

Background

- SNDS (*Système National des Données de Santé*) is the French Nationwide Healthcare System Database
 - Covers about 99% of the French population (about 66.6 million persons) from birth (or immigration) to death (or emigration)
 - Merges information of reimbursed claims (SNIIRAM) linked to the national hospital-discharge summary (PMSI) and the national death registry (CépiDC)
 - Includes: demographics; physician or paramedical visits (without medical indication); dispensed drugs and medical devices; lab tests (without results); chronic medical conditions; hospitalization with ICD10 codes for primary, linked and associated diagnoses, date and duration; reimbursed procedures
- EGB (*Echantillon Généraliste de Bénéficiaires*) is the 1/97th random permanent representative sample of SNDS.
 - 810 000 persons
 - Planned 20-year longitudinal data
 - Same structure than SNDS
- EGB is an ideal candidate for assessing the possibility to apply the OMOP CDM to the SNDS

Methods

1. **Extraction of an EGB sample** (extraction of all the myocardial infarctions between 2009 and 2014)
2. **Data management** of the 74 resulting datasets to 14 transitional ones
 - Socio demographic data
 - Medical visits
 - Chronic medical conditions
 - Hospitalization diagnoses
 - Drug dispensings
 - Lab tests
 - Outpatient procedures
 - Inpatient procedures
3. **ETL design**
 - Theoretical mapping of the variables from EGB to OMOP CDM v5 (Figure 1.)
 - Assessment of the mapping from source codes and thesauri to OMOP standardized vocabulary (Table 1.)
4. **Effective mapping of data**
 - Variables mapping
 - Generation of the treatment duration for the *drug_exposure* table according to drug dispensings data
 - Codes mapping whenever present in OMOP vocabulary

Results

ETL design

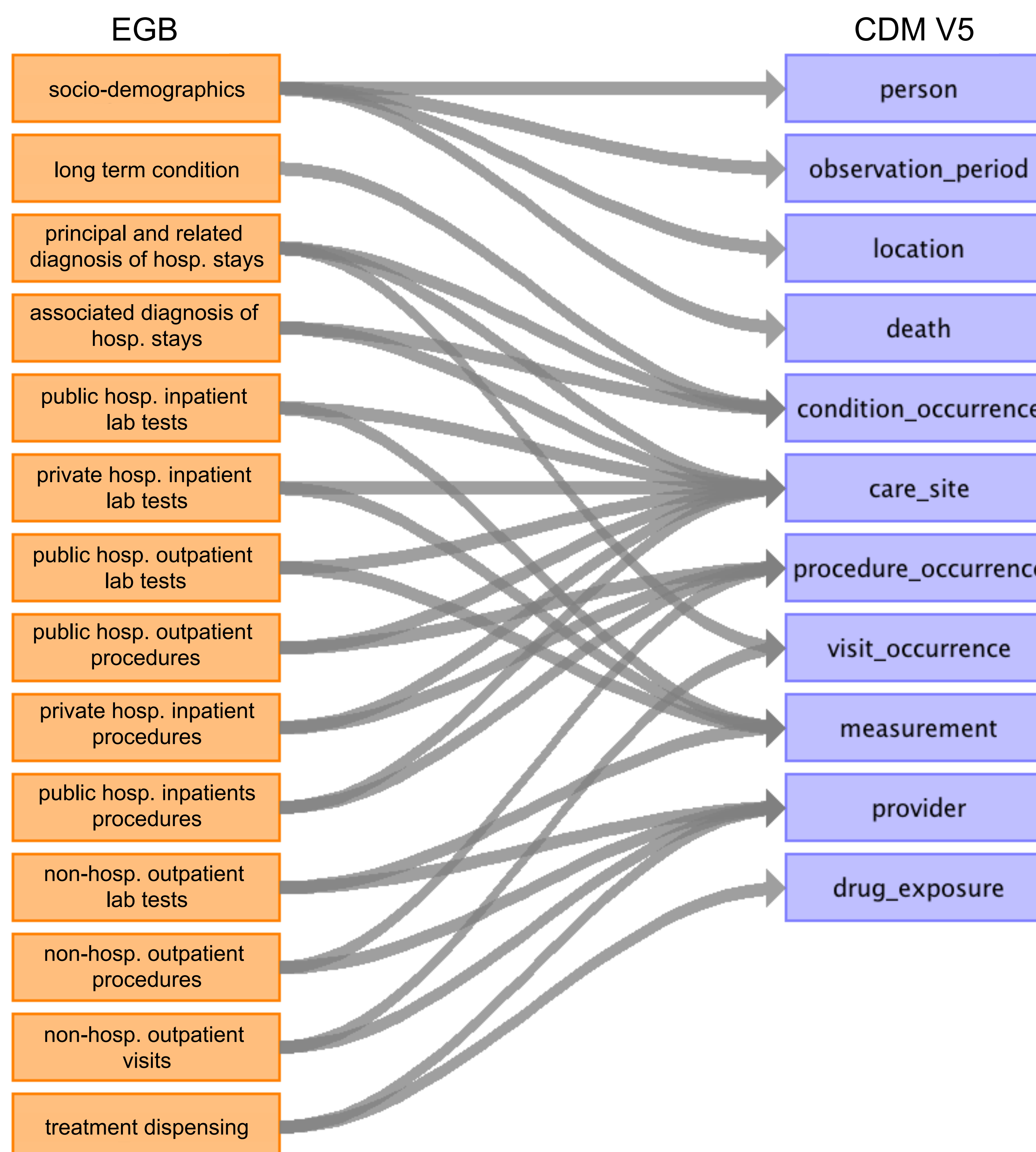


Figure 1. EGB data Mapping Approach to CDM v5 (clinical part only, without devices)

Effective mapping of data

- To date, 4 data tables have been generated
 - *person*
 - *location*
 - *drug_exposure*
 - *observation_period*
- Drug treatment duration for the *drug_exposure* table has been generated using medians of treatment duration and taking into account the 3-month drug boxes.

- The theoretical mapping of the 14 transitional tables corresponding to the clinical part of the EGB has been realized through *WhiteRabbit* and *RabbitInHat* softwares (Figure 1.)

Table 1. Presence of EGB and SNDS vocabularies in the OMOP vocabularies

Domains	EGB / SNDS vocabularies	Presence in the OMOP vocabularies
Condition	ICD10	Yes
Drug	ATC	Yes
	CIP	No
	UCD	No
Device	LPP	No
Procedure	CCAM	No
Measurement	NABM	No

ATC: Anatomical Therapeutic Chemical; **CCAM:** Classification Commune des Actes Médicaux; **CIP:** Code Identifiants de Présentations; **ICD:** International Classification of Diseases; **LPP:** Liste des Produits et Prestations; **NABM:** Nomenclature des Actes de Biologie Médicale; **UCD:** Unité Commune de Dispensation

- Most of the national nomenclatures used in France are not included in the OMOP vocabularies
 - CIP is used to identify drug boxes dispensed in community pharmacies
 - UCD is used to identify expensive drugs dispensed in the hospitals
 - LPP is mainly used to code reimbursed medical devices
 - CCAM is used to code outpatient and inpatient procedures
 - NABM is used to code inpatient and outpatient lab tests

Conclusions

- Most of the standardized clinical data tables can be generated following OHDSI Extract Transform and Load (ETL) process. However the mapping of the French terminologies for procedures, lab tests, drugs and devices to OMOP Standard Vocabularies is required to complete fully the transformation of the clinical part of EGB and SNDS and enable the use of the whole set of OHDSI tools.

