****

**OMOP Common Data Model (CDM V5.0)**

**Health Cost and Utilization Project (HCUP) – Nationwide Inpatient Sample (NIS)**

**Mapping Specification**

**Version 1.5**

**11-November-2016**

Martijn Schuemie

Erica Voss

Patrick Ryan

Chris Knoll

Contents

[Introduction 3](#_Toc466987732)

[Source Data Mapping Approach 3](#_Toc466987733)

[Data Mapping 4](#_Toc466987734)

[Table name: person 4](#_Toc466987735)

[Table name: visit\_occurrence 7](#_Toc466987736)

[Table name: observation\_period 9](#_Toc466987737)

[Table name: location 10](#_Toc466987738)

[Table name: care\_site 12](#_Toc466987739)

[Table name: condition\_occurrence 13](#_Toc466987740)

[Table name: death 16](#_Toc466987741)

[Table name: procedure\_occurrence 17](#_Toc466987742)

[Table name: measurement 22](#_Toc466987743)

[Table name: observation 25](#_Toc466987744)

[Table name: cdm\_source 28](#_Toc466987745)

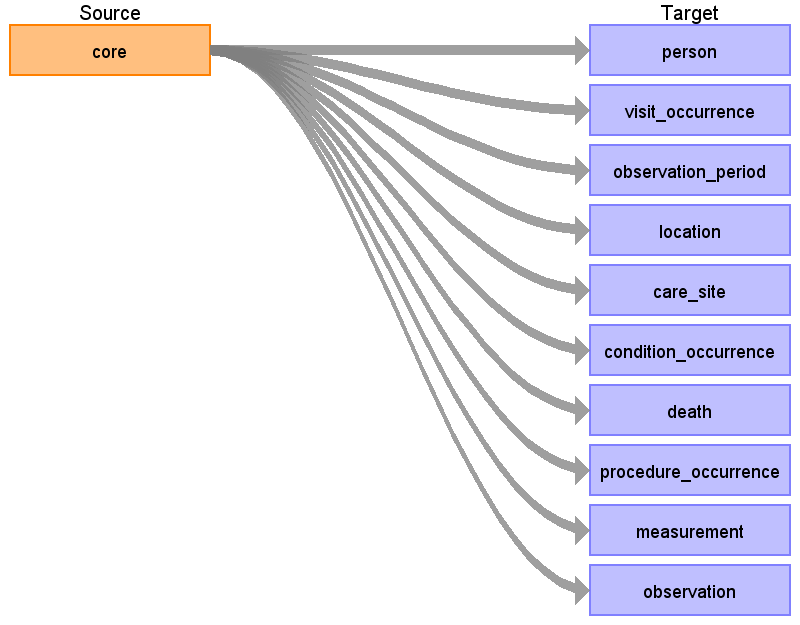
[Table name: cdm\_domain\_meta 29](#_Toc466987746)

# Introduction

This document reflects the requirements, assumptions, business rules and transformations for the Healthcare Cost and Utilization Project (HCUP) - Nationwide Inpatient Sample (NIS) implementation of the Common Data Model Version 5.0 (CDM) as implemented by Janssen.

The HCUP databases contain hospital data, collection of this data is sponsored by the Agency for Healthcare Research and Quality (AHRQ). The NIS HCUP data includes diagnoses, procedures, discharge status, demographics, and charges for hospital care in the United States, regardless of payer beginning in 1988. IT enables research on cost and quality of health services, medical practice patterns, access, and outcomes of treatments at the national, State, and local market levels.

# Source Data Mapping Approach

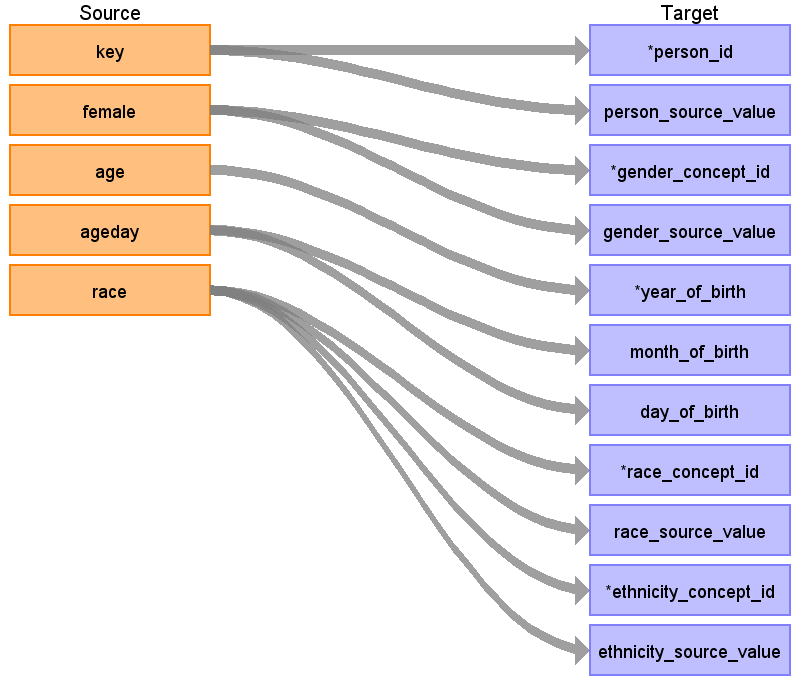


# Data Mapping

## Table name: person

Reading from core

Each record in the core table describes one visit, but we can't link multiple visits of the same person, so considering each visit a unique person.

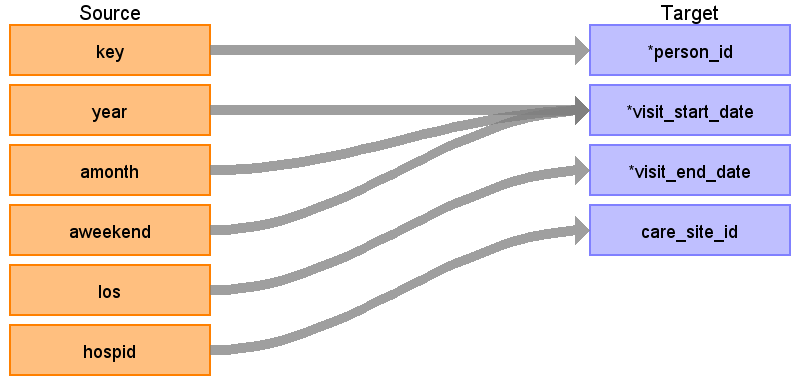


|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| person\_id | key |  | In 2013 format, ‘key’ is called ‘key\_nis’ |
| person\_source\_value | key |  | In 2013 format, ‘key’ is called ‘key\_nis’ |
| gender\_concept\_id | female | 1 = female (8532) 0 = male (8507) else = unknown (0) |  |
| gender\_source\_value | female |  |  |
| year\_of\_birth | age | if age > 0 then year\_of\_birth = year(visit\_start\_date) - age  else if ageday = NULL and age\_neonate = 1 then year\_of\_birth = year(visit\_start\_date)  else if (ageday = NULL and age\_neonate = 0) or age\_day < 0 then year\_of\_birth = year(visit\_start\_date – 180 days)  else year\_of\_birth + month\_of\_birth + day\_of\_birth = visit\_start\_date – ageday  If patient experiences age is >110 drop patient. |  |
| month\_of\_birth | ageday | see year\_of\_birth logic. Only filled if age = 0, else NULL |  |
| day\_of\_birth | ageday | see year\_of\_birth logic. Only filled if age = 0, else NULL |  |
| race\_concept\_id | race | See mapping table:  source\_value label concept\_id  1 White 8527  2 Black 8516  3 Other 8522  4 Pacific Islander 8557  5 Native American 8657  6 Other 0  If missing, also use 0 |  |
| race\_source\_value | race |  |  |
| ethnicity\_concept\_id | race | if race = 3, set to 38003563 (Hispanic), else 0 |  |
| ethnicity\_source\_value | race | if race = 3, put that value here, else 0 (note: will be stored in race\_source\_value) |  |
| time\_of\_birth |  |  |  |
| location\_id |  |  |  |
| provider\_id |  |  |  |
| care\_site\_id |  |  |  |
| gender\_source\_concept\_id |  |  |  |
| race\_source\_concept\_id |  |  |  |
| ethnicity\_source\_concept\_id |  |  |  |

## Table name: visit\_occurrence

Reading from core

Each record in the core table describes a single visit.



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| visit\_occurrence\_id |  |  | Autogenerate |
| person\_id | key |  | In 2013 format, ‘key’ is called ‘key\_nis’ |
| visit\_concept\_id |  |  | 9201 |
| visit\_start\_date | year  amonth  aweekend | visit date = year + amonth + ? (first day of month that is correct weekday (weekday or weekend))  if amonth < 1 then amonth = random month  if aweekend < 0 then aweekend = 0 |  |
| visit\_start\_time |  |  |  |
| visit\_end\_date | los | if los < 1 set to 0 (same day stay): start\_date = end\_date | There are some bogus values, but they are very infrequent. |
| visit\_end\_time |  |  |  |
| visit\_type\_concept\_id |  |  | 44818517 (Visit derived from encounter on claim) |
| provider\_id |  |  |  |
| care\_site\_id | hospid |  |  |
| visit\_source\_value |  |  |  |
| visit\_source\_concept\_id |  |  |  |

## Table name: observation\_period

Reading from core

Same logic as visit\_occurrence table (visit = observation\_period)

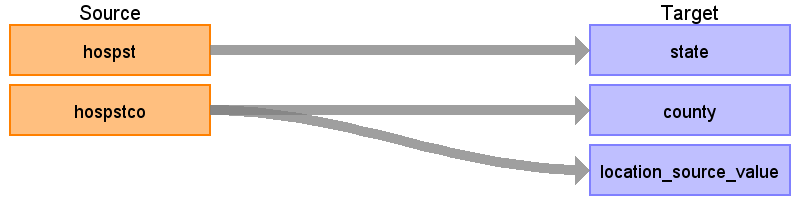
Generated

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| observation\_period\_id |  |  |  |
| person\_id |  |  |  |
| observation\_period\_start\_date |  |  |  |
| observation\_period\_end\_date |  |  |  |
| period\_type\_concept\_id |  |  | 44814724 Period covering healthcare encounters |

## Table name: location

Do not create a location record if both hospst and hospstco fields are null.

Reading from core



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| location\_id |  |  | For every combination of hospst + hospstco, a new location\_id is generated. hoststco = NULL and hospstco = -9999 are considered to be the same. |
| address\_1 |  |  |  |
| address\_2 |  |  |  |
| city |  |  |  |
| state | hospst |  |  |
| zip |  |  |  |
| county | hospstco | Lookup in FIPS table http://www.census.gov/geo/reference/docs/codes/national\_county.txt County name is truncated to 20 characters (note: the truncation has been verified to not create duplicate county names) |  |
| location\_source\_value | hospstco |  |  |

## Table name: care\_site

Reading from core



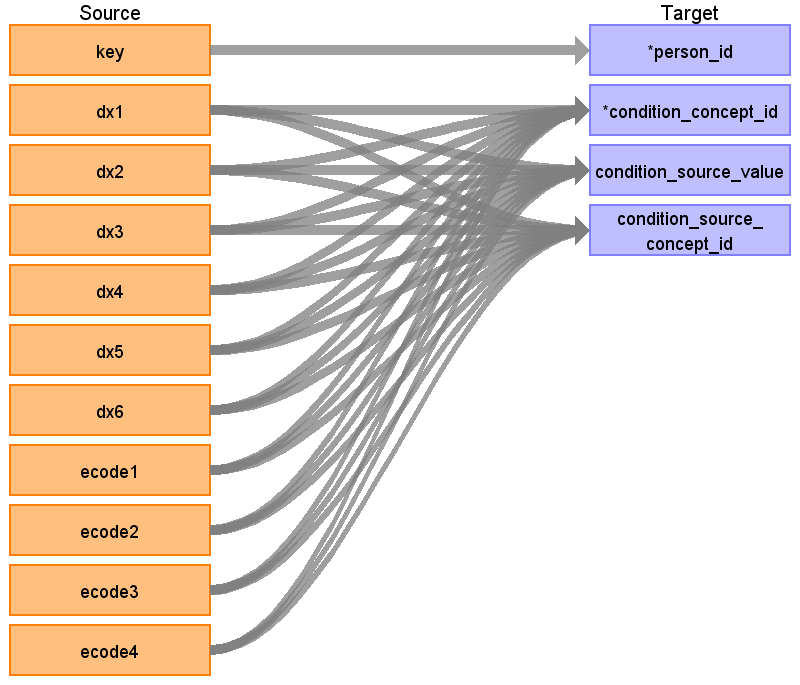
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| care\_site\_id | hospid |  |  |
| care\_site\_name |  |  |  |
| place\_of\_service\_concept\_id |  |  | 9201 |
| location\_id |  |  | From location table based on hospst + hospstco. null value when hospst and hospstco are both null. |
| care\_site\_source\_value | hospid |  |  |
| place\_of\_service\_source\_value |  |  |  |

## Table name: condition\_occurrence

Reading from core

Each records in the core table has 0, 1, or multiple diagnose codes. One condition\_occurrence record will be created for each code.

Use domain\_id attribute to see which dx and ecode map to conditions. If any of the diagnose fields maps to concept 4014295 (Single live birth) and the patient is either male or younger than 12, the code is discarted.



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| condition\_occurrence\_id |  |  | Autogenerate |
| person\_id | key |  | In 2013 format, ‘key’ is called ‘key\_nis’ |
| condition\_concept\_id | dx1  dx2  dx3  dx4  dx5  dx6  ecode1  ecode2  ecode3  ecode4 |  |  |
| condition\_start\_date |  |  | Use computed visit\_start\_date |
| condition\_end\_date |  |  |  |
| condition\_type\_concept\_id |  |  | 38000184 – 38000198 for dx1-dx25, respectively  if ecode then 38000184-38000185 |
| stop\_reason |  |  |  |
| provider\_id |  |  |  |
| visit\_occurrence\_id |  |  | From generated visit\_occurrence\_id |
| condition\_source\_value | dx1  dx2  dx3  dx4  dx5  dx6  ecode1  ecode2  ecode3  ecode4 |  |  |
| condition\_source\_concept\_id | dx1  dx2  dx3  dx4  dx5  dx6  ecode1  ecode2  ecode3  ecode4 |  | Lookup concept ID of source codes |

## Table name: death

Reading from core

Create a death record if the 'died' field has value '1'.



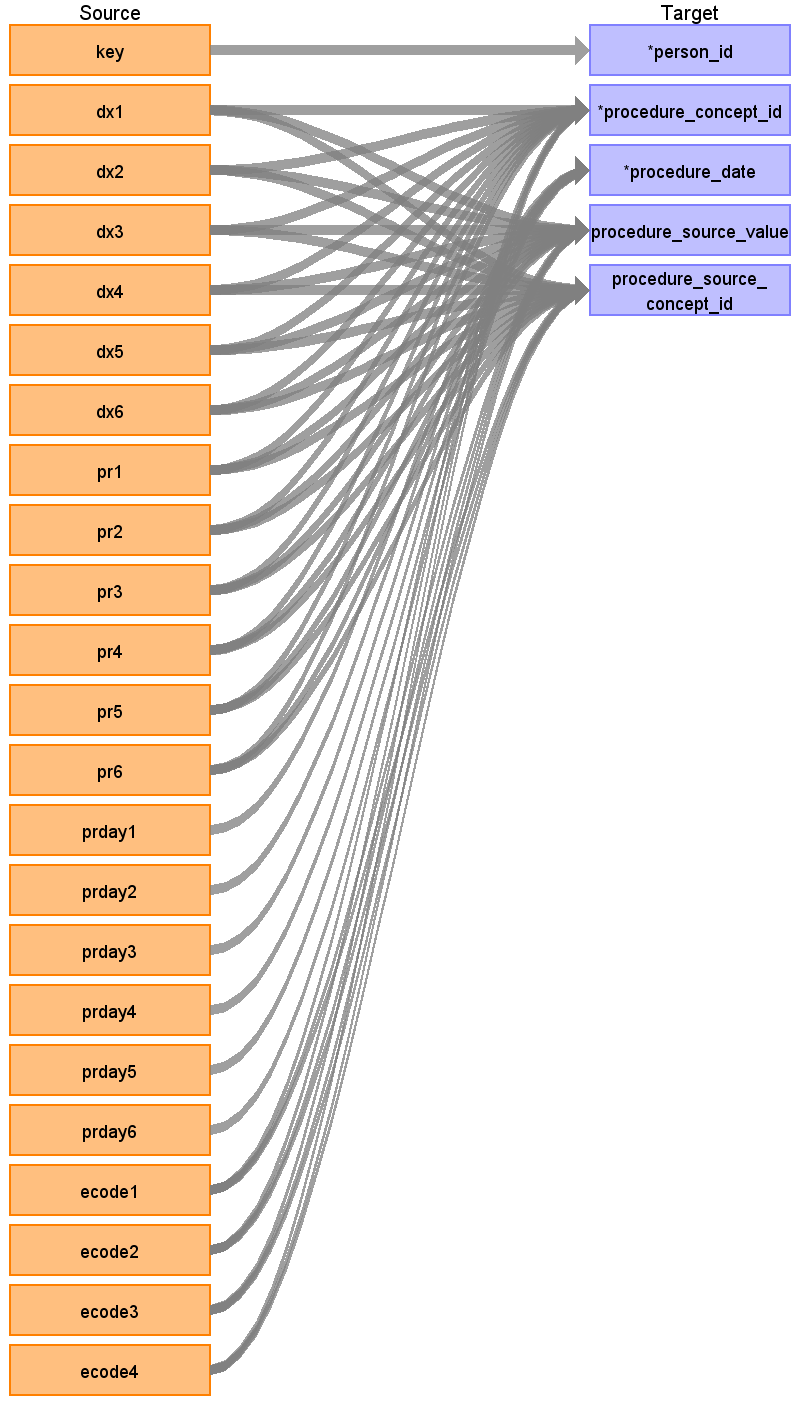
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| person\_id | key |  | In 2013 format, ‘key’ is called ‘key\_nis’ |
| death\_date |  |  | visit\_end\_date |
| death\_type\_concept\_id |  |  | 38003566 Medical claim discharge status "Died" |
| cause\_concept\_id |  |  |  |
| cause\_source\_value |  |  |  |
| cause\_source\_concept\_id |  |  |  |

## Table name: procedure\_occurrence

Reading from core

Each row in the core table can have 0, 1, or multiple procedure cores. One procedure\_occurrence record will be created for each code.

Additionally, some DX and ECODEs map to standard concepts in the 'procedure' domain and should also have records here.

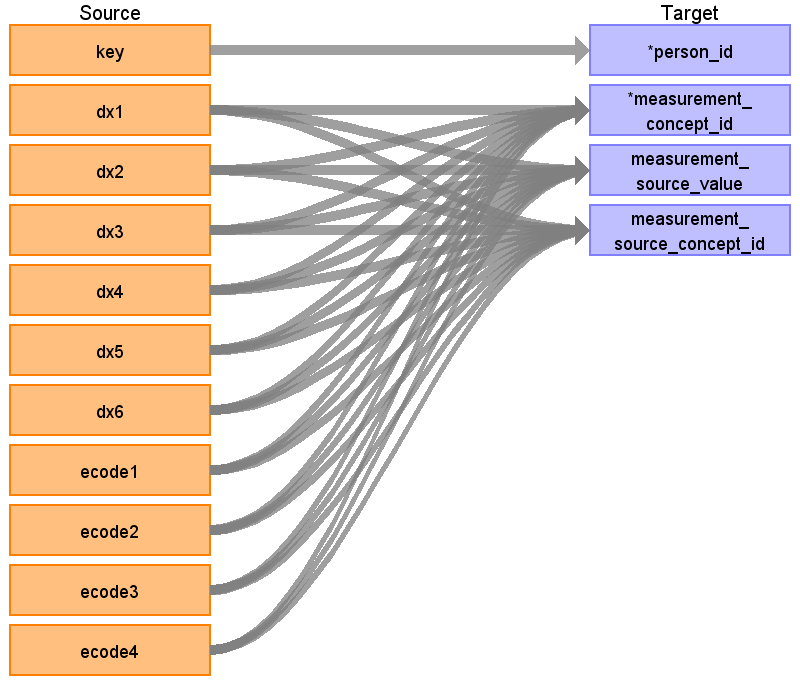


|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| procedure\_occurrence\_id |  |  | Autogenerate |
| person\_id | key |  | In 2013 format, ‘key’ is called ‘key\_nis’ |
| procedure\_concept\_id | dx1  dx2  dx3  dx4  dx5  dx6  ecode1  ecode2  ecode3  ecode4  pr1  pr2  pr3  pr4  pr5  pr6 |  |  |
| procedure\_date | prday1  prday2  prday3  prday4  prday5  prday6 | When using pr field then:  if prdayx < 1 then visit\_start\_date, else (visit\_start\_date + prdayx)  if prdayx > los then remove procedure  When using dx field use visit\_start\_date |  |
| procedure\_type\_concept\_id |  |  | 38000251 to 38000265, for pr1 through pr15  38000184 to 38000198 for dx1-dx15, respectively  if ecode then 38000184-38000185 |
| modifier\_concept\_id |  |  |  |
| quantity |  |  |  |
| provider\_id |  |  |  |
| visit\_occurrence\_id |  |  | From visit occurrence table |
| procedure\_source\_value | dx1  dx2  dx3  dx4  dx5  dx6  ecode1  ecode2  ecode3  ecode4  pr1  pr2  pr3  pr4  pr5  pr6 |  |  |
| procedure\_source\_concept\_id | dx1  dx2  dx3  dx4  dx5  dx6  ecode1  ecode2  ecode3  ecode4  pr1  pr2  pr3  pr4  pr5  pr6 |  | from visit occurrence table |
| qualifier\_source\_value |  |  |  |

## Table name: measurement

Reading from core

Some DX and ECODEs map to concepts in the 'measurement' domain and should have records here.

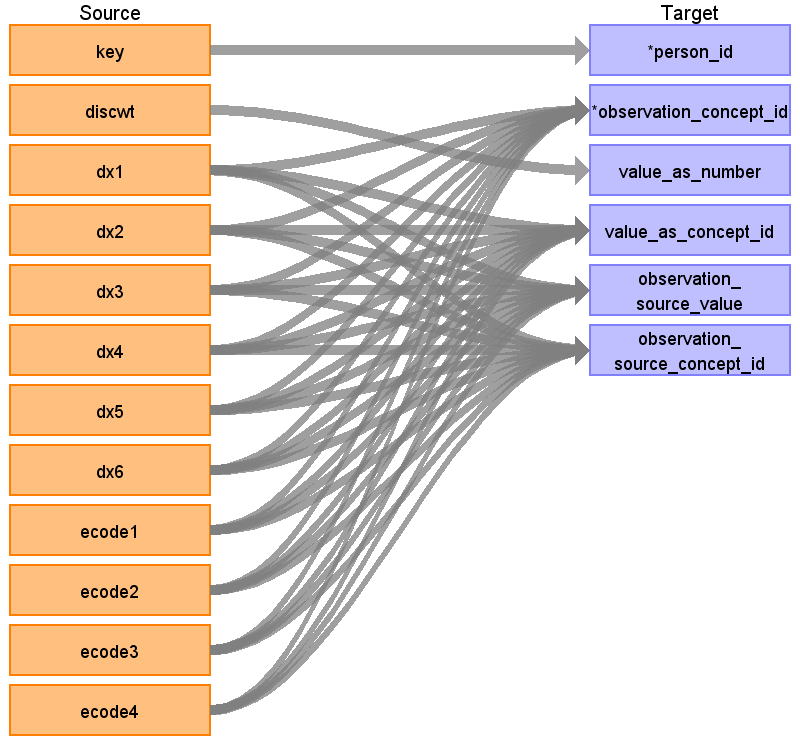


|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| measurement\_id |  |  | Autogenerate |
| person\_id | key |  | In 2013 format, ‘key’ is called ‘key\_nis’ |
| measurement\_concept\_id | dx1  dx2  dx3  dx4  dx5  dx6  ecode1  ecode2  ecode3  ecode4 |  |  |
| measurement\_date |  |  |  |
| measurement\_time |  |  |  |
| measurement\_type\_concept\_id |  |  | 38000184 – 38000198 for dx1-dx15, respectively  if ecode then 38000184-38000185 |
| operator\_concept\_id |  |  |  |
| value\_as\_number |  |  |  |
| value\_as\_concept\_id |  |  | Some codes map to specific value concepts through a 'Maps to value' relationship, else 4181412 (present) |
| unit\_concept\_id |  |  |  |
| range\_low |  |  |  |
| range\_high |  |  |  |
| provider\_id |  |  |  |
| visit\_occurrence\_id |  |  | from visit occurrence table |
| measurement\_source\_value | dx1  dx2  dx3  dx4  dx5  dx6  ecode1  ecode2  ecode3  ecode4 |  |  |
| measurement\_source\_concept\_id | dx1  dx2  dx3  dx4  dx5  dx6  ecode1  ecode2  ecode3  ecode4 |  | Look up concept ID of source codes. |
| unit\_source\_value |  |  |  |
| value\_source\_value |  |  |  |

## Table name: observation

Reading from core

Some DX and ECODEs map to concepts in the 'observation' domain and should have records here. Also, the sample weights (per discharge) are copied to the observation table.



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| observation\_id |  |  | Autogenerate |
| person\_id | key |  | In 2013 format, ‘key’ is called ‘key\_nis’ |
| observation\_concept\_id | dx1  dx2  dx3  dx4  dx5  dx6  ecode1  ecode2  ecode3  ecode4 |  |  |
| observation\_date |  |  |  |
| observation\_time |  |  |  |
| observation\_type\_concept\_id |  | 38000184 – 38000198 for dx1-dx15, respectively  if ecode then 38000184-38000185  if discwt then 900000003 |  |
| value\_as\_number | discwt | For each row in the core table, create an observation with value\_as\_number discwt. These field should also be populated: person\_id, observation\_date (visit\_start\_date), observation\_type\_concept\_id, observation\_source\_value. | Discharge weight |
| value\_as\_string |  |  |  |
| value\_as\_concept\_id | dx1  dx2  dx3  dx4  dx5  dx6  ecode1  ecode2  ecode3  ecode4 |  | Some codes map to specific value concepts through a 'Maps to value' relationship, else 45877994 (Yes) |
| qualifier\_concept\_id |  |  |  |
| unit\_concept\_id |  |  |  |
| provider\_id |  |  |  |
| visit\_occurrence\_id |  |  |  |
| observation\_source\_value | dx1  dx2  dx3  dx4  dx5  dx6  ecode1  ecode2  ecode3  ecode4 | If sample weight then ‘DISCWT’ |  |
| observation\_source\_concept\_id | dx1  dx2  dx3  dx4  dx5  dx6  ecode1  ecode2  ecode3  ecode4 |  | Look up concepts for source codes. |
| unit\_source\_value |  |  |  |
| qualifier\_source\_value |  |  |  |

## Table name: cdm\_source

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| cdm\_source\_name |  | HCUP National (Nationwide) Inpatient Sample |  |
| cdm\_source\_abbreviation |  | HCUP |  |
| source\_description |  | The HCUP National Inpatient Sample database represents hospital data, collected under sponsorship of the Agency for Healthcare Research and Quality (AHRQ). The data includes diagnoses, procedures, discharge status, demographics, and charges for hospital care in the United States, regardless of payer. Information on drugs and devices is not included. The HCUP data is based on hospital visits, with no information linking multiple visits of the same patient together. Cost information has not yet been included in the CDM. |  |
| source\_documentation\_reference |  | https://www.hcup-us.ahrq.gov/nisoverview.jsp |  |
| cdm\_etl\_reference |  | https://github.com/OHDSI/JCdmBuilder |  |
| cdm\_version |  | 5.0.1 |  |

## Table name: cdm\_domain\_meta

|  |  |  |
| --- | --- | --- |
| **TABLE NAME** | **DOMAIN\_ID** | **DESCRIPTION** |
| PERSON | Person | For every visit an entry has been created in the person table, because we are not able to link multiple visits from the same patient together. For most children under 1 year old at admission the exact date of birth could be derived, for all others only the age in years and hence the approximate year of birth is available. |
| OBSERVATION\_PERIOD | Observation Period | Observation periods are an exact copy of the visit occurrences; the patient is only assumed to be observed during their hospital visit. |
| CARE\_SITE | Care Site | One records has been created for each hospital in the database. |
| VISIT\_OCCURRENCE | Visit | Only the calendar month of the admission date is available, the day of the month has been imputed based on whether the admission date was classified as weekday or weekend. Visit end date is defined as visit start date + length of stay. |
| LOCATION | Location | For every county found in the database a location record has been created. |
| DEATH | Death | Data on in-hospital death was captured for almost all patients. |
| CONDITION\_OCCURRENCE | Condition | Principal and secondary diagnoses during hospitalization. Also included are external cause of injury codes. The condition start date is assumed to be the visit start date. |
| PROCEDURE\_OCCURRENCE | Procedure | Procedures performed during hospitalization. For procedures the exact day of the procedure (relative to the visit start date) is available. Some of the procedure occurrences are derived from diagnose codes using the vocabulary. |
| MEASUREMENT | Measurement | Some of the diagnose codes mapped to the measurement domain and were recorded here. |
| OBSERVATION | Observation | Some of the diagnose codes mapped to the observation domain and were recorded here. Also, the sample weights (per discharge) are recorded in the observation table. |