Data Dictionary for CMS Medicare Data

Why Were the Data Extracted and Transformed?

- Our hypothetical healthcare company – Acme Healthcare – decided to integrate CMS Medicare claims data with clinical data in their EMR system
    - Executive leaders believed that integrating the data would help them evaluate the quality of the claims data
- The Acme Healthcare analysts exported the claims data and transformed the data structure so that it could be loaded into tables within their EMR database
    - The data were put into new tables (see summary below), and some of the fields were renamed. For example, the fields in the CHRONIC table come from fields in the CMS files that has SP_...
    - In general, the prefix "**SP_**" was removed from the fields.
    - Analysts can search the fields in the DE-SynPUF codebook to gain more information about the original claims data

Where Do the Original Data Come From? Is there a Codebook?

- The original data come from the **CMS 2008-2010 Data Entrepreneurs' Synthetic Public Use File (DE-SynPUF).**
    - These CMS data are administrative claims data
- Although the DE-SynPUF data are synthetic to protect patient identity, the data fields and values are very similar to the real-world claims that were used to create the datasets.
    - For this assignment, it is reasonable to assume that these are real-world claims data
- You can look at the CMS codebook using the following link
    - https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/Downloads/SynPUF_Codebook.pdf

Summary of New Tables

Data Extracted from DE-SynPUF

- **PATIENT** – has one unique row per patient. Table includes demographic and Medicare-related fields
- **CHRONIC** -- provides information about chronic diseases for patients
- **ORDER_MED** – Medications ordered for patients
- **PAT_ENC** – Includes all the encounters that patients experienced
- **PAT_ENC_HSP** – Is a special table for inpatient or hospital encounters
- **PAT_ENC_DX** – Documents all the diagnoses as defined by ICD-9 codes that occurred for specific patient encounters
- **PAT_ENC_PR** – Documents all the diagnoses as defined by ICD-9 codes that occurred for specific patient encounters

Grouper Tables

- **ICD_PR** – Provides labels for the procedures as defined by ICD-9 procedure codes that occurred for specific patient encounters
- **ICD_DX** -- Provides labels for the diagnoses as defined by ICD-9 diagnosis codes that occurred for specific patient encounters
- **RX_NORM** – The RX_NORM file that includes NDC codes and associated labels and information about the specific drugs

Provider Table

- **NPPES_SAMPLE** -- This is a database of active National Provider Identifiers (NPI). Healthcare providers who want to bill for services for Medicare patients must obtain an NPI
- The NPPES table cannot be linked to the DE-SynPUF tables

How Do Can the Tables be Linked?

- The main field to link the tables is **PAT_ID**. This is a unique value for each unique patient, thus you should use this to link tables based on patients.
- Another important field for some of the tables is **PAT_ENC_ID**. This is a unique identifier that defines a specific patient encounter (or visit) to the health system.
  - For some analyses you may want to join data based on patients, whereas in others you may want to use the encounter ID.
- The NPPES table cannot be linked to the other tables

Fields and Values

PATIENT

- **PAT_ID** – patient ID
  - Unique values and used as key for joining tables
- **BIRTH_DATE** – birth date of patient
- **DEATH_DATE** -- death date of patient. If null, no evidence that the patient has died
- **SEX_CD** -- Gender.   1=male, 2=female
- **RACE_CD** – Race. 1=white, 2=black, 3=Others, 4=Hispanic
- **ESRN_IND** -- End stage renal disease Indicator.   0=does not have, 1=has ESRD
- **STATE_CD** -- State
  - 01=AL, 02=AK, 03=AZ, 04=AR, 05=CA, 06=CO, 07=CT, 08=DE, 09=DC, 10=FL, 11=GA, 12=HI, 13=ID, 14=IL, 15=IN, 16=IA, 17=KS, 18=KY, 19=LA, 20=ME, 21=MD, 22=MA, 23=MI, 24=MN, 25=MS, 26=MO, 27=MT, 28=NE, 29=NV, 30=NH, 31=NJ, 32=NM, 33=NY, 34=NC, 35=ND, 36=OH, 37=OK, 38=OR, 39=PA, 41=RI, 42=SC, 43=SD, 44=TN, 45=TX, 46=UT, 47=VT, 49=VA, 50=WA, 51=WV, 52=WI, 53=WY, 54=OTHERS (PUERTO RICO, VIRGIN ISLANDS, AFRICA, ASIA OR CALIFORNIA; INSTITUTIONAL PROVIDER OF SERVICES (IPS) ONLY, CANADA & ISLANDS, CENTRAL AMERICA AND WEST INDIES, EUROPE, MEXICO, OCEANIA, PHILIPPINES, SOUTH AMERICA, U.S. POSSESSIONS, AMERICAN SAMOA, GUAM, SAIPAN OR NORTHERN MARIANAS, TEXAS; INSTITUTIONAL PROVIDER OF SERVICES (IPS) ONLY, NORTHERN MARIANAS, GUAM, UNKNOWN.)

- **COUNTY_CD** – County code
- **HI_CVRAGE_TOT_MONS** -- Total number of months of Medicare part A coverage for the beneficiary
- **SMI_CVRAGE_TOT_MONS** -- Total number of months of Medicare part B coverage for the beneficiary
- **HMO_CVRAGE_TOT_MONS** -- Total number of months of HMO coverage for the beneficiary
- **PLAN_CVRG_MOS_NUM** -- Total number of months of Medicare part D plan coverage for the beneficiary

Chronic

- **PAT_ID** – patient ID
  - Unique values and used as key for joining tables
- **Chronic diseases** (1 = Yes, the patient has diseases, 2=No, the patient does not have the disease)
  - ALZHEIMER,
  - CHF,
  - CHRONIC_KIDNEY_DISEASE,
  - CANCER,
  - COPD,
  - DEPRESSION,
  - DIABETES,
  - ISCHEMIC_HEART,
  - OSTEOPOROSIS,
  - ARTHRITIS

ORDER_MED

- **PAT_ID** – patient ID
  - Unique values and used as key for joining tables
- **RX_ENC** -- Pharmacy encounter ID
- **CONTACT_DATE** – Date of the pharmacy encounter
- **NDC_CD** -- National Drug Code
- **QTY_DSPNSD_NUM** -- Quantity dispensed
- **DAYS_SUPLY_NUM** – Days supply
- **PTNT_PAY_AMT** – Patient payment amount
- **TOT_RX_CST_AMT** – Gross drug cost

PAT_ENC

- **PAT_ID** – Patient ID.   Unique values and used as key for joining tables
- **PAT_ENC_ID** – Patient encounter (visit id). Join to encounter ID in the DX and procedure tables
- **ADMIT_DIAGNOSIS** - Claim Admitting Diagnosis Code
- **CONTACT_DATE** – Encounter contact date
- **DEPARTMENT_ID** – Department ID
- **PROV_ATTENDING_ID** – Attending provider ID
- **PROV_OPERATING_ID** – Operating provider ID

PAT_ENC_HSP

- **PAT_ID** -- patient ID
    - o   Unique values and used as key for joining tables
- **PAT_ENC_ID** – Patient encounter (visit id). Join to encounter ID in the DX and procedure tables
- **ADMIT_DIAGNOSIS** - Claim Admitting Diagnosis Code
- **ADMISSION_DATE** – Inpatient admission date
- **DISCHARGE_DATE** -- Inpatient discharge date
- **DEPARTMENT_ID** – Department ID
- **PROV_ATTENDING_ID** – Attending provider ID
- **PROV_OPERATING_ID** – Operating provider ID

PAT_ENC_DX

- **PAT_ID** -- patient ID
    - o Unique values and used as key for joining tables
- **PAT_ENC_ID** – Patient encounter (visit id). Join to encounter ID in PAT_ENC and PAT_ENC_HSP
- **ICD_9** – International Classification of Disease Code version 9

PAT_ENC_PR

- **PAT_ID** -- patient ID
    - o   Unique values and used as key for joining tables
- **PAT_ENC_ID** – Patient encounter (visit id). Join to encounter ID in PAT_ENC and PAT_ENC_HSP
- **TYPE** – Identifies if the code on the row is ICD-9 procedure code of a Medicare HCPCS code
- **PROC_CD** – ICD-9 procedure of HCPCS code

NPPES_SAMPLE

- Acme analysts provided 60,000 rows of NPI data in the table called "NPPES_SAMPLE".
- These data come from the Provider Data National Plan and Provider Enumeration System (NPPES)
    - o   This is a database of active National Provider Identifiers (NPI). Healthcare providers who want to bill for services for Medicare patients must obtain an NPI
- Learn more about NPI data here:
    - o   https://npiregistry.cms.hhs.gov/
- See the data dictionary:
    - o   https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/NationalProvIdentStand/Downloads/Data_Dissemination_File-Readme.pdf