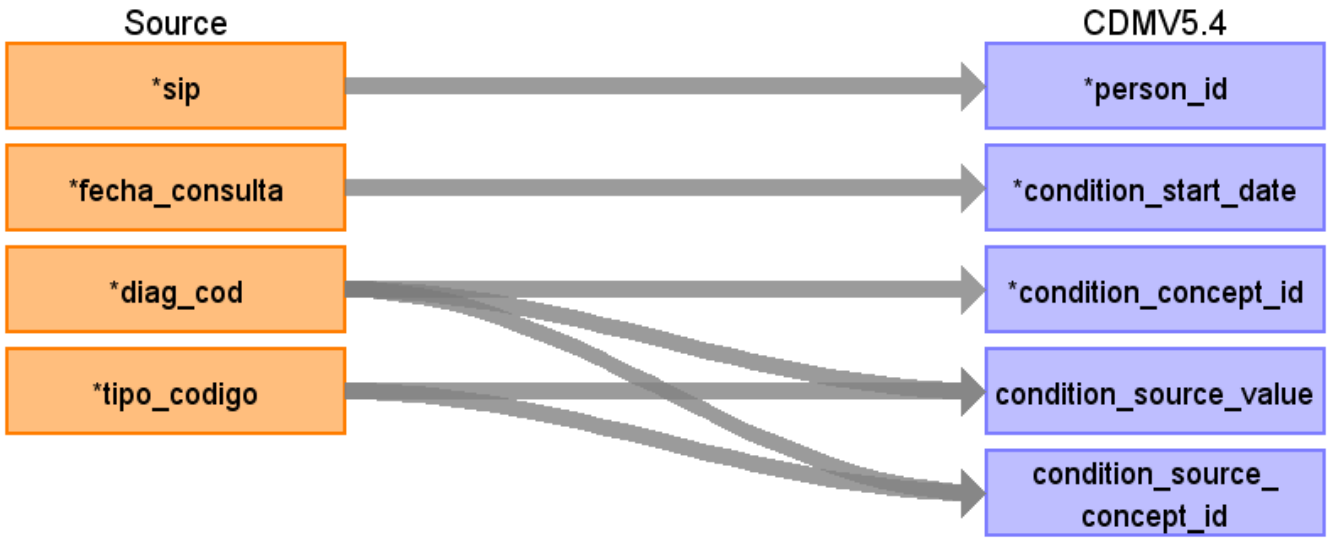


Figure 9: pcv to condition_occurrence mapping diagram

Table 10: pcv to condition_occurrence mapping description

Destination Field	Source Field	Logic	Comment
condition_occurrence_id			Autogenerate. When in the same visit there are duplicate conditions, they will be collapsed.
person_id	*sip*		

Destination Field	Source Field	Logic	Comment
condition_concept_id	*diag_cod*		Standardized CONCEPT ID from ICD9CM or ICD10ES codes.
condition_start_date	*fecha_consulta*		
condition_start_datetime	*fecha_consulta*		
condition_end_date			NA_Date_
condition_end_datetime			NA_Date_
condition_type_concept_id			In **PCV** there are primary care visits. The Concept ID is '32834: EHR outpatient note'.
condition_status_concept_id			When source table is **PCV**, **CEX**, **AED**, or **DIAGNOSES**, the *condition_status_concept_id* is '32893: Confirmed diagnosis'.
stop_reason			NA_character_
provider_id			NA_integer_
visit_occurrence_id			Retrieve the *visit_occurrence_id* from the intermediate table **from_pcv_to_visit_occurrence**.
visit_detail_id			NA_integer_
condition_source_value	*tipo_codigo*, *diag_cod*		The ICD9CM or ICD10ES code. *tipo_codigo* flags if the code is ICD9CM or ICD10ES.
condition_source_concept_id	*tipo_codigo*, *diag_cod*		ICD9CM or ICD10ES CONCEPT ID.
condition_status_source_value			PCV

3.1.4.2 From cex to condition_occurrence

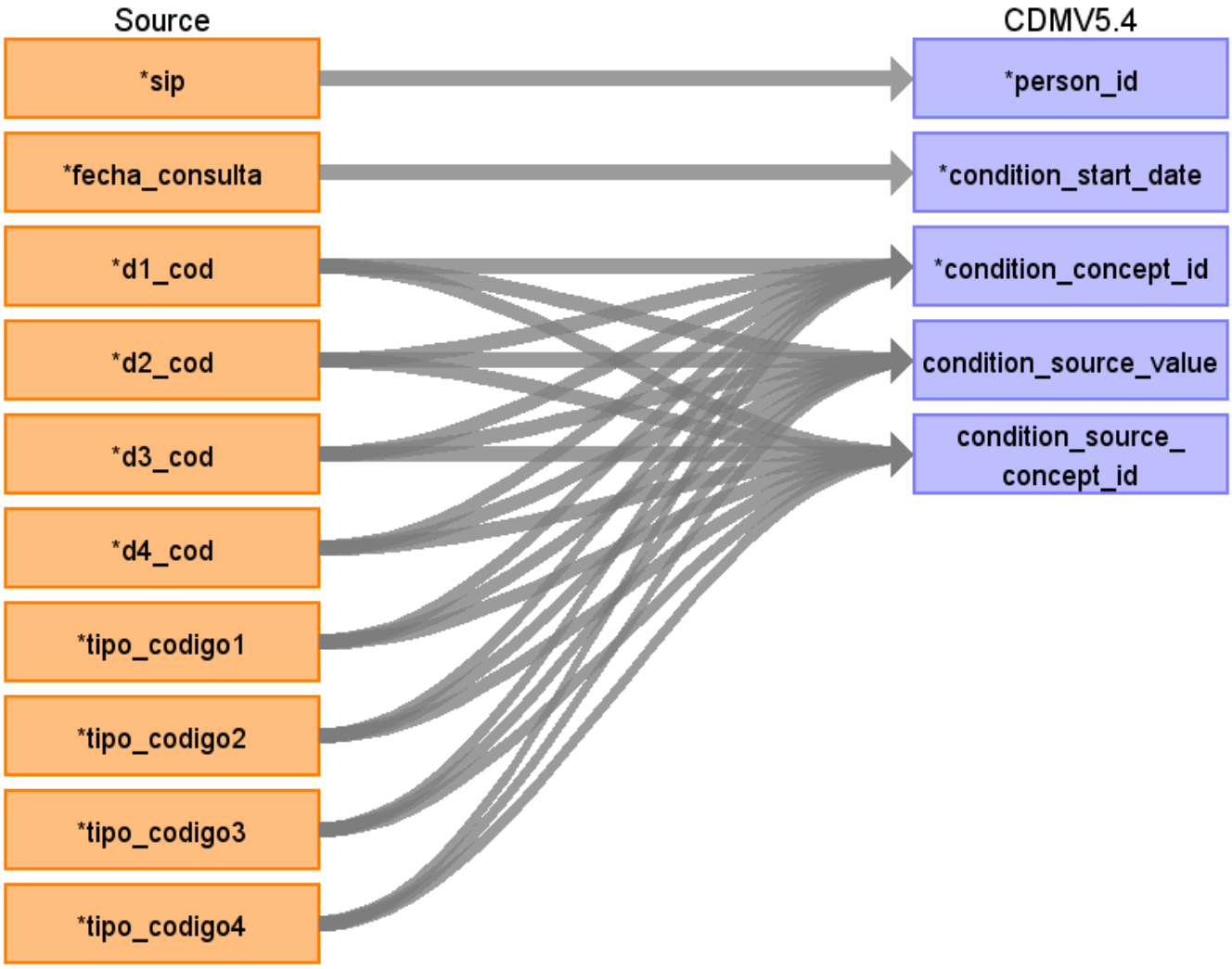


Figure 10: cex to condition_occurrence mapping diagram

Table 11: cex to condition_occurrence mapping description

Destination Field	Source Field	Logic	Comment
condition_occurrence_id			Autogenerate. When in the same visit there are duplicate conditions, they will be collapsed.
person_id	*sip*		
condition_concept_id	*d1_cod*, *d2_cod*, *d3_cod*, *d4_cod*	In each **CEX** visit there are up to 4 diagnosis codes. Each	Standardized CONCEPT ID from ICD9CM or ICD10ES codes.

Destination Field	Source Field	Logic	Comment
		distinct code is mapped in a different row.	
condition_start_date	*fecha_consulta*		
condition_start_datetime	*fecha_consulta*		
condition_end_date			NA_Date_
condition_end_datetime			NA_Date_
condition_type_concept_id			In **CEX** there are specialist care visits. The Concept ID is '32834: EHR outpatient note'.
condition_status_concept_id			When source table is **PCV**, **CEX**, **AED**, or **DIAGNOSES**, the *condition_status_concept_id* is '32893: Confirmed diagnosis'.
stop_reason			NA_character_
provider_id			NA_integer_
visit_occurrence_id			Retrieve the 'visit_occurrence_id' from the intermediate table **from_cex_to_visit_occurrence**.
visit_detail_id			NA_integer_
condition_source_value	*tipo_codigo1*, *tipo_codigo2*, *tipo_codigo3*, *tipo_codigo4*, *d1_cod*, *d2_cod*, *d3_cod*, *d4_cod*		The ICD9CM or ICD10ES code. *tipo_codigo* flags if the code is ICD9CM or ICD10ES.
condition_source_concept_id	*tipo_codigo1*, *tipo_codigo2*, *tipo_codigo3*, *tipo_codigo4*, *d1_cod*, *d2_cod*, *d3_cod*, *d4_cod*		ICD9CM or ICD10ES CONCEPT ID.
condition_status_source_value			CEX

3.1.4.3 From mbds to condition_occurrence

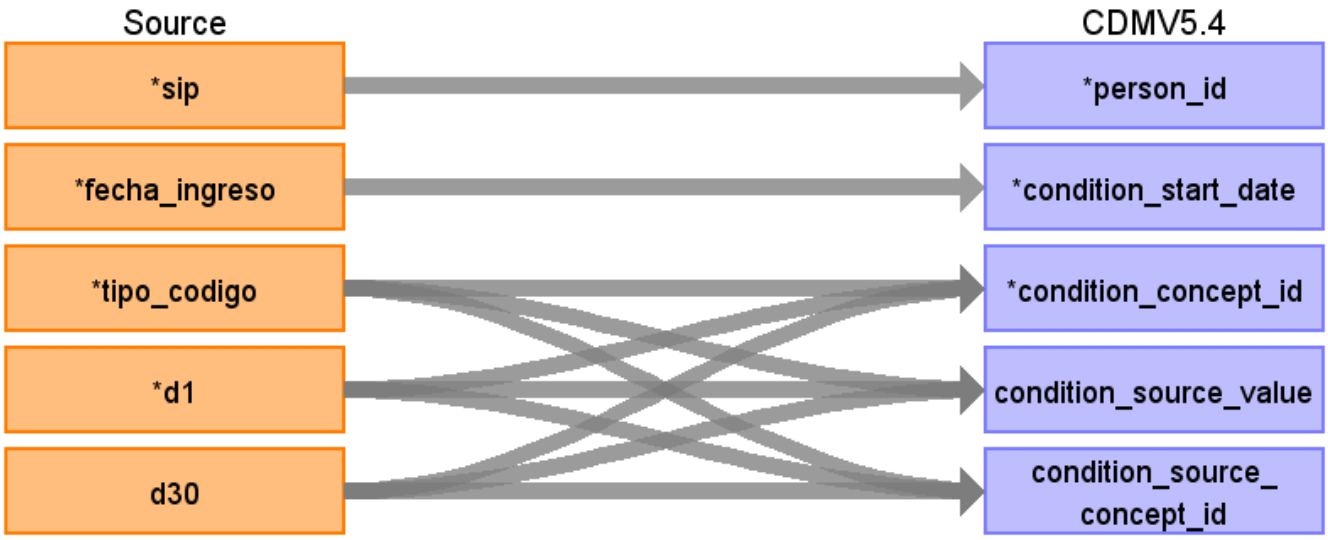


Figure 11: mbds to condition_occurrence mapping diagram

Table 12: mbds to condition_occurrence mapping description

Destination Field	Source Field	Logic	Comment
condition_occurrence_id			Autogenerate. When in the same visit there are duplicate conditions, they will be collapsed.
person_id	*sip*		
condition_concept_id	from *d1* to *d30*	In each mbds visit there are up to 30 diagnosis codes. Each distinct code is mapped in a different row.	Standardized CONCEPT ID from ICD9CM or ICD10ES codes. There are up to 30 diagnostic codes. However, for the sake of simplicity only *d1* and *d30* are depicted in the diagram.

Destination Field	Source Field	Logic	Comment
condition_start_date	*fecha_ingreso*		
condition_start_datetime	*fecha_ingreso*		
condition_end_date			NA_Date_
condition_end_datetime			NA_Date_
condition_type_concept_id			In **MBDS** there are hospital discharge summaries. The Concept ID is '32824: EHR discharge summary'.
condition_status_concept_id			When source table is **MBDS**, the condition_status_concept_id is '32903: Primary discharge diagnosis' (when the concept comes from *d1*) and '32909: Secondary discharge diagnosis' when the concept comes from *d2*:*d30*).
stop_reason			NA_character_
provider_id			NA_integer_
visit_occurrence_id			Retrieve the visit_occurrence_id from the intermediate table **from_mbds_to_visit_occurrence**.
visit_detail_id			NA_integer_
condition_source_value	*tipo_codigo*, *d1*:*d30*		The ICD9CM or ICD10ES code. tipo_codigo flags if the code is ICD9CM or ICD10ES.
condition_source_concept_id	*tipo_codigo*, *d1*:*d30*		ICD9CM or ICD10ES CONCEPT ID.
condition_status_source_value			MBDS

3.1.4.4 From aed to condition_occurrence

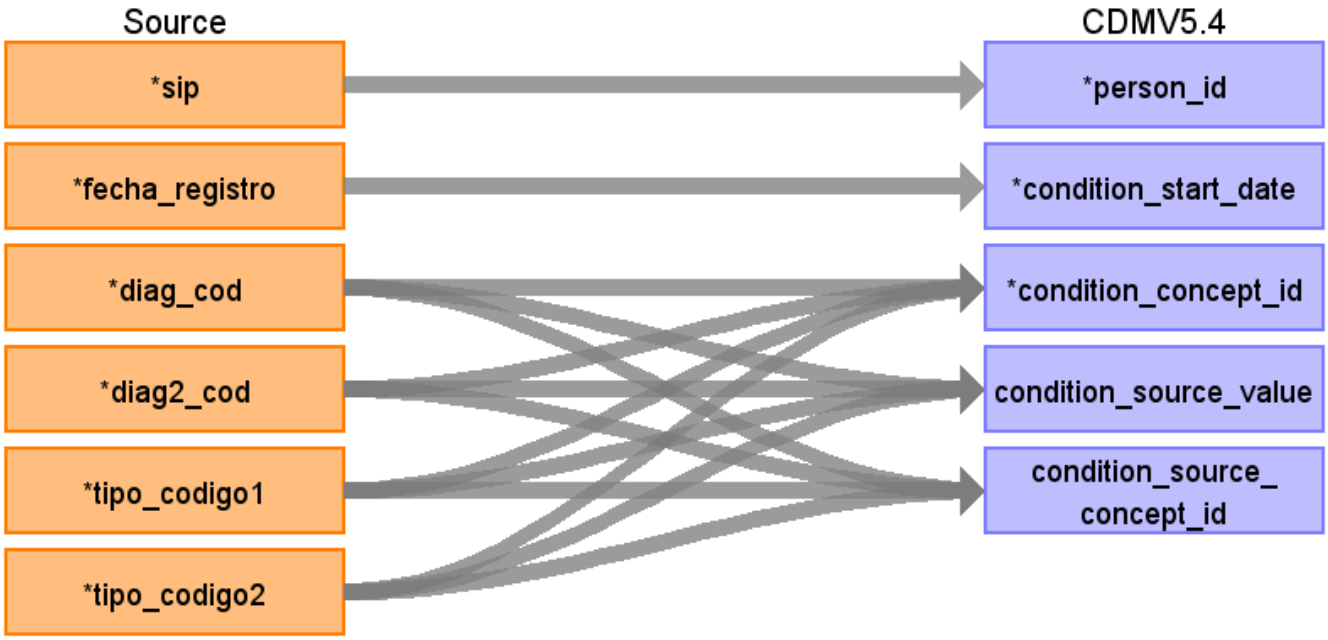


Figure 12: aed to condition_occurrence mapping diagram

Table 13: aed to condition_occurrence mapping description

Destination Field	Source Field	Logic	Comment
condition_occurrence_id			Autogenerate. When in the same visit there are duplicate conditions, they will be collapsed.
person_id	*sip*		
condition_concept_id	*diag_cod*, *diag2_cod*	In each aed visit there are up to 2 diagnosis codes. Each distinct code is mapped in a different row.	Standardized CONCEPT ID from ICD9CM or ICD10ES codes.
condition_start_date	*fecha_registro*		
condition_start_datetime	*fecha_registro*		
condition_end_date			NA_Date_

Destination Field	Source Field	Logic	Comment
condition_end_datetime			NA_Date_
condition_type_concept_id			**AED** are emergency visits. The Concept ID is '32826: EHR emergency room note'.
condition_status_concept_id			When source table is **PCV**, **CEX**, **AED**, or **DIAGNOSES**, the *condition_status_concept_id* is '32893: Confirmed diagnosis'.
stop_reason			NA_character_
provider_id			NA_integer_
visit_occurrence_id			Retrieve the 'visit_occurrence_id' from the intermediate table **from_aed_to_visit_occurrence**.
visit_detail_id			NA_integer_
condition_source_value	*tipo_codigo1*, *tipo_codigo2*, *diag_cod*, *diag2_cod*		The ICD9CM or ICD10ES code. *tipo_codigo* flags if the code is ICD9CM or ICD10ES.
condition_source_concept_id	*tipo_codigo1*, *tipo_codigo2*, *diag_cod*, *diag2_cod*		ICD9CM or ICD10ES CONCEPT ID.
condition_status_source_value			AED

3.1.4.5 From diagnoses to condition_occurrence

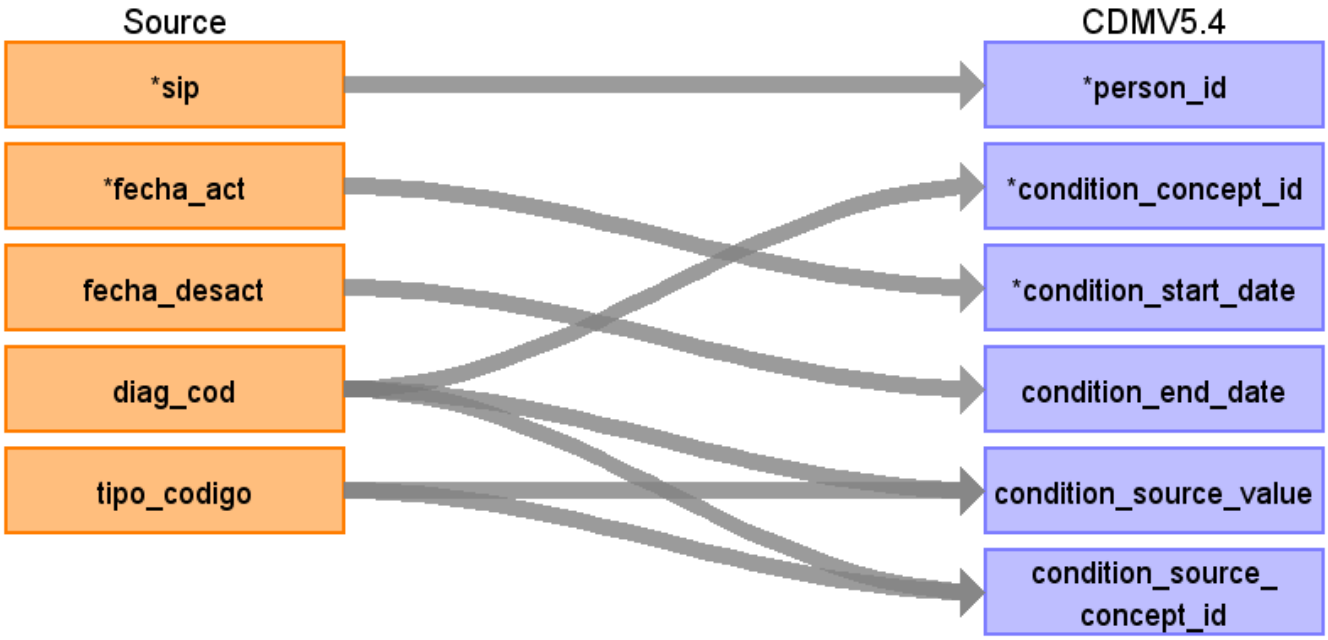


Figure 13: diagnoses to condition_occurrence mapping diagram

Table 14: diagnoses to condition_occurrence mapping description

Destination Field	Source Field	Logic	Comment
condition_occurrence_id			Autogenerate. When in the same visit there are duplicate conditions, they will be collapsed.
person_id	*sip*		
condition_concept_id	*diag_cod*		Standardized CONCEPT ID from ICD9CM or ICD10ES codes.
condition_start_date	*fecha_act*		
condition_start_datetime	*fecha_act*		
condition_end_date	*fecha_desact*		*condition_end_date* only is captured when the source table is **DIAGNOSES**.
condition_end_datetime	*fecha_desact*		
condition_type_concept_id			**DIAGNOSES** are confirmed diagnoses. The Concept ID is '32817: EHR'.

Destination Field	Source Field	Logic	Comment
condition_status_concept_id			When source table is **PCV**, **CEX**, **AED**, or **DIAGNOSES**, the *condition_status_concept_id* is '32893: Confirmed diagnosis'.
stop_reason			NA_character_
provider_id			NA_integer_
visit_occurrence_id			NA_integer_
visit_detail_id			NA_integer_
condition_source_value	*tipo_codigo*, *diag_cod*		The ICD9CM or ICD10ES code. *tipo_codigo* flags if the code is ICD9CM or ICD10ES.
condition_source_concept_id	*tipo_codigo*, *diag_cod*		ICD9CM or ICD10ES CONCEPT ID.
condition_status_source_value			**DIAGNOSES**

3.1.4.6 From mdr to condition_occurrence

From **MDR** we are creating pregnancy-related condition_occurrences in order to create pregnancy eras.

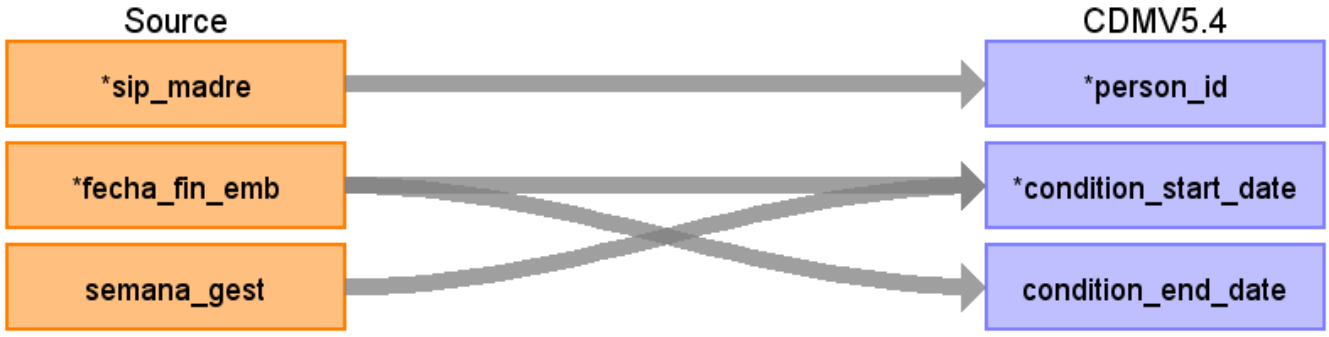


Figure 14: mdr to condition_occurrence mapping diagram

Table 15: mdr to condition_occurrence mapping description

Destination Field	Source Field	Logic	Comment
condition_occurrence_id			Autogenerate
person_id	*sip*		
condition_concept_id			'4336958: Term pregnancy'.
condition_start_date	*fecha_fin_emb*, *semana_gest*	*fecha_fin_emb* %m-% weeks(*semana_gest*)	
condition_start_datetime	*fecha_fin_emb*, *semana_gest*	*fecha_fin_emb* %m-% weeks(*semana_gest*)	
condition_end_date	*fecha_fin_emb*		
condition_end_datetime	*fecha_fin_emb*		
condition_type_concept_id			'32879: Registry'.
condition_status_concept_id			'32906: Resolved condition'.
stop_reason			livebirth
provider_id			NA_integer_
visit_occurrence_id			NA_integer_
visit_detail_id			NA_integer_
condition_source_value			NA_character_
condition_source_concept_id			NA_integer_
condition_status_source_value			MDR

3.1.4.7 From eos to condition_occurrence

From **EOS** we are creating pregnancy-related condition_occurrences in order to create pregnancy eras.

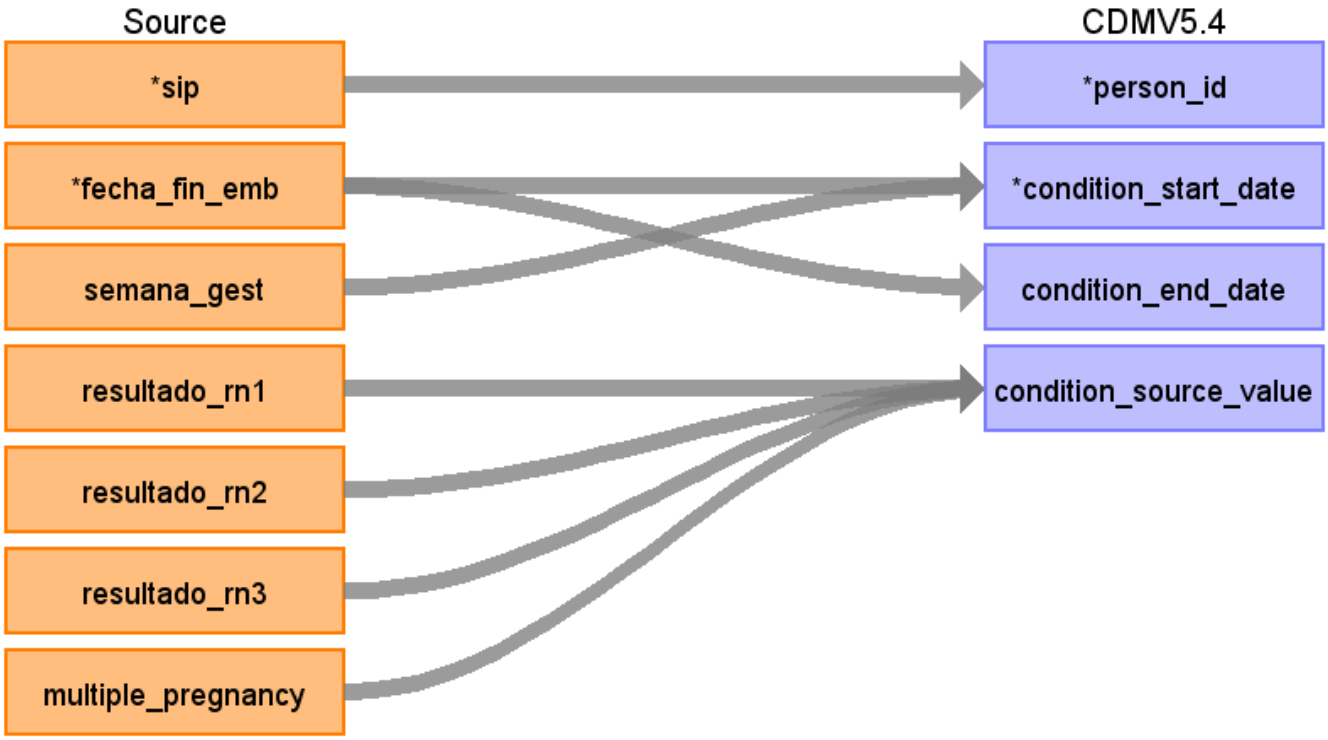


Figure 15: eos to condition_occurrence mapping diagram

Table 16: eos to condition_occurrence mapping description

Destination Field	Source Field	Logic	Comment
condition_occurrence_id			Autogenerate. Whe are duplicate condi collapsed.
person_id	*sip*		
condition_concept_id			'4142469: Pregnanc
condition_start_date	*fecha_fin_emb*, *semana_gest*	*fecha_fin_emb* %m-% weeks(*semana_gest*)	
condition_start_datetime	*fecha_fin_emb*, *semana_gest*	*fecha_fin_emb* %m-% weeks(*semana_gest*)	
condition_end_date	*fecha_fin_emb*		
condition_end_datetime	*fecha_fin_emb*		
condition_type_concept_id			In the **EOS** there ar through EHR. The C

Destination Field	Source Field	Logic	Comment
condition_status_concept_id			'32906: Resolved co
stop_reason			End of pregnancy
provider_id			NA_integer_
visit_occurrence_id			NA_integer_
visit_detail_id			NA_integer_
condition_source_value	*multiple_pregnancy, resultado_rn1, resultado_rn2, resultado_rn3*		Result of pregnancy {*multiple_pregnancy* {*resultado_rn1*}/{*resu*
condition_source_concept_id			NA_integer
condition_status_source_value			EOS

3.1.5 Target table: procedure_occurrence

The target table **procedure_occurrence** is populated with the information from the source table **MBDS**.

3.1.5.1 From mbds to procedure_occurrence

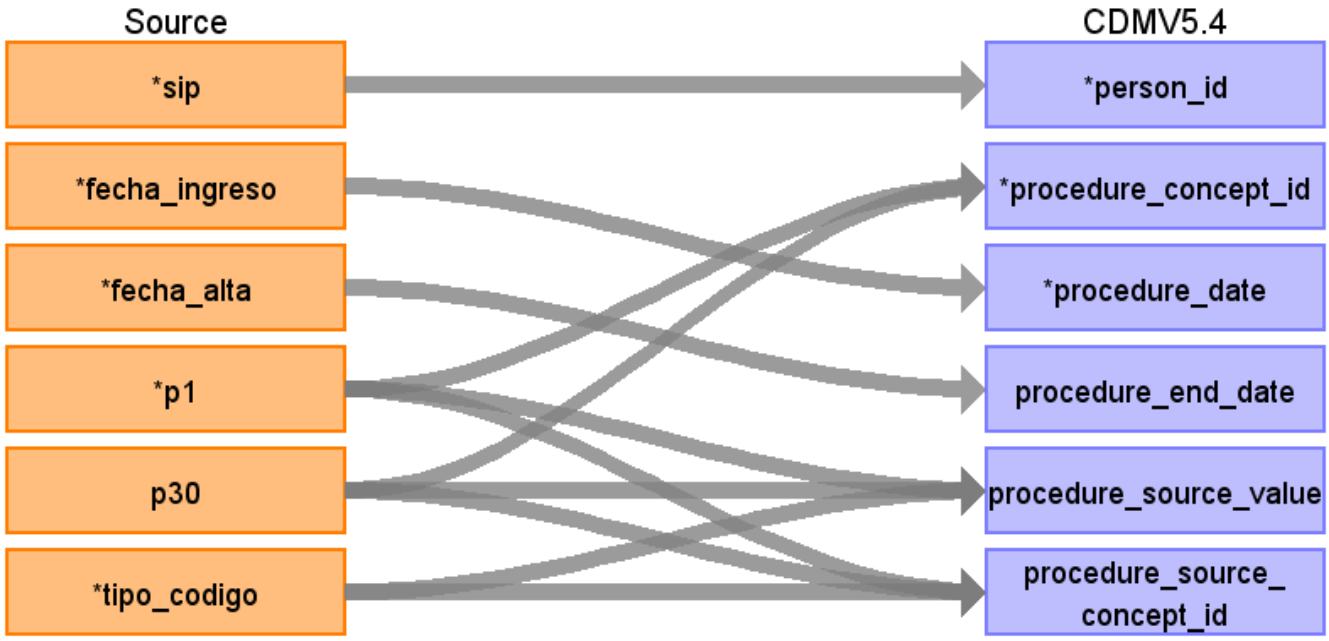


Figure 16: mbds to procedure_occurrence mapping diagram

Table 17: mbds to procedure_occurrence mapping description

Destination Field	Source Field	Logic	Comment
procedure_occurrence_id			
person_id	*sip*		
procedure_concept_id	from *p1* to *p30*		Standardized Concept ID from ICD9CM or ICD10ES procedure code. There are up to 30 procedure codes. However, for the sake of simplicity only *p1* and *p30* are depicted in the diagram.
procedure_date	*fecha_ingreso*		Procedures, usually take place in the same day. The accurate procedure_date is some date between *fecha_ingreso* and *fecha_alta*.
procedure_datetime	*fecha_ingreso*		
procedure_end_date	*fecha_alta*		Procedures, usually take place in the same day. The accurate procedure_date is some date between *fecha_ingreso* and *fecha_alta*.
procedure_end_datetime	*fecha_alta*		
procedure_type_concept_id			'32824: EHR discharge summary'.
modifier_concept_id			NA_integer_
quantity			NA_integer_
provider_id			NA_integer_
visit_occurrence_id			Use the intermediate table **from_mbds_to_visit_occurrence**.
visit_detail_id			NA_integer_
procedure_source_value	*tipo_codigo*, *p1:p30*		The ICD9CM or ICD10ES code. *tipo_codigo* flags if the code is ICD9CM or ICD10ES

Destination Field	Source Field	Logic	Comment
procedure_source_concept_id	*tipo_codigo, p1:p30*		The CONCEPT ID from the ICD9CM or ICD10ES code
modifier_source_value			MBDS

3.1.6 Target table: death

The target table **death** is populated with the information from the source tables **SIP**, and **PMR**.

3.1.6.1 From sip to death

Figure 17: sip to death mapping diagram

Table 18: sip to death mapping description

Destination Field	Source Field	Logic	Comment
person_id	*sip*		
death_date	*fecha_defuncion*		
death_datetime	*fecha_defuncion*		
death_type_concept_id			'32848: Government report'.
cause_concept_id			NA_integer_
cause_source_value			NA_character_
cause_source_concept_id			NA_integer_

3.1.6.2 From pmr to death

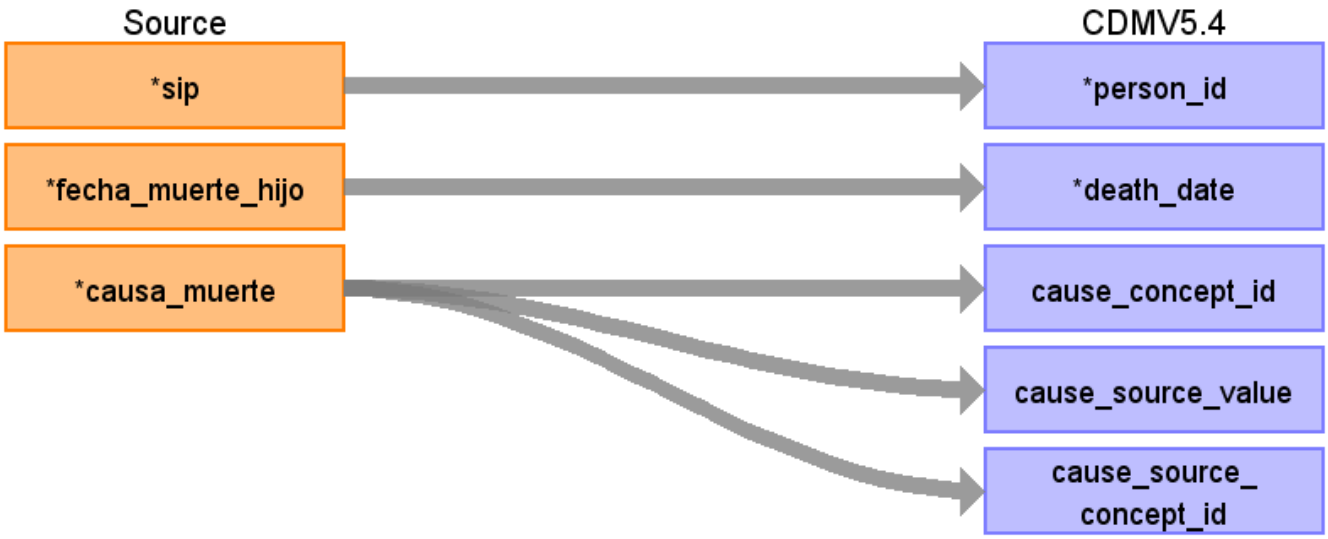


Figure 18: pmr to death mapping diagram

Table 19: pmr to death mapping description

Destination Field	Source Field	Logic	Comment
person_id	*sip*		
death_date	*fecha_muerte_hijo*		
death_datetime	*fecha_muerte_hijo*		
death_type_concept_id		'32879: Registry'.	
cause_concept_id	*causa_muerte*		the cause of death is an standardized code (SNOMED) from *causa_muerte*.
cause_source_value	*causa_muerte*		The *cause_source_value* is *causa_muerte*, which is an ICD9CM or ICD10ES code.
cause_source_concept_id	*causa_muerte*		The *cause_source_concept_id* is the corresponding

Destination Field	Source Field	Logic	Comment
			concept_id from the ICD9CM or ICD10ES code. Otherwise NA_integer_.

> **Important**
>
> For now, it is not possible to perform the ETL process for the '**From pmr to death**' information. The sex of the newborn is currently unavailable, and in accordance with the OMOP conventions, these newborn records cannot be stored in the **person** table.

3.1.7 Target table: drug_exposure

The target table **drug_exposure** is populated with the information from the source tables **GAIA**, and **SIV**.

3.1.7.1 From gaia to drug_exposure

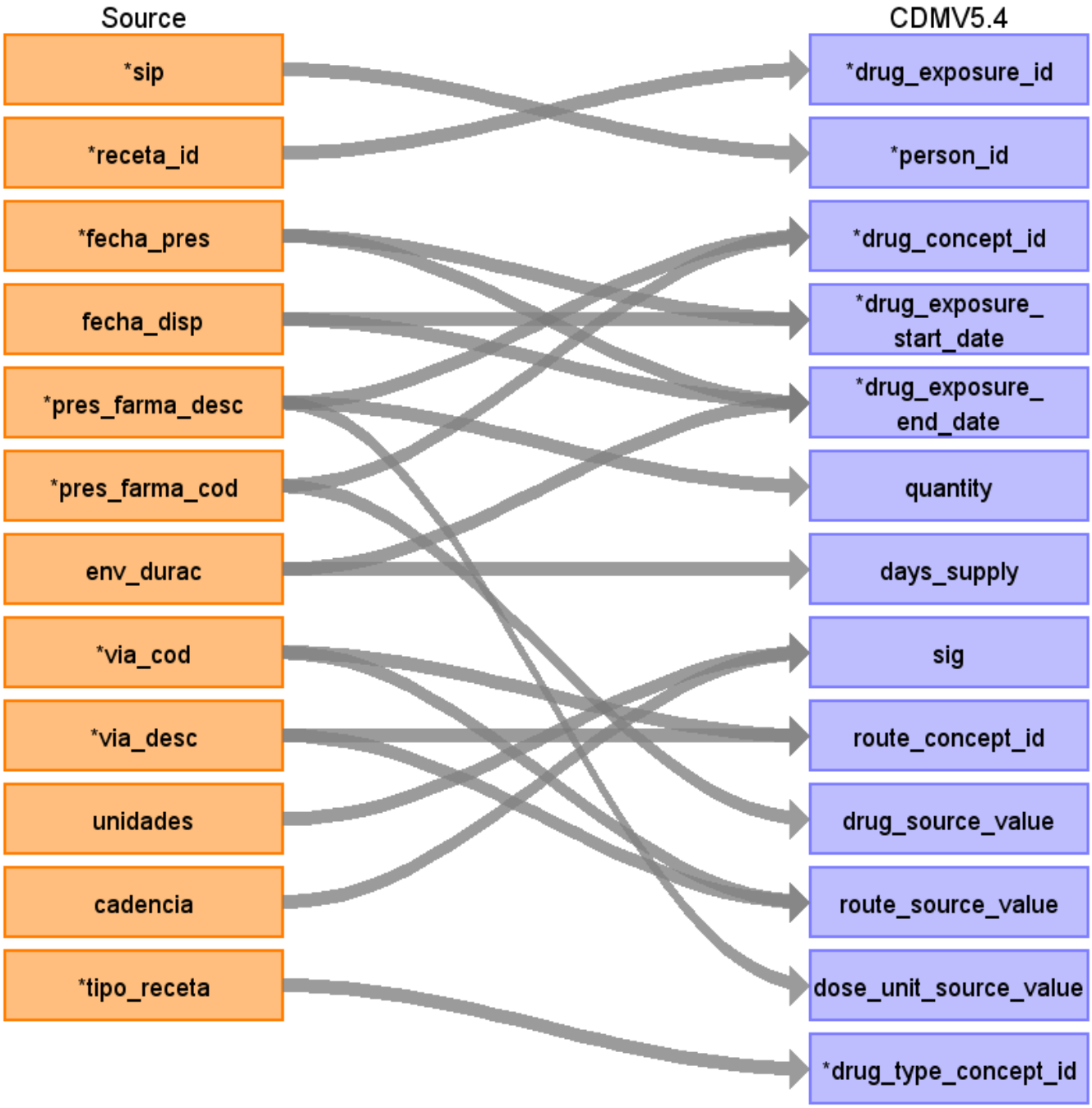


Figure 19: gaia to drug_exposure mapping diagram

Table 20: gaia to drug_exposure mapping description

Destination Field	Source Field	Logic	Comment
drug_exposure_id	*receta_id*		Autogenerate an integer for each unique *receta_id*.
person_id	*sip*		

Destination Field	Source Field	Logic	Comment
drug_concept_id	*pres_farma_cod*, *pres_farma_desc*		Obtain Standard CONCEPT ID using *pres_farma_desc* and *pres_farma_cod* with the help of the USAGI tool.
drug_exposure_start_date	*fecha_disp*, *fecha_pres*		When !is.na(*fecha_disp*), the *drug_exposure_start_date* is *fecha_disp*. When is.na(*fecha_disp*), the *drug_exposure_start_date* is *fecha_pres* (although this is not an actual exposure, as the patient do not filled the prescription. However, could be useful for assessing prescription patterns or patient adherence).
drug_exposure_start_datetime	*fecha_disp*, *fecha_pres*		
drug_exposure_end_date			*drug_exposure_end_date* is calculated as *drug_exposure_start_date* %m+% days(*env_durac*).
drug_exposure_end_datetime			
verbatim_end_date			NA_Date_
drug_type_concept_id			When !is.na(*fecha_disp*), the CONCEPT ID is 32825: EHR dispensing record. When is.na(*fecha_disp*), the CONCEPT ID is 32838: EHR prescription.
stop_reason			NA_character_
refills			NA_integer_
quantity	*pres_farma_desc*		Extracted from *pres_farma_desc*.

Destination Field	Source Field	Logic	Comment
days_supply	*env_durac*		
sig	*unidades*, *cadencia*		Dosage using *unidades* and *cadencia* variables. The structure is: '{*unidades*} form(s) each {*cadencia*} hours'.
route_concept_id	*via_cod*		Standardized CONCEPT ID route code.
lot_number			0
provider_id			NA_integer_
visit_occurrence_id			
visit_detail_id			NA_integer_
drug_source_value	*pres_farma_cod*		
drug_source_concept_id			NA_integer_
route_source_value	*via_desc*		
dose_unit_source_value		extract dose unit from *pres_farma_desc*.	

3.1.7.2 From siv to drug_exposure

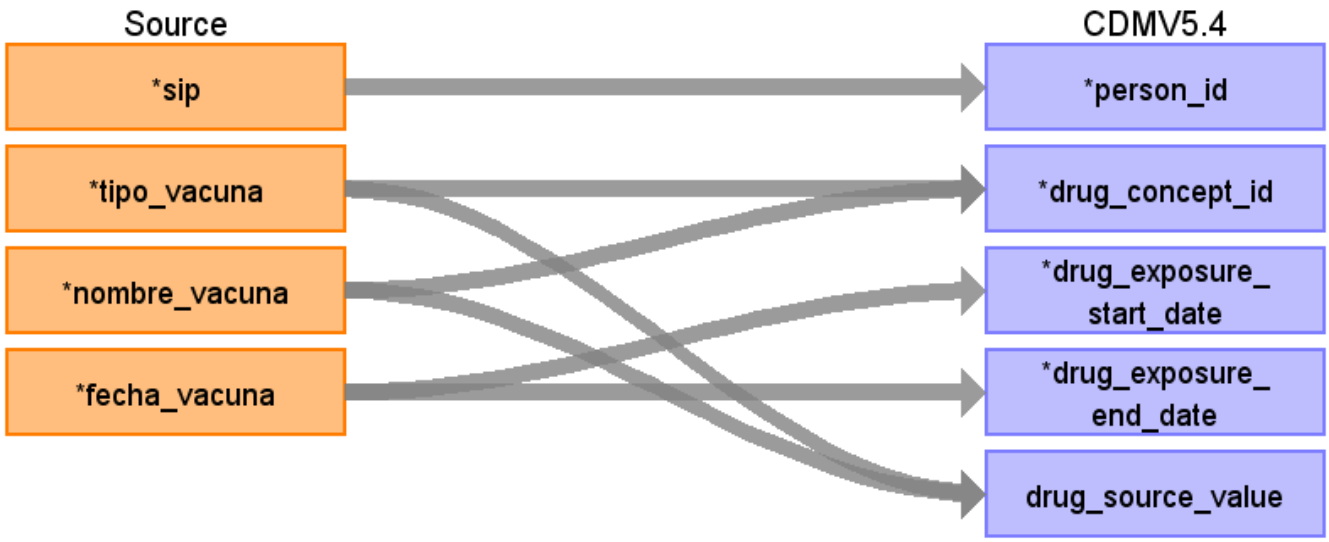


Figure 20: siv to drug_exposure mapping diagram

Table 21: siv to drug_exposure mapping description

Destination Field	Source Field	Logic	Comment
drug_exposure_id			
person_id	*sip*		
drug_concept_id			ingredient, or drug comp, or branded drug Standard CONCEPT ID
drug_exposure_start_date	*fecha_vacuna*		
drug_exposure_start_datetime	*fecha_vacuna*		
drug_exposure_end_date	*fecha_vacuna*		
drug_exposure_end_datetime	*fecha_vacuna*		
verbatim_end_date			NA_Date_
drug_type_concept_id	*tipo_vacuna*, *nombre_vacuna*		the CONCEPT ID is '32818: EHR administration record'.
stop_reason			NA_character_
refills			0
quantity			1
days_supply			0
sig			NA_character_
route_concept_id			The *route_concept_id* is '4302612: Intramuscular' (we will check if some vaccine has another administration route different to intramuscular).

Destination Field	Source Field	Logic	Comment
lot_number			0
provider_id			NA_integer_
visit_occurrence_id			NA_integer_
visit_detail_id			NA_integer_
drug_source_value	*tipo_vacuna*, *nombre_vacuna*		ingredient, or drug comp, or branded
drug_source_concept_id	*tipo_vacuna*, *nombre_vacuna*		ingredient, or drug comp, or branded CONCEPT ID
route_source_value			Intramuscular
dose_unit_source_value			NA_character_

3.1.8 Target table: care_site

In the Valencia region, the public health care is divided into 24 health departments. The departments, by their part, contain one or more hospitals and are divided into health basic areas.

The target table **care_site** is populated with the information from the source tables **SIP** and **MBDS**.

3.1.8.1 From sip to care_site

In the **SIP** table, we populate the departments and the health basic areas.

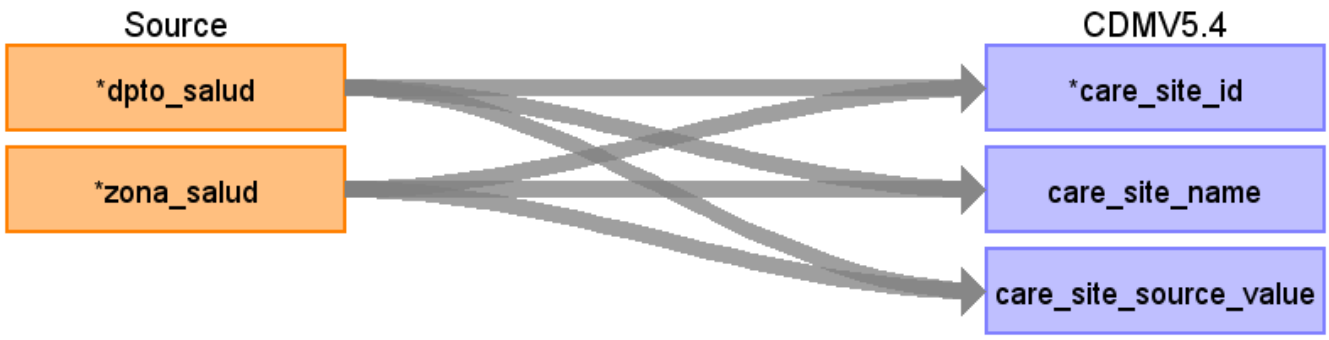


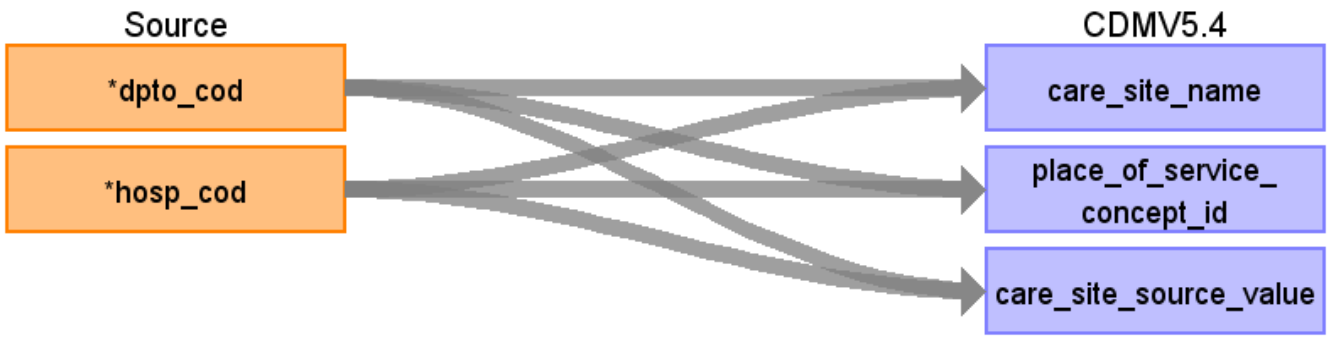
Figure 21: sip to care_site mapping diagram

Table 22: sip to care_site mapping description

Destination Field	Source Field	Logic	Comment
care_site_id	*dpto_salud*, *zona_salud*		Autogenerate
care_site_name	*dpto_salud*, *zona_salud*	Paste the *dpto_salud* (department) and the *zona_salud* (basic health area) names.	In this field it is indicated the 'name' of the department and the basic health area (separated with a semicolon symbol ';') assigned to each individual.
place_of_service_concept_id			The CONCEPT ID is '38004226: Ambulatory Health Service Clinic / Center'.
location_id			NA_integer_
care_site_source_value	*dpto_salud*, *zona_salud*	Paste the *dpto_salud* (department) and the *zona_salud* (basic health area) codes.	In this field it is indicated the original 'codes' and 'names' of the department and the basic health area assigned to each individual. The structure is: {*dpto_salud_cod*}: {*dpto_salud_name*}; {*zona_salud_cod*}: {*zona_salud_name*}.
place_of_service_source_value			Zona básica de salud.

3.1.8.2 From mbds to care_site

In the **MBDS** table, we have the hospital information. There is a master table with the corresponding names and codes of all the hospitals within VID. This master table is independent of the study/project extraction. Therefore, it will be used in order to map the care_site information related to hospitals.



mbds to care_site mapping diagram

Table 23: mbds to care_site mapping description

Destination Field	Source Field	Logic	Comment
care_site_id	*dpto_cod, hosp_cod*		Autogenerate
care_site_name	*dpto_cod, hosp_cod*		In this field it is indicated the 'name' of the hospital.
place_of_service_concept_id			The CONCEPT ID is '38004515: Hospital'.
location_id			NA_integer_
care_site_source_value	*hosp_cod*		In this field it is indicated the 'code' of the hospital.
place_of_service_source_value			Hospital

3.1.9 Target table: fact_relationship

The target table **fact_relationship** is populated with the information from the source table **MDR**.

3.1.9.1 From mdr to fact_relationship

This table cannot be populated because we don't have information about the sex of the children, and we cannot create a person entry for them. We have included this section in anticipation of the possibility of obtaining the sex of the newborns in the future.

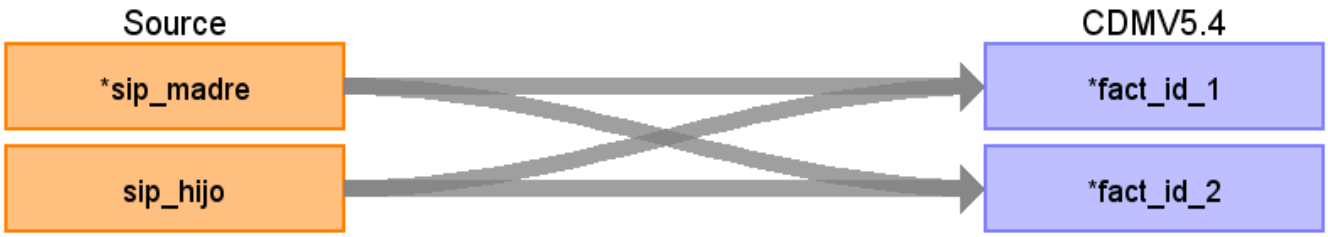


Figure 22: mdr to fact_relationship mapping diagram

Use the standard concepts in order to obtain the bidirectional association: Person, 1, Person, 2, mother of. Person, 2, Person, 1, child of.

Table 24: mdr to fact relationship mapping description

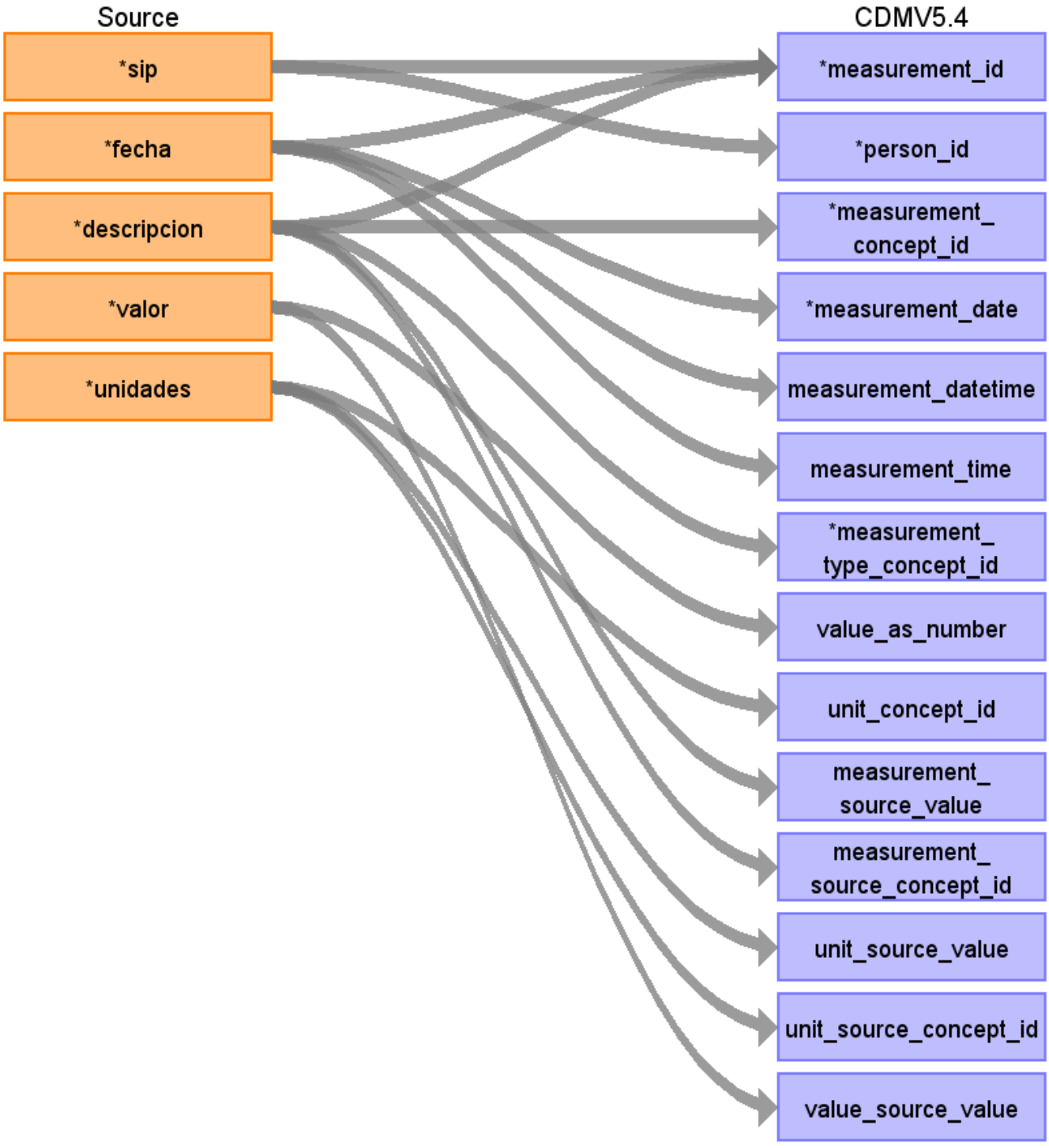
Destination Field	Source Field	Logic	Comment
domain_concept_id_1			'32879 (Registry)'.
fact_id_1	*sip_madre*, *sip_hijo*		*person_id* of 'person1'/'person2'.
domain_concept_id_2			'32879 (Registry)'.
fact_id_2	*sip_hijo*, *sip_madre*		*person_id* of 'person1'/'person2'.
relationship_concept_id			For each pair of related relationships, 'Mother: 4248584' and 'Child: 4285883'.

3.1.10 Target table: measurement

The target table **measurement** is populated with the information from the source tables **TESTS** and **REDMIVA**.

3.1.10.1 From tests to measurement

The **TESTS** table contains information on diastolic blood pressure (mmHg), systolic blood pressure (mmHg), height (cm), weight (kg) and BMI (kg/m^2).



tests to measurement mapping diagram

Table 25: tests to measurement mapping description

Destination Field	Source Field	Logic	Comment
measurement_id	*sip, fecha, descripcion*		Autogenerate
person_id	*sip*		

measurement_concept_id	*descripcion*		Depending on the *descripcion* variable in the source table as follows:
			`- Tensión arterial diastólica: 3012888 (Diastolic blood pressure).`
			
			`- Tensión Arterial Sistólica: 3004249 (Systolic blood pressure).`
			
			`- Talla: 3015514 (Body height -- standing).`
			
			`- Peso/Peso Medio: 3025315 (Body weight).`
			
			`- Índice de Masa Corporal: 3038553 (Body mass index (BMI) [Ratio]).`

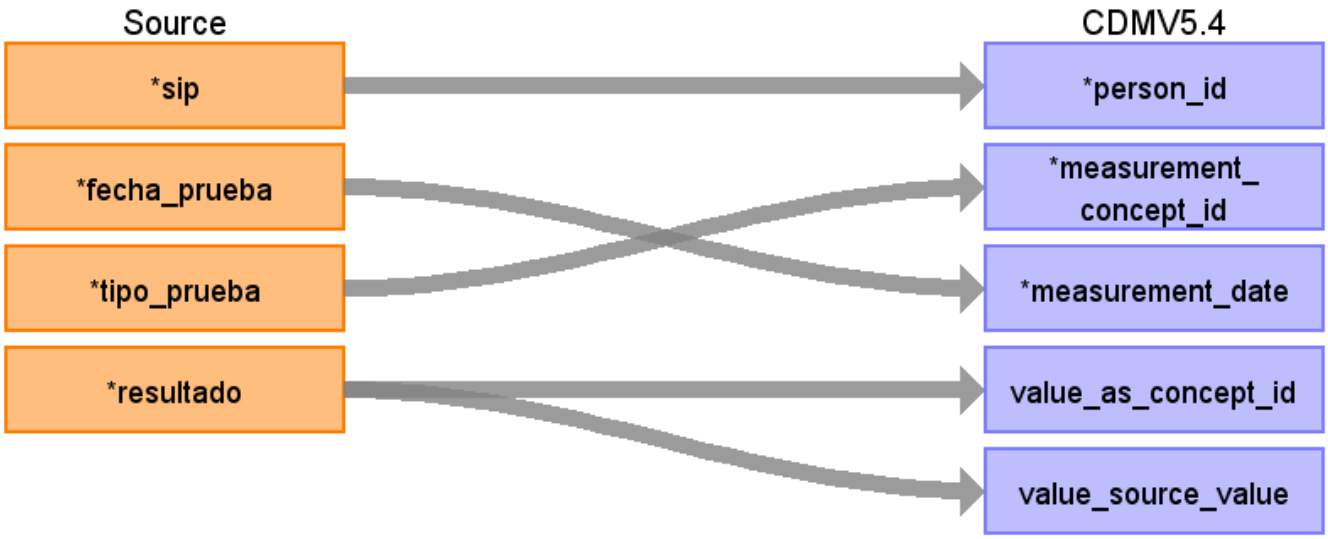
measurement_date	*fecha*		
measurement_datetime	*fecha*		
measurement_time	*fecha*		
measurement_type_concept_id	*descripcion*		For all records, '32836: EHR physical examination', except for BMI which is '32880: Standard algorithm'.
operator_concept_id			0

Destination Field	Source Field	Logic	Comment
value_as_number	*valor*		0
value_as_concept_id			NA_integer_

unit_concept_id	*unidades*		Depending on the *unidades* variable in the source table as follows:
			`- cm: 8582 (centimeter).`
			`- kg: 9529 (kilogram).`
			`- kg/m2: 9531 (kilogram per square meter).`
			`- mmHg: 8876 (millimeter mercury column).`

range_low			0
range_high			0
provider_id			NA_integer_
visit_occurrence_id			NA_integer_
visit_detail_id			NA_integer_
measurement_source_value	*descripcion*		
measurement_source_concept_id	*descripcion*		
unit_source_value	*unidades*		
unit_source_concept_id	*unidades*		
value_source_value	*valor*		
measurement_event_id			NA_integer_
meas_event_field_concept_id			0

3.1.10.2 From redmiva to measurement



redmiva to measurement mapping diagram

Table 26: redmiva to measurement mapping description

Destination Field	Source Field	Logic	Comment
measurement_id			Autogenerate
person_id	*sip*		

Destination Field	Source Field	Logic	Comment
measurement_concept_id	*tipo_prueba*		When *tipo_prueba* is (PCR) in the source table, the concept id is '586310: Measurement of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Genetic material using Molecular method), while when is antigen tests is mapped to '37310257: Measurement of Severe acute respiratory syndrome coronavirus 2 antigen'.
measurement_date	*fecha_prueba*		
measurement_datetime	*fecha_prueba*		
measurement_time	*fecha_prueba*		
measurement_type_concept_id			'32856: Lab'.
operator_concept_id			0
value_as_number			0
value_as_concept_id	*resultado*		'Positive: 45884084'.
unit_concept_id			0
range_low			0
range_high			0
provider_id			NA_integer_
visit_occurrence_id			NA_integer_

Destination Field	Source Field	Logic	Comment
visit_detail_id			NA_integer_
measurement_source_value			0
measurement_source_concept_id			0
unit_source_value			0
unit_source_concept_id			0
value_source_value	*resultado*		Positivo
measurement_event_id			NA_integer_
meas_event_field_concept_id			0

3.1.11 Instance table: cdm_source

The CDM_SOURCE table contains detail about the source database and the process used to transform the data into the OMOP Common Data Model.

Table 27: cdm_source table description

Destination Field	Logic	Value in the last instance
cdm_source_name	The name of the CDM instance.	Valencia Health System Integrated Da CONSIGN project
cdm_source_abbreviation	The abbreviation of the CDM instance.	vid_consign
cdm_holder	The holder of the CDM instance.	FISABIO-HSRP
source_description	The description of the CDM instance.	Instance of the CONSIGN project util
source_documentation_reference	The reference to the source documentation.	https://github.com/frasansa/VID2OM
cdm_etl_reference	Version of the ETL script used. e.g. link	https://github.com/frasansa/VID2OM

Destination Field	Logic	Value in the last instance
	to the Git release.	
source_release_date	The date the data was extracted from the source system.	2022-01-01
cdm_release_date	The date the ETL script was completed. Typically this is after the source_release_date.	2024-06-25
cdm_version	Version of the OMOP CDM used as string. e.g. v5.4	5.4
cdm_version_concept_id	The Concept Id representing the version of the CDM.	798878
vocabulary_version	Version of the OMOP standardised vocabularies loaded	v5.0 31-AUG-23

3.2 Part II Overview

In the Part II of the ETL we perform the following steps:

1. The content of the **condition_occurrence**, **procedure_occurrence**, and **drug_exposure** tables, whose standardized concepts domain was different from their respective table, has been relocated. To achieve this, the target tables **device_exposure** and **observation** have been created.

2. The *sig* and *drug_exposure_end_date* of the **drug_exposure** table are revised for some drugs of interest.

3. The target table **source_to_concept_map** has been created to store the source-to-concept mapping.

3.2.1 Target table: device_exposure

The target table **device_exposure** has been created to include rows from **drug_exposure** and **procedure_occurrence** where the standardized domain is 'Device'.

3.2.1.1 From drug_exposure to device_exposure

The rows in the **device_exposure** table that originate from the **drug_exposure** table are populated using the following code snippet:

```
# from_drug_exposure_vid_0_to_device_exposure--------------------------------
from_drug_exposure_vid_0_to_device_exposure <- drug_exposure_vid_0 |>
  filter(drug_concept_id %in% device_drug_exposure_concept_ids) |>
  arrange(drug_exposure_id) |>
  transmute(
    device_exposure_id = 1:n(),
    person_id = person_id,
    device_concept_id = drug_concept_id,
    device_exposure_start_date = drug_exposure_start_date,
    device_exposure_start_datetime = drug_exposure_start_datetime,
    device_exposure_end_date = drug_exposure_end_date,
    device_exposure_end_datetime = drug_exposure_end_datetime,
    device_type_concept_id = drug_type_concept_id,
    unique_device_id = NA_integer_,
    production_id = NA_integer_,
    quantity = 1,
    provider_id = NA_integer_,
    visit_occurrence_id = NA_integer_,
    visit_detail_id = NA_integer_,
    device_source_value = drug_source_value,
    device_source_concept_id = drug_source_concept_id,
    unit_concept_id = 0,
    unit_source_value = 0,
    unit_source_concept_id = 0
  )
```

3.2.1.2 From procedure_occurrence to device_exposure

The rows in the **device_exposure** table that originate from the **procedure_occurrence** table are populated using the following code snippet:

```
# from_procedure_occurrence_vid_0_to_device_exposure--------------------------
from_procedure_occurrence_vid_0_to_device_exposure <- procedure_occurrence_vid_0 |>
  filter(procedure_concept_id %in% device_procedure_occurrence_concept_ids) |>
  arrange(procedure_occurrence_id) |>
  transmute(
    device_exposure_id = 1:n() + last_drug_exposure_vid_0_device_id,
    person_id = person_id,
    device_concept_id = procedure_concept_id,
    device_exposure_start_date = procedure_date,
    device_exposure_start_datetime = procedure_datetime,
    device_exposure_end_date = procedure_end_date,
    device_exposure_end_datetime = procedure_end_datetime,
    device_type_concept_id = procedure_type_concept_id,
    unique_device_id = NA_integer_,
```

```
    production_id = NA_integer_,
    quantity = 1,
    provider_id = NA_integer_,
    visit_occurrence_id = NA_integer_,
    visit_detail_id = NA_integer_,
    device_source_value = procedure_source_value,
    device_source_concept_id = procedure_source_concept_id,
    unit_concept_id = 0,
    unit_source_value = 0,
    unit_source_concept_id = 0
  )
```

3.2.2 Target table: observation

The target table **observation** has been created to include rows from **condition_occurrence** and **drug_exposure** where the standardized domain is 'Observation'.

3.2.2.1 From condition_occurrence to observation

The rows in the **observation** table that originate from the **condition_occurrence** table are populated using the following code snippet:

```
# from_condition_occurrence_vid_0_to_observation----------------------------
from_condition_occurrence_vid_0_to_observation <- condition_occurrence_vid_0_observat
  transmute(
    observation_id = 1:n(),
    person_id = person_id,
    observation_concept_id = condition_concept_id,
    observation_date = condition_start_date,
    observation_datetime = condition_start_datetime,
    observation_type_concept_id = condition_type_concept_id,
    value_as_number = 0,
    value_as_string = NA_character_,
    value_as_concept_id = value_as_concept_id,
    qualifier_concept_id = 0,
    unit_concept_id = 0,
    provider_id = NA_integer_,
    visit_occurrence_id = visit_occurrence_id,
    visit_detail_id = NA_integer_,
    observation_source_value = condition_source_value,
    observation_source_concept_id = condition_source_concept_id,
    unit_source_value = NA_character_,
    qualifier_source_value = NA_character_,
    value_source_value = NA_character_,
    observation_event_id = NA_integer_,
    obs_event_field_concept_id = 0
  )
```

3.2.2.2 From drug_exposure to observation

The rows in the **observation** table that originate from the **drug_exposure** table are populated using the following code snippet:

```
# from_drug_exposure_vid_0 to observation---------------------------------
from_drug_exposure_vid_0_to_observation <- drug_exposure_vid_0 |>
  filter(drug_concept_id %in% observation_drug_concept_ids) |>
  transmute(
    observation_id = 1:n() + last_condition_occurrence_vid_0_observation_id,
    person_id = person_id,
    observation_concept_id = drug_concept_id,
    observation_date = drug_exposure_start_date,
    observation_datetime = drug_exposure_start_datetime,
    observation_type_concept_id = drug_type_concept_id,
    value_as_number = 0,
    value_as_string = NA_character_,
    value_as_concept_id = 0,
    qualifier_concept_id = 0,
    unit_concept_id = 0,
    provider_id = NA_integer_,
    visit_occurrence_id = visit_occurrence_id,
    visit_detail_id = NA_integer_,
    observation_source_value = drug_source_value,
    observation_source_concept_id = drug_source_concept_id,
    unit_source_value = NA_character_,
    qualifier_source_value = NA_character_,
    value_source_value = NA_character_,
    observation_event_id = NA_integer_,
    obs_event_field_concept_id = 0
  )
```

3.2.3 Target table: source_to_concept_map

The source to concept map table is recommended for use in ETL processes to maintain local source codes which are not available as Concepts in the Standardized Vocabularies, and to establish mappings for each source code into a Standard Concept as target_concept_ids that can be used to populate the Common Data Model tables. The SOURCE_TO_CONCEPT_MAP table is no longer populated with content within the Standardized Vocabularies published to the OMOP community.

3.3 Part III Overview

In this part, the derived tables **condition_era** and **drug_era** are created.

3.3.1 Derived table: condition_era

A Condition Era is defined as a span of time when the Person is assumed to have a given condition. Similar to Drug Eras, Condition Eras are chronological periods of Condition Occurrence. Combining individual Condition Occurrences into a single Condition Era serves two purposes:

It allows aggregation of chronic conditions that require frequent ongoing care, instead of treating each Condition Occurrence as an independent event.

It allows aggregation of multiple, closely timed doctor visits for the same Condition to avoid double-counting the Condition Occurrences. For example, consider a Person who visits her Primary Care Physician (PCP) and who is referred to a specialist. At a later time, the Person visits the specialist, who confirms the PCP's original diagnosis and provides the appropriate treatment to resolve the condition. These two independent doctor visits should be aggregated into one Condition Era.

3.3.2 Derived table: drug_era

A Drug Era is defined as a span of time when the Person is assumed to be exposed to a particular active ingredient. A Drug Era is not the same as a Drug Exposure: Exposures are individual records corresponding to the source when Drug was delivered to the Person, while successive periods of Drug Exposures are combined under certain rules to produce continuous Drug Eras.

4 VOCABULARY MAPPING

Some vocabularies available in ATHENA and some not available in ATHENA have been used for mapping.

4.1 Vocabularies available in ATHENA

This section provides an overview of the vocabulary mapping step. The Table 28 shows the source code vocabularies available in the database, included in ATHENA and downloaded in csv tables.

Table 28: Source vocabularies

Vocabulary	Reference Link	Description	Data Domains
ICD9CM	ICD9CM	International Classification of Diseases, 9th revision, Clinical Modification.	*condition_occurrence*

Vocabulary	Reference Link	Description	Data Domains
ICD10ES	ICD10ES	International Classification of Diseases, 10th revision, Clinical Modification, Spanish Edition.	*condition_occurrence*

4.2 Vocabularies not available in ATHENA

In addition to the ATHENA vocabularies, the vocabularies *VID Drugs*, *VID Routes* and *VID Vaccines* have been used. Their implementation can be found in the script 2_02_ETL_Implementation_part_2_CONSIGN.qmd.

APPENDIX 1. SOURCE DATA MODEL

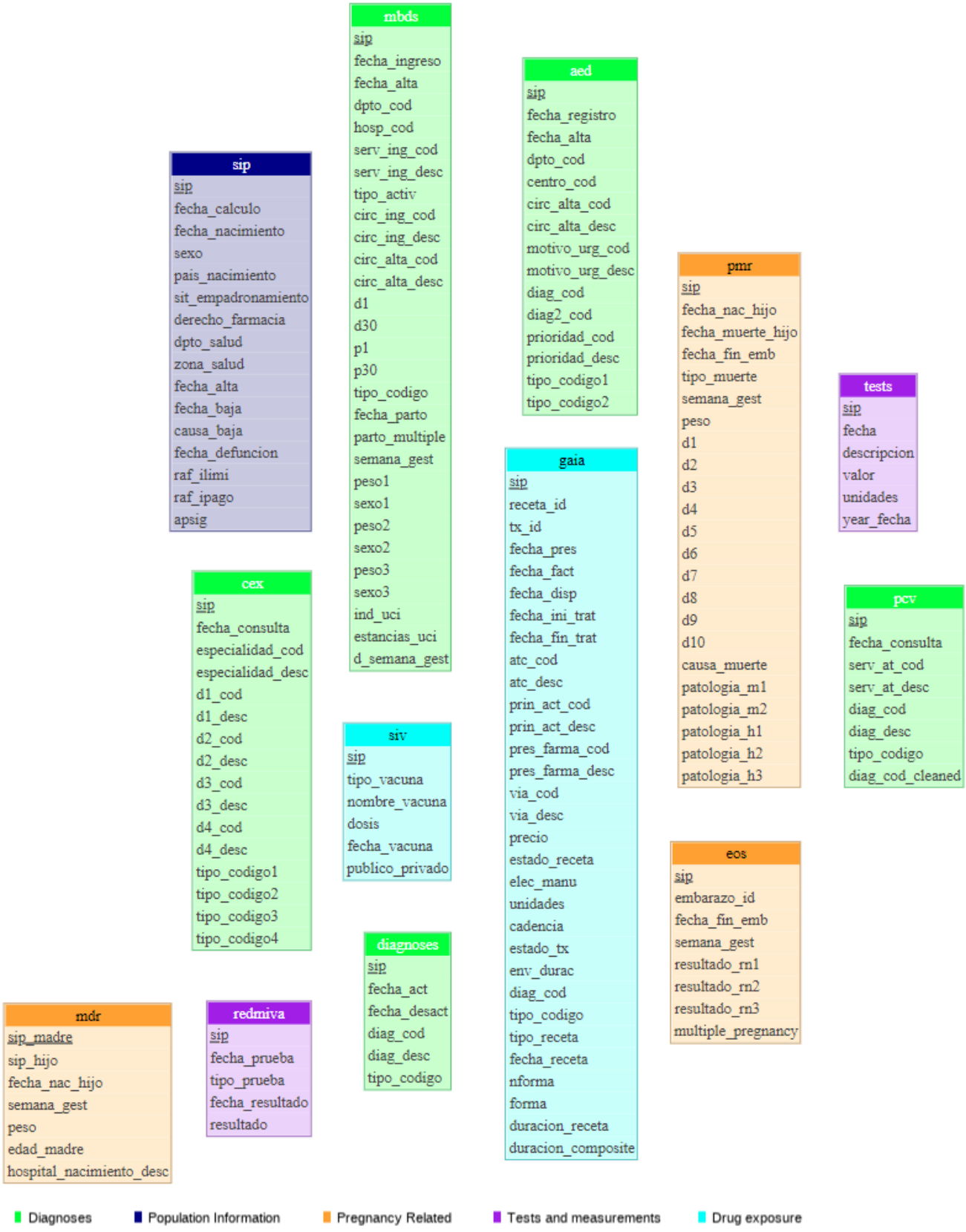


Figure 23: VID data model diagram

APPENDIX 2. DATA DICTIONARY

4.3 Tables

Table 29: Source tables description

Source Table	English Name	Description
SIP	Population Information System	Population and social information of the VID population (such as sex, birth date, income, etc.). A record is created when anyone, resident or foreigner (e.g. tourists), contacts the system. Everyone is assigned an ID that is linkable across the tables. The table is updated each year and there are information from 2008 to current date. This table is used for cohort definition/creation and it is also used to identify deaths.
PCV	Primary Care Visit	Information of the primary care visits (general practice)
CEX	Speciality Visit	Information of the specialist care visits
MBDS	Minimum Basic Data Set	Hospital admission minimum basic data set. Records are triggered by hospital admissions and capture the information about anyone who has an admission, regardless of their residency status.
AED	Accident and Emergency Department	Information of the hospital Accident and Emergency Department visits. The **AED** visits that led to hospitalization can be linked with the **MBDS**.

Source Table	English Name	Description
DIAGNOSES	Diagnoses	Information about the active (and non-active) diagnoses of the population.
GAIA	Pharmaceutical Information	**GAIA** contains the information about Pharmaceutical information. It is the result of the combination about 3 tables (prescription, dispensing and treatment episodes). Each prescription has an individual prescription ID that permits to link prescription and dispensing information. Prescriptions are grouped into treatment episodes that also have a treatment ID.
SIV	Vaccines Information System	**SIV** contains the information about the vaccines. Records of the administration of any type of vaccine are found (those of the usual calendar, those specific of pregnancy, flu campaign, vaccines against the COVID-19, etc.). The database contains more than 50 million of vaccines (records).
MDR	Metabolic Disease Register	Metabolic disease register records. As the register contains the information of all livebirths it can be considered as a birth register. It allows to link the mother person id with the newborn person id.
PMR	Perinatal Mortality Register	Perinatal mortality register records. It contains the information about fetal deaths occurred from 21 gestational weeks and newborn deaths produced to 28 days after birth.

Source Table	English Name	Description
EOS	Electronic Obstetric Sheet	Electronic obstetric sheet. It contains the information about the pregnancy follow-up. It is used in order to detect spontaneous abortions (and to confirm births and stillbirths).
TESTS	Tests	Tests results and measurements, such as blood pressure, height, weight and body mass index (bmi).
CONG	Congenital anomalies	In this base are collected the information about congenital anomalies (This table is not mapped into the OMOP CDM yet).
REDMIVA	Microbiological Surveillance Network	It contains the information about COVID-19 test results.

4.4 Fields

Table 30: Source fields description

Source Table	Field	Type	Description
SIP	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
SIP	*fecha_calculo*	DATE	calculation date (year of the information)
SIP	*fecha_nacimiento*	DATE	birth date
SIP	*sexo*	VARCHAR	sex
SIP	*pais_nacimiento*	VARCHAR	country of birth (INE code + name)

Source Table	Field	Type	Description
SIP	*sit_empadronamiento*	VARCHAR	census situation
SIP	*derecho_farmacia*	VARCHAR	pharmacy rights
SIP	*dpto_salud*	VARCHAR	health department
SIP	*zona_salud*	VARCHAR	health zone
SIP	*fecha_alta*	DATE	activation date
SIP	*fecha_baja*	DATE	deactivation date
SIP	*causa_baja*	VARCHAR	deactivation cause
SIP	*fecha_defuncion*	DATE	defunction date
SIP	*raf_ilimi*	INT	copayment maximum limit
SIP	*raf_ipago*	VARCHAR	copayment percentage category
SIP	*apsig*	VARCHAR	multicomponent sociodemographic code
PCV	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
PCV	*fecha_consulta*	DATE	date of the visit
PCV	*serv_at_cod*	VARCHAR	diagnosis code
PCV	*serv_at_desc*	VARCHAR	diagnosis description
PCV	*diag_cod*	VARCHAR	contact type code
PCV	*diag_desc*	VARCHAR	contact type description
PCV	*tipo_codigo*	VARCHAR	diagnosis code vocabulary
CEX	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
CEX	*fecha_consulta*	DATE	date of the visit
CEX	*especialidad_cod*	VARCHAR	speciality code

Source Table	Field	Type	Description
CEX	*especialidad_desc*	VARCHAR	especiality description
CEX	*d1_cod*	VARCHAR	contact type
CEX	*d1_desc*	VARCHAR	diagnosis code 1
CEX	*d2_cod*	VARCHAR	diagnosis description 1
CEX	*d2_desc*	VARCHAR	diagnosis code 2
CEX	*d3_cod*	VARCHAR	diagnosis description 2
CEX	*d3_desc*	VARCHAR	diagnosis code 3
CEX	*d4_cod*	VARCHAR	diagnosis description 3
CEX	*d4_desc*	VARCHAR	diagnosis code 4
CEX	*tipo_codigo1*	VARCHAR	diagnosis description 4
CEX	*tipo_codigo2*	VARCHAR	diagnosis code 1 vocabulary
CEX	*tipo_codigo3*	VARCHAR	diagnosis code 2 vocabulary
CEX	*tipo_codigo4*	VARCHAR	diagnosis code 3 vocabulary
MBDS	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
MBDS	*fecha_ingreso*	DATE	date of the hospitalisation admission
MBDS	*fecha_alta*	DATE	date of the hospitalisation discharge
MBDS	*dpto_cod*	VARCHAR	health department code
MBDS	*hosp_cod*	INT	hospital code

Source Table	Field	Type	Description
MBDS	*serv_ing_cod*	VARCHAR	admission service code
MBDS	*serv_ing_desc*	VARCHAR	admission service description
MBDS	*tipo_activ*	VARCHAR	activity type: ambulatory or overnight
MBDS	*circ_ing_cod*	INT	admission circumstances code
MBDS	*circ_ing_desc*	VARCHAR	admission circumstances description
MBDS	*circ_alta_cod*	INT	discharge circumstances code
MBDS	*circ_alta_desc*	VARCHAR	discharge circumstances code
MBDS	*d1*	VARCHAR	main diagnosis of the admission (d1)
MBDS	*d2*	VARCHAR	secondary diagnosis (d2)
MBDS	*d3*	VARCHAR	secondary diagnosis (d3)
MBDS	*d4*	VARCHAR	secondary diagnosis (d4)
MBDS	*d5*	VARCHAR	secondary diagnosis (d5)
MBDS	*d6*	VARCHAR	secondary diagnosis (d6)
MBDS	*d7*	VARCHAR	secondary diagnosis (d7)
MBDS	*d8*	VARCHAR	secondary diagnosis (d8)
MBDS	*d9*	VARCHAR	secondary diagnosis (d9)

Source Table	Field	Type	Description
MBDS	*d10*	VARCHAR	secondary diagnosis (d10)
MBDS	*d11*	VARCHAR	secondary diagnosis (d11)
MBDS	*d12*	VARCHAR	secondary diagnosis (d12)
MBDS	*d13*	VARCHAR	secondary diagnosis (d13)
MBDS	*d14*	VARCHAR	secondary diagnosis (d14)
MBDS	*d15*	VARCHAR	secondary diagnosis (d15)
MBDS	*d16*	VARCHAR	secondary diagnosis (d16)
MBDS	*d17*	VARCHAR	secondary diagnosis (d17)
MBDS	*d18*	VARCHAR	secondary diagnosis (d18)
MBDS	*d19*	VARCHAR	secondary diagnosis (d19)
MBDS	*d20*	VARCHAR	secondary diagnosis (d20)
MBDS	*d21*	VARCHAR	secondary diagnosis (d21)
MBDS	*d22*	VARCHAR	secondary diagnosis (d22)
MBDS	*d23*	VARCHAR	secondary diagnosis (d23)
MBDS	*d24*	VARCHAR	secondary diagnosis (d24)
MBDS	*d25*	VARCHAR	secondary diagnosis (d25)
MBDS	*d26*	VARCHAR	secondary diagnosis (d26)

Source Table	Field	Type	Description
MBDS	*d27*	VARCHAR	secondary diagnosis (d27)
MBDS	*d28*	VARCHAR	secondary diagnosis (d28)
MBDS	*d29*	VARCHAR	secondary diagnosis (d29)
MBDS	*d30*	VARCHAR	secondary diagnosis (d30)
MBDS	*p1*	VARCHAR	main procedure in the admission (p1)
MBDS	*p2*	VARCHAR	secondary procedure (p2)
MBDS	*p3*	VARCHAR	secondary procedure (p3)
MBDS	*p4*	VARCHAR	secondary procedure (p4)
MBDS	*p5*	VARCHAR	secondary procedure (p5)
MBDS	*p6*	VARCHAR	secondary procedure (p6)
MBDS	*p7*	VARCHAR	secondary procedure (p7)
MBDS	*p8*	VARCHAR	secondary procedure (p8)
MBDS	*p9*	VARCHAR	secondary procedure (p9)
MBDS	*p10*	VARCHAR	secondary procedure (p10)
MBDS	*p11*	VARCHAR	secondary procedure (p11)
MBDS	*p12*	VARCHAR	secondary procedure (p12)
MBDS	*p13*	VARCHAR	secondary procedure (p13)

Source Table	Field	Type	Description
MBDS	*p14*	VARCHAR	secondary procedure (p14)
MBDS	*p15*	VARCHAR	secondary procedure (p15)
MBDS	*p16*	VARCHAR	secondary procedure (p16)
MBDS	*p17*	VARCHAR	secondary procedure (p17)
MBDS	*p18*	VARCHAR	secondary procedure (p18)
MBDS	*p19*	VARCHAR	secondary procedure (p19)
MBDS	*p20*	VARCHAR	secondary procedure (p20)
MBDS	*p21*	VARCHAR	secondary procedure (p21)
MBDS	*p22*	VARCHAR	secondary procedure (p22)
MBDS	*p23*	VARCHAR	secondary procedure (p23)
MBDS	*p24*	VARCHAR	secondary procedure (p24)
MBDS	*p25*	VARCHAR	secondary procedure (p25)
MBDS	*p26*	VARCHAR	secondary procedure (p26)
MBDS	*p27*	VARCHAR	secondary procedure (p27)
MBDS	*p28*	VARCHAR	secondary procedure (p28)
MBDS	*p29*	VARCHAR	secondary procedure (p29)
MBDS	*p30*	VARCHAR	secondary procedure (p30)

Source Table	Field	Type	Description
MBDS	*tipo_codigo*	VARCHAR	diagnosis code vocabulary
MBDS	*fecha_parto*	DATE	labor date
MBDS	*parto_multiple*	INT	multiple labor
MBDS	*semana_gest*	INT	gestational age (in weeks)
MBDS	*peso1*	INT	newborn1 weight (in g)
MBDS	*sexo1*	VARCHAR	sex of newborn1
MBDS	*peso2*	INT	newborn1 weight (in g)
MBDS	*sexo2*	VARCHAR	sex of newborn2
MBDS	*peso3*	INT	newborn1 weight (in g)
MBDS	*sexo3*	VARCHAR	sex of newborn3
AED	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
AED	*fecha_registro*	DATE	date of emergency room visit record
AED	*fecha_alta*	DATE	date of emergency room discharge
AED	*dpto_cod*	INT	health department code
AED	*centro_cod*	INT	centre code
AED	*circ_alta_cod*	INT	discharge circumstances code
AED	*circ_alta_desc*	VARCHAR	discharge circumstances code
AED	*motivo_urg_cod*	INT	emergency admission code

Source Table	Field	Type	Description
AED	*motivo_urg_desc*	VARCHAR	emergency admission description
AED	*diag_cod*	VARCHAR	diagnosis code 1
AED	*diag2_cod*	VARCHAR	diagnosis code 2
AED	*tipo_codigo1*	VARCHAR	diagnosis code 1 vocabulary
AED	*tipo_codigo2*	VARCHAR	diagnosis code 2 vocabulary
AED	*prioridad_cod*	INT	priority code
AED	*prioridad_desc*	VARCHAR	priority description
DIAGNOSES	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
DIAGNOSES	*fecha_act*	DATE	date of diagnosis activation
DIAGNOSES	*fecha_desact*	DATE	date of diagnosis deactivation
DIAGNOSES	*diag_cod*	VARCHAR	diagnosis code
DIAGNOSES	*diag_desc*	VARCHAR	diagnosis description
DIAGNOSES	*tipo_codigo*	VARCHAR	diagnosis code vocabulary
PRES	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
PRES	*receta_id*	VARCHAR	pseudonymised prescription id, which links prescription and dispensing information
PRES	*tx_id*	VARCHAR	pseudonymised treatment id, which links prescription and treatment information

Source Table	Field	Type	Description
PRES	*fecha_pres*	DATE	prescription date
PRES	*atc_cod*	VARCHAR	level 4 (5 digits) or level 5 (7 digits) atc code
PRES	*atc_desc*	VARCHAR	level 4 (5 digits) or level 5 (7 digits) atc code
PRES	*prin_act_cod*	VARCHAR	active ingredient code
PRES	*prin_act_desc*	VARCHAR	active ingredient description
PRES	*pres_farma_cod*	INT	pharmaceutical presentation code
PRES	*pres_farma_desc*	VARCHAR	pharmaceutical presentation description
PRES	*via_cod*	VARCHAR	route of administration code
PRES	*via_desc*	VARCHAR	route of administration description
PRES	*precio*	REAL	cost of the product (in euros)
PRES	*estado_receta*	VARCHAR	prescription state
PRES	*receta_abucasis*	VARCHAR	electronic or manual prescription
FACT	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
FACT	*receta_id*	VARCHAR	pseudonymised prescription id, which links prescription and dispensing information
FACT	*fecha_fact*	DATE	billing dispensing date (year and moth)

Source Table	Field	Type	Description
FACT	*fecha_disp*	DATE	dispensing date (year, month and day)
FACT	*atc_cod*	VARCHAR	level 4 (5 digits) or level 5 (7 digits) atc code
FACT	*atc_desc*	VARCHAR	level 4 (5 digits) or level 5 (7 digits) atc code
FACT	*prin_act_cod*	VARCHAR	active ingredient code
FACT	*prin_act_desc*	VARCHAR	active ingredient description
FACT	*pres_farma_cod*	INT	pharmaceutical presentation code
FACT	*pres_farma_desc*	VARCHAR	pharmaceutical presentation description
FACT	*via_cod*	VARCHAR	route of administration code
FACT	*via_desc*	VARCHAR	route of administration description
RELE	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
RELE	*receta_id*	VARCHAR	pseudonymised prescription id, which links prescription and dispensing information
RELE	*fecha_fact*	DATE	billing dispensing date (year and moth)
RELE	*fecha_disp*	DATE	dispensing date (year, month and day)

Source Table	Field	Type	Description
RELE	*atc_cod*	VARCHAR	level 4 (5 digits) or level 5 (7 digits) atc code
RELE	*atc_desc*	VARCHAR	level 4 (5 digits) or level 5 (7 digits) atc code
RELE	*prin_act_cod*	VARCHAR	active ingredient code
RELE	*prin_act_desc*	VARCHAR	active ingredient description
RELE	*pres_farma_cod*	INT	pharmaceutical presentation code
RELE	*pres_farma_desc*	VARCHAR	pharmaceutical presentation description
RELE	*via_cod*	VARCHAR	route of administration code
RELE	*via_desc*	VARCHAR	route of administration description
TX	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
TX	*tx_id*	VARCHAR	pseudonymised treatment id, which links prescription and treatment information
TX	*unidades*	VARCHAR	dosing units
TX	*cadencia*	INT	dosing (in hours)
TX	*estado_tx*	VARCHAR	treatment state
TX	*fecha_ini_trat*	DATE	date of treatment start
TX	*fecha_fin_trat*	DATE	date of treatment end

Source Table	Field	Type	Description
TX	*atc_cod*	VARCHAR	level 4 (5 digits) or level 5 (7 digits) atc code
TX	*atc_desc*	VARCHAR	level 4 (5 digits) or level 5 (7 digits) atc code
TX	*prin_act_cod*	VARCHAR	active ingredient code
TX	*prin_act_desc*	VARCHAR	active ingredient description
TX	*pres_farma_cod*	INT	pharmaceutical presentation code
TX	*pres_farma_desc*	VARCHAR	pharmaceutical presentation description
TX	*via_cod*	VARCHAR	route of administration code
TX	*via_desc*	VARCHAR	route of administration description
TX	*diag_cod*	VARCHAR	diagnosis code for the treatment
TX	*tipo_codigo*	VARCHAR	diagnosis code vocabulary
GAIA	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
GAIA	*receta_id*	VARCHAR	pseudonymised prescription id, which links prescription and dispensing information
GAIA	*tx_id*	VARCHAR	pseudonymised treatment id, which links prescription and treatment information

Source Table	Field	Type	Description
GAIA	*fecha_pres*	DATE	prescription date
GAIA	*fecha_fact*	DATE	billing dispensing date (year and moth)
GAIA	*fecha_disp*	DATE	dispensing date (year, month and day)
GAIA	*fecha_ini_trat*	DATE	date of treatment start
GAIA	*fecha_fin_trat*	DATE	date of treatment end
GAIA	*atc_cod*	VARCHAR	level 4 (5 digits) or level 5 (7 digits) atc code
GAIA	*atc_desc*	VARCHAR	level 4 (5 digits) or level 5 (7 digits) atc code
GAIA	*prin_act_cod*	VARCHAR	active ingredient code
GAIA	*prin_act_desc*	VARCHAR	active ingredient description
GAIA	*pres_farma_cod*	INT	pharmaceutical presentation code
GAIA	*pres_farma_desc*	VARCHAR	pharmaceutical presentation description
GAIA	*via_cod*	VARCHAR	route of administration code
GAIA	*via_desc*	VARCHAR	route of administration description
GAIA	*precio*	REAL	value of the product (in euros)
GAIA	*estado_receta*	VARCHAR	prescription state
GAIA	*elec_manu*	VARCHAR	electronic or manual prescription
GAIA	*unidades*	VARCHAR	dosing units

Source Table	Field	Type	Description
GAIA	*cadencia*	INT	dosing (in hours)
GAIA	*estado_trat*	VARCHAR	treatment state
GAIA	*env_durac*	REAL	'in origin' estimation of the prescription duration
GAIA	*diag_cod*	VARCHAR	diagnosis code for the treatment
GAIA	*tipo_codigo*	VARCHAR	diagnosis code vocabulary
GAIA	*tipo_receta*	INT	information available of the prescription: prescription, dispensing or both
GAIA	*fecha_receta*	DATE	date of the prescription calculated by FISABIO-HSRP
GAIA	*nforma*	INT	number of product forms
GAIA	*forma*	VARCHAR	type of product forms
GAIA	*duracion_receta*	REAL	estimation of the prescription duration performed by FISABIO-HSRP
GAIA	*duracion_composite*	VARCHAR	when available 'env_durac', otherwise 'duracion_receta'
SIV	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
SIV	*tipo_vacuna*	VARCHAR	type of vaccine (COV-2, Flu, etc.)
SIV	*nombre_vacuna*	VARCHAR	vaccine brand name
SIV	*dosis*	INT	dose number
SIV	*fecha_vacuna*	DATE	vaccination date

Source Table	Field	Type	Description
SIV	*publico_privado*	VARCHAR	payer of the vaccine (public or private)
MDR	*sip_madre*	VARCHAR	pseudonymised id number (unique for each patient) of the mother
MDR	*sip_hijo*	VARCHAR	pseudonymised id number (unique for each patient) of the newborn
MDR	*fecha_fin_emb*	DATE	date of the birth (and end of pregnancy)
MDR	*semana_gest*	INT	gestational age (in weeks)
MDR	*peso*	INT	newborn weight (in g)
MDR	*edad_madre*	INT	mother age (in years)
MDR	*hospital_nacimiento_cod*	INT	birth hospital code
MDR	*hospital_nacimiento_desc*	VARCHAR	birth hospital name
MDR	*hospital_muestra_cod*	INT	results hospital code
MDR	*hospital_muestra_desc*	VARCHAR	results hospital name
PMR	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
PMR	*tipo_muerte*	VARCHAR	type of death (neonatal or fetal)
PMR	*fecha_muerte_hijo*	DATE	date of newborn/fetus death
PMR	*fecha_nac_hijo*	DATE	date of newborn birth
PMR	*semana_gest*	INT	gestational age (in weeks)
PMR	*peso*	INT	newborn weight (in g)
PMR	*d1*	VARCHAR	diagnosis code 1

Source Table	Field	Type	Description
PMR	*d2*	VARCHAR	diagnosis code 2
PMR	*d3*	VARCHAR	diagnosis code 3
PMR	*d4*	VARCHAR	diagnosis code 4
PMR	*d5*	VARCHAR	diagnosis code 5
PMR	*d6*	VARCHAR	diagnosis code 6
PMR	*d7*	VARCHAR	diagnosis code 7
PMR	*d8*	VARCHAR	diagnosis code 8
PMR	*d9*	VARCHAR	diagnosis code 9
PMR	*d10*	VARCHAR	diagnosis code 10
PMR	*causa_muerte*	VARCHAR	death cause
PMR	*patologia_m1*	VARCHAR	mother patology1
PMR	*patologia_m2*	VARCHAR	mother patology2
PMR	*patologia_h1*	VARCHAR	newborn patology1
PMR	*patologia_h2*	VARCHAR	newborn patology2
PMR	*patologia_h3*	VARCHAR	newborn patology3
EOS	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
EOS	*embarazo_id*	VARCHAR	pseudonymised pregnancy id number (unique for each pregnancy)
EOS	*fecha_visita_emb*	DATE	date of record
EOS	*semana_gest*	INT	gestational age (in weeks)
EOS	*fecha_fin_emb*	DATE	date of event
EOS	*resultado_rn1*	VARCHAR	event type of the first child delivered: birth, spontaneous abortion or stillbirth

Source Table	Field	Type	Description
EOS	*resultado_rn2*	VARCHAR	event type of the second (if apply) child delivered: birth, spontaneous abortion or stillbirth
EOS	*resultado_rn3*	VARCHAR	event type of the third (if apply) child delivered: birth, spontaneous abortion or stillbirth
EOS	*multiple_pregnancy*	VARCHAR	number of fetuses in the pregnancy: singleton, twins or triplets
TESTS	*sip*	VARCHAR	pseudonymised id number (unique for each patient) of the mother
TESTS	*fecha*	DATE	date of test or measurement
TESTS	*descripcion*	VARCHAR	type of test or measurement
TESTS	*valor*	FLOAT	value of test or measurement
TESTS	*unidades*	VARCHAR	units of result or measurement
CONG	*sip*	VARCHAR	pseudonymised id number (unique for each patient) of the mother
CONG	*numloc*	VARCHAR	local ID
CONG	*birth_date*	DATE	date of birth
CONG	*sex*	VARCHAR	sex
CONG	*nbrbaby*	INT	number of babies/fetuses delivered

Source Table	Field	Type	Description
CONG	*nbrmalf*	INT	number of malformed in multiple set
CONG	*type*	INT	type of Birth
CONG	*weight*	INT	birth weight
CONG	*semana_gest*	INT	length of gestation in completed weeks
CONG	*death_date*	DATE	date of death
CONG	*datemo*	DATE	date of birth of mother
CONG	*agemo*	INT	age of mother at delivery
CONG	*totpreg*	VARCHAR	total number of previous pregnancies
CONG	*whendisc*	INT	when discovered
CONG	*condisc*	INT	condition at discovery
CONG	*agedisc*	INT	if prenatally diagnosed, gestational age at discovery
CONG	*firstpre*	INT	first positive prenatal test
CONG	*sp_firstpre*	VARCHAR	specify first prenatal test in text if coded 7 ('other test positive')
CONG	*syndrome*	VARCHAR	syndrome
CONG	*sp_syndrome*	VARCHAR	specify syndrome
CONG	*malfo1*	VARCHAR	malformation 1 ICD code
CONG	*sp_malfo1*	VARCHAR	specify malformation 1
CONG	*malfo2*	VARCHAR	malformation 2 ICD code

Source Table	Field	Type	Description
CONG	*sp_malfo2*	VARCHAR	specify malformation 2
CONG	*malfo3*	VARCHAR	malformation 3 ICD code
CONG	*sp_malfo3*	VARCHAR	specify malformation 3
CONG	*malfo4*	VARCHAR	malformation 4 ICD code
CONG	*sp_malfo4*	VARCHAR	specify malformation 4
CONG	*malfo5*	VARCHAR	malformation 5 ICD code
CONG	*sp_malfo5*	VARCHAR	specify malformation 5
CONG	*malfo6*	VARCHAR	malformation 6 ICD code
CONG	*sp_malfo6*	VARCHAR	specify malformation 6
CONG	*malfo7*	VARCHAR	malformation 7 ICD code
CONG	*sp_malfo7*	VARCHAR	specify malformation 7
CONG	*malfo8*	VARCHAR	malformation 8 ICD code
CONG	*sp_malfo8*	VARCHAR	Specify malformation 8
CONG	*presyn*	INT	prenatal diagnosis for syndrome
CONG	*premal1*	INT	prenatal diagnosis for malformation 1
CONG	*premal2*	INT	prenatal diagnosis for malformation 2
CONG	*premal3*	INT	prenatal diagnosis for malformation 3

Source Table	Field	Type	Description
CONG	*premal4*	INT	prenatal diagnosis for malformation 4
CONG	*premal5*	INT	prenatal diagnosis for malformation 5
CONG	*premal6*	INT	prenatal diagnosis for malformation 6
CONG	*premal7*	INT	prenatal diagnosis for malformation 7
CONG	*premal8*	INT	prenatal diagnosis for malformation 8
CONG	*illbef1*	INT	illness before pregnancy 1
CONG	*illbef2*	INT	illness before pregnancy 2
CONG	*illdur1*	INT	illness during pregnancy 1
CONG	*illdur2*	INT	illness during pregnancy 2
CONG	*folic_g14*	INT	folic acid supplementation
CONG	*firsttri*	INT	first trimester medication
CONG	*drugs1*	VARCHAR	drug 1
CONG	*sp_drugs1*	VARCHAR	specify drug exposure 1
CONG	*drugs2*	VARCHAR	drug 2
CONG	*sp_drugs2*	VARCHAR	specify drug exposure 2
CONG	*drugs3*	VARCHAR	drug 3
CONG	*sp_drugs3*	VARCHAR	specify drug exposure 3
CONG	*drugs4*	VARCHAR	drug 4

Source Table	Field	Type	Description
CONG	*sp_drugs4*	VARCHAR	specify drug exposure 4
CONG	*drugs5*	VARCHAR	drug 5
CONG	*sp_drugs5*	VARCHAR	specify drug exposure 5
CONG	*extra_drugs*	VARCHAR	extra drugs
CONG	*pre_live*	INT	number of livebirths born before
CONG	*datedisc*	DATE	date of the discovery
CONG	*soinfo1*	INT	source of information 1
CONG	*soinfo2*	INT	source of information 2
CONG	*soinfo3*	INT	source of information 3
CONG	*soinfo4*	INT	source of information 4
CONG	*tot_malf*	INT	number total of malformations
CONG	*datedisc2*	DATE	date of the discovery cleaned
CONG	*death_date2*	DATE	death date without 'X' characters
CONG	*fecha_fin_emb*	DATE	date of end of pregnancy
CONG	*tipo_fin_emb*	VARCHAR	type of end of pregnancy
REDMIVA	*sip*	VARCHAR	pseudonymised id number (unique for each patient)
REDMIVA	*tipo_prueba*	VARCHAR	test type: Antigen or PCR
REDMIVA	*fecha_prueba*	DATE	date of the test

Source Table	Field	Type	Description
REDMIVA	*fecha_resultado*	DATE	date of the result
REDMIVA	*resultado*	VARCHAR	result of the test