



# TriNetX

# DATA DICTIONARY

June 2019

This data dictionary applies to TriNetX's de-identified dataset. Access to and utilization of this dataset is subject to entering into a data use agreement with TriNetX. Any use, disclosure or publication of the dataset or any derivatives thereof will be governed by the terms, conditions and restrictions set forth in the Dataset Terms of Use.

# Table of Contents

---

TriNetX Data Overview .....	3
Data Tables Relationship .....	5
Patient Demographic Table .....	6
Encounter Table .....	7
Diagnosis Table .....	8
Procedure Table .....	9
Medication Table .....	10
Lab Result Table .....	11
Vital Sign Table .....	12
Tumor Properties Table .....	13
Oncology Treatment Table .....	14
Tumor Table .....	15
Chemotherapy Lines of Treatment Table .....	16
Genomics Table .....	17
Cohort Details Table .....	18
Dataset Details Table .....	19
Patient Cohort Table .....	20
Standardized Terminology Table .....	21

# TriNetX Data Overview

---

TriNetX datasets provide researchers access to de-identified patient data from networks of healthcare organizations (HCO) and other data providers.

Below are a set of questions and answers that describe the overall characteristics of TriNetX data.

## **What kind of data comes in a TriNetX dataset?**

TriNetX datasets are comprised of clinical patient data such as demographics, diagnoses, procedures, labs, and medications. This is commonly referred to as real-world data (RWD).

The data in TriNetX datasets are:

- Primarily from HCOs electronic medical record (EMR) systems
- Collected for the primary purpose of providing care to patients

The data in TriNetX datasets are not:

- Claims data, data primarily collected for the purposes of billing
- Data collected for the purposes of randomized clinical trials

## **Where does the data in a TriNetX dataset originate?**

Data in TriNetX datasets comes from HCOs and other data providers. The data these entities provide primarily comes from:

- EMR systems
  - Structured data
  - Unstructured data processed by Natural Language Processing (NLP) technology
- Cancer registries
- Other sources (e.g., genomic data from third party genomic testing labs)

## **What are the characteristics of the HCOs that provide TriNetX with data?**

The majority of the HCOs are large academic medical institutions with both inpatient and outpatient facilities. Most of these HCOs are adult acute-care hospitals with multiple facilities and locations. All HCOs are currently located within the United States.

HCOs provide TriNetX with both inpatient and outpatient data. The data they provide is representative of the entire patient population at the HCO. Most HCOs provide an average of seven years of historical data.

## **How is data transformed from its original source?**

TriNetX typically receives data from HCOs and other data providers in one of two ways:

1. TriNetX ingests data directly from an HCO's research repository (e.g., i2b2) into the TriNetX environment
2. An HCO or data provider sends TriNetX data extracts in the form of CSV files

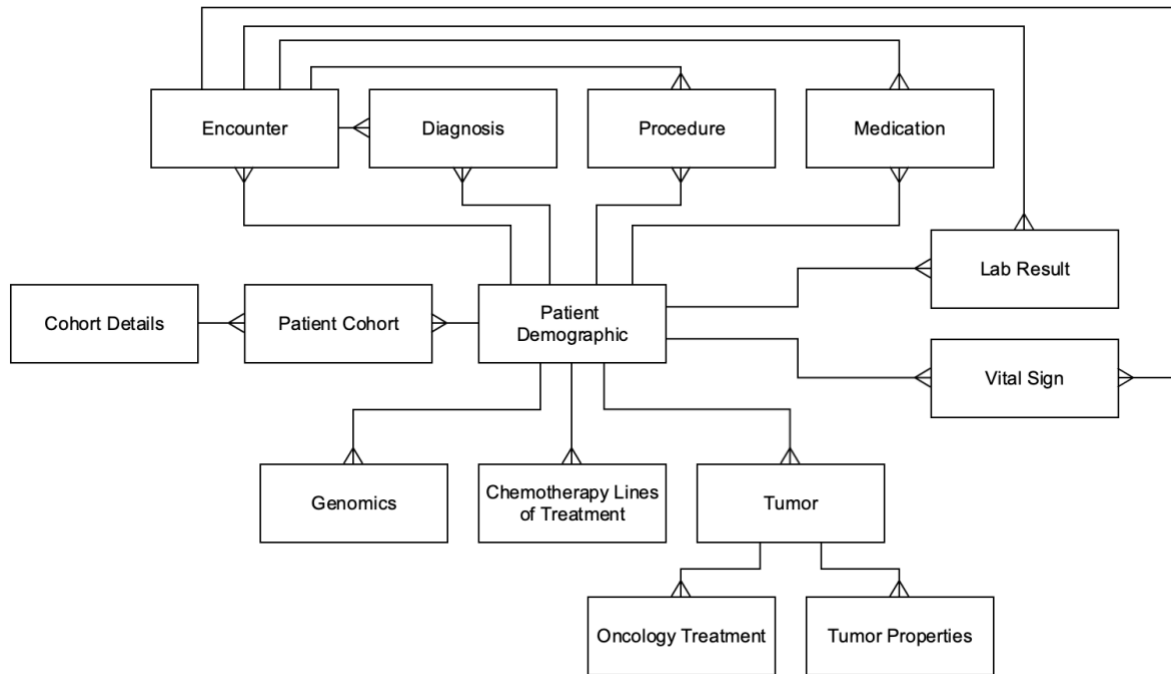
TriNetX maps the data to a standard and controlled set of clinical terminologies. The data is then transformed into a proprietary data schema. This transformation process includes an extensive data quality assessment that includes 'data cleaning' that rejects records that don't meet the TriNetX quality standards.

**How fresh (up to date) is the data?**

One of the distinguishing characteristics of the TriNetX dataset is that it is continuously refreshed. HCOs and other data providers update their data at various times with over 80% refreshing in 1, 2, or 4-week frequency intervals. The average lag time for an HCO's source data refresh is one month.

## Data Tables Relationship

---



## Patient Demographic Table

---

Data Element	Data Type	Sample Data	Description
<b>patient_id</b>	VARCHAR	123456789	The unique ID for the patient (de-identified).
<b>sex</b>	VARCHAR	M	The biological sex of the patient. Possible values are M, F, Unknown.
<b>race</b>	VARCHAR	White	The race of the patient. Possible values are American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, Unknown.
<b>ethnicity</b>	VARCHAR	Hispanic	The ethnicity (cultural background) of the patient. Possible values are Hispanic or Latino, Not Hispanic or Latino, Unknown.
<b>year_of_birth</b>	BIGINT	1958	The birth year of the patient. May be blank if the birth year occurred more than 90 years before the year the dataset was created.
<b>age_at_death</b>	BIGINT	57	The age at the time of the patient's death. If a patient's age at death is above 90, the age at death is rounded to 90 in order to protect patient privacy.
<b>postal_code</b>	VARCHAR		The postal code of the patient. Only available on Diamond network data.

## Encounter Table

---

Data Element	Data Type	Sample Data	Description
<b>encounter_id</b>	VARCHAR	987654321	The unique ID for the encounter (de-identified).
<b>patient_id</b>	VARCHAR	123456789	The unique ID for the patient (de-identified).
<b>start_date</b>	DATETIME (YYYYMMDD)	20110315	The date the encounter began.
<b>end_date</b>	DATETIME (YYYYMMDD)	20110318	The date the encounter ended.
<b>type</b>	VARCHAR	AMB	The care setting of the encounter. Possible values are Ambulatory (AMB), Emergency (EMER), Field (FLD), Home Health (HH), Inpatient Encounter (IMP), Inpatient Acute (ACUTE), Inpatient Non-acute (NONAC), Observation (OBSENC), Pre-admission (PRENC), Short Stay (SS), Virtual (VR). These values are based on HL7 v3 Value Set ActEncounterCode.
<b>derived_by_TriNetX</b>	BOOLEAN	T	Flag that indicates whether the encounter start or end date was derived by TriNetX. Possible values are T for TRUE and F for FALSE.

## Diagnosis Table

---

Data Element	Data Type	Sample Data	Description
<b>patient_id</b>	VARCHAR	123456789	The unique ID for the patient (de-identified).
<b>encounter_id</b>	VARCHAR	987654321	The unique ID for the encounter (de-identified).
<b>code_system</b>	VARCHAR	ICD-10-CM	The name of the code system in which this diagnosis is coded. Possible code systems are ICD-9-CM, ICD-10-CM.
<b>code</b>	VARCHAR	E11	The diagnosis code.
<b>date</b>	DATETIME (YYYYMMDD)	20110315	The date the diagnosis was recorded.
<b>derived_by_TriNetX</b>	BOOLEAN	T	Flag that indicates whether the diagnosis was derived by TriNetX. Possible values are T for TRUE and F for FALSE.



## Procedure Table

---

Data Element	Data Type	Sample Data	Description
<b>patient_id</b>	VARCHAR	123456789	The unique ID for the patient (de-identified).
<b>encounter_id</b>	VARCHAR	987654321	The unique ID for the encounter (de-identified).
<b>code_system</b>	VARCHAR	ICD-10-PCS	The name of the code system in which this procedure is coded. Possible code systems are ICD-9-CM, ICD-10-PCS, CPT.
<b>code</b>	VARCHAR	03CJ0ZZ	The procedure code.
<b>date</b>	DATETIME (YYYYMMDD)	20150314	The date the procedure was recorded.
<b>derived_by_TriNetX</b>	BOOLEAN	T	Flag that indicates whether the procedure was derived by TriNetX. Possible values are T for TRUE and F for FALSE.

## Medication Table

---

Data Element	Data Type	Sample Data	Description
<b>patient_id</b>	VARCHAR	123456789	The unique ID for the patient (de-identified).
<b>encounter_id</b>	VARCHAR	987654321	The unique ID for the encounter (de-identified).
<b>code_system</b>	VARCHAR	RxNorm	The name of the code system in which this medication is coded. The code system is RxNorm.
<b>code</b>	VARCHAR	26225	The medication code.
<b>start_date</b>	DATETIME (YYYYMMDD)	20120914	The date the medication order, prescription, or administration was recorded.
<b>route</b>	VARCHAR	Oral Product	The route of administration. Possible values are Drug implant, Inhalant, Injectable, Intraperitoneal, Nasal, Ophthalmic, Oral, Otic, Rectal, Topical, Urethral, Vaginal, Unknown.
<b>brand</b>	VARCHAR	Zofran	The medication brand.
<b>strength</b>	VARCHAR	4 mg	The medication strength.
<b>derived_by_TriNetX</b>	BOOLEAN	T	Flag that indicates whether the medication was derived by TriNetX. Possible values are T for TRUE and F for FALSE.

## Lab Result Table

---

Data Element	Data Type	Sample Data	Description
<b>patient_id</b>	VARCHAR	123456789	The unique ID for the patient (de-identified).
<b>encounter_id</b>	VARCHAR	987654321	The unique ID for the encounter (de-identified).
<b>code_system</b>	VARCHAR	LOINC	The name of the code system in which this lab observation is coded. The code system is LOINC.
<b>code</b>	VARCHAR	2885-2	The code representing the lab test.
<b>date</b>	DATETIME (YYYYMMDD)	20120914	The date the test result was recorded.
<b>lab_result_num_val</b>	DECIMAL	7	The lab result for numeric results.
<b>derived_by_TriNetX</b>	BOOLEAN	T	Flag that indicates whether the lab result was derived by TriNetX. Possible values are T for TRUE and F for FALSE.

## Vital Sign Table

---

Data Element	Data Type	Sample Data	Description
<b>patient_id</b>	VARCHAR	123456789	The unique ID for the patient (de-identified).
<b>encounter_id</b>	VARCHAR	987654321	The unique ID for the encounter (de-identified).
<b>code_system</b>	VARCHAR	LOINC	The name of the code system in which this vital sign is coded. The code system is LOINC.
<b>code</b>	VARCHAR	8302-2	The code representing the vital sign.
<b>date</b>	DATETIME (YYYYMMDD)	20120914	The date the vital sign was recorded.
<b>value</b>	VARCHAR	72	The value of this vital sign.
<b>derived_by_TriNetX</b>	BOOLEAN	T	Flag that indicates whether the vital sign was derived by TriNetX. Possible values are T for TRUE and F for FALSE.

## Tumor Properties Table

Data Element	Data Type	Sample Data	Description
<b>patient_id</b>	VARCHAR	123456789	The unique ID for the patient (de-identified).
<b>diagnosis_date</b>	DATETIME (YYYYMMDD)	20120914	The date of the original primary cancer diagnosis.
<b>observation_date</b>	DATETIME (YYYYMMDD)	20121114	The date the property was recorded.
<b>tumor_site_code_system</b>	VARCHAR	ICD-O	The name of the code system in which the tumor site is coded. The code system is ICD-O.
<b>tumor_site_code</b>	VARCHAR	C50	The tumor site code.
<b>morphology_code_system</b>	VARCHAR	ICD-O	The name of the code system in which morphology is coded. The code system is ICD-O.
<b>morphology_code</b>	VARCHAR	8500/3	The morphology code.
<b>tumor_property_code_system</b>	VARCHAR	TriNetX – Tumor Property	The name of the code system in which the tumor property is coded. The code system is TriNetX. This is a code system created by TriNetX for oncology specific factors.
<b>tumor_property_code</b>	VARCHAR	CSF07-Colon 060	The code that indicates the type of tumor property.
<b>derived_by_TriNetX</b>	BOOLEAN	T	Flag that indicates whether the tumor property was derived by TriNetX. Possible values are T for TRUE and F for FALSE.

## Oncology Treatment Table

Data Element	Data Type	Sample Data	Description
<b>patient_id</b>	VARCHAR	123456789	The unique ID for the patient (de-identified).
<b>diagnosis_date</b>	DATETIME (YYYYMMDD)	20120914	The date of the primary cancer diagnosis.
<b>tumor_site_code_system</b>	VARCHAR	ICD-O	The name of the code system in which the tumor site is coded. The code system is ICD-O.
<b>tumor_site_code</b>	VARCHAR	C50	The tumor site code.
<b>morphology_code_system</b>	VARCHAR	ICD-O	The name of the code system in which morphology is coded. The code system is ICD-O.
<b>morphology_code</b>	VARCHAR	8500/3	The morphology code.
<b>oncology_treatment_start_date</b>	DATETIME (YYYYMMDD)	20121001	The start date of the course of oncology treatment.
<b>oncology_treatment_code_system</b>	VARCHAR	TriNetX – Oncology Treatment	The name of the code system in which the oncology treatment is coded. The code system is TriNetX. This is a code system created by TriNetX for oncology treatment.
<b>oncology_treatment_code</b>	VARCHAR	1390 1	The code for the oncology treatment.
<b>derived_by_TriNetX</b>	BOOLEAN	T	Flag that indicates whether the oncology treatment was derived by TriNetX. Possible values are T for TRUE and F for FALSE.

## Tumor Table

Data Element	Data Type	Sample Data	Description
<b>patient_id</b>	VARCHAR	123456789	The unique ID for the patient (de-identified).
<b>diagnosis_date</b>	DATETIME (YYYYMMDD)	20120914	The date of the original primary cancer diagnosis.
<b>observation_date</b>	DATETIME (YYYYMMDD)	20121114	The date the property was recorded.
<b>tumor_site_code_system</b>	VARCHAR	ICD-O	The name of the code system in which the tumor site is coded. The code system is ICD-O.
<b>tumor_site_code</b>	VARCHAR	C50	The tumor site code.
<b>morphology_code_system</b>	VARCHAR	ICD-O	The name of the code system in which morphology is coded. The code system is ICD-O.
<b>morphology_code</b>	VARCHAR	8500/3	The morphology code.
<b>stage_code_system</b>	VARCHAR	TriNetX – Oncology Stage	The name of the code system in which the tumor stage is coded. The code system is TriNetX. This is a code system created by TriNetX for tumor stages.
<b>stage_code</b>	VARCHAR	2b	The code for the tumor stage.
<b>derived_by_TriNetX</b>	BOOLEAN	T	Flag that indicates whether the tumor entry was derived by TriNetX. Possible values are T for TRUE and F for FALSE.

## Chemotherapy Lines of Treatment Table

---

Data Element	Data Type	Sample Data	Description
<b>patient_id</b>	VARCHAR	123456789	The unique ID for the patient (de-identified).
<b>start_date</b>	DATETIME (YYYYMMDD)	20150314	The date the chemotherapy line of treatment was determined to start.
<b>line</b>	BIGINT	1	The sequential order of chemotherapy regimens. Possible values are 1, 2, 3, 4, or 5 with 1 = the first regimen and 5 = the last regimen. These lines are derived by TriNetX.
<b>derived_by_TriNetX</b>	BOOLEAN	T	Flag that indicates whether the chemotherapy line of treatment was derived by TriNetX. Possible values are T for TRUE and F for FALSE.



## Genomics Table

---

Data Element	Data Type	Sample Data	Description
<b>patient_id</b>	VARCHAR	123456789	The unique ID for the patient (de-identified).
<b>code_system</b>	VARCHAR	HGVS	The name of the code system in which genomic data is coded. The syntax of the code conforms to HGVS.
<b>code</b>	VARCHAR	BRAF p.V600E c.1799T>A	Variant description.
<b>test_date</b>	DATETIME (YYYYMMDD)	20120914	The date the genetic test was recorded.
<b>derived_by_TriNetX</b>	BOOLEAN	T	Flag that indicates whether the genomic data was derived by TriNetX. Possible values are T for TRUE and F for FALSE.

## Cohort Details Table

---

Data Element	Data Type	Sample Data	Description
<b>cohort_name</b>	VARCHAR	Diabetes women aged 18-45	The name of the cohort included in the dataset.
<b>cohort_number</b>	BIGINT	1	The number of the cohort included in the dataset.
<b>total_patient_records</b>	BIGINT	20,000	The total number of patient records in the cohort in the dataset.

## Dataset Details Table

---

Data Element	Data Type	Sample Data	Description
<b>total_number_unique_patients</b>	BIGINT	19,000	The total number of unique patient records across multiple cohorts in the dataset. A patient's record could be in a single cohort multiple times if the patient visited more than one HCO that contributed data to a cohort.
<b>total_number_HCOs</b>	BIGINT	7	The total number of healthcare organizations contributing data to the dataset.
<b>date_created</b>	DATETIME (YYYYMMDD)	20180316	The date the dataset was created.

## Patient Cohort Table

---

Data Element	Data Type	Sample Data	Description
<b>patient_id</b>	VARCHAR	123456789	The unique ID for the patient (de-identified).
<b>cohort_name</b>	VARCHAR	Diabetes women aged 18-45	The name of the cohort in which the patient's record is included.
<b>cohort_number</b>	BIGINT	1	The number of the cohort in which the patient's record is included.

## Standardized Terminology Table

---

Data Element	Data Type	Sample Data	Description
<b>code_system</b>	VARCHAR	RxNORM	The name of the code system in which the data element is coded.
<b>code</b>	VARCHAR	1191	The code for the data element.
<b>code_description</b>	VARCHAR	Aspirin	The textual description of the data element.
<b>path</b>	VARCHAR	N0000010574/N0000029132/ N0000029133/N0000029135/ 1191	The terms the data element is mapped to and the path in which those terms exist.
<b>unit</b>	VARCHAR	inches	The unit of measurement for a code value. This field only applies to codes in the Lab Result table and the Vital Sign table.