

Indiana University / Regenstrief Institute Empirical Data Characterization Summary

Table IA. Demographic Summary

This table contains general descriptive information about the patients in the DEMOGRAPHIC table. These patients may or may not be represented in other CDM tables.

| | N | % | Source table |
|----------------------|-----------|------|--------------------|
| Patients | 2,138,794 | | DEM_L3_N |
| Age | | | DEM_L3_AGEYRSDIST1 |
| Mean | 40 | | |
| Median | 38 | | |
| Age group | | | DEM_L3_AGEYRSDIST2 |
| 0-4 | 118,168 | 5.5 | |
| 5-14 | 279,507 | 13.0 | |
| 15-21 | 181,155 | 8.5 | |
| 22-64 | 1,165,075 | 54.5 | |
| 65+ | 394,889 | 18.4 | |
| Missing | 0 | 0.0 | |
| Hispanic | | | DEM_L3_HISPDIST |
| N (No) | 0 | 0.0 | |
| Y (Yes) | 83,756 | 3.9 | |
| Missing or Refused | 2,055,038 | 96.1 | |
| Sex | | | DEM_L3_SEXDIST |
| F (Female) | 1,137,779 | 53.2 | |
| M (Male) | 993,246 | 46.4 | |
| Missing or Ambiguous | 7,769 | 0.4 | |
| Race | | | DEM_L3_RACEDIST |
| White | 1,514,912 | 70.8 | |
| Non-White | 337,603 | 15.8 | |
| Missing or Refused | 286,279 | 13.4 | |

Table excludes records with values outside of CDM specifications.

Cell counts between 1 and the low-cell count threshold are displayed as BT and treated as missing.

Sums which include BT values are marked with an asterisk (*), either accompanying the sum of all above-threshold values or alone if all values are BT.

Table IB. Potential Pools of Patients

This table illustrates the number of patients meeting different inclusion criteria and supports data checks 3.04 and 3.05. Data check exceptions are highlighted in red and must be corrected.

| Metric | Metric Description | Result | Source table |
|---|--|-----------|-----------------------|
| Potential pool of patients for observational studies | Number of unique patients with at least 1 ED, EI, IP, or AV encounter within the past 5 years | 1,818,284 | ENC_L3_DASH2 |
| Potential pool of patients for trials | Number of unique patients with at least 1 ED, EI, IP, or AV encounter within the past 1 year | 895,675 | ENC_L3_DASH2 |
| Potential pool of patients for studies requiring data on diagnoses, vital measures and (a) medications or (b) medications and lab results | Number of unique patients with at least 1 ED, EI, IP, or AV encounter and DIAGNOSIS and VITAL records within the past 5 years | 225,502 | XTBL_L3_DASH1 |
| | Number of unique patients with at least 1 encounter and DIAGNOSIS, VITAL, and PRESCRIBING or DISPENSING records within the past 5 years | 134,170 | XTBL_L3_DASH2 |
| | Number of unique patients with at least 1 encounter and DIAGNOSIS, VITAL, PRESCRIBING or DISPENSING, and LAB_RESULT_CM records within the past 5 years | 115,118 | XTBL_L3_DASH3 |
| Patients with diagnosis data | Percentage of patients with encounters who have at least 1 diagnosis | 89% | ENC_L3_N; DIA_L3_N |
| Patients with procedure data | Percentage of patients with encounters who have at least 1 procedure | 92% | ENC_L3_N; PRO_L3_N |

Table excludes records with values outside of CDM specifications.

Cell counts between 1 and the low-cell count threshold are displayed as BT and treated as missing.

Sums which include BT values are marked with an asterisk (*), either accompanying the sum of all above-threshold values or alone if all values are BT.

Table IC. Height, Weight, and Body Mass Index

This table contains descriptive statistics and frequencies of VITAL table measurements.

| | Result | % | Source table |
|------------------------------------|---------|------|----------------|
| Height measurements | | | VIT_L3_HT_DIST |
| Records | 557,590 | | |
| Height (inches), mean | 67 | | |
| Height (inches), median | 66 | | |
| Weight measurements | | | VIT_L3_WT_DIST |
| Records | 682,081 | | |
| Weight (lbs.), mean | 191 | | |
| Weight (lbs.), median | 182 | | |
| Body Mass Index (BMI) measurements | | | VIT_L3_BMI |
| Records | 377,310 | | |
| BMI <=25 | 176,463 | 46.8 | |
| BMI 26-30 | 70,400 | 18.7 | |
| BMI >=31 | 130,447 | 34.6 | |

Table excludes records with null or missing values.

Cell counts between 1 and the low-cell count threshold are displayed as BT and treated as missing.

Sums which include BT values are marked with an asterisk (*), either accompanying the sum of all above-threshold values or alone if all values are BT.

Table ID. Records, Patients, Encounters, and Date Ranges by Table

This table contains summary counts and date ranges by table. The Encounters column shows the number of distinct ENCOUNTERIDS in each table. ENCOUNTERID is required for ENCOUNTER, DIAGNOSIS, and PROCEDURES and optional for VITAL, PRESCRIBING and DISPENSING. Missingness for ENCOUNTERID is depicted in Table IVD.

| Table | Records | Patients | Encounters | Data Range | | | Source Tables |
|---------------|------------|-----------|------------|----------------|----------|----------|---------------------------|
| | | | | Field name | Min | Max | |
| DEMOGRAPHIC | 2,138,794 | 2,138,794 | --- | BIRTH_DATE | 1926_Sep | 2024_Oct | DEM_L3_N; XTBL_L3_DATES |
| ENROLLMENT | 1,037,211 | 1,037,211 | --- | ENR_START_DATE | 2010_Jan | 2016_Nov | ENR_L3_N; XTBL_L3_DATES |
| ENCOUNTER | 32,898,282 | 2,086,387 | 32,898,282 | ADMIT_DATE | 2010_Jan | 2016_Dec | ENC_L3_N; XTBL_L3_DATES |
| DIAGNOSIS | 58,916,690 | 1,858,460 | 16,781,709 | ADMIT_DATE | 2010_Jan | 2016_Nov | DIA_L3_N; XTBL_L3_DATES |
| PROCEDURES | 33,124,706 | 1,927,952 | 29,181,955 | ADMIT_DATE | 2010_Jan | 2016_Dec | PRO_L3_N; XTBL_L3_DATES |
| DEATH | 43,802 | 43,802 | --- | DEATH_DATE | 1615_Oct | 2017_Jan | DEATH_L3_N; XTBL_L3_DATES |
| VITAL | 3,077,926 | 228,037 | 843,766 | MEASURE_DATE | 2009_Nov | 2015_Feb | VIT_L3_N; XTBL_L3_DATES |
| LAB_RESULT_CM | 21,041,024 | 1,173,623 | 6,113,496 | RESULT_DATE | 2009_Oct | 2016_Dec | LAB_L3_N; XTBL_L3_DATES |
| PRESCRIBING | 1,938,218 | 130,682 | 1,756,108 | RX_START_DATE | 2009_Sep | 2015_Feb | PRES_L3_N; XTBL_L3_DATES |
| DISPENSING | 76,596,701 | 945,282 | --- | DISPENSE_DATE | 2009_Sep | 2016_Dec | DISP_L3_N; XTBL_L3_DATES |

Cell counts between 1 and the low-cell count threshold are displayed as BT and treated as missing.

Table IE. Records Per Table By Encounter Type

This table contains record counts by encounter type for the ENCOUNTER, DIAGNOSIS, and PROCEDURES tables.

| Encounter Type | ENCOUNTER | | DIAGNOSIS | | PROCEDURES | |
|-----------------------------------|----------------|------|----------------|------|----------------|------|
| | N | % | N | % | N | % |
| AV (Ambulatory Visit) | 28,041,377 | 85.2 | 41,283,949 | 70.1 | 29,138,235 | 88.0 |
| ED (Emergency Dept) | 2,916,867 | 8.9 | 7,169,749 | 12.2 | 1,674,892 | 5.1 |
| EI (ED to IP Stay) | 0 | | 0 | | 0 | |
| IP (Inpatient Hospital Stay) | 1,182,200 | 3.6 | 9,469,876 | 16.1 | 2,305,657 | 7.0 |
| IS (Non-acute Institutional Stay) | 0 | | 0 | | 0 | |
| OA (Other Ambulatory Visit) | 0 | | 0 | | 0 | |
| Missing, NI, UN or OT | 757,838 | 2.3 | 993,116 | 1.7 | 5,922 | 0.0 |
| Total | 32,898,282 | | 58,916,690 | | 33,124,706 | |
| Source table | ENC_L3_ENCTYPE | | DIA_L3_ENCTYPE | | PRO_L3_ENCTTPE | |

Table IG. Lab Results

This table illustrates the number of records and number of unique patients for each LAB_NAME, and the percentage of patients in the ENCOUNTER table who have these results. Although there is not a required relationship between the ENCOUNTER and LAB_RESULT_CM tables, patients with encounters are the most relevant denominator for this table. LAB_NAME will be NULL or missing for lab results which do not correspond to a common lab measure.

| LAB_NAME | Records | Percentage of records in the LAB_RESULT_CM table | Patients | Percentage of patients in the ENCOUNTER table | Source tables |
|-----------------|------------|--|-----------|---|----------------------|
| A1C | 1,091,471 | 5.2 | 374,817 | 18.0 | LAB_L3_NAME;ENC_L3_N |
| CK | 852,711 | 4.1 | 215,169 | 10.3 | LAB_L3_NAME;ENC_L3_N |
| CK_MB | 524,983 | 2.5 | 149,971 | 7.2 | LAB_L3_NAME;ENC_L3_N |
| CK_MBI | 954 | 0.0 | 729 | 0.0 | LAB_L3_NAME;ENC_L3_N |
| CREATININE | 7,256,758 | 34.5 | 993,858 | 47.6 | LAB_L3_NAME;ENC_L3_N |
| HGB | 7,742,037 | 36.8 | 1,047,889 | 50.2 | LAB_L3_NAME;ENC_L3_N |
| INR | 1,653,658 | 7.9 | 273,155 | 13.1 | LAB_L3_NAME;ENC_L3_N |
| LDL | 1,092,916 | 5.2 | 444,130 | 21.3 | LAB_L3_NAME;ENC_L3_N |
| TROP_I | 825,536 | 3.9 | 212,365 | 10.2 | LAB_L3_NAME;ENC_L3_N |
| TROP_T_QL | 0 | | 0 | 0.0 | LAB_L3_NAME;ENC_L3_N |
| TROP_T_QN | 0 | | 0 | 0.0 | LAB_L3_NAME;ENC_L3_N |
| NULL or missing | 0 | | 0 | 0.0 | LAB_L3_NAME;ENC_L3_N |
| Total | 21,041,024 | | 3,712,083 | | |

Table excludes records with values outside of CDM specifications.
 Cell counts between 1 and the low-cell count threshold are displayed as BT and treated as missing.

Table IIID. Encounters Per Visit and Per Patient

This table shows the number of encounters, patients, visits (unique combinations of PATID, ADMIT_DATE, ENC_TYPE, and PROVIDER_ID), encounters per patient, and encounters per visit by encounter type and supports Data Check 2.04. A high number of encounters per visit may signal potential redundancy or duplication. Data check exceptions are highlighted in blue and should be investigated.

| Encounter Type | Encounters | Patients | Encounters per Patient | Encounters with known PROVIDERID | Visit (unique combinations of PATID, ENC_TYPE, ADMIT_DATE, and PROVIDERID) | Encounters with known PROVIDERID per visit | Source table |
|-----------------------------------|------------|-----------|------------------------|----------------------------------|--|--|----------------|
| AV (Ambulatory Visit) | 28,041,377 | 1,829,657 | 15.3 | 28,041,377 | 28,041,377 | 1.00 | ENC_L3_ENCTYPE |
| ED (Emergency Dept) | 2,916,867 | 846,954 | 3.4 | 2,916,867 | 2,916,867 | 1.00 | ENC_L3_ENCTYPE |
| EI (ED to IP Stay) | 0 | 0 | | 0 | 0 | | ENC_L3_ENCTYPE |
| IP (Inpatient Hospital Stay) | 1,182,200 | 526,857 | 2.2 | 1,182,200 | 1,182,200 | 1.00 | ENC_L3_ENCTYPE |
| IS (Non-acute Institutional Stay) | 0 | 0 | | 0 | 0 | | ENC_L3_ENCTYPE |
| OA (Other Ambulatory Visit) | 0 | 0 | | 0 | 0 | | ENC_L3_ENCTYPE |
| Missing, NI, UN or OT | 757,838 | 152,050 | 5.0 | 757,838 | 757,838 | 1.00 | ENC_L3_ENCTYPE |

Cell counts between 1 and the low-cell count threshold are displayed as BT and treated as missing.

Sums which include BT values are marked with an asterisk (*), either next to the sum of the above-threshold values or alone if all values are BT.