Sentinel Data Quality Metrics for Electronic Health Records

Protocol Definitions

Prepared by the Sentinel Operations Center

Version 1.0

January 20, 2022

The Sentinel System is sponsored by the U.S. Food and Drug Administration (FDA) to proactively monitor the safety of FDA-regulated medical products and complements other existing FDA safety surveillance capabilities. The Sentinel System is one piece of FDA's Sentinel Initiative, a long-term, multi-faceted effort to develop a national electronic system. Sentinel Collaborators include Data and Academic Partners that provide access to healthcare data and ongoing scientific, technical, methodological, and organizational expertise. The Sentinel Operations Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number 75F40119D10037.

Sentinel Data Quality Metrics for Electronic Health Records

Table of Contents

History of Modifications

Version	Date	Modification	Author
1.0.0	01/20/2022	Original Public Version	Sentinel Operations Center

1 Introduction

The U.S. Food & Drug Administration (FDA) uses the Sentinel System for pharmaceutical safety surveillance and regulation. As part of an ongoing initiative to broaden and improve data sources, the FDA has prioritized creation of a Real-World Evidence (RWE) Data Enterprise.[1] An essential part of this data enterprise is the incorporation of Electronic Health Record (EHR) data from about 10 million lives.

Historically, the Sentinel System has been primarily composed of administrative claims data to allow for nearly complete longitudinal capture of patient healthcare data as a consequence of enrollment in health plans. However, claims data has known limitations, including data lag, lack of detailed (granular) inpatient data, and lack of structured clinical measurements such as body mass index. Yet, in the U.S., these types of data elements are often scattered across multiple EHR systems maintained by different healthcare organizations without interoperability, which makes longitudinal analyses challenging. In addition to many EHRs, the current public health data landscape includes many EHR-based common data models to increase interoperability.[2][3]

Sentinel's data characterization and quality procedures are foundational to declaring data fit-for-purpose. They have evolved over a decade and are a gold standard for data checking. Yet, they are primarily based on data formatted to the Sentinel Common Data Model including checks for completeness, validity, accuracy, integrity, and consistency.[4] To assess fit-for-purposeness among EHR-based databases that are candidates for inclusion in the RWE Data Enterprise, these data quality metrics needed adaptation.

With the support of the Department of Health and Human Services, Assistant Secretary for Planning and Evaluation (ASPE), and FDA, Sentinel previously created a harmonized data characterization toolkit to address this challenge.[5] The toolkit was built on the work done by MG Kahn et al.[6] and established a framework to collect data quality standards and reporting metrics, a process for collating measure results, and software to analyze and compare results by data source. The initial steps in utilizing this approach are to author data quality metrics and assess the measures generated when executing the metric against a data source.

In collaboration with FDA and other clinical thought leaders, we have defined a set of data quality metrics, with a specific focus on efficiently assessing and comparing fitness-for-purpose in EHR data. Although previous work describes data quality in three categories: (1) Conformance, (2) Completeness, and (3) Plausibility, the metrics defined in this document focus

---

[1] Gottlieb, S. (2018, June 10). FDA budget matters: Notes on data and Real World Evidence. U.S. Food and Drug Administration. https://www.fda.gov/news-events/fda-voices/fda-budget-matters-cross-cutting-data-enterprise-real-world-evidence

[2] Weeks, J., & Pardee, R. (2019). Learning to share health care data: a brief timeline of influential common data models and distributed health data networks in US health care research. eGEMs, 7(1).

[3] Birkhead, G. S., Klompas, M., & Shah, N. R. (2015). Uses of electronic health records for public health surveillance to advance public health. Annual review of public health, 36, 345-359.

[4] Sentinel Operations Center. (2017, February 27). Sentinel Data Quality Assurance Practices: Compliance With "Guidance for Industry and FDA Staff: Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data." https://sentinelinitiative.org/sites/default/files/data/distributed-database/Sentinel_DataQAPractices_Memo.pdf

[5] See Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data. (2018, September 14). https://www.sentinelinitiative.org/methods-data-tools/methods/standardization-and-querying-data-quality-metrics-and-characteristics

[6] Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., ... & Schilling, L. (2016). A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. Egems, 4(1).

on the latter two. They do not require conformance to a standard data model since conformance quality checks "describe the compliance of the representation of data against internal or external formatting, relational, or computational definitions."[7]

This document lists quality metrics alongside rationale for inclusion and technical instruction for implementing to produce measurable results.

2 Glossary of Terms

The following terms will be referenced in subsequent sections of this document.

Medical fact – a unit of utilization that represents a medical observation on a patient including, but not limited to, the following types: diagnosis; procedure; medication (ordered or dispensed); laboratory test: vital sign; clinical observation; patient-reported outcomes; "lifestyle factor," e.g. smoking, alcohol consumption, recreational drug use, BMI, height, weight, etc.

Encounter – a unit of utilization that represents a medical interaction between a patient and provider. The uniqueness of an encounter is dependent on care setting, type of service, and site/source system. For example, the conversion of an interaction in the Ambulatory (AV) or Emergency (ED) setting to the Inpatient (IP) or Institutional Stay (IS) through an admission would be represented by multiple encounters, as would a transfer from one facility to another. A well visit or sick visit to a PCP would represent a single encounter regardless of whether multiple providers/clinicians are part of the interaction. While encounters in billed claims systems will generally include at least one medical fact, this is not always the case in EHR systems.

Setting – represents the medical setting where a medical fact or encounter took place. For the purposes of this assessment, the following types of settings will be utilized:

- Ambulatory Visit (AV): Includes outpatient clinics, physician offices, same day/ambulatory surgery centers, urgent care facilities, and other same-day ambulatory hospitals. Excludes emergency department.
- Emergency department (ED): Includes ED visits that become inpatient stays. Excludes urgent care facilities and observation stays.
- Inpatient (IP): Includes same-day hospital discharges, hospital transfers, and acute hospital care where the discharge is after the admission date. Excludes observation stays.
- Observation Stay (OS): "Hospital outpatient services given to help the doctor decide if the patient needs to be admitted as an inpatient or can be discharged. Observations services may be given in the emergency department or another area of the hospital." Definition from Medicare, CMS Product No. 11435, https://www.medicare.gov/Pubs/pdf/11435.pdf.
- Non-Acute Institutional Stay (IS): Includes hospice, skilled nursing facility (SNF), rehab center, nursing home, residential, overnight non-hospital dialysis, and other non-hospital stays.
- Telehealth (TH): Includes telemedicine or virtual visits, which can be conducted via video, phone, or other means.

[7] Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., ... & Schilling, L. (2016). A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. Egems, 4(1).

- Other Ambulatory Visit (OA)– includes non-overnight AV encounters such as at-home hospice, home health, skilled nursing, and other non-hospital visits. May also include "only" visits for pharmacy, labs, images, etc., that represent events outside a face-to-face visit.
- Other (OT) – all other settings not specified in this documentation
- Unknown (UN)- indicates value of unknown provided in system.
- Missing (MS)– indicates this information is not provided or available in system.

Completeness Metrics – metrics that assess missing data at one or more moments in time and are unrelated to the plausibility or conformance of the data values.

Plausibility Metrics – metrics that focus on data values being "truthful" or "believable." Unlike completeness metrics, a data value being present does not alone define quality. Extreme values, values that contradict related values, and values that are out of line with real-world expectations are key concerns for plausibility metrics that may indicate issues in data capture or transformation.

Consistency Metrics – in the Kahn framework, consistency metrics are a form of temporal plausibility. For the purposes of Sentinel data, consistency is a key component to data quality initiatives for ongoing research data partners. Identifying and understanding changes in values over time in both variables that are time-varying and intended to be static is critical.

3 Fact-Based Quality Metrics

3.1 Missing Demographics

3.1.1 Description

This metric provides rates of missingness for persons with recorded medical facts during a measurement period. Missingness is assessed in demographic data elements that include date of birth, age, sex, race, and ethnicity. Results are stratified by the source of the demographic element.

3.1.2 Rationale

The review of missing demographic information is used to assess data completeness. A significant proportion of missing demographic data could prevent accurate patient characterization or cohort capture.

3.1.3 Technical Methods

A site may elect to provide documentation and omit portions of this reporting if (1) one of the demographic elements evaluated is known to be entirely populated or if (2) information needed to capture the metric, e.g. calculated age or date of demographic update/fact, is not available.

1. Select the measure population of patients by identifying those with at least one medical fact dated between 1/1/2015 and 12/31/2019 inclusive. Exclude patients identified via scheduled or canceled medical fact if information to do so is available at site.
2. Link to separate fact tables as necessary to attach demographic elements.
    a. Include variables representing the following elements: date of birth, sex, race, and ethnicity.
    b. If duplicates exist, retain one record per distinct patient using the following logic:
        i. If date of demographic update/demographic fact date available, select single record with **latest** update/demographic fact date. If duplicates

remain, select single record per patient at random. Record source of demographic date as "DEMOGRAPHIC".

ii. Otherwise, select single record with **earliest** medical fact date. If duplicates remain, select single record per patient at random. Record source of demographic date as "FACT".

3. Group data to count overall distinct patients and those with missing demographic elements. See Appendix for output file format.

3.2 Utilization in 65+ Population

3.2.1 Description

This metric quantifies medical utilization among patients aged 65+ during a measurement period. Utilization is assessed relative to an index date that is defined, per patient, as their earliest fact date and then measured in 1, 3, and 5+ year intervals. Results are stratified by the year of the index medical fact. Persons without utilization are included in results but will present with a missing index year.

3.2.2 Rationale

These results will be used to assess completeness. There is an expectation that persons aged 65 or older will have at least one healthcare encounter with a diagnosis or procedural fact after turning 65 years of age due to an increase in chronic and/or acute conditions.

3.2.3 Technical Methods

1. Select the measure population as **distinct patients** aged 65 years or more between 1/1/2015 and 12/31/2019 inclusive.
2. Exclude patients with death record between 1/1/2015 and 12/31/2019 inclusive.
3. Assign a single index date per patient by selecting the earliest fact date recorded between 1/1/2015 and 12/31/2019 inclusive. Please note:
    a. Scheduled or canceled medical facts should be excluded when information to do so is available at site.
    b. If no fact identified during this period, index date should be set to missing.
4. Utilize the index date to identify and flag persons with at least **one diagnosis or procedure medical fact** within 1, 3, and 5 years following. Note that:
    a. Scheduled or canceled medical facts should be excluded when information to do so is available at site.
    b. This logic does not apply to persons without an index date. Flag values for these patients should default to 0/not true.
5. Group data by year of index date and time interval evaluated to get the count of distinct patients per stratum alongside the count of distinct patients with a recorded medical fact. Note patients without an index event should be included in these results, but with a missing year value. See Appendix for output file format.

3.3 Infant Utilization

3.3.1 Description

This metric captures medical utilization rates in newborns. Utilization is measured in 6-, 12- and 24-month increments relative to date of birth. Results are stratified by year of birth.

3.3.2 Rationale

It is expected that infants will have at least one healthcare encounter associated with a diagnosis or procedural fact in the first two years of life due to the recommended schedule of vaccinations. Lack of this information may be a signal that data is not complete.

3.3.3 Technical Methods

1. Select the measure population as **distinct patients** with a date of birth between **1/1/2015** and **12/31/2019** inclusive.
2. Exclude patients with death record between 1/1/2015 and 12/31/2019 inclusive.
3. Use each patient's date of birth as the index to identify and flag patients with at least one recorded **diagnosis or procedure medical fact** in 6-, 12- and 24-month increments. Please note that scheduled or canceled medical facts should be excluded if information to do so is available at site.
4. Group data by year of birth to get the count of distinct patients with medical facts according to file format in Appendix.

3.4 Counts by Diagnosis Code

3.4.1 Description

This metric lists the most common diagnosis codes recorded at a site over time. Diagnosis codes are limited to the first three characters. Output is limited to the 20 most common records per year according to distinct patient counts.

3.4.2 Rationale

This metric will be used to assess plausibility. Results should be interpreted in context with known characteristics of a site to identify unexpected results that may be indicative of an issue with data capture and/or translation.

3.4.3 Technical Methods

1. Select all **diagnosis medical facts for ICD codes** dated between **1/1/2016** and **12/31/2019** inclusive. Exclude scheduled or canceled medical facts if information to do so is available at site.
2. Limit each diagnosis code to the first three characters, not including punctuation.
3. Group data by year of medical fact and diagnosis code substring to get the count of distinct patients and total facts associated with each stratum.
4. Per each stratum, retain the 20 records with high value of distinct patient counts. See Appendix for output file format.

3.5 Invalid Age

3.5.1 Description

This metric provides rates of invalid age in medical facts reported during a measurement period. Invalid age categories include persons less than 0 or greater than 120 years. Results are stratified by year and include the count of total facts and distinct patients.

3.5.2 Rationale

This information will be used to assess plausibility. While it is possible that the presence of facts for either age group is valid, a high count taken into consideration with a site's known population characteristics may be indicative of a data capture or translation issue.

3.5.3 Technical Methods

1. Define the measure population as patients with at least one medical fact, including all types, dated between **1/1/2015** and **12/31/2019** inclusive. Exclude scheduled or canceled medical facts if information to do so is available at site.
2. Link to separate facts tables and process as necessary to:
    a. Assign or calculate patient age in years at time of fact.
    b. Limit records to those for patients aged < 0 (exclude missing) or > 120 years.
3. Group data by year of fact and age category to get the count of distinct patients and total facts associated with each stratum. See Appendix for output file format.

3.6 Post-Death Events

3.6.1 Description

This metric provides rates of events post-death. The population includes distinct patients with a date of death during a measurement period. Post-death events measured include medical facts and births. Results are stratified by year of death.

3.6.2 Rationale

Medical facts recorded after death may be valid in specific situations. Date of birth following death would be invalid in all contexts. The results of this metric will be used in concert with known site, and population attributes to assess plausibility.

3.6.3 Technical Methods

1. Define the measure population as distinct patients with date of death between 1/1/2010 and 12/31/2019 inclusive.
2. Link to separate fact tables as necessary and use each patient's date of death as the index to identify patients with date of birth and/or one or more medical facts following death.
3. Group data by year of death to get the count of distinct deceased patients and those with post-death births and/or facts. See Appendix for output file format.

3.7 Fact Types by Year

3.7.1 Description

This metric provides the count of medical facts reported during a measurement period stratified by medical fact type and year.

3.7.2 Rationale

These rates will be reviewed to measure plausibility relative to clinical and statistical expectations, as well as consistency over time.

3.7.3 Technical Methods

1. Select medical facts dated between 1/1/2015 and 12/31/2019 inclusive. Exclude scheduled or canceled medical facts if information to do so is available at site.
2. Categorize each fact into one of the following types:
    a. PRO - Procedure
    b. DIAG - Diagnosis
    c. ORX - Ordered Medication
    d. DRX - Dispensed Medication
    e. LAB - Laboratory Test (includes order and result)

f. VIT - Vital Sign (include all types)
g. OTH – Other
h. UNK – Unknown

3. Group data by year of fact date and type to get the count of distinct patients and total facts per stratum. See Appendix for output file format.

4 Encounter-Based Quality Metrics

The concept of an encounter and its ability to accurately and consistently capture specific data elements may be critical to certain research aims. However, it is known that this unit may not be available or consistently populated at all data sites. As such, the encounter-based metrics described in this section should be limited to data where encounter is defined in the source system, as opposed to being manufactured during transformation.

4.1 Missing Encounter Data Elements

4.1.1 Description

This metric provides the count of distinct encounters within a measurement period and those without specific, populated data elements. Encounters are grouped by setting and stratified by year. Data elements evaluated include admission/start dates and discharge/end dates.

4.1.2 Rationale

Missing encounter data rates can be used to assess data completeness and plausibility. A significant proportion of missing information may affect one's ability to accurately capture cohorts or identify events within an EHR system. Further, unexpected rates by setting may reflect errors in data capture and/or transformation.

4.1.3 Technical Methods

1. Select encounters with start dates between 1/1/2015 and 12/31/2019 inclusive or missing date information.
2. Link to separate fact tables and process data as necessary to attach the following data elements: setting, admission/start date, and discharge/end date.
3. Group the value of encounter setting into one of the following categories, noting if setting is not available/null. This value should be set to MS (missing):
    a. AV - Ambulatory Visit
    b. ED - Emergency Department
    c. IP - Inpatient Stay
    d. OS - Observation Stay
    e. IS - Institutional Stay
    f. TH – Telehealth
    g. OA – Other Ambulatory Visit
    h. OT - Other
    i. UN - Unknown
    j. MS - Missing
4. Group data by setting to get the total count of distinct encounters and those with a missing data element. See Appendix for output file format.

4.2 Post-Discharge Facts

4.2.1 Description

This metric provides the counts of inpatient encounters with and without diagnostic and/or procedural facts recorded in days 3, 7, 30, and 90 post-discharge. Results are stratified by year of encounter.

4.2.2 Rationale

These results will be used to assess data completeness and plausibility. Unexpected rates may indicate data is not adequate for identifying and characterizing a cohort or that there is an issue with data capture and/or transformation.

4.2.3 Technical Methods

1. Select inpatient encounters with start dates between 1/1/2015 and 12/31/2019 inclusive.
2. Link to separate dimension(s) and process data as necessary to identify and flag encounters with diagnosis or procedure facts within 3, 7, 30, and 90-days following discharge.
3. Separately group data by year of discharge date to get the total count of distinct encounters and those with a fact recorded in each evaluation period.
4. See Appendix for output file format.

4.3 Encounter Attributes by Year

4.3.1 Description

This metric includes all encounters captured during a measurement period grouped by setting to provide basic statistics for length of stay and associated medical facts. Results are stratified by the month and year of encounter.

4.3.2 Rationale

This information can be used to measure data consistency and plausibility. Deviations from statistical expectations may indicate the need for additional investigation, while the presence or absence of specific information in a specific encounter setting may indicate an issue with data capture or transformation.

4.3.3 Technical Methods

1. Process data to isolate encounters with start/admission dates between 1/1/2015 and 12/31/2019 inclusive. Note if a start or admission date is not available, a different, most appropriate date field should be utilized.
2. Link to separate fact tables and process data as necessary to:
    a. Define encounter setting using the following categories:
        i. AV - Ambulatory Visit
        ii. ED - Emergency Department
        iii. IP - Inpatient Stay
        iv. OS - Observation Stay
        v. IS - Institutional Stay
        vi. TH – Telehealth
        vii. OA-Other Ambulatory
        viii. OT - Other
        ix. UN - Unknown

x. MS – Missing

b. Calculate length of stay, noting a missing value is permissible when either discharge date/end date or admission date/start date is missing.
c. Identify and flag encounters linked to one or more medical facts (all types included), procedural fact, and diagnostic fact.

3. Group data by encounter month-year and setting to calculate the following per stratum:
    a. Distinct patients
    b. Distinct encounters
    c. Distinct encounters without an associated fact including all types combine, procedural only, and diagnostic only
    d. Minimum, median, maximum, mean facts (include all types) per encounter
    e. Minimum, median, maximum, mean procedural facts per encounter
    f. Minimum, median, maximum, mean diagnostic facts per encounter
    g. Distinct encounters with missing length of stay
    h. Minimum, median, maximum, mean length of stay (excluding missing values)

See Appendix for file format.

5 Appendix

5.1 Missing Demographics

Table 1 – File format for reporting missing demographics measure

Field	Description	Type	Population/ Mapping	Example
1	Demographic Variable	Character varying (20)	dob = date of birth sex race ethnicity	dob
2	Source of demographic update	Character varying (10)	fact other	fact
3	Count of distinct patients	Integer		345, 557
4	Count of distinct patients with null demographic element	Integer		455

5.2 Utilization in 65 + Population

Table 2- File format for reporting utilization in cohort aged 65 years or greater measure

Field	Description	Type	Population/ Mapping	Example
1	Year of fact	Date	Missing, 2015-2019	2019
2	Utilization window	Character varying (20)	0-1 year 0-3 years 0-5 years	0-1 years
3	Total distinct patients	Integer		500
4	Count of distinct patients with fact within utilization window	Integer		100

5.3 Infant Utilization

Table 3- File format for reporting infant utilization

Field	Description	Type	Population/ Mapping	Example
1	Year of birth	Date	ccyy	2019
2	Utilization window	Character varying (20)	0-6 months 0-12 months 0-24 months	0-6 months
3	Total distinct patients	Integer		34554
4	Count of distinct patients with fact within utilization window	Integer		33221

5.4 Counts by Diagnosis Code

Table 4- File format for reporting counts by diagnosis code

Field	Description	Type	Population/ Mapping	Example
1	Diagnosis fact year	Date	ccyy	2019
2	First three alphanumeric characters of diagnosis code	Character varying (5)		J18
3	Fact count	Integer		57000
4	Distinct patient count	Integer		56881

5.5 Invalid Age

Table 4- File format for the invalid age measure

Field	Description	Type	Population/ Mapping	Example
1	Fact year	Date	ccyy	2019
2	Invalid age category	Character varying (50)	< 0 years >= 120 years	<0 years

Field	Description	Type	Population/ Mapping	Example
3	Total facts	Integer		250
4	Distinct patient count	Integer		235

5.6 Post-Death Events

Table 5- File format for the post-death events measure

Field	Description	Type	Population/ Mapping	Example
1	Year of death	Date	ccyy	2019
2	Distinct patient count	Integer		10000
3	Distinct patients with date of birth following death	Integer		2
4	Distinct patients with 1 or more facts following death	Integer		89

5.7 Fact Types by Year

Table 6- File format for reporting fact types counts by year

Field	Description	Type	Population/Mapping	Example
1	Year of fact	Date	ccyy	2019
2	Fact type	Character varying (5)	PRO - Procedure DIAG - Diagnosis ORX - Ordered Medication DRX - Dispensed Medication LAB - Laboratory Test VIT - Vital Sign OTH – Other UNK - Unknown	ORX
3	Distinct patient count	Integer		550436
4	Fact count	Integer		2344789

5.8 Missing Encounter Data

Table 7- File format for the missing encounter data measure

Field	Description	Type	Population/Mapping	Example
1	Encounter setting	Character varying (20)	AV - Ambulatory Visit ED - Emergency Department IP - Inpatient Stay OS - Observation Stay IS - Institutional Stay TH – Telehealth OA – Other Ambulatory Visit OT - Other UN - Unknown MS - Missing	AV
2	Data element evaluated	Character varying (20)	adate = admission/start date ddate = discharge/end date	adate
3	Total distinct encounters	Integer		50000
4	Count of distinct encounter where variable value null	Integer		345557

5.9 Post-Discharge Facts

Table 8- File format for post-discharge facts reporting

Field	Description	Type	Population/Mapping	Example
1	Year of inpatient encounter	Date	ccyy	2019
2	Total distinct inpatient encounters	Integer		50000
3	Distinct inpatient encounters with diagnosis or procedure in days 1-3 following discharge	Integer		37000
4	Distinct inpatient encounters with diagnosis or procedure In days 1-7 following discharge	Integer		42000
5	Distinct inpatient encounters with diagnosis or procedure in days 1-30 following discharge	Integer		35000
6	Distinct inpatient encounters with diagnosis or procedure in days 1-90 following discharge	Integer		25000

5.10 Encounter Attributes by Year

Table 9- File format for the encounter attributes by year measure

Field	Description	Type	Population/Mapping	Example
1	Encounter month and year	Date	mmccyy	012019
2	Encounter setting	Character varying (20)	AV - Ambulatory Visit ED - Emergency Department IP - Inpatient Stay OS - Observation Stay IS - Institutional Stay TH – Telehealth OA – Other Ambulatory Visit OT - Other UN - Unknown MS - Missing	AV
3	Distinct patient count	Integer		500000
4	Count of distinct encounters	Integer		750000
5	Count of distinct encounters with medical fact	Integer		750000
6	Minimum medical fact counts	Integer		1
7	Maximum medical fact count	Integer		366
8	Median medical fact count	Integer		5
9	Mean medical fact count	Integer		6
10	Count of encounters with procedure fact	Integer		72000
11	Minimum procedure fact count	Integer		0
12	Maximum procedure fact count	Integer		3546
13	Median procedure fact count	Integer		3
14	Mean procedure fact count	Integer		3
15	Count of encounters with diagnosis fact	Integer		747000
16	Minimum diagnosis fact count	Integer		0
17	Maximum diagnosis fact count	Integer		3546
18	Median diagnosis fact count	Integer		3
19	Mean diagnosis fact count	Integer		3
20	Count encounters with missing length of stay	Integer		458

Field	Description	Type	Population/Mapping	Example
21	Minimum length of stay	Integer		-3
22	Maximum length of stay	Integer		366
23	Median length of stay	Integer		1
24	Mean length of stay	Integer		1