FISABIO-HSRP OMOP ETL design v1.1

AUTHOR
Francisco Sanchez-Saez

PUBLISHED
April 13, 2023

DOCUMENT HISTORY

Version Number	Date	Changes
1.0	2023-02-23	
1.1	2023-04-06	The pipeline of FISABIO-HSRP for the ETL VID-OMOP has been updated: the curation of the source tables before the transformation process has been enhanced In addition, a source data quality check has been implemented to distinguish between potential issues in the source data to potential issues in the transformation to the OMOP CDM process. Particularly, the next changes has been done: 1. The source data dictionary has been updated: • The **GAIA** table, which is a processed table has been redisigned, and the source tables that make it up has been added (**pres**, **fact**, **rele**, and **tx**). • The **CONG** table, which contains the information about congenital anomalies, has been added. • A new column: *Mandatory*, which indicates if it is mandatory to extract a variable in the source table, has been added. 2. Some variable names of the source tables has been updated at the data dictionary and also in the data mapping section: • In **CEX**: ◦ '*fecha*' has been updated to '*fecha_consulta*'. • In **AED**: ◦ '*momento_reg*' has been updated to '*fecha_reg*'. ◦ '*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*'.

Version Number	Date	Changes

- '*tipo_fin_rn3*' has been updated to '*resultado_rn3*'.
- '*multiple_pregnancy*' variable has been removed as this information can be obtained knowing if '*resultado_rn2*' is present.

- In **MDR**:
  - '*sip*' has been updated to '*sip_madre*'.

3. The figure of the Appendix 1 has been modified.

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. 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. These are: SIP, PCV, CEX, MBDS, AED, DIAGNOSES, GAIA, SIV, MDR, PMR, EOS, and REDMIVA.

2 TECHNICAL INFRASTRUCTURE

The ETL will be performed in R 4.2.2 using a machine with 80 GB of RAM. Data transformation will be done using *Tidyverse* (*dplyr*) and *data.table* packages. The target tables will be saved during the process in *.csv* files. Once the quality of the ETL will be assessed, the tables will be stored in a *PostgreSQL* database.

The VID extractions are obtained specifically for each study. Therefore, we will develop the ETL pipeline in a particular study. The selected study has been 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. Then, in order to validate the generalization of the pipeline, the ETL will be performed in another project (Opioids project). In the Opioids project the extraction is done for another setting (2 143 683 individuals), and for another period (2010-2018). It is worth to note the Opioids project does not contain all the tables of the ETL design (the pregnancy and Covid related tables are not used).

In order to test the logic of the ETL design, first we will do the implementation in a sample of 1,000 women of the Consign study and we will perform several unit tests. When the results will be satisfactory, we will do the implementation in the whole cohort. Finally, we will repeat this procedure in the Opioids project.

3 DATA MAPPING

The data mapping has been designed with the help of Rabbit-in-a-hat.

3.1 Overview

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

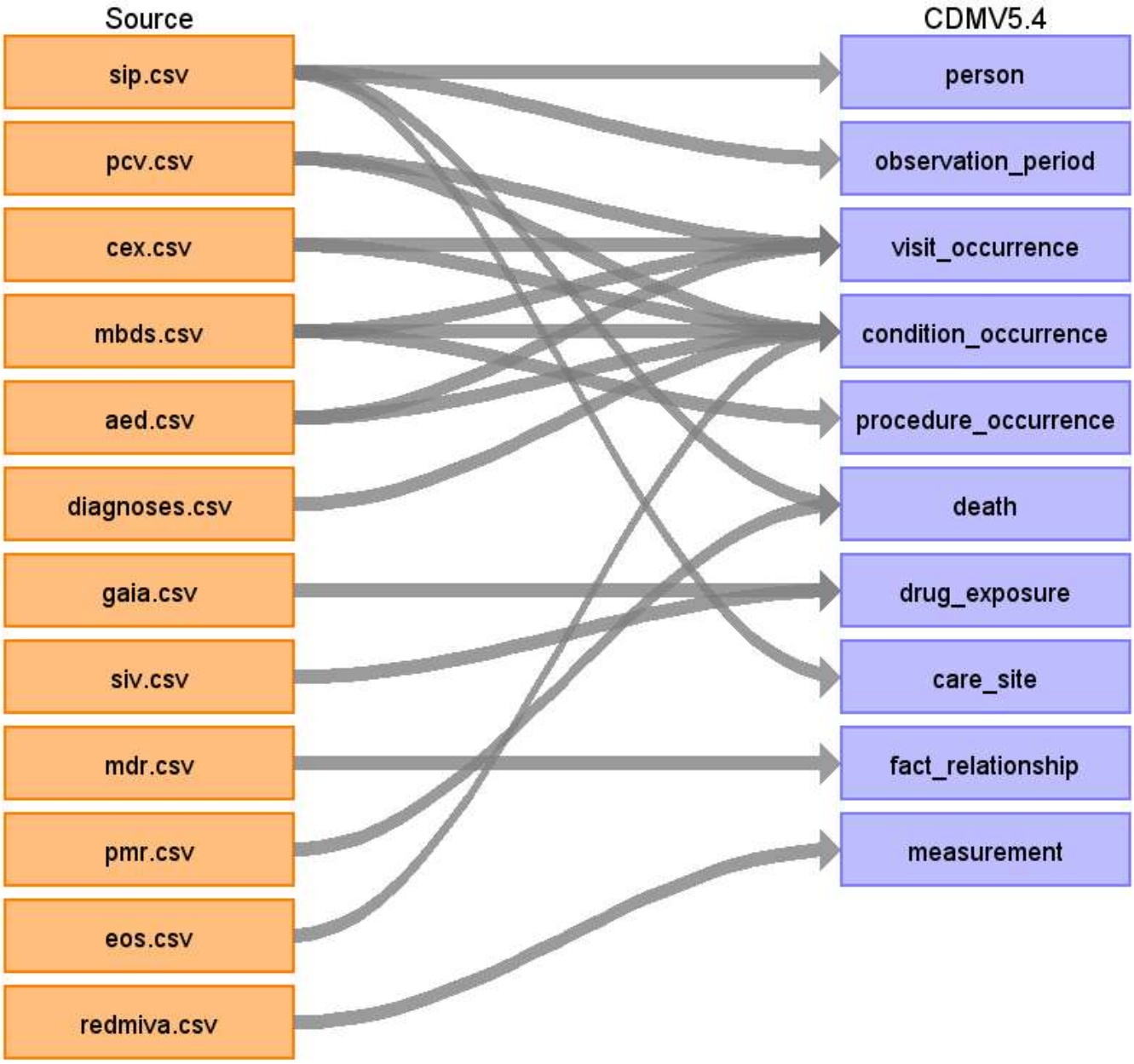


Figure 2: Overview of VID-OMOP CDM v5.4 ETL diagram

3.2 Target table: person

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

3.2.1 From sip to person

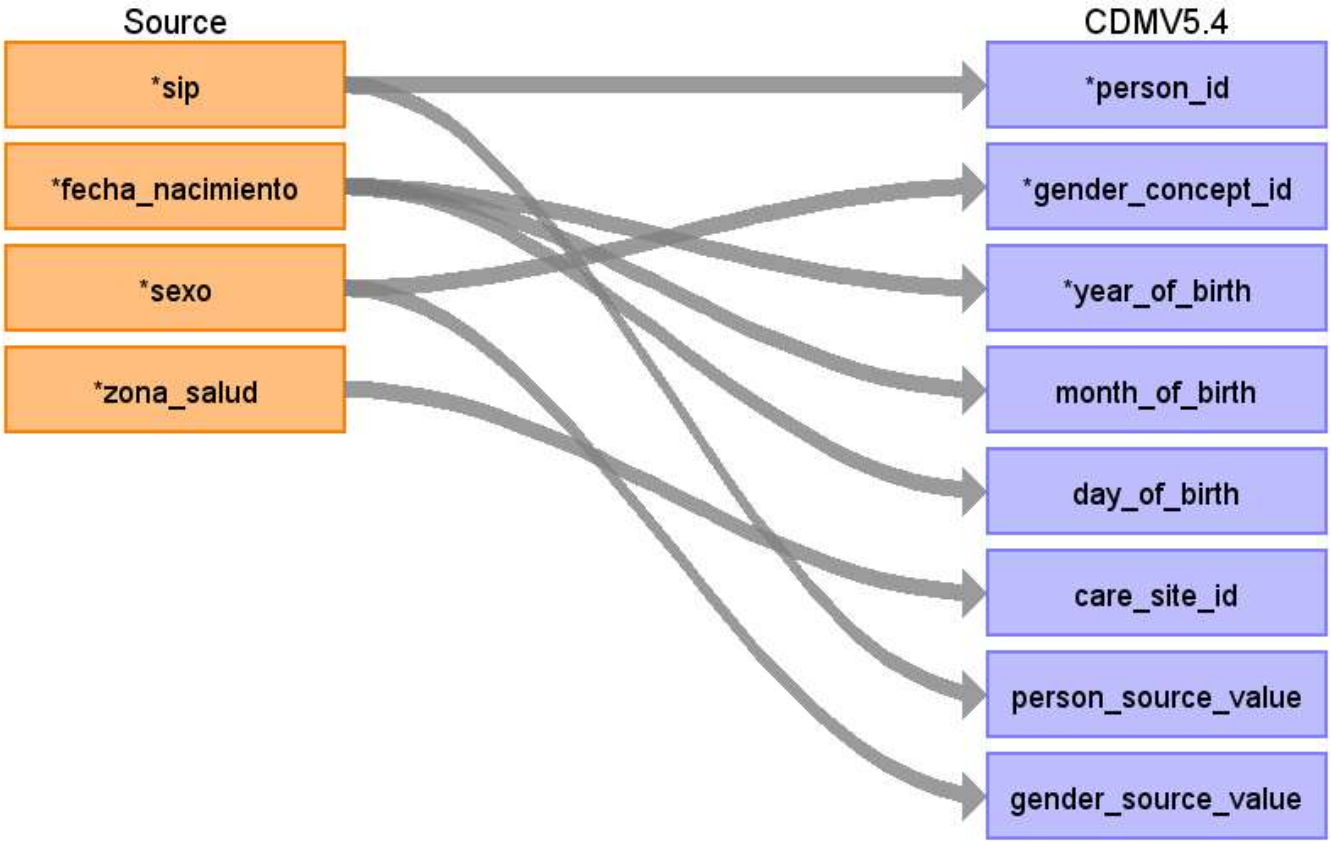


Figure 3: sip to person mapping diagram

Table 1: 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 == 'H' ~ 8507, sexo == 'M' ~ 8532, T ~ NA)	When sexo is H, the gender_concept_id is 8507: Male. When sexo is M the gender_concept_id is 8532: Female. Otherwise, drop person: after the mapping of the gender_concept_id, filter the table with: person |> filter(!is.na(gender_concept_i
year_of_birth	fecha_nacimiento	as.numeric(str_sub(fecha_nacimiento,1,4))	fecha_nacimiento is a Date format variable (YYYY-mm-d

Destination Field	Source Field	Logic	Comment
month_of_birth	fecha_nacimiento	as.numeric(str_sub(fecha_nacimiento,6,7))	fecha_nacimiento is a Date format variable (YYYY-mm-d
day_of_birth	fecha_nacimiento	as.numeric(str_sub(fecha_nacimiento,9,10))	fecha_nacimiento is a Date format variable (YYYY-mm-d
birth_datetime			NULL
race_concept_id			0
ethnicity_concept_id			0
location_id			NULL
provider_id			NULL
care_site_id	zona_salud	zona_salud is converted to a care_site_id	
person_source_value	sip		
gender_source_value	sexo		
gender_source_concept_id			0
race_source_value			NULL
race_source_concept_id			0
ethnicity_source_value			NULL
ethnicity_source_concept_id			0

3.3 Target table: observation_period

The target table **observatio_period** is populated with the information from the source table **sip**.

3.3.1 From sip to observation_period

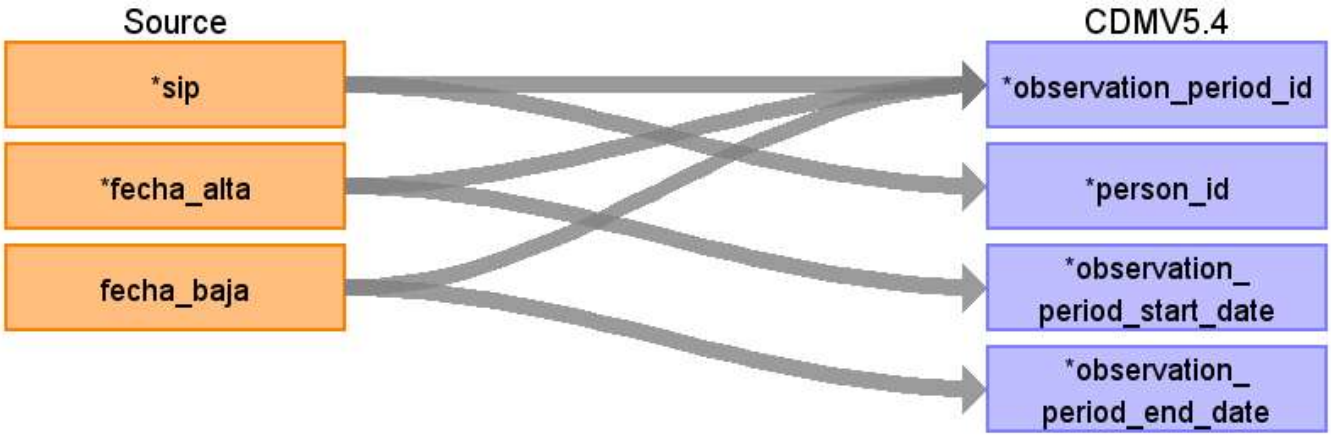


Figure 4: sip to observation_period mapping diagram

Table 2: 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), so the category that best fits is 32827: EHR encounter record.

3.4 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:

- pcv_to_visit_occurrence.
- cex_to_visit_occurrence.
- mbds_to_visit_occurrence.
- aed_to_visit_occurrence.

3.4.1 From pcv to visit_occurrence

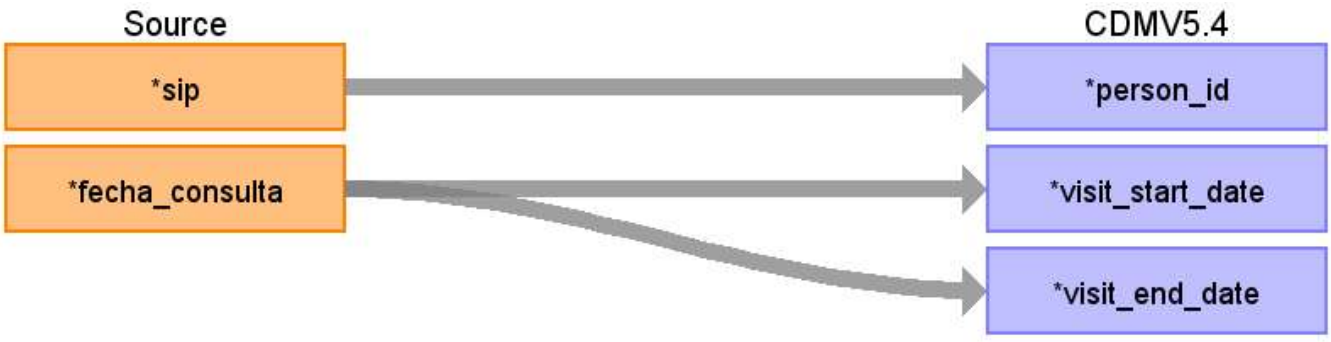


Figure 5: pcv to visit_occurrence mapping diagram

Table 3: 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		

Destination Field	Source Field	Logic	Comment
visit_concept_id			PCV are primary care visits. The Concept ID is 9202: Outpatient Visit.
visit_start_date	fecha_consulta		
visit_start_datetime			NULL
visit_end_date	fecha_consulta		
visit_end_datetime			NULL
visit_type_concept_id			PCV are primary care visits. The Concept ID is 32834: EHR outpatient note.
provider_id			NULL
care_site_id			NULL
visit_source_value			PCV
visit_source_concept_id			0
admitted_from_concept_id			0
admitted_from_source_value			NULL
discharged_to_concept_id			0
discharged_to_source_value			NULL
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.

3.4.2 From cex to visit_occurrence

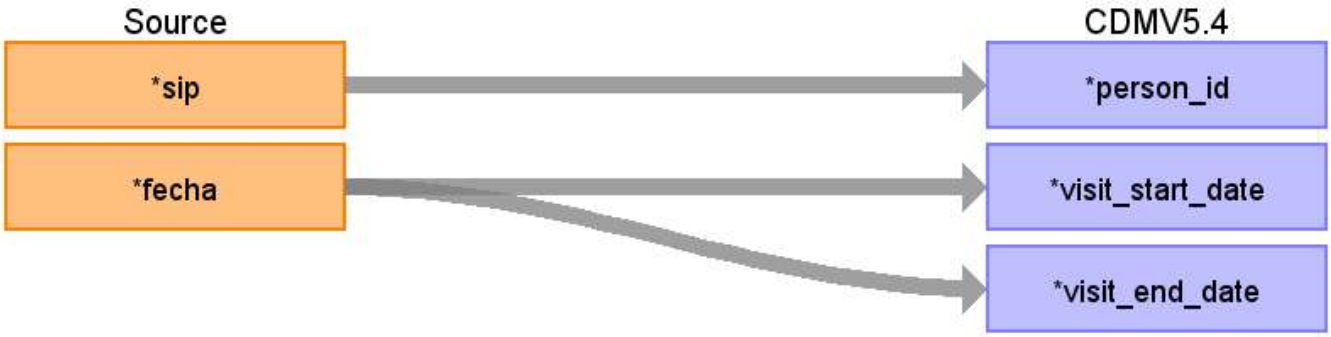


Figure 6: cex to visit_occurrence mapping diagram

Table 4: 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			CEX are specialist care visits. The Concept ID is 9202: Outpatient Visit.
visit_start_date	fecha_consulta		
visit_start_datetime			NULL
visit_end_date	fecha_consulta		
visit_end_datetime			NULL
visit_type_concept_id			CEX are specialist care visits. The Concept ID is 32834: EHR outpatient note.
provider_id			NULL
care_site_id			NULL
visit_source_value			CEX
visit_source_concept_id			0
admitted_from_concept_id			0
admitted_from_source_value			NULL
discharged_to_concept_id			0

Destination Field	Source Field	Logic	Comment
discharged_to_source_value			NULL
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.

3.4.3 From mbds to visit_occurrence

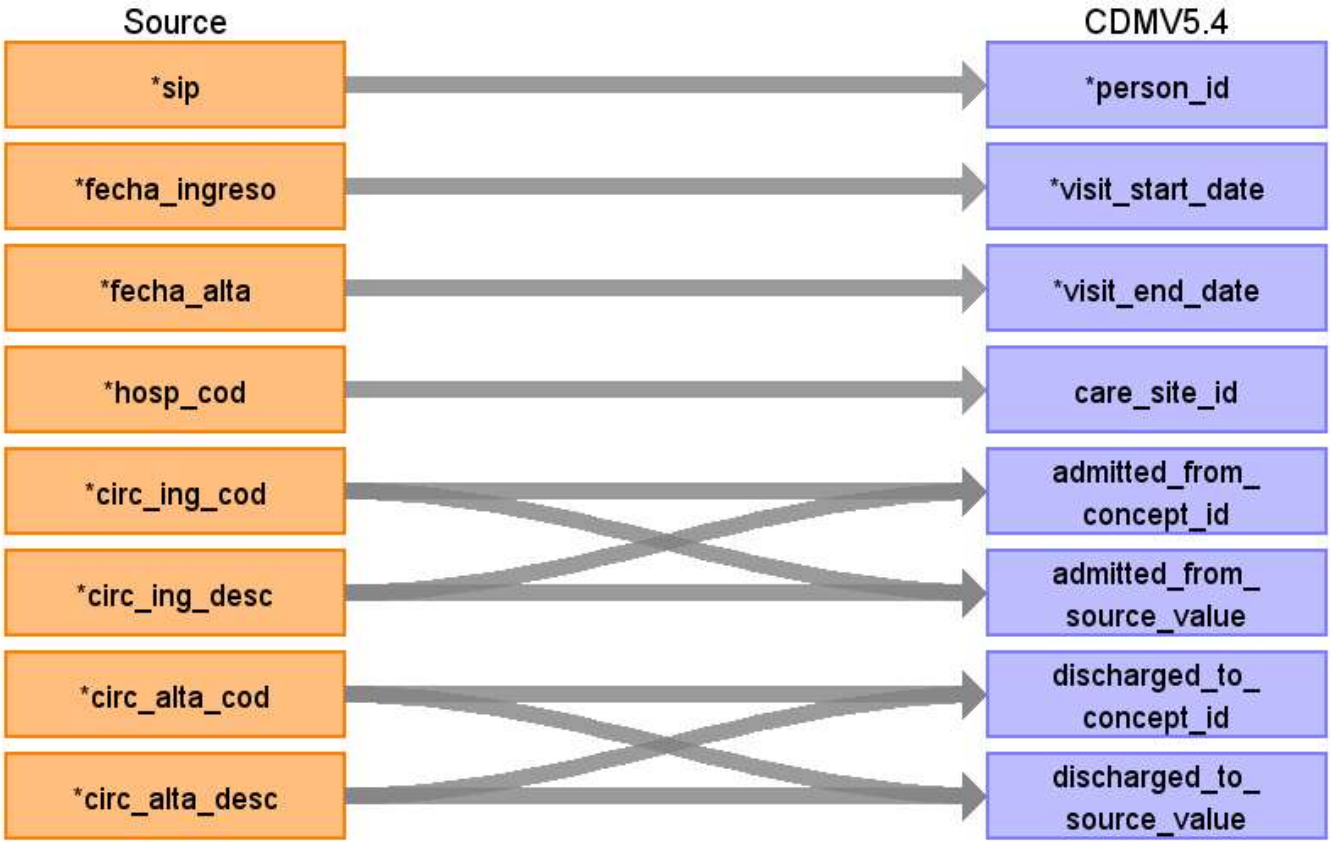


Figure 7: mbds to visit_occurrence mapping diagram

Table 5: mbds to visit_occurrence mapping description

Destination Field	Source Field	Logic	Comment

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			MBDS are hospital admissions. The Concept ID is 8717: Inpatient Hospital.
visit_start_date	fecha_ingreso		
visit_start_datetime			NULL
visit_end_date	fecha_alta		
visit_end_datetime			NULL
visit_type_concept_id			MBDS are hospital discharge summaries. The Concept ID is 32824: EHR discharge summary.
provider_id			NULL
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 Standadardized CONCEPT ID.
admitted_from_source_value	circ_ing_cod, circ_ing_desc		source admission code + description.
discharged_to_concept_id	circ_alta_cod, circ_alta_desc		discharge Standadardized CONCEPT ID.
discharged_to_source_value	circ_alta_cod, circ_alta_desc		source discharge code + description
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.

3.4.4 From aed to visit_occurrence

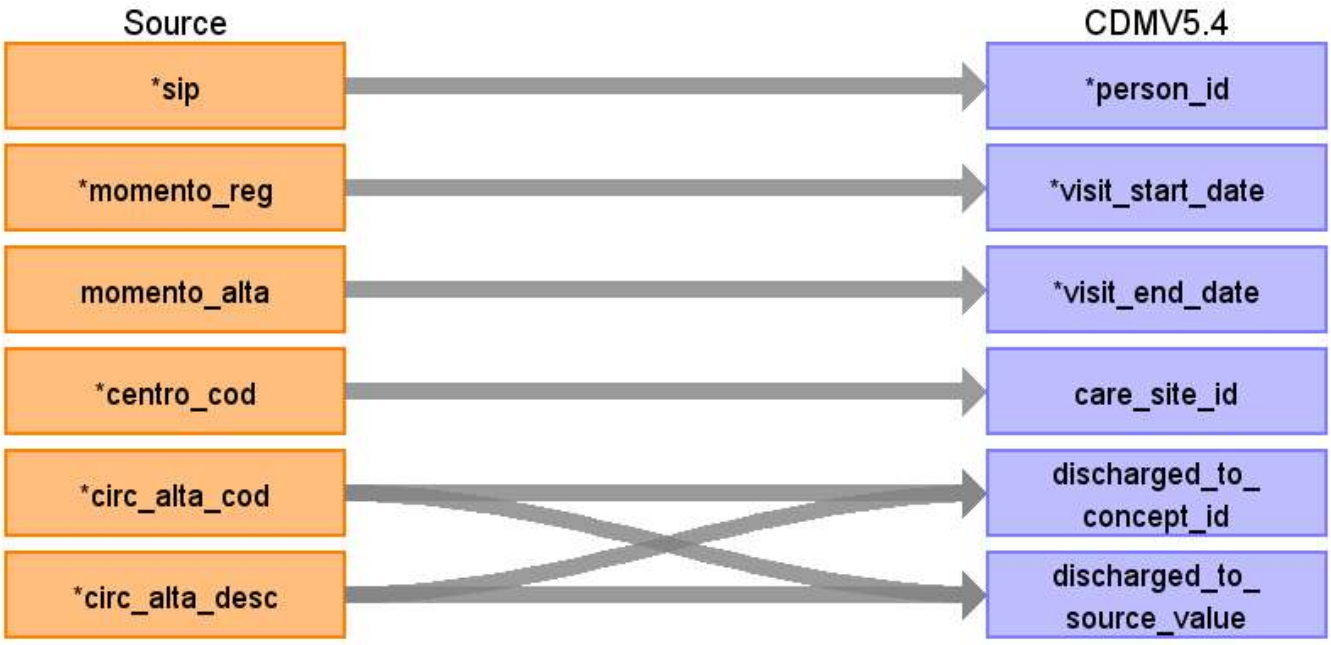


Figure 8: aed to visit_occurrence mapping diagram

Table 6: 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			AED are emergency visits. The Concept ID is 9203: Emergency Room Visit.
visit_start_date	fecha_registro		
visit_start_datetime			NULL
visit_end_date	fecha_alta		
visit_end_datetime			NULL

Destination Field	Source Field	Logic	Comment
visit_type_concept_id			AED are emergency visits. The Concept ID is 32826: EHR emergency room note.
provider_id			NULL
care_site_id			NULL
visit_source_value			AED
visit_source_concept_id			0
admitted_from_concept_id			0
admitted_from_source_value			NULL
discharged_to_concept_id	circ_alta_cod, circ_alta_desc		discharge Standadardized CONCEPT ID.
discharged_to_source_value	circ_alta_cod, circ_alta_desc		source discharge code + description
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.

3.5 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**.

3.5.1 From pcv to condition_occurrence

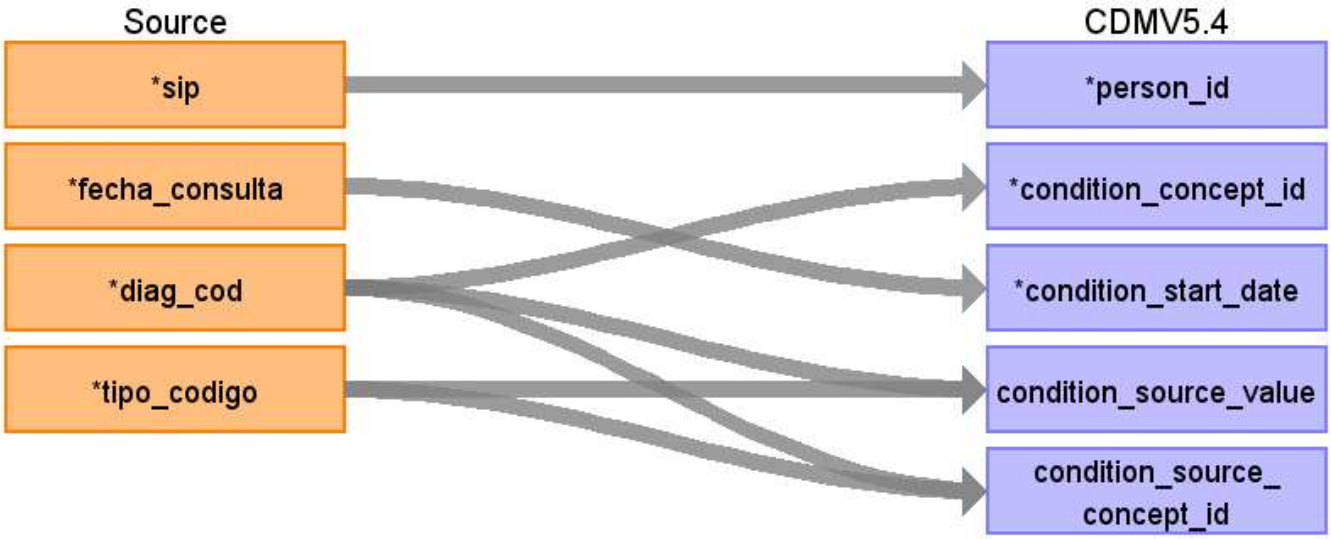


Figure 9: pcv to condition_occurrence mapping diagram

Table 7: 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		
condition_concept_id	diag_cod		Standardized CONCEPT ID from ICD9 or ICD10 codes.
condition_start_date	fecha_consulta		
condition_start_datetime			NULL
condition_end_date			
condition_end_datetime			NULL
condition_type_concept_id			PCV 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			NULL

Destination Field	Source Field	Logic	Comment
provider_id			NULL
visit_occurrence_id			Retrieve the visit_occurrence_id from the appropriate intermediate table (*pcv_to_visit_occurrence*).
visit_detail_id			0
condition_source_value	tipo_codigo, diag_cod		The ICD9 or ICD10 code. tipo_codigo flags if the code is ICD9 or ICD10.
condition_source_concept_id	tipo_codigo, diag_cod		ICD9 or ICD10 CONCEPT ID.
condition_status_source_value			PCV

3.5.2 From cex to condition_occurrence

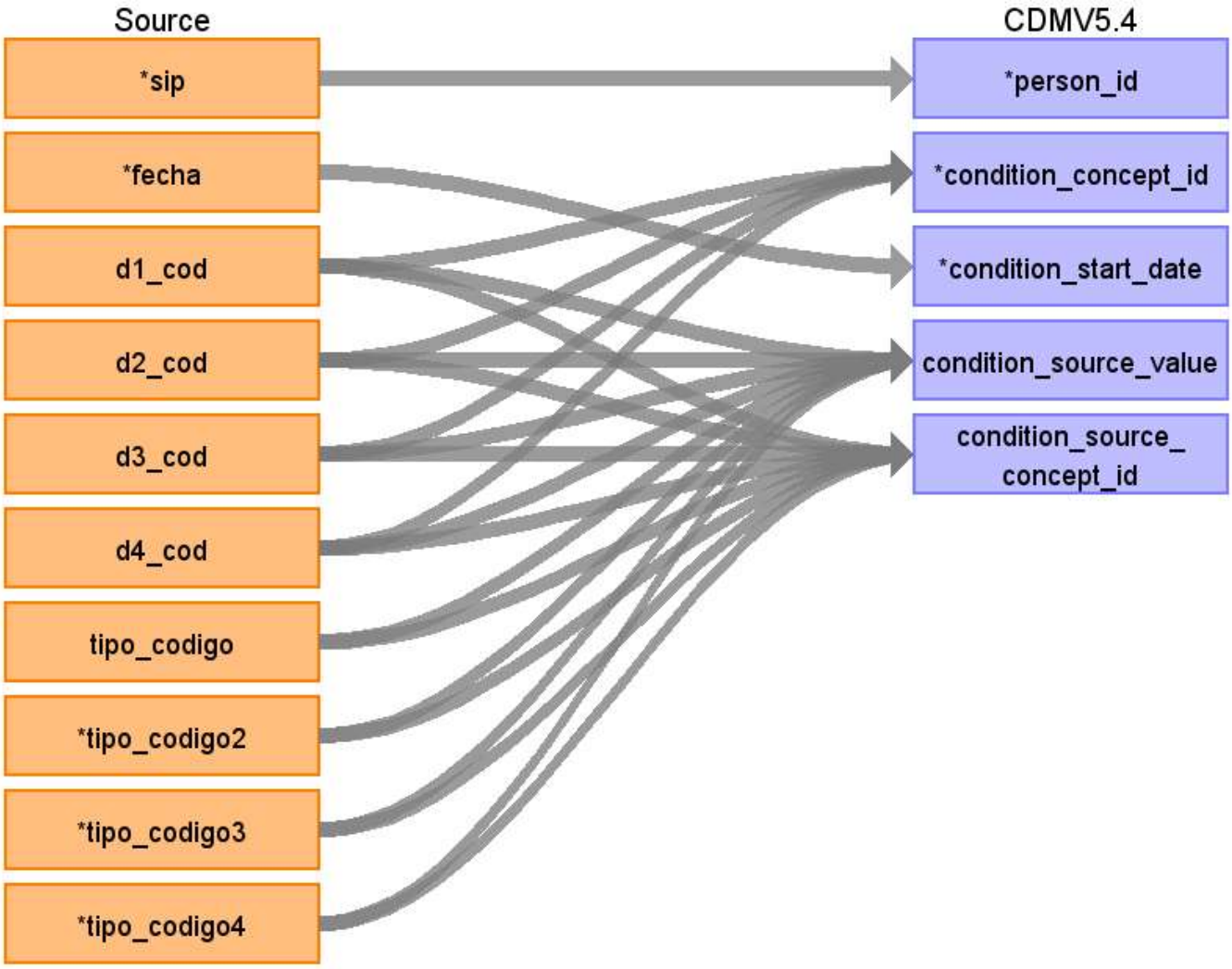


Figure 10: cex to condition_occurrence mapping diagram

Table 8: 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 distinct code is mapped in	Standardized CONCEPT ID from ICD9 or ICD10 codes.

Destination Field	Source Field	Logic	Comment
		a different row.	
condition_start_date	fecha_consulta		
condition_start_datetime			NULL
condition_end_date			
condition_end_datetime			NULL
condition_type_concept_id			CEX 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			NULL
provider_id			NULL
visit_occurrence_id			Retrieve the visit_occurrence_id from the appropriate intermediate table (*cex_to_visit_occurrence*).
visit_detail_id			0
condition_source_value	tipo_codigo, tipo_codigo2, tipo_codigo3, tipo_codigo4, d1_cod, d2_cod, d3_cod, d4_cod		The ICD9 or ICD10 code. tipo_codigo flags if the code is ICD9 or ICD10.
condition_source_concept_id	tipo_codigo, tipo_codigo2, tipo_codigo3, tipo_codigo4, d1_cod, d2_cod, d3_cod, d4_cod		ICD9 or ICD10 CONCEPT ID.
condition_status_source_value			CEX

3.5.3 From mbds to condition_occurrence

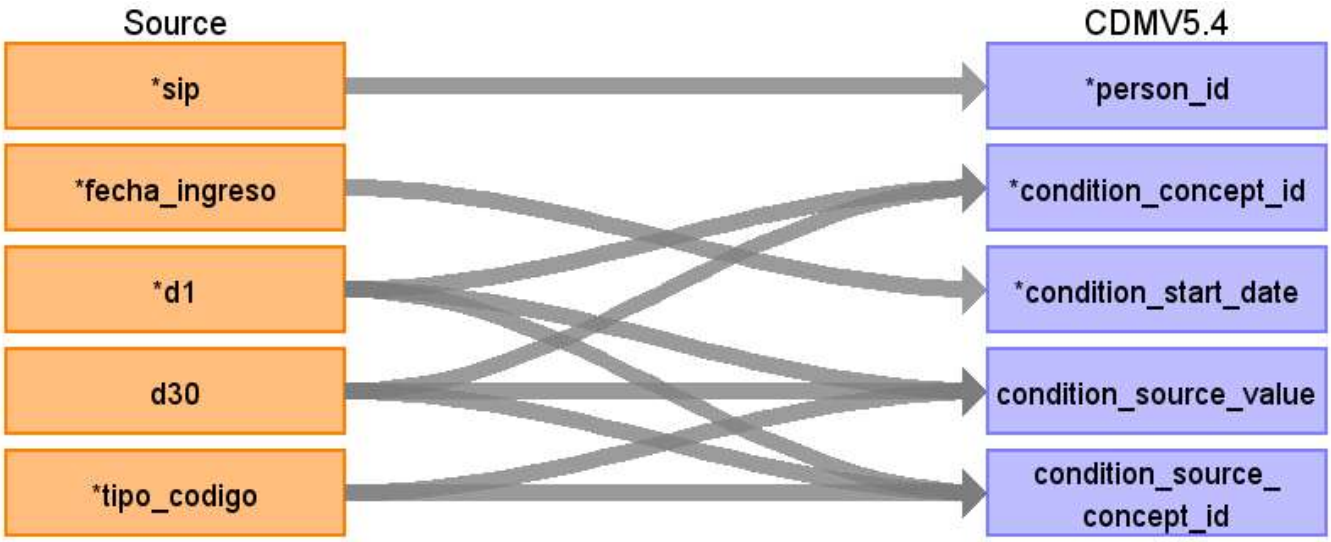


Figure 11: mbds to condition_occurrence mapping diagram

Table 9: 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 ICD9 or ICD10 codes. There are up to 30 diagnostic codes. However, for the sake of simplicity only d1 and d30 are depicted in the diagram.
condition_start_date	fecha_ingreso		
condition_start_datetime			NULL

Destination Field	Source Field	Logic	Comment
condition_end_date			
condition_end_datetime			NULL
condition_type_concept_id			MBDS are hospital discharge summaries. The Concept ID is 32824: EHR EHR discharge summary.
condition_status_concept_id			When source table is MBDS, the condition_status_concept_id are 32903: Primary discharge diagnosis (when the concept comes from d1) and 32909: Secondary discharge diagnosis when the concept comes from d2:d30).
stop_reason			NULL
provider_id			NULL
visit_occurrence_id			Retrieve the visit_occurrence_id from the appropriate intermediate table (*mbds_to_visit_occurrence*).
visit_detail_id			0
condition_source_value	tipo_codigo, d1:d30		The ICD9 or ICD10 code. tipo_codigo flags if the code is ICD9 or ICD10.
condition_source_concept_id	tipo_codigo, d1:d30		ICD9 or ICD10 CONCEPT ID.
condition_status_source_value			MBDS

3.5.4 From aed to condition_occurrence

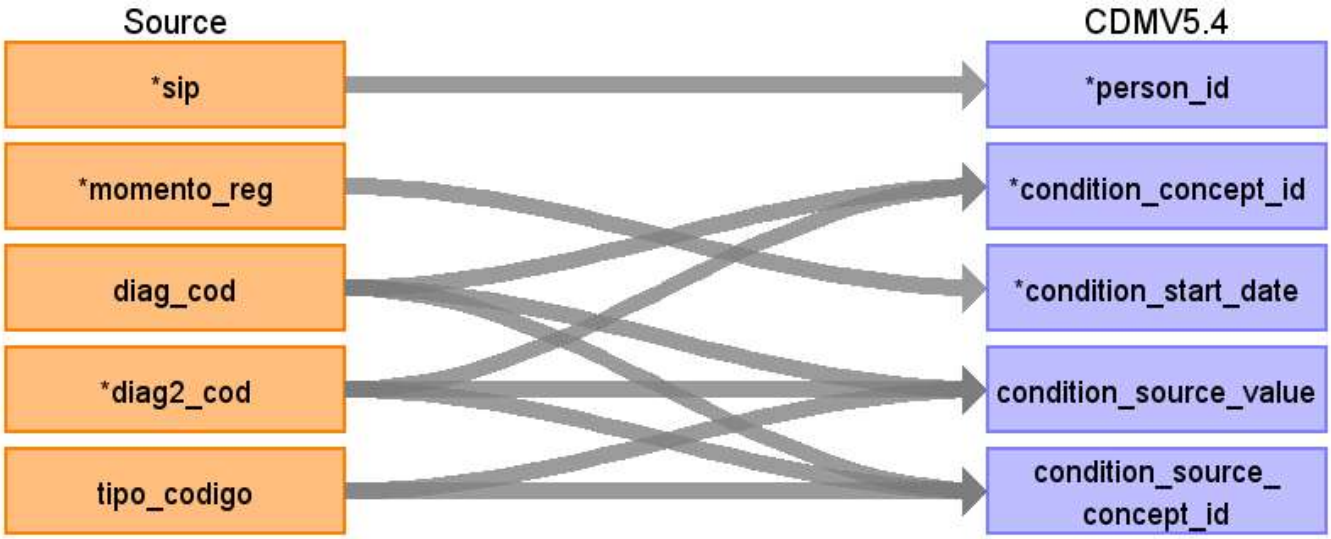


Figure 12: aed to condition_occurrence mapping diagram

Table 10: 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 ICD9 or ICD10 codes.
condition_start_date	fecha_registro		
condition_start_datetime			NULL
condition_end_date			
condition_end_datetime			NULL
condition_type_concept_id			AED are emergency visits. The Concept ID is 32826: EHR emergency room note.

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			NULL
provider_id			NULL
visit_occurrence_id			Retrieve the visit_occurrence_id from the appropriate intermediate table (*aed_to_visit_occurrence*).
visit_detail_id			0
condition_source_value	tipo_codigo, diag_cod, diag2_cod		The ICD9 or ICD10 code. tipo_codigo flags if the code is ICD9 or ICD10.
condition_source_concept_id	tipo_codigo, diag_cod, diag2_cod		ICD9 or ICD10 CONCEPT ID.
condition_status_source_value			AED

3.5.5 From diagnoses to condition_occurrence

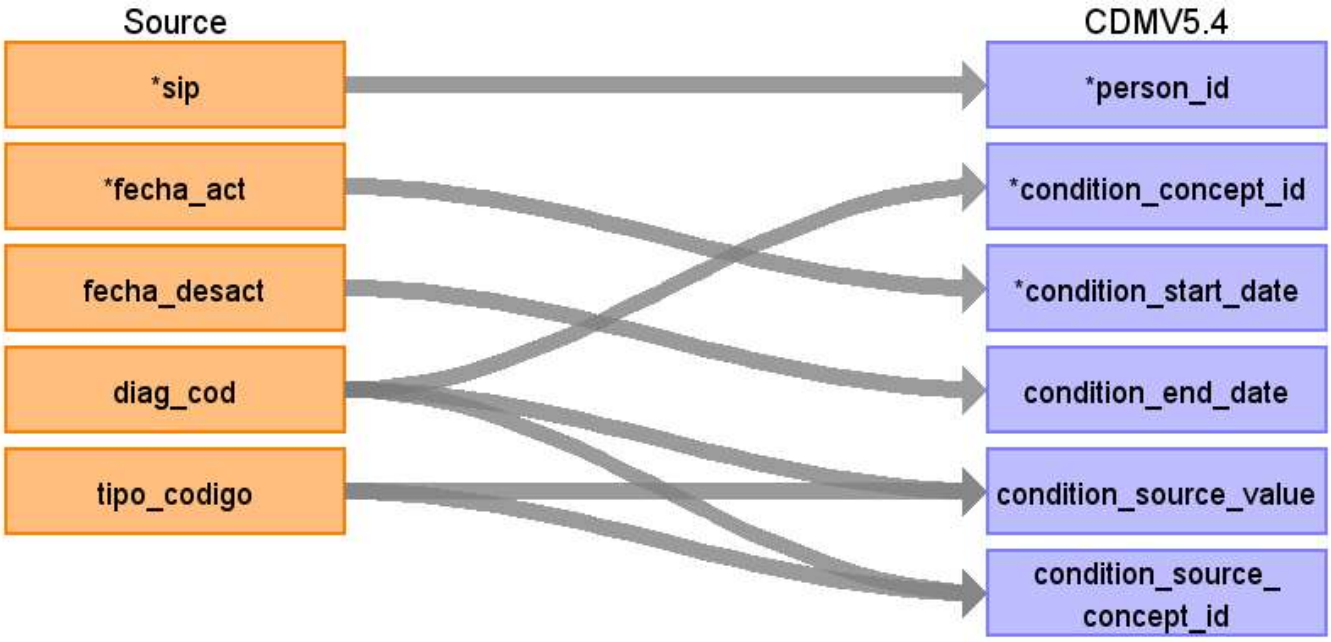


Figure 13: diagnoses to condition_occurrence mapping diagram

Table 11: 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 ICD9 or ICD10 codes.
condition_start_date	fecha_act		
condition_start_datetime			NULL
condition_end_date	fecha_desact		condition_end_date only is captured when the source table is DIAGNOSES.
condition_end_datetime			NULL
condition_type_concept_id			DIAGNOSES are confirmed diagnoses. The Concept ID is 32817: EHR.
condition_status_concept_id			When source table is PCV, CEX, AED, or DIAGNOSES, the condition_status_concept_id is 32893: Confirmed diagnosis.

Destination Field	Source Field	Logic	Comment
stop_reason			NULL
provider_id			NULL
visit_occurrence_id			NULL
visit_detail_id			0
condition_source_value	tipo_codigo, diag_cod		The ICD9 or ICD10 code. tipo_codigo flags if the code is ICD9 or ICD10.
condition_source_concept_id	tipo_codigo, diag_cod		ICD9 or ICD10 CONCEPT ID.
condition_status_source_value			DIAGNOSES

3.5.6 From eos to condition_occurrence

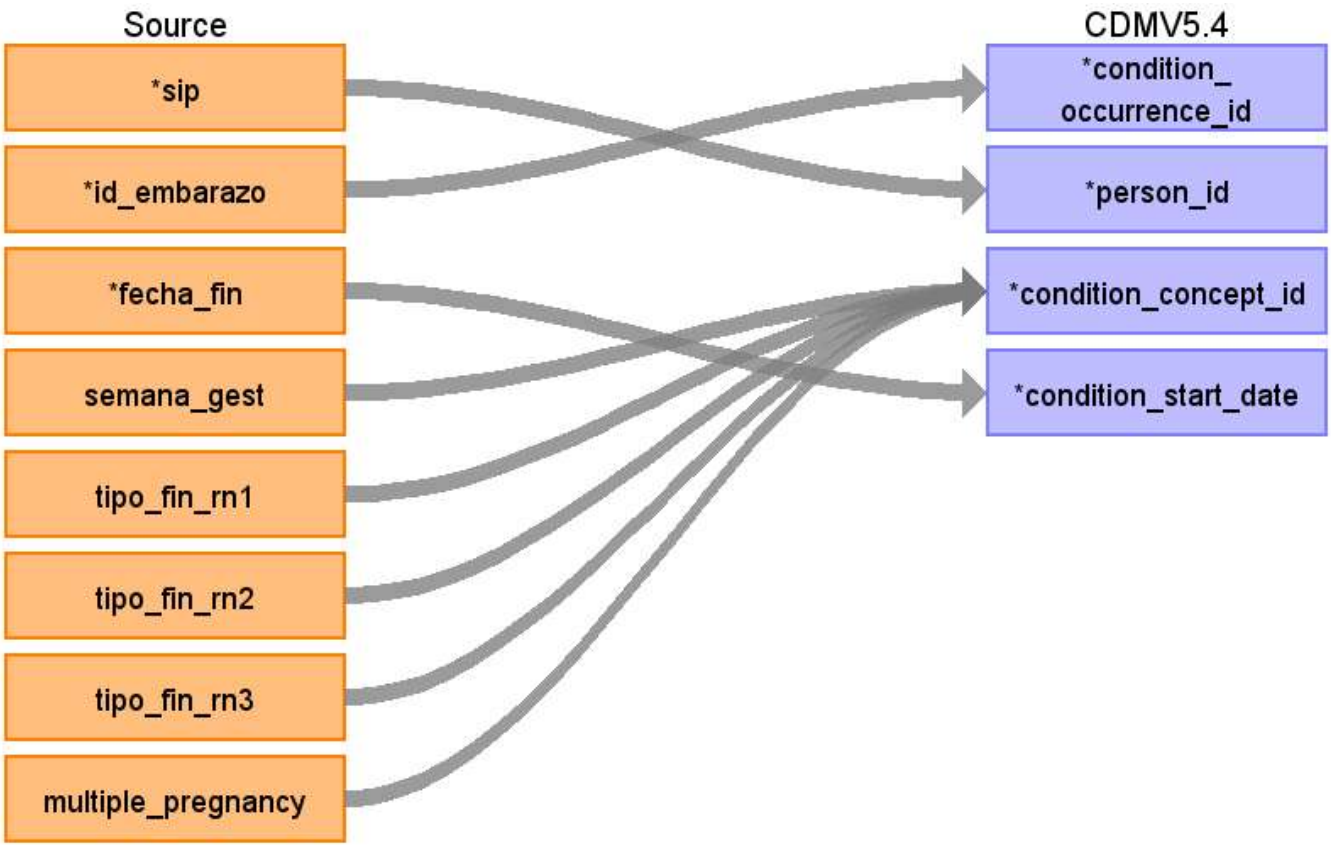


Figure 14: eos to condition_occurrence mapping diagram

Table 12: eos 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	resultado_rn1, semana_gest, resultado_rn2, resultado_rn3		Standardized code for each condition obtained from EOS. Maybe some values fit better in measurement or observation. Further investigation is still required.
condition_start_date	fecha_fin		
condition_start_datetime			NULL
condition_end_date			
condition_end_datetime			NULL
condition_type_concept_id			In the EOS there are information obtained through EHR. The Concept ID is 32817: EHR.
condition_status_concept_id			32893: Confirmed diagnosis.
stop_reason			NULL
provider_id			NULL
visit_occurrence_id			NULL
visit_detail_id			0
condition_source_value			The source value
condition_source_concept_id			When possible ICD9 or ICD10 CONCEPT ID, otherwise 0.
condition_status_source_value			EOS

3.6 Target table: procedure_occurrence

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

3.6.1 From mbds to procedure_occurrence

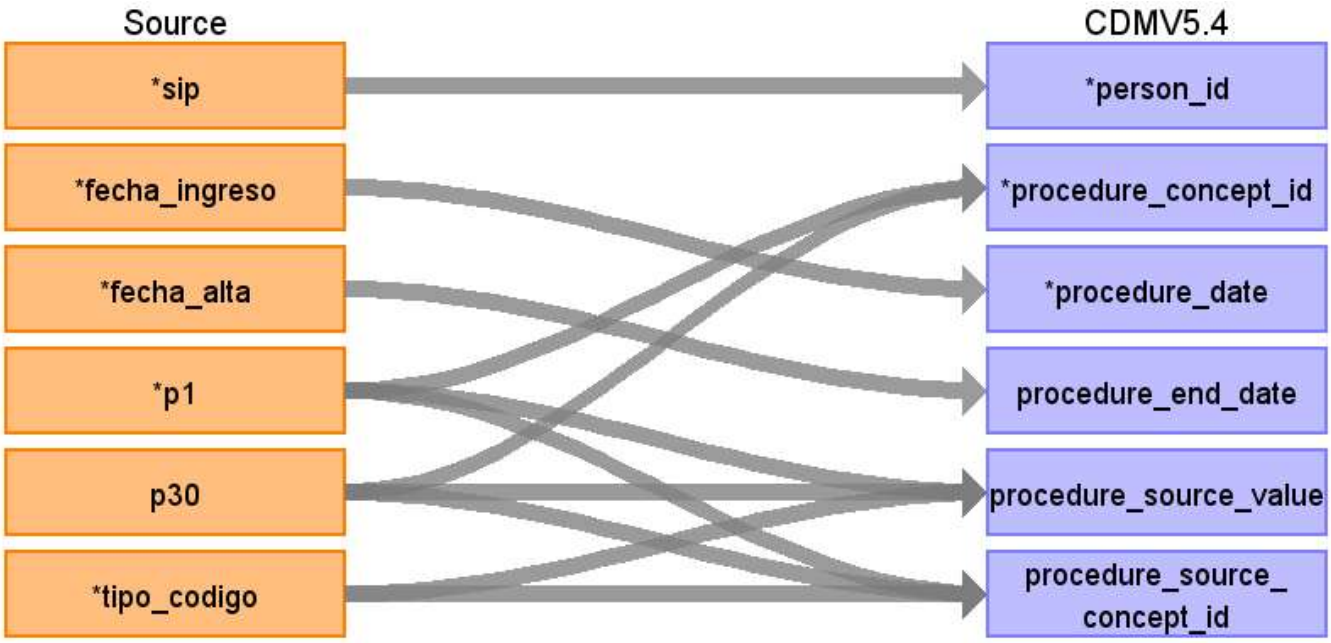


Figure 15: mbds to procedure_occurrence mapping diagram

Table 13: 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 ICD9 or ICD10 procedure code. There are up to 30 procedure codes. However, for the sake of simplicity only p1 and p30 are depicted in the diagram.

Destination Field	Source Field	Logic	Comment
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			NULL
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			NULL
procedure_type_concept_id			32824: EHR discharge summary
modifier_concept_id			NULL
quantity			NULL
provider_id			NULL
visit_occurrence_id			Use the intermediate table mbds_to_visit_occurrence.
visit_detail_id			0
procedure_source_value	tipo_codigo, p1:p30		The ICD9 or ICD10 code. tipo_codigo flags if the code is ICD9 or ICD10
procedure_source_concept_id	tipo_codigo, p1:p30		The CONCEPT ID from the ICD9 or ICD10 code
modifier_source_value			NULL

3.7 Target table: death

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

3.7.1 From sip to death

Figure 16: sip to death mapping diagram

Table 14: sip to death mapping description

Destination Field	Source Field	Logic	Comment
person_id	sip		
death_date	fecha_defuncion		
death_datetime			NULL
death_type_concept_id			32848: Government report
cause_concept_id			NULL
cause_source_value			NULL
cause_source_concept_id			NULL

3.7.2 From pmr to death

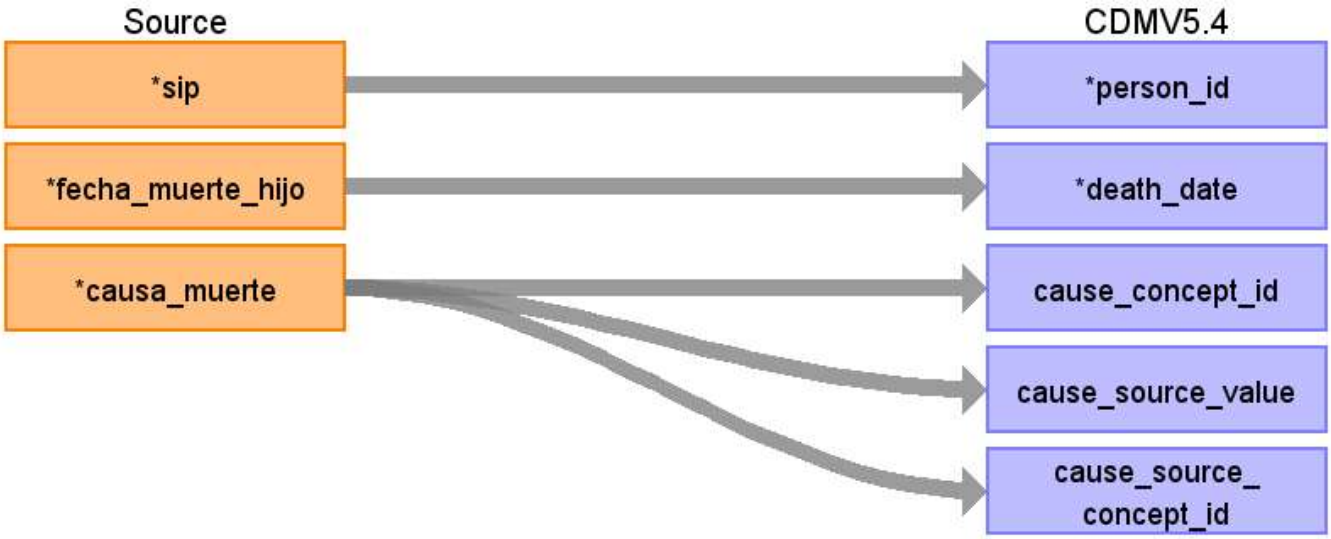


Figure 17: pmr to death mapping diagram

Table 15: pmr to death mapping description

Destination Field	Source Field	Logic	Comment
person_id	sip		
death_date	fecha_muerte_hijo		
death_datetime			NULL
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 ICD9 or ICD10 code.
cause_source_concept_id	causa_muerte		The cause_source_concept_id is the corresponding concept_id from the ICD9 or ICD10 code. Otherwise NULL.

3.8 Target table: drug_exposure

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

3.8.1 From gaia to drug_exposure

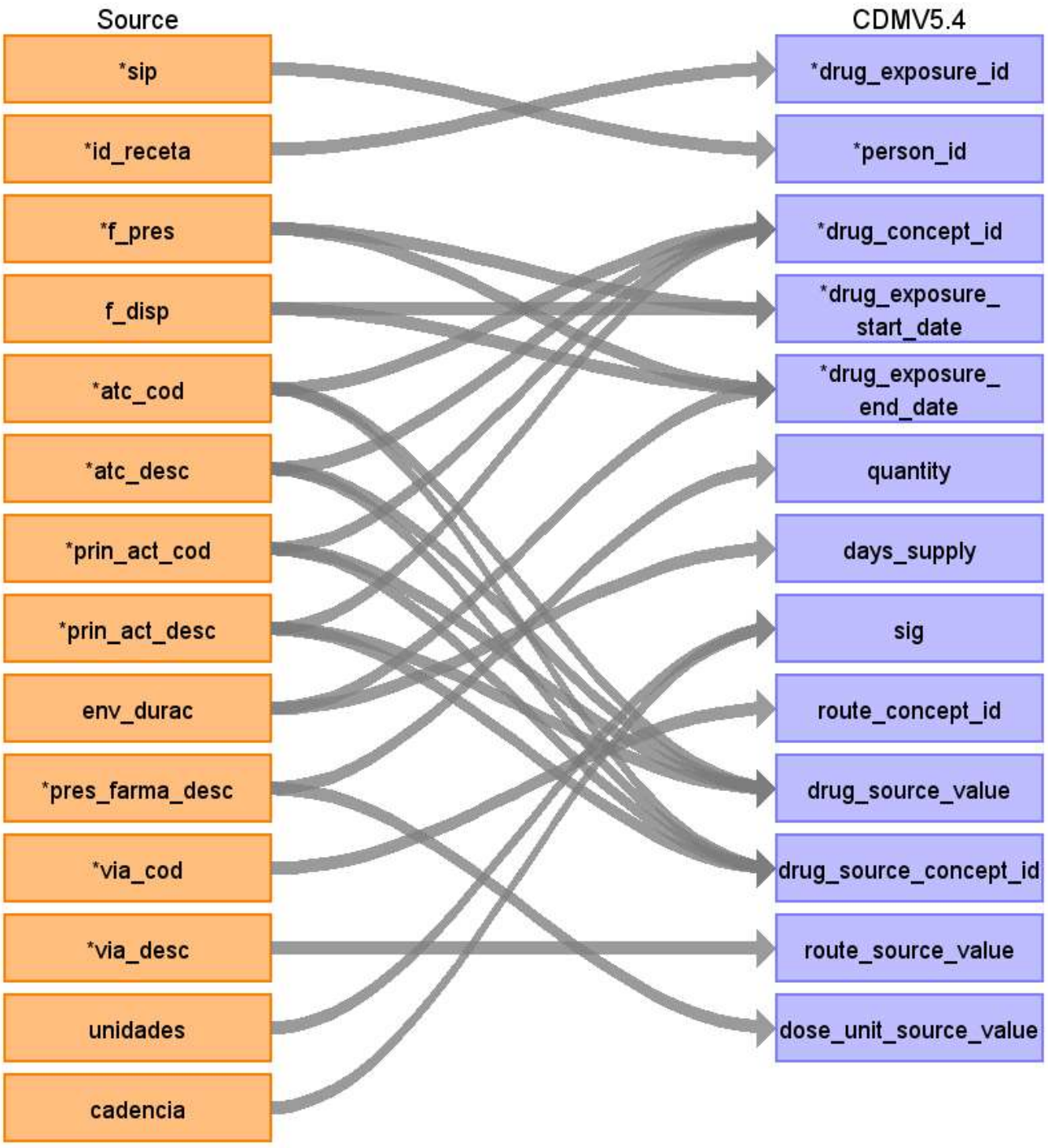


Figure 18: gaia to drug_exposure mapping diagram

Table 16: 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		
drug_concept_id	prin_act_cod, prin_act_desc, atc_cod, atc_desc		Obtain Standard CONCEPT ID using *prin_act_cod*, *prin_act_desc*, *atc_cod*, and *atc_desc*.
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			NULL
drug_exposure_end_date			drug_exposure_end_date is calculated as drug_exposure_start_date %m+% days(env_durac).
drug_exposure_end_datetime			NULL
verbatim_end_date			NULL
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			NULL
refills			NULL
quantity	pres_farma_desc		Extracted from *pres_farma_desc*.

Destination Field	Source Field	Logic	Comment
days_supply	env_durac		
sig	unidades, cadencia		Dosage as 'unidades' units each 'cadencia' hours.
route_concept_id	via_cod		Standardized CONCEPT ID route code.
lot_number			0
provider_id			0
visit_occurrence_id			
visit_detail_id			
drug_source_value	atc_cod, atc_desc		ATC or drug ingredient description.
drug_source_concept_id	atc_cod		ATC or drug ingredient code CONCEPT ID.
route_source_value	via_cod, via_desc		
dose_unit_source_value	pres_farma_desc	extract dose unit from *pres_farma_desc*.	Extracted from *pres_farma_desc*.

3.8.2 From siv to drug_exposure

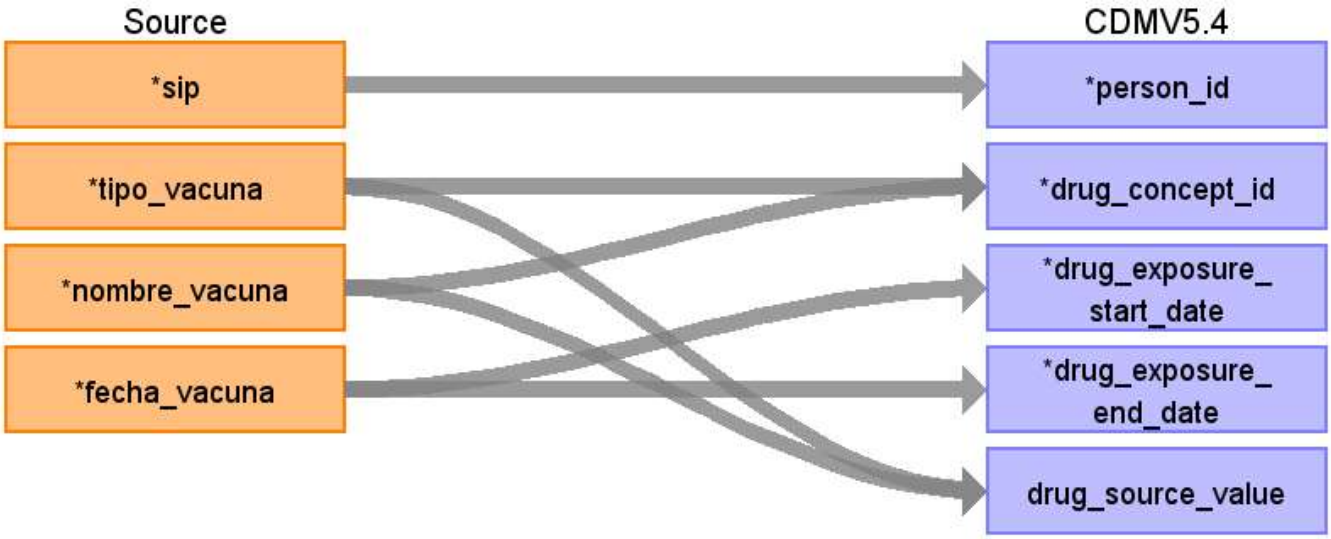


Figure 19: siv to drug_exposure mapping diagram

Table 17: 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			NULL
drug_exposure_end_date	fecha_vacuna		
drug_exposure_end_datetime			NULL
verbatim_end_date			NULL
drug_type_concept_id	tipo_vacuna, nombre_vacuna		the CONCEPT ID is 32818: EHR administration record.
stop_reason			NULL
refills			0
quantity			1
days_supply			0

Destination Field	Source Field	Logic	Comment
sig			NULL
route_concept_id			The route_concept_id is 4302612: Intramuscular (we will check if some vaccine has another administration route different to intramuscular).
lot_number			0
provider_id			0
visit_occurrence_id			NULL
visit_detail_id			0
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			NULL

3.9 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**.

3.9.1 From sip to care_site

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

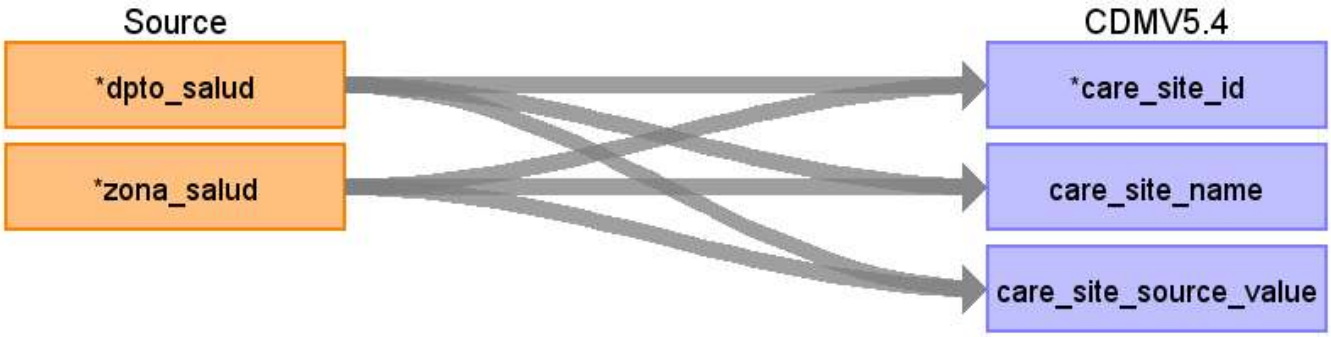


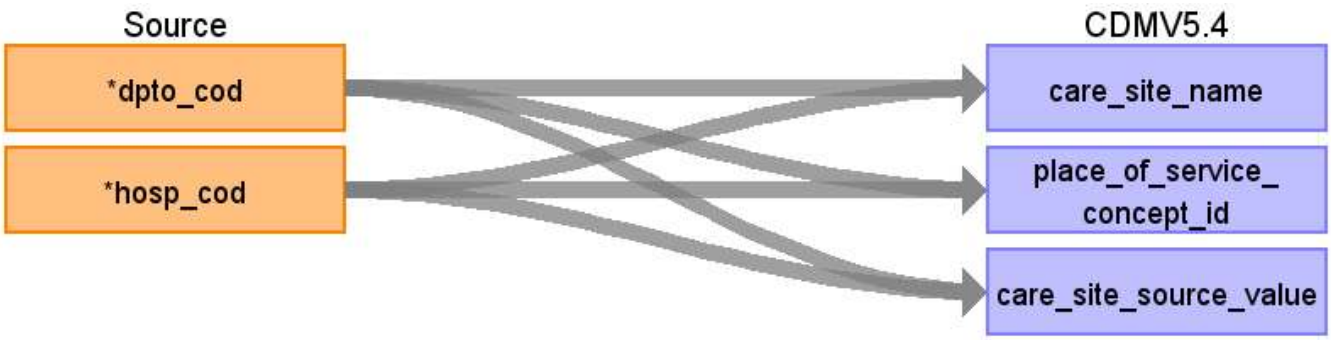
Figure 20: sip to care_site mapping diagram

Table 18: 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 or the basic health area assigned to each individual.
place_of_service_concept_id			When the row refers to a department the CONCEPT ID is 38004226: Ambulatory Health Service Clinic / Center.
location_id			NULL
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 **code** of the department and the basic health area assigned to each individual.
place_of_service_source_value			When the row refers to a department the source_value is 'zona básica de salud'.

3.9.2 From mbds to care_site

In the *mbds* table, we populate the hospitals.



mbds to care_site mapping diagram

Table 19: 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			NULL
care_site_source_value	dpto_cod, hosp_cod		In this field it is indicated the **code** of the department and the hospital.
place_of_service_source_value			Hospital

3.10 Target table: fact_relationship

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

3.10.1 From mdr to fact_relationship

Figure 21: 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.

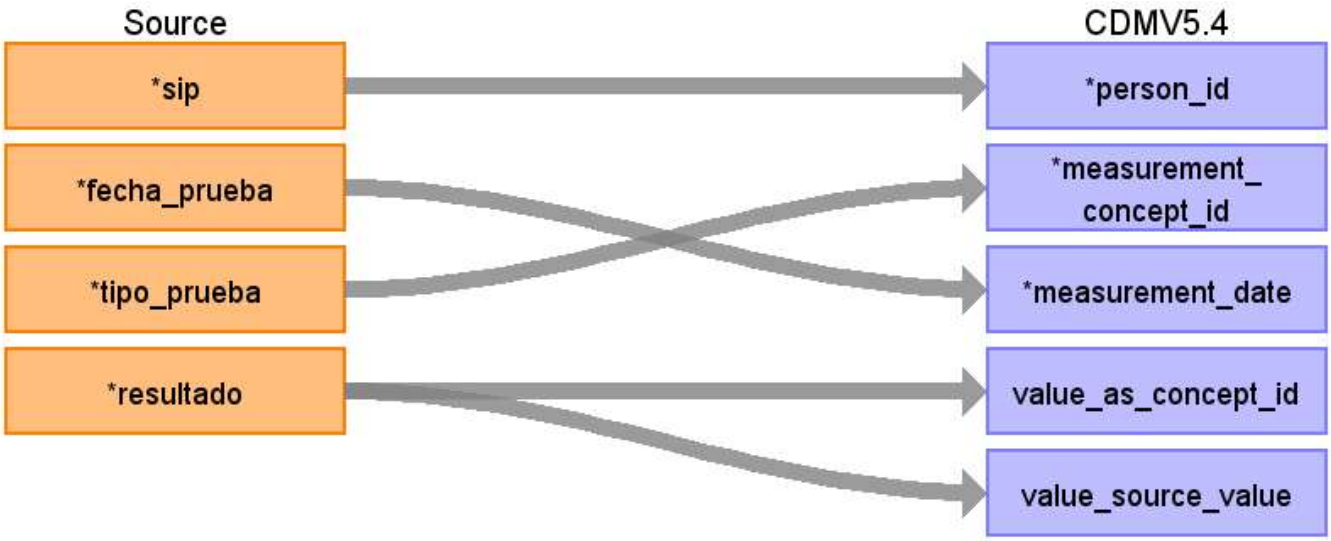
Table 20: mdr to fact relationship mapping description

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

3.11 Target table: measurement

The target table **measurement** is populated with the information from the source table **redmiva**.

3.11.1 From redmiva to measurement



redmiva to measurement mapping diagram

Table 21: redmiva to measurement mapping description

Destination Field	Source Field	Logic	Comment
measurement_id			Autogenerate
person_id	sip		
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		
value_as_concept_id	resultado		Positive: 45884084
value_source_value	resultado		Positivo
measurement_datetime			NULL

Destination Field	Source Field	Logic	Comment
measurement_time			NULL
measurement_type_concept_id			32856: Lab
operator_concept_id			0
value_as_number			
unit_concept_id			0
range_low			NULL
range_high			NULL
provider_id			0
visit_occurrence_id			NULL
visit_detail_id			0
measurement_source_value			NULL
measurement_source_concept_id			NULL
unit_source_value			NULL
unit_source_concept_id			NULL
measurement_event_id			NULL
meas_event_field_concept_id			NULL

4 VOCABULARY MAPPING

In this section an overview is provided of the vocabulary mapping step. The following table shows the source code vocabularies that are present in the database. All the source vocabularies are included in ATHENA and are downloaded into csv tables.

Table 22: Source vocabularies

Vocabulary	Reference Link	Description	Data Domains
ICD9CM	ICD9CM	International Classification of Diseases, 9th revision, Clinical Modification.	condition_occurrence
ICD10ES	ICD10ES	International Classification of Diseases, 10th revision, Clinical Modification, Spanish Edition.	condition_occurrence
ATC	ATC codes	ATC classification system. In the ATC classification system, the active substances are classified in a hierarchy with five different levels. The system has fourteen main anatomical/pharmacological groups or 1st levels. Each ATC main group is divided into 2nd levels which could be either pharmacological or therapeutic groups. The 3rd and 4th levels are chemical, pharmacological or therapeutic subgroups and the 5th level is the chemical substance. The 2nd, 3rd and 4th levels are often used to identify pharmacological subgroups when that is considered more appropriate than therapeutic or chemical subgroups.	drug_exposure

5 NEXT STEPS

- **Mapping of source tables and vocabularies.** we will map all data sources in stepwise fashion, being the planned order the same as the order presented in the data set description section.
- **Technical ETL Development. - Setting up of Infrastructure.**
- **(Technical) testing of the ETL.**
- **Replicate the ETL pipeline in the Opioids project data extraction.** In order to check that the ETL works for another setting of the VID database.
- **Data Quality Assessment** (In an iterative manner).
- **Completion of the data catalogue.** This task will be performed after the data quality assessment.
- **Inspection Report.** This task is externalized by an SME.

A slight delay of one month in the project has been produces, therefore, the new estimated timelines for the project milestones are:

Table 23: Milestones estimated timelines

Milestone	Estimated Timelines
ETL definition document	Month 4
ETL Implemented and Infrastructure Operational	Month 8
Database catalogue entry following final inspection by certified SME, final report	Month 10

APPENDIX 1. SOURCE DATA MODEL

Figure 22: VID data model diagram

APPENDIX 2. DATA DICTIONARY

5.1 Tables

Table 24: 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.
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	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.
MDR	Metabolic Disease 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.
PMR	Perinatal Mortality Register	Electronic obstetric sheet. It contains the information about the pregnancy follow-up. It is used in order to detect spontaneous

Source Table	English Name	Description
		abortions (and to confirm births and stillbirths).
EOS	Electronic Obstetric Sheet	Tests, such as pregnancy test, ordered by a physician.
REDMIVA	Microbiological Surveillance Network	It contains the information about COVID-19 test results.

5.2 Fields

Table 25: 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)
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

Source Table	Field	Type	Description
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
CEX	especialidad_desc	VARCHAR	speciality description
CEX	tipo_contacto	VARCHAR	contact type
CEX	d1_cod	VARCHAR	diagnosis code 1
CEX	d1_desc	VARCHAR	diagnosis description 1
CEX	d2_cod	VARCHAR	diagnosis code 2
CEX	d2_desc	VARCHAR	diagnosis description 2
CEX	d3_cod	VARCHAR	diagnosis code 3
CEX	d3_desc	VARCHAR	diagnosis description 3
CEX	d4_cod	VARCHAR	diagnosis code 4
CEX	d4_desc	VARCHAR	diagnosis description 4

Source Table	Field	Type	Description
CEX	tipo_codigo	VARCHAR	diagnosis code 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	VARCHAR	health department name
MBDS	serv_ing_cod	INT	hospital code
MBDS	serv_ing_desc	VARCHAR	hospital name
MBDS	tipo_activ	VARCHAR	admission service code
MBDS	circ_ing_cod	VARCHAR	admission service description
MBDS	circ_ing_desc	VARCHAR	activity type: ambulatory or overnight
MBDS	circ_alta_cod	INT	admission circumstances code
MBDS	circ_alta_desc	VARCHAR	admission circumstances description
MBDS	d1	INT	discharge circumstances code
MBDS	d2	VARCHAR	discharge circumstances code
MBDS	d3	VARCHAR	main diagnosis of the admission (d1)
MBDS	d4	VARCHAR	secondary diagnosis (d2)
MBDS	d5	VARCHAR	secondary diagnosis (d3)
MBDS	d6	VARCHAR	secondary diagnosis (d4)
MBDS	d7	VARCHAR	secondary diagnosis (d5)
MBDS	d8	VARCHAR	secondary diagnosis (d6)
MBDS	d9	VARCHAR	secondary diagnosis (d7)
MBDS	d10	VARCHAR	secondary diagnosis (d8)
MBDS	d11	VARCHAR	secondary diagnosis (d9)

Source Table	Field	Type	Description
MBDS	d12	VARCHAR	secondary diagnosis (d10)
MBDS	d13	VARCHAR	secondary diagnosis (d11)
MBDS	d14	VARCHAR	secondary diagnosis (d12)
MBDS	d15	VARCHAR	secondary diagnosis (d13)
MBDS	d16	VARCHAR	secondary diagnosis (d14)
MBDS	d17	VARCHAR	secondary diagnosis (d15)
MBDS	d18	VARCHAR	secondary diagnosis (d16)
MBDS	d19	VARCHAR	secondary diagnosis (d17)
MBDS	d20	VARCHAR	secondary diagnosis (d18)
MBDS	d21	VARCHAR	secondary diagnosis (d19)
MBDS	d22	VARCHAR	secondary diagnosis (d20)
MBDS	d23	VARCHAR	secondary diagnosis (d21)
MBDS	d24	VARCHAR	secondary diagnosis (d22)
MBDS	d25	VARCHAR	secondary diagnosis (d23)
MBDS	d26	VARCHAR	secondary diagnosis (d24)
MBDS	d27	VARCHAR	secondary diagnosis (d25)
MBDS	d28	VARCHAR	secondary diagnosis (d26)
MBDS	d29	VARCHAR	secondary diagnosis (d27)
MBDS	d30	VARCHAR	secondary diagnosis (d28)
MBDS	p1	VARCHAR	secondary diagnosis (d29)
MBDS	p2	VARCHAR	secondary diagnosis (d30)
MBDS	p3	VARCHAR	main procedure in the admission (p1)
MBDS	p4	VARCHAR	secondary procedure (p2)
MBDS	p5	VARCHAR	secondary procedure (p3)
MBDS	p6	VARCHAR	secondary procedure (p4)

Source Table	Field	Type	Description
MBDS	p7	VARCHAR	secondary procedure (p5)
MBDS	p8	VARCHAR	secondary procedure (p6)
MBDS	p9	VARCHAR	secondary procedure (p7)
MBDS	p10	VARCHAR	secondary procedure (p8)
MBDS	p11	VARCHAR	secondary procedure (p9)
MBDS	p12	VARCHAR	secondary procedure (p10)
MBDS	p13	VARCHAR	secondary procedure (p11)
MBDS	p14	VARCHAR	secondary procedure (p12)
MBDS	p15	VARCHAR	secondary procedure (p13)
MBDS	p16	VARCHAR	secondary procedure (p14)
MBDS	p17	VARCHAR	secondary procedure (p15)
MBDS	p18	VARCHAR	secondary procedure (p16)
MBDS	p19	VARCHAR	secondary procedure (p17)
MBDS	p20	VARCHAR	secondary procedure (p18)
MBDS	p21	VARCHAR	secondary procedure (p19)
MBDS	p22	VARCHAR	secondary procedure (p20)
MBDS	p23	VARCHAR	secondary procedure (p21)
MBDS	p24	VARCHAR	secondary procedure (p22)
MBDS	p25	VARCHAR	secondary procedure (p23)
MBDS	p26	VARCHAR	secondary procedure (p24)
MBDS	p27	VARCHAR	secondary procedure (p25)
MBDS	p28	VARCHAR	secondary procedure (p26)
MBDS	p29	VARCHAR	secondary procedure (p27)
MBDS	p30	VARCHAR	secondary procedure (p28)
MBDS	tipo_codigo	VARCHAR	secondary procedure (p29)

Source Table	Field	Type	Description
MBDS	dpto_desc	VARCHAR	secondary procedure (p30)
MBDS	hosp_desc	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
AED	motivo_urg_desc	VARCHAR	emergency admission description
AED	diag_cod	VARCHAR	main diagnosis code
AED	diag2_cod	VARCHAR	secondary diagnosis code
AED	tipo_codigo	VARCHAR	diagnosis code vocabulary
AED	prioridad_cod	INT	priority code

Source Table	Field	Type	Description
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
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	elec_manu	VARCHAR	electronic or manual prescription

Source Table	Field	Type	Description
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)
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)
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

Source Table	Field	Type	Description
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
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
GAIA	fecha_pres	DATE	prescription date

Source Table	Field	Type	Description
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
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

Source Table	Field	Type	Description
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
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_nac_hijo	DATE	date of the birth
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

Source Table	Field	Type	Description
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
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

Source Table	Field	Type	Description
EOS	resultado_rn1	VARCHAR	event type of the first child delivered: birth, spontaneous abortion or stillbirth
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
CONG	sip_madre	VARCHAR	pseudonymised id number (unique for each patient) of the mother
CONG	nacidos_vivos	REAL	livebirth number
CONG	fecha_nacimiento_hijo	DATE	child's date of birth
CONG	semana_gest	REAL	gestational age (in weeks)
CONG	n_hijos_parto	REAL	number of newborns in the pregnancy
CONG	sexo	INT	sex of the newborn
CONG	peso	REAL	newborn weight (in g)
CONG	nbrmalf	INT	type of malformations
CONG	fecha_muerte_hijo	DATE	newborn death date
CONG	fecha_dx_anomalia	DATE	date of the anomaly detection
CONG	dx_anomalia	VARCHAR	anomaly code
CONG	dx_vivo_muerto	VARCHAR	diagnosis when aliver or death
CONG	tipo_nacimiento	VARCHAR	type of birth
CONG	tot_malf	INT	total number of malformations
CONG	sindrome	VARCHAR	syndrome code
CONG	sindrome_desc	VARCHAR	syndrome description
CONG	malfo_cod1	VARCHAR	malformation code 1
CONG	malfo_desc1	VARCHAR	malformation description 1
CONG	malfo_cod2	VARCHAR	malformation code 2
CONG	malfo_desc2	VARCHAR	malformation description 2

Source Table	Field	Type	Description
CONG	malfo_cod3	VARCHAR	malformation code 3
CONG	malfo_desc3	VARCHAR	malformation description 3
CONG	malfo_cod4	VARCHAR	malformation code 4
CONG	malfo_desc4	VARCHAR	malformation description 4
CONG	malfo_cod5	VARCHAR	malformation code 5
CONG	malfo_desc5	VARCHAR	malformation description 5
CONG	malfo_cod6	VARCHAR	malformation code 6
CONG	malfo_desc6	VARCHAR	malformation description 6
CONG	malfo_cod7	VARCHAR	malformation code 7
CONG	malfo_desc7	VARCHAR	malformation description 7
CONG	malfo_cod8	VARCHAR	malformation code 8
CONG	malfo_desc8	VARCHAR	malformation description 8
CONG	semana_gest_dx	REAL	gestational age (in weeks) at the moment of the diagnosis
CONG	presyn	INT	type of syndrome diagnosis (1:pre-birth, 2:post-birth, 3:partially pre-birth, 9:unknown)
CONG	premal1	INT	type of malformation diagnosis 1 (1:pre-birth, 2:post-birth, 3:partially pre-birth, 9:unknown)
CONG	premal2	INT	type of malformation diagnosis 2 (1:pre-birth, 2:post-birth, 3:partially pre-birth, 9:unknown)
CONG	premal3	INT	type of malformation diagnosis 3 (1:pre-birth, 2:post-birth, 3:partially pre-birth, 9:unknown)
CONG	premal4	INT	type of malformation diagnosis 4 (1:pre-birth, 2:post-birth, 3:partially pre-birth, 9:unknown)
CONG	premal5	INT	type of malformation diagnosis 5 (1:pre-birth, 2:post-birth, 3:partially pre-birth, 9:unknown)
CONG	premal6	INT	type of malformation diagnosis 6 (1:pre-birth, 2:post-birth, 3:partially pre-birth, 9:unknown)

Source Table	Field	Type	Description
CONG	premal7	INT	type of malformation diagnosis 7 (1:pre-birth, 2:post-birth, 3:partially pre-birth, 9:unknown)
CONG	premal8	INT	type of malformation diagnosis 8 (1:pre-birth, 2:post-birth, 3:partially pre-birth, 9:unknown)
CONG	ill_bef1	VARCHAR	illness before the pregnancy 1
CONG	ill_bef2	VARCHAR	illness before the pregnancy 2
CONG	ill_dur1	VARCHAR	illness during the pregnancy 1
CONG	ill_dur2	VARCHAR	illness during the pregnancy 2
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
REDMIVA	fecha_resultado	DATE	date of the result
REDMIVA	resultado	VARCHAR	result of the test