Common Data Model (CDM) Specification, Version 1.0

Released by the Data Standards, Security and Network Infrastructure (DSSNI) Task Force on May 30, 2014

1. Table of Contents

2. PCORnet Common Data Model (CDM): Overview and Development

2.1. License and Use

The PCORnet data model is freely available for use. An open-source license will be selected by PCORI and included in the next release of the data model. The PCORnet Distributed Research Network (DRN) and its infrastructure, including the Common Data Model (CDM), is overseen and guided by the PCORnet Data Standards, Security, and Networking Infrastructure Task Force (DSSNI) Task Force.

The PCORnet CDM is based on the Mini-Sentinel Common Data Model (MSCDM; www.mini-sentinel.org) and has been informed by other distributed initiatives such as the HMO Research Network, the Vaccine Safety Datalink, various AHRQ Distributed Research Network projects, and the ONC Standards & Interoperability Framework Query Health Initiative.

2.2. Overview

PCORnet is developing the PCORnet DRN design to be a "…functional distributed research network that facilitates multi-site patient–centered research across the Clinical Data Research Networks (CDRNs), Patient-Powered Research Networks (PPRNs), and other interested contributors. The distributed network will enable the conduct of observational research and clinical trials while allow each participating organization to maintain physical and operational control over its data." [DSSNI charter, 2014]

The PCORnet CDM is the foundation of the PCORnet DRN. Guiding Principles for the PCORnet DRN and CDM are included in the DSSNI Charter. The audience for this document includes current and future partners, and other stakeholders.

For more information about PCORnet, please visit http://pcornet.org

2.3. Design and Modeling for the PCORnet CDM

Prioritization of Analytic Functionality

PCORnet CDM Guiding Principle #5 states,

> "Documentation will be clear and transparent so that its contents are understandable to all contributors. The CDM will be intuitive and easy for analysts and investigators to use. Investigators and analysts with prior experience using research data will not need additional skills or knowledge to use the CDM."

This is expressed in the CDM design through prioritization of **analytic** functionality, and a parsimonious approach based upon analytic utility. At times, this results in decisions that are not based in relational database modeling principles such as normalization. The model is designed to facilitate routine and rapid execution of distributed complex analytics. To meet this design requirement, some fields are duplicated across multiple tables to support faster analytic operations for distributed querying. The PCORnet CDM is based on the FDA Mini-Sentinel CDM. This allows PCORnet to more easily leverage the large array of analytic tools and expertise developed for the MSCDM, including data characterization approaches and the various tools for complex distributed analytics.

Primary/Foreign Keys

Database programmers will notice that fields used as primary/foreign keys, especially PATID and ENCOUNTERID, are specified as TEXT instead of NUMBER. This approach, informed by prior experience in developing large-scale multi-site distributed networks, makes it easier to implement than requiring new key generation that could impact database management within source systems. Each organization is encouraged to use a consistent format for primary keys and foreign keys based upon their existing source data. Not all tables in the PCORnet CDM have a primary key specified (such as the DIAGNOSIS and PROCEDURE tables), but data partners are permitted to include their own primary or surrogate keys in implementation of the CDM. For example, this might include the INSTANCE_NUM in the i2b2 Data Repository Cell.

Date Formatting

Because the PCORnet CDM is intended to support multiple Relational Database Management Systems (RDBMS), date format consistency is an issue, given that most RDBMS's have platform-specific native date representation. To address this issue, the PCORnet CDM will adopt the ISO 8601 standard of date and time representation, which is platform-agnostic. The CDM will always separate date fields and time fields for consistency, and employ a naming convention of suffix "_DATE" or "_TIME".

Missing or Unknown Data Values

The PCORnet CDM will have use the HL7 conventions of "Null Flavors" (http://hl7.org/implement/standards/fhir/v3/NullFlavor/) as a basis for representing missing or unknown values. Specifically, we will populate null values as follows:

1. **A data field is not present** in the source system. (populate with null)
2. A data field is present in the source system, **but the source value is null or blank**. (populate with NI = No Information)
3. A data field is present in the source system, but the source value **explicitly denotes an unknown value**. (populate with UN = Unknown)
4. A data field is present in the source system, but the source value **cannot be mapped to the CDM**. (populate with OT = Other)

Source Data Consistency

The CDM does not include data consistency rules or edits, such as upper and lower limits of numeric values. The value recorded in the originating source system should be the value populated in the CDM, even if the value is outside a normally acceptable limit. Inclusion of all originating data, without modification, supports data characterization and better data provenance.

Decisions about inclusion (or censoring) of outlier values will be made as part of each <u>analysis or query</u>, allowing for these decisions to be driven by appropriateness for each individual analysis.

"Raw" Fields
The data model uses a convention for "raw data fields" That are optional fields for storing the originating source value of a field, prior to mapping into PCORnet CDM value set. It may also be used for source-specific ontologies.

The "RAW" fields are intended to support data provenance and facilitate quality control checking by local implementation, if desired. These fields will have a naming convention of prefix "RAW_". We will not include these fields in the Entity-Relationship (ER) diagram.

Case Sensitivity
The PCORnet CDM recognizes that some relational database management systems RDBMS) do not have case sensitive object naming. However, if the CDM is implemented in a platform that is case sensitive, please always name objects in UPPERCASE (for example TABLE_NAME and FIELD_NAME). Distributed programs will assume that all table names and field names are in UPPERCASE. Categorized value sets will also use UPPERCASE.

Additional Fields
PCORnet sites are welcome to include additional fields in their local CDM implementation that will assist with transformation or clarity.

2.4.   Development Notes

- Because the PCORnet DRN has independent objectives and priorities, the PCORnet CDM will not reuse an existing data model, but will develop a stand-alone PCORnet CDM based on existing data models, as appropriate.
- The model was informed by results from the PCORnet DSSNI Preliminary Partner Survey (also known as the "Tech Survey") completed in December 2013 and January 2014.
- Recommendations from the PCORnet CDM Working Group, from discussions held in May 2014, have been incorporated.
- The PCORnet CDM priority data domains and implementation approach are based on PCORI needs, planned future capabilities, and the data sources and expertise of the PCORnet partners.

As stated in PCORnet CDM Guiding Principle #2,

> "It is expected that not all CDRNs and PPRNs will have data needed to populate all parts of the PCORnet CDM. It is the responsibility of the CDRNs and PPRNs to communicate availability of each data domain and element."

The PCORnet CDM will be implemented in phases. This will allow incorporation of new data domains and fields throughout the life of the project, building based on PCORnet needs, lessons learned from use, and data availability.

The implementation process for each PCORnet site should be documented for review by the DSSNI Task Force. The DSSNI Task Force will provide a template for documentation and provide support and consultation as implementation questions arise.

2.5. Future State and Revision Process

The PCORnet CDM will be updated and revised regularly. Revisions will be based on lessons learned during implementation and PCORnet priorities. Based on existing PCORnet priorities, we expect that the next release of the PCORnet CDM will include **medications ordered**, **medications dispensed**, **laboratory result values**, and selected **patient-reported outcomes (PRO)/ information**. The PCORnet DSSNI Task Force is working with the PRO Task Force and other stakeholders to determine the best ways to incorporate PROs and other pertinent patient-reported information, and to include those in subsequent releases of the CDM.

2.6. Comments on Protected Health Information (PHI)

The CDM will contain some of the 18 elements that define PHI under HIPAA, including encounter dates and date of birth. However, these dates will remain under the control of the institutions that already maintain PHI. To maximize analytic flexibility and allow for all types of analyses, complete and exact dates should be included in the CDM. Distributed analytic programs will use the date fields for analysis, but will generate results that contain the minimum necessary information to address the question. The results returned to the requester will typically be aggregated and not include any PHI. Queries that generate results sets with PHI (eg, a person-level analysis under an IRB, with all necessary data agreements in place) will be clearly flagged as such and will only be distributed with the appropriate approvals clearly documented. As with all distributed queries, sites should review all results before release.
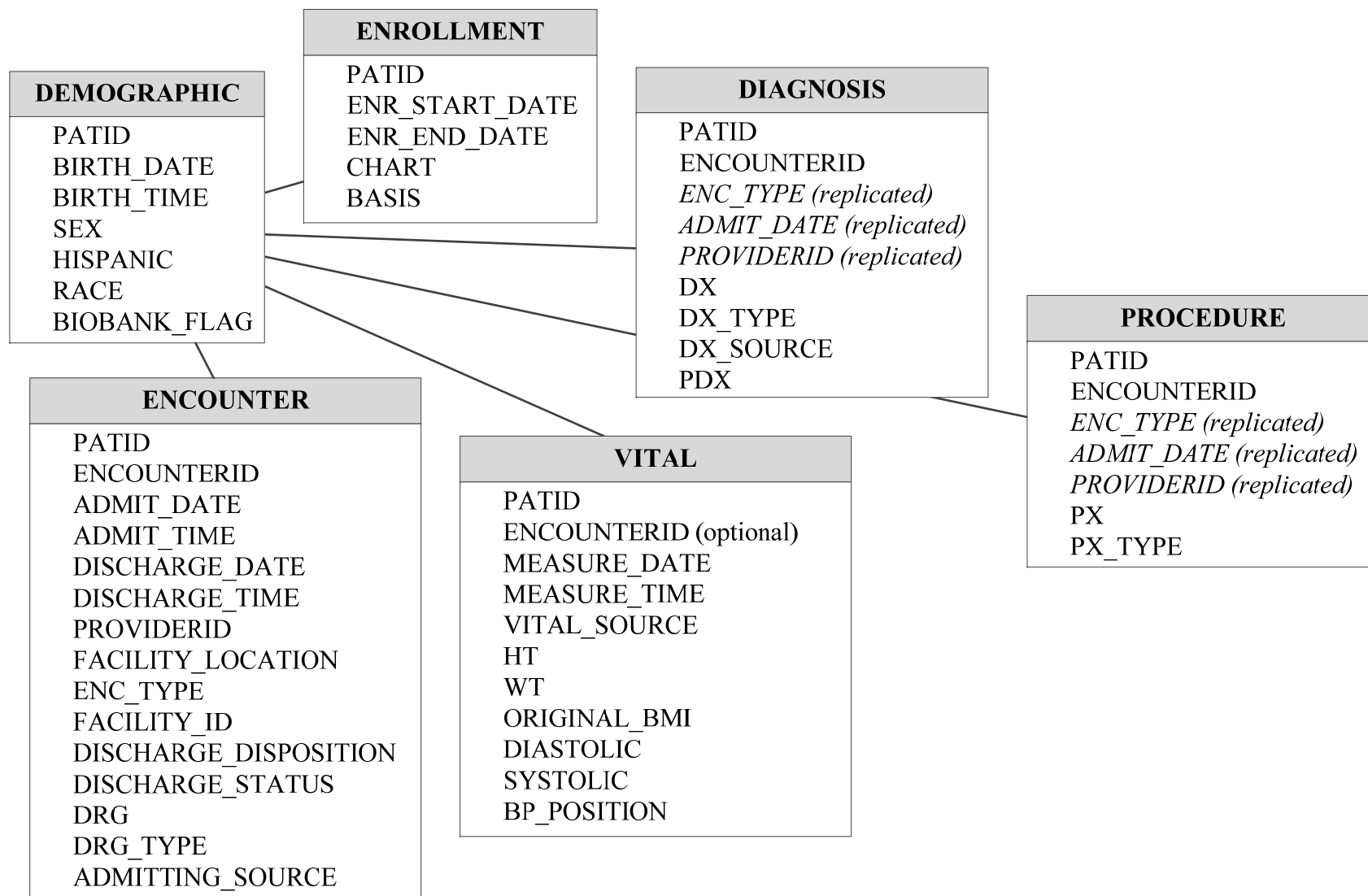
The necessary "cross-walks" between the arbitrary identifiers included in the CDM and their originating data are not specified in the scope of the CDM, but are expected to be maintained by each data partner.

- PATID is a pseudoidentifier with a consistent crosswalk to the true identifier retained by the source site. For analytical data sets requiring patient-level data, only the pseudoidentifier is used to link across all information belonging to a patient.
- The PATID pseudoidentifier should not be a true identifier. It is not appropriate to use Medical Record Identifiers (MRNs) for this purpose because MRN is a direct patient identifier.
- Locally maintained "mapping tables" are tables necessary to implement so that each data partner has the ability to map arbitrary identifiers back to the originating data and patient.
- These mapping tables are not part of the PCORnet DRN.

Mapping tables for implementation of the CDM v1.0 should include (but are not limited to):

- PATID crosswalk
- PROVIDER crosswalk

3. Individual Table Specifications

ENROLLMENT

PATID
ENR_START_DATE
ENR_END_DATE
CHART
BASIS

DEMOGRAPHIC

PATID
BIRTH_DATE
BIRTH_TIME
SEX
HISPANIC
RACE
BIOBANK_FLAG

DIAGNOSIS

PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
DX
DX_TYPE
DX_SOURCE
PDX

PROCEDURE

PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
PX
PX_TYPE

ENCOUNTER

PATID
ENCOUNTERID
ADMIT_DATE
ADMIT_TIME
DISCHARGE_DATE
DISCHARGE_TIME
PROVIDERID
FACILITY_LOCATION
ENC_TYPE
FACILITY_ID
DISCHARGE_DISPOSITION
DISCHARGE_STATUS
DRG
DRG_TYPE
ADMITTING_SOURCE

VITAL

PATID
ENCOUNTERID (optional)
MEASURE_DATE
MEASURE_TIME
VITAL_SOURCE
HT
WT
ORIGINAL_BMI
DIASTOLIC
SYSTOLIC
BP_POSITION

3.1. Table: DEMOGRAPHIC

Description:
- The DEMOGRAPHIC table contains one record per patient.
- The most recently available information should be populated for BIRTH_DATE, SEX, and other characteristics. If these attributes have been updated in the patient record, please use the most recent value.

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
PATID	TEXT(x)	.	Arbitrary person-level identifier used to link across tables. PATID is a pseudoidentifier with a consistent crosswalk to the true identifier retained by the source data partner. For analytical data sets requiring patient-level data, only the pseudoidentifier is used to link across all information belonging to a patient. The PATID must be unique within the data source being queried. Creating a unique identifier within a CDRN would be beneficial and acceptable. The PATID is not the basis for linkages across data partners.	MSCDM
BIRTH_DATE	TEXT(10): Format as YYYY-MM-DD	.	Date of birth.	MSCDM with modified data type Source of date format: ISO 8601
BIRTH_TIME	TEXT(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	.	Time of birth.	PCORnet Source of time format: ISO 8601

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
SEX	TEXT(2)	A = Ambiguous F = Female M = Male NI = No information UN = Unknown OT = Other	Administrative sex.	MSCDM with modified field size and value set Source: Administrative Sex (HL7) http://phinvads.cdc.gov/vads/ViewValueSet.action?id=06D34BBC-617F-DD11-B38D-00188B398520
HISPANIC	TEXT(2)	Y = Yes N = No NI = No information UN = Unknown OT = Other	A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.	MSCDM with modified field size and value set Compatible with "OMB Hispanic Ethnicity" (Hispanic or Latino, Not Hispanic or Latino)
RACE	TEXT(2)	01 = American Indian or Alaska Native 02 = Asian 03 = Black or African American 04 = Native Hawaiian or Other Pacific Islander 05 = White 06 = Multiple race 07 = Refuse to answer NI = No information UN = Unknown OT = Other	Please use only one race value per patient. **Details of categorical definitions:** American Indian or Alaska Native: A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment. Asian: A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. Black or African American: A person having origins in any of the black racial groups of Africa.	MSCDM with modified field size and value set Original value set is based upon U.S. Office of Management and Budget (OMB) standard, and is compatible with the 2010 U.S. Census

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
			Native Hawaiian or Other Pacific Islander: A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. White: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.	
BIOBANK_FLAG	TEXT(1)	Y = Yes N = No	Flag to indicate that one or more biobanked specimens are stored and available for research use. Examples of biospecimens could include plasma, urine, or tissue. If biospecimens are available, locally maintained "mapping tables" would be necessary to map between the DEMOGRAPHIC record and the originating biobanking system(s). If no known biobanked specimens are available, this field should be marked "No".	
RAW_SEX	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet
RAW_HISPANIC	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet
RAW_RACE	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet

3.2.   Table: ENROLLMENT

Description:
- The ENROLLMENT table has a start/stop structure that contains records for continuous enrollment periods.
- "Enrollment" is an insurance-based concept that defines a period during which all medically-attended events are expected to be observed. For partners that do not have enrollment information for some of their patients, other approaches for identifying periods during which complete medical capture is expected can be used.
- This table is designed to identify periods during which a person is expected to have complete data capture.
- Members with medical coverage, drug coverage, or both should be included.
- A record is expected to represent a unique combination of PATID, ENR_START_DATE, and BASIS.
- A break in enrollment (of at least one day) or a change in the chart abstraction flag should generate a new record.
- The ENROLLMENT table provides an important analytic basis for identifying periods during which medical care should be observed, for calculating person-time, and for inferring the meaning of unobserved care (ie, if care is not observed, it likely did not happen).

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
PATID	TEXT(x)	.	Arbitrary person-level identifier used to link across tables.	MSCDM
ENR_START_DATE	TEXT(10): Format as YYYY-MM-DD	.	Date of the beginning of the enrollment period. If the exact date is unknown, use the first day of the month. For implementation of the CDM, a long span of longitudinal data is desirable; however, especially for historical data more than a decade old, the appropriate beginning date should be determined by the data partner's knowledge of the validity and usability of the data. More specific guidance can be provided through implementation discussions.	MSCDM with modified field name and data type Source of date format: ISO 8601
ENR_END_DATE	TEXT(10): Format as YYYY-MM-DD	.	Date of the end of the enrollment period. If the exact date is unknown, use the last day of the month.	MSCDM with modified field name and data type Source of date format: ISO 8601
CHART	TEXT(1)	Y = Yes N = No	Chart abstraction flag is intended to answer the question, "Are you able to request (or review) charts for this person?" This flag does not address chart availability. Mark as "Yes" if there are no contractual or other restrictions between you and the individual (or sponsor) that would prohibit you from requesting any chart for this member.	MSCDM

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
			Note: This field is most relevant for health insurers that can request charts from affiliated providers. This field allows exclusion of patients from studies that require chart review to validate exposures and/or outcomes. It identifies patients for whom charts are **never** available and for whom the chart can never be requested.	
ENR_BASIS	TEXT(1)	I = Insurance G = Geography A = Algorithmic E = Encounter-based	When insurance information is not available but complete capture can be asserted some other way, please identify the basis on which complete capture is defined. Additional information on the approach identified will be required from each data partner. ENR_BASIS is a property of the time period defined. A patient can have multiple entries in the table. **Details of categorical definitions:** Insurance: The start and stop dates are based upon the concept of enrollment with health plan insurance Geography: An assertion of complete data capture between the start and end dates based upon geographic characteristics, such as regional isolation Algorithmic: An assertion of complete data capture between the start and end dates, based on a locally developed or applied algorithm, often using multiple criteria Encounter-based: The start and stop dates are populated from the earliest-observed encounter and latest-observed encounter.	PCORnet Based upon the HMORN VDW

3.3. Table: ENCOUNTER

Description:
- The ENCOUNTER Table contains one record per PATID and ENCOUNTERID (which reflects a unique combination of PATID, ADMIT_DATE, PROVIDERID and ENC_TYPE).
- For the special situation of Emergency Department (ED) encounters: We would expect separate, unique records for ED and inpatient, even if ED leads to an inpatient stay.
- The encounter table should include information on interactions between patients and providers. Each diagnosis and procedure recorded during the encounter should have a separate record in the Diagnosis or Procedure Tables.
- Multiple visits to the same provider on the same day may be considered one encounter (especially if defined by a reimbursement basis); if so, the ENCOUNTER record should be associated with all diagnoses and procedures that were recorded during those visits.
- Visits to different providers on the same day, such as a physician appointment that leads to a hospitalization, would generally be considered multiple encounters in the source system.
- Rollback or voided transactions and other adjustments should be processed before populating this table.
- Note: Although "Expired" is represented in both DISCHARGE_DISPOSITION and DISCHARGE_STATUS, this overlap represents the reality that both fields are captured in hospital data systems but there is variation in which field is best populated.

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
PATID	TEXT(x)	.	Arbitrary person-level identifier used to link across tables.	MSCDM
ENCOUNTERID	TEXT(x)	.	Arbitrary encounter-level identifier. Used to link across tables, including the ENCOUNTER, DIAGNOSIS, and PROCEDURE tables.	MSCDM
ADMIT_DATE	TEXT(10): Format as YYYY-MM-DD	.	Encounter or admission date.	MSCDM with modified field name and data type Source of date format: ISO 8601
ADMIT_TIME	TEXT(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	.	Encounter or admission time.	PCORnet Source of time format: ISO 8601

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
DISCHARGE_DATE	TEXT(10): Format as YYYY-MM-DD	.	Discharge date. Should be populated for all Inpatient Hospital Stay (IP) and Non-Acute Institutional Stay (IS) encounter types. May be populated for Emergency Department (ED) encounter types. Should be missing for ambulatory visit (AV or OA) encounter types.	MSCDM with modified field name and data type

Source of date format: ISO 8601 |
| DISCHARGE_TIME | TEXT(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute | . | Discharge time. | PCORnet

Source of time format: ISO 8601 |
| PROVIDERID | TEXT(x) | . | Provider code for the provider who is most responsible for this encounter. For encounters with multiple providers choose one so the encounter can be linked to the diagnosis and procedure tables. As with the PATID, the provider code is a pseudoidentifier with a consistent crosswalk to the real identifier. | MSCDM |
| FACILITY_LOCATION | TEXT(3) | . | Geographic location (3 digit zip code). Should be left blank if not recorded in source system. | MSCDM |
| ENC_TYPE | TEXT(2) | AV = Ambulatory Visit
ED = Emergency Department
IP = Inpatient Hospital Stay
IS = Non-Acute Institutional Stay
OA = Other Ambulatory Visit
NI = No information
UN = Unknown
OT = Other | Encounter type.

Details of categorical definitions:
Ambulatory Visit: Includes visits at outpatient clinics, physician offices, same day/ambulatory surgery centers, urgent care facilities, and other same-day ambulatory hospital encounters, but excludes emergency department encounters.

Emergency Department (ED): Includes ED encounters that become inpatient stays (in which case inpatient stays would be a separate encounter). Excludes urgent care visits. ED claims should be pulled before hospitalization claims to ensure that ED with subsequent admission won't be rolled up in the hospital event. | MSCDM with modified value set |

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
			Inpatient Hospital Stay: Includes all inpatient stays, including: same-day hospital discharges, hospital transfers, and acute hospital care where the discharge is after the admission date. Non-Acute Institutional Stay: Includes hospice, skilled nursing facility (SNF), rehab center, nursing home, residential, overnight non-hospital dialysis and other non-hospital stays. Other Ambulatory Visit: Includes other non-overnight AV encounters such as hospice visits, home health visits, skilled nursing facility visits, other non-hospital visits, as well as telemedicine, telephone and email consultations. May also include "lab only" visits (when a lab is ordered outside of a patient visit), "pharmacy only" (e.g., when a patient has a refill ordered without a face-to-face visit), "imaging only", etc.	
FACILITYID	TEXT(x)	.	Arbitrary local facility code that identifies the hospital or clinic. Used for chart abstraction and validation. FACILITYID can be a true identifier, or a pseudoidentifier with a consistent crosswalk to the true identifier retained by the source data partner.	MSCDM with modified field name
DISCHARGE_DISPOSITION	TEXT(2)	A = Discharged alive E = Expired NI = No information UN = Unknown OT = Other	Vital status at discharge. Should be populated for Inpatient Hospital Stay (IP) and Non-Acute Institutional Stay (IS) encounter types. May be populated for Emergency Department (ED) encounter types. Should be missing for ambulatory visit (AV or OA) encounter types.	MSCDM with modified field size and value set
DISCHARGE_STATUS	TEXT(2)	AF = Adult Foster Home AL = Assisted Living Facility AM = Against Medical Advice	Discharge status. Should be populated for Inpatient Hospital Stay (IP) and Non-Acute Institutional Stay (IS) encounter types. May be populated for Emergency Department (ED) encounter types. Should be missing for ambulatory visit (AV or OA) encounter types.	MSCDM with modified value set

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
		AW = Absent without leave EX = Expired HH = Home Health HO = Home / Self Care HS = Hospice IP = Other Acute Inpatient Hospital NH = Nursing Home (Includes ICF) RH = Rehabilitation Facility RS = Residential Facility SH = Still In Hospital SN = Skilled Nursing Facility NI = No information UN = Unknown OT = Other		
DRG	TEXT(3)	.	3-digit Diagnosis Related Group (DRG). Should be populated for IP and IS encounter types. May be populated for ED encounter types. Should be missing for AV or OA encounters. Use leading zeroes for codes less than 100. The DRG is used for reimbursement for inpatient encounters. It is a Medicare requirement that combines diagnoses into clinical concepts for billing. Frequently used in observational data analyses.	MSCDM
DRG_TYPE	TEXT(2)	01 = CMS-DRG (old system) 02 = MS-DRG (current system) NI = No information UN = Unknown OT = Other	DRG code version. MS-DRG (current system) began on 10/1/2007. Should be populated for IP and IS encounter types. May be populated for ED encounter types. Should be missing for AV or OA encounters.	MSCDM with modified field size and value set

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
ADMITTING_SOURCE	TEXT(2)	AF = Adult Foster Home AL = Assisted Living Facility AV = Ambulatory Visit ED = Emergency Department HH = Home Health HO = Home / Self Care HS = Hospice IP = Other Acute Inpatient Hospital NH = Nursing Home (Includes ICF) RH = Rehabilitation Facility RS = Residential Facility SN = Skilled Nursing Facility NI = No information UN = Unknown OT = Other	Admitting source. Should be populated for Inpatient Hospital Stay (IP) and Non-Acute Institutional Stay (IS) encounter types. May be populated for Emergency Department (ED) encounter types. Should be missing for ambulatory visit (AV or OA) encounter types.	MSCDM with modified value set
RAW_ ENC_TYPE	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet
RAW_ DISCHARGE_DISPOSITION	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet
RAW_ DISCHARGE_STATUS	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet
RAW_ DRG_TYPE	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet
RAW_ ADMITTING_SOURCE	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet

3.4. Table: DIAGNOSIS

Description:
- DIAGNOSIS should capture all uniquely recorded diagnoses for all encounters.
- If a patient has multiple diagnoses associated with one encounter, then there would be one record in this table for each diagnosis.
- ENCOUNTERID is not optional for DIAGNOSIS and PROCEDURE. The definitions of the DIAGNOSIS and PROCEDURE tables are dependent upon a healthcare context; therefore, the ENCOUNTER basis is necessary.
- Data in this table are expected to be from healthcare-mediated processes and reimbursement drivers. This can include both technical/facility billing and professional billing.
- We recognize that, in many cases, these diagnoses may only be related to the **treatment** of the patient during the specific encounter.
- Please do not include diagnoses generated from problem lists, which are a different workflow and data-generating activity.
- If a local ontology is used, but cannot be mapped to a standard ontology such as ICD-9-CM, DX_TYPE should be populated as "Other".

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
PATID	TEXT(x)	.	Arbitrary person-level identifier. Used to link across tables.	MSCDM
ENCOUNTERID	TEXT(x)	.	Arbitrary encounter-level identifier. Used to link across tables.	MSCDM
ENC_TYPE	TEXT(2)	AV = Ambulatory Visit ED = Emergency Department IP = Inpatient Hospital Stay IS = Non-Acute Institutional Stay OA = Other Ambulatory Visit NI = No information UN = Unknown OT = Other	Please note: This is a field replicated from the ENCOUNTER table. See the ENCOUNTER table for definitions.	MSCDM with modified value set
ADMIT_DATE	TEXT(10): Format as YYYY-MM-DD	.	Please note: This is a field replicated from the ENCOUNTER table. See the ENCOUNTER table for definitions.	MSCDM with modified field name and data type Source of date format: ISO 8601
PROVIDERID	TEXT(x)	.	Please note: This is a field replicated from the ENCOUNTER table. See the ENCOUNTER table for definitions.	MSCDM

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
DX	TEXT(18)	.	Diagnosis code. Leading zeroes and different levels of decimal precision are permissible in this field. Please populate the exact textual value of this diagnosis code, but remove source-specific suffixes and prefixes. Other codes should be listed as recorded in the source data.	MSCDM
DX_TYPE	TEXT(2)	09 = ICD-9-CM 10 = ICD-10-CM 11 = ICD-11-CM SM = SNOMED CT NI = No information UN = Unknown OT = Other	Diagnosis code type. We provide values for ICD and SNOMED code types. Other code types will be added as new terminologies are more widely used. Please note: The "Other" category is meant to identify internal use ontologies and codes.	MSCDM with modified field name
DX_SOURCE	TEXT(2)	AD = Admitting DI = Discharge FI = Final IN = Interim NI = No information UN = Unknown OT = Other	Classification of diagnosis source. We include these categories to allow some flexibility in implementation. The context is to capture available diagnoses recorded during a specific encounter. It is not necessary to populate interim diagnoses unless readily available. Ambulatory encounters would generally be expected to have a source of "Final."	PCORnet
PDX	TEXT(2)	P = Principal S = Secondary X = Unable to Classify NI = No information UN = Unknown OT = Other	Principal discharge diagnosis flag. Relevant only on IP and IS encounters. For ED, AV, and OA encounter types, mark as X = Unable to Classify. (Billing systems do not require a primary diagnosis for ambulatory visits (eg, professional services)) One principal diagnosis is expected, although in some instances more than one diagnosis may be flagged as principal.	MSCDM with modified field size and value set
RAW_DX	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet
RAW_DX_TYPE	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
RAW_ DX_SOURCE	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet
RAW_ PDX	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet

3.5. Table: PROCEDURE

Description:
- The PROCEDURE Table contains one record per unique combination of PATID, ENCOUNTERID, PX, and PX_TYPE. This table should capture all uniquely recorded procedures for all encounters.
- If a patient has multiple procedures ordered during one encounter, then there would be one record in this table for each procedure.
- ENCOUNTERID is not optional for DIAGNOSIS and PROCEDURE. The definitions of the DIAGNOSIS and PROCEDURE tables are dependent upon a healthcare context; therefore, the ENCOUNTER basis is necessary.
- Healthcare mediated process and reimbursement driver.
- Can include both technical/facility billing and professional billing.
- Only billed procedures should be included in the PROCEDURE table. The ORDER concept may be incorporated into future phases of the CDM.
- If a local ontology is used, but cannot be mapped to a standard ontology such as ICD-9-CM, PX_TYPE should be populated as "Other".

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
PATID	TEXT(x)	.	Arbitrary person-level identifier. Used to link across tables.	MSCDM
ENCOUNTERID	TEXT(x)	.	Arbitrary encounter-level identifier. Used to link across tables.	MSCDM
ENC_TYPE	TEXT(2)	AV = Ambulatory Visit ED = Emergency Department IP = Inpatient Hospital Stay IS = Non-Acute Institutional Stay OA = Other Ambulatory Visit NI = No information UN = Unknown OT = Other	Please note: This is a field replicated from the ENCOUNTER table. See ENCOUNTER table for definitions.	MSCDM with modified field name and value set

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
ADMIT_DATE	TEXT(10): Format as YYYY-MM-DD	.	Please note: This is a field replicated from the ENCOUNTER table. See ENCOUNTER table for definitions.	MSCDM with modified field name and data type Source of date format: ISO 8601
PROVIDERID	TEXT(x)	.	Please note: This is a field replicated from the ENCOUNTER table. See ENCOUNTER table for definitions.	MSCDM
PX	TEXT(11)	.	Procedure code.	MSCDM
PX_TYPE	TEXT(2)	09 = ICD-9-CM 10 = ICD-10-PCS 11 = ICD-11-PCS C2 = CPT Category II C3 = CPT Category III C4 = CPT-4 (i.e., HCPCS Level I) H3 = HCPCS Level III HC = HCPCS (i.e., HCPCS Level II) LC = LOINC ND = NDC RE = Revenue NI = No information UN = Unknown OT = Other	Procedure code type. We include a number of code types for flexibility, but the basic requirement that the code refer to a medical procedure remains. Revenue codes are a standard concept in Medicare billing and can be useful for defining care settings. If those codes are available they can be included. Medications administered by clinicians can be captured in billing data and Electronic Health Records (EHRs) as HCPCS procedure codes. Administration (infusion) of chemotherapy is an example. We are now seeing NDCs captured as part of procedures because payers are demanding it for payment authorization. Inclusion of this code type enables those data partners that capture the NDC along with the procedure to include the data. Please note: The "Other" category is meant to identify internal use ontologies and codes.	MSCDM with modified field name and value set
RAW_PX	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet
RAW_PX_TYPE	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet

3.6. Table: VITAL

Description:
- The Vital Signs Table contains one record per result/entry.
  - Please note: Multiple measurements per encounter can be populated (for example, 3 blood pressure readings); each measurement would be a separate record.
- Both healthcare and non-healthcare settings.
- Does not include direct feeds from devices.

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
PATID	TEXT(x)	.	Arbitrary person-level identifier. Used to link across tables.	MSCDM
ENCOUNTERID	TEXT(x)	.	Arbitrary encounter-level identifier. This is an optional relationship; the ENCOUNTERID should be present if the vitals were measured as part of healthcare delivery.	PCORnet
MEASURE_DATE	TEXT(10): Format as YYYY-MM-DD	.	Date of vitals measure.	MSCDM with modified data type

Source of date format: ISO 8601 |
| MEASURE_TIME | TEXT(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute | . | Time of vitals measure. | MSCDM with modified data type

Source of time format: ISO 8601 |
| VITAL_SOURCE | TEXT(2) | PR = Patient-reported
HC = Healthcare delivery setting
NI = No information
UN = Unknown
OT = Other | Please note: The "Patient-reported" category can include reporting by patient's family or guardian | PCORnet |
| HT | NUMBER(8) | . | Height (in inches) measured by standing. Only populated if measure was taken on this date. If missing, leave blank. Decimal precision is permissible. | MSCDM |

Field Name	Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Source
WT	NUMBER(8)	.	Weight (in pounds). Only populated if measure was taken on this date. If missing, leave blank. Decimal precision is permissible.	MSCDM
DIASTOLIC	NUMBER(4)	.	Diastolic blood pressure (in mmHg). Only populated if measure was taken on this date. If missing, leave blank. Only report 1 reading per encounter.	MSCDM
SYSTOLIC	NUMBER(4)	.	Systolic blood pressure (in mmHg). Only populated if measure was taken on this date. If missing, leave blank. Only report 1 reading per encounter.	MSCDM
ORIGINAL_BMI	NUMBER(8)	.	BMI if calculated in the source system. Important: **Please do not calculate BMI during CDM implementation**. This field should only reflect originating source system calculations, if height and weight are not stored in the source.	PCORnet
BP_POSITION	TEXT(2)	01 = Sitting 02 = Standing 03 = Supine NI = No information UN = Unknown OT = Other	Position for orthostatic blood pressure. Leave blank if blood pressure was not measured.	MSCDM with modified field name, field size, and value set
RAW_ VITAL_SOURCE	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet
RAW_ DIASTOLIC	TEXT(x)	.	Optional field for originating value of field, prior to formatting into the PCORnet CDM.	PCORnet
RAW_ SYSTOLIC	TEXT(x)	.	Optional field for originating value of field, prior to formatting into the PCORnet CDM.	PCORnet
RAW_ BP_POSITION	TEXT(x)	.	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet

4. Glossary of Terms

Abbreviation	Term	Definition
CDM	Common Data Model	A CDM standardizes the definition, format and content of data across participating data partners so that standardized applications, tools and methods can be applied.
CPT	Current Procedural Terminology	CPT (Current Procedural Terminology) is a set of detailed, standardized codes used primarily to identify medical services and procedures. These services may be ordered or provided by physicians and other health care professionals. Examples of CPT codes are 90653 (Influenza vaccine, inactivated, subunit, adjuvanted, for intramuscular use) and 70551(Magnetic resonance imaging (MRI), brain; without contrast material). CPT codes are a subset of the HCPCS coding system. CPT and HCPCS codes are relevant to research because they enable users to see which patients received a procedure of interest, for example, the number of patients at XYZ health system who received 2 or more MRIs in 2013.
DRN	Distributed Research Network	An approach to multi-site research that allows secure analysis of separate data resources held by data partners behind their firewalls. In a DRN there is no central data warehouse. Each data resource held locally consists of data collected, captured, or otherwise obtained by the local health system.
EHR	Electronic Health Record	A repository of electronic information about an individual's health status and health care. EHRs contain much of the same information that is found in a patient's (paper) medical chart, but because the records are digitized, the data can be viewed, transmitted, and/or integrated across settings (e.g. inpatient hospital, office) and between different health care providers (e.g. primary care physicians, specialists) and can capture far more extensive information. EHRs may contain administrative and billing data, patient demographics, progress notes, vital signs, medical histories, diagnoses, medications, immunization records, allergies, radiology images, laboratory and other test results, and much more.
ETL	Extract, Transform, and Load	ETL is a process in which programmers extract data from one or more data sources, transform the data to fit certain requirements or specifications, and then load the data into a desired location. In the context of PCORnet, programmers at the various CDRNs and PPRNs will extract the data needed to populate the PCORnet Common Data Model from the data sources which house the necessary information, transform their data to fit into the Common Data Model, and then load that transformed data into a defined location.
HCPCS	Healthcare Common Procedure Coding System	HCPCS (Healthcare Common Procedure Coding System) is a set of codes used to identify tasks and procedures performed by a health care practitioner. They include Level 1 codes (CPT codes) and Level II codes (codes that primarily represent non-physician services such as ambulance, durable medical equipment, and home health services). Examples of Level II codes are EO430 (Portable gaseous oxygen system, purchase; includes regulator, flowmeter, humidifier, cannula or mask, and tubing) and A0429 (Ambulance service, basic life support, emergency transport). These codes are maintained by the American Medical Association (Level I/CPT) and the Centers for Medicare and Medicaid (Level II).

HMORN VDW	HMO Research Network Virtual Data Warehouse	HMORN (HMO Research Network) is a network comprised of research centers based in multiple health care systems. The VDW (Virtual Data Warehouse) is the Common Data Model used by the HMORN research network. It contains data on enrollment, vital signs, pharmacy, and other standardized data elements. Link to the VDM: http://www.hmoresearchnetwork.org/en/Tools%20&%20Materials/VDW/
MSN	Mini-Sentinel Network	A distributed research network sponsored by the US Food and Drug Administration (FDA) to help monitor the safety of FDA-regulated medical products such as medications and vaccines.
MSCDM	Mini-Sentinel Common Data Model	The Common Data Model used by the Mini-Sentinel Network. Link to the document library: http://www.mini-sentinel.org/data_activities/distributed_db_and_data/default.aspx Link to the Mini-Sentinel Common Data Model, v4.0: www.mini-sentinel.org/work_products/Data_Activities/Mini-Sentinel_Common-Data-Model.pdf
ICD	International Classification of Diseases	ICD (International Classification of Diseases) is the standard terminology used to classify diseases and other health problems recorded on many types of health and vital records including death certificates and health records. ICD code sets include ICD-9-CM (Ninth International Classification of Diseases-Clinical Modification),ICD-10-CM (Tenth International Classification of Diseases-Clinical Modification), and ICD-10-PCS (Tenth International Classification of Diseases-Procedure Coding System). ICD codes are primarily used to record diagnoses, but a subset (ICD-9-CM codes volume 3 and ICD-10-PCS codes) are used to record procedures performed in an inpatient setting. These codes are maintained by the World Health Organization. See WHO ICD for more information. Most health care providers in the United States are required to switch from ICD-9 to ICD-10 by October 1, 2014. ICD codes often used by researchers to identify patients with a condition of interest.
PHI	Personal Health Information or Protected Health Information	PHI ((Personal Health Information or Protected Health Information) is information about patients that is protected from inappropriate disclosure under the privacy and security mandates of the Health Insurance Portability and Accountability Act of 1996 and subsequent related legislation.
	PopMedNet TM	An open-source software application which enables simple creation, operation, and governance of distributed health data networks. http://www.popmednet.org/
	Source data	Data residing in operational healthcare systems such as electronic health record systems, claims systems, patient registries, and databases containing survey data.

| SNOMED CT | Systematized Nomenclature of Medicine-Clinical Terms | SNOMED (Systematized Nomenclature of Medicine-Clinical Terms) codes are used in electronic medical records to document diagnoses, procedures, drugs, and other health care interactions. These codes are maintained by International Health Terminology Standards Development Organization. The primary unit of measurement in SNOMED is called a concept. Each concept has a concept unique identifier (CUID), and concepts are related to one another in multiple ways. For example, concepts pertaining to asthma include COI55877 (Allergic asthma), CO451607 (Asthma management plan given) and C0455544 (history of asthma). |
| | Value set | Value sets define which values are acceptable for a given field in a database. Some value sets may only define the format & length (e.g. a zip code is a 5 digit numeric value), and others may establish a predefined list of values (e.g. patient satisfaction ratings of "Very Dissatisfied", "Dissatisfied", "Neither dissatisfied nor satisfied", "Satisfied", and "Very satisfied"). |

5. History of Releases and Modifications

Note on version conventions: Major releases are denoted with whole number incrementation (eg, v1.0, v2.0, v3.0). Minor releases are denoted with decimal incrementation (eg, v1.1, v1.2) and will be used for bug fixes and minor adjustments.

Version	Date of Release	Description of Release
v1.0	2014-05-30	The DSSNI Task Force thanks the many individuals who provided thoughtful feedback, comments, and suggestions for this first release of the PCORnet CDM. A special thanks to members of the task force who volunteered to serve on the CDM working group.