



Common Data Model (CDM) Specification, Version 6.1

1. Table of Contents

- 1. Table of Contents..... 1
- 2. Overview of the PCORnet Common Data Model (CDM) 4
 - 2.1. License and Use4
 - 2.2. Introduction4
 - 2.3. History of Releases and Modifications4
 - 2.4. Overview Diagram10
 - 2.5. Implementation Expectations10
- 3. Design of the CDM11
 - 3.1. Special Topics for CDM Modeling11
 - 3.2. Development Notes14
 - 3.3. Comments on Protected Health Information (PHI)16
 - 3.4. The Continuum of Medication-related Data Domains17
- 4. Implementation Guidance18
- 5. Table Summaries (Core Tables)22
- 6. Core CDM Table Specifications33
 - 6.1. Table: DEMOGRAPHIC33
 - 6.2. Table: ENROLLMENT39
 - 6.3. Table: ENCOUNTER42
 - 6.4. Table: DIAGNOSIS54
 - 6.5. Table: PROCEDURES62
 - 6.6. Table: VITAL68
 - 6.7. Table: DISPENSING76
 - 6.8. Table: LAB_RESULT_CM80
- Implementation Guidance Reference Table 1: Laboratory Results & LOINC Codes 92
- Implementation Guidance Reference Table 2: Laboratory Results and CPT Codes 92

Implementation Guidance Reference Table 3: Laboratory Standard Abbreviations.....	92
6.9. Table: CONDITION	93
6.10. Table: PRO_CM.....	98
CDM Reference Table: PRO Common Measures	111
6.11. Table: PRESCRIBING.....	112
Implementation Guidance Reference Table 4: Ordering of RxNorm Term Types.....	119
Implementation Guidance Reference Table 4a: RxNorm Term Types with examples	121
6.12. Table: PCORNET_TRIAL	122
6.13. Table: DEATH.....	127
6.14. Table: DEATH_CAUSE.....	129
6.15. Table: MED_ADMIN.....	131
6.16. Table: PROVIDER	136
6.17. Table: OBS_CLIN	138
6.18. Table: OBS_GEN	145
6.19. Table: HASH_TOKEN	153
6.20. Table: LDS_ADDRESS_HISTORY	159
6.21. Table: IMMUNIZATION	165
6.22. Table: HARVEST	171
6.23. Table: LAB_HISTORY	187
7. Supplemental Table Specifications	193
7.1. Supplemental Table: PRIVATE_DEMOGRAPHIC.....	193
7.2. Supplemental Table: PRIVATE_ADDRESS_HISTORY	201
7.3. Supplemental Table: PRIVATE_ADDRESS_GEOCODE	208

Important Links and References

The PCORnet CDM documentation can be accessed online at: <https://pcornet.org/data/>

Updates made in the most recent releases are highlighted in green (major release) and yellow (minor release) to assist with visually scanning the document.

Note to programmers: The separate “CDM parseable file” is more helpful for direct use in implementation, and contains the complete table specifications. The Value Set Referene File, which contains the expanded value sets for a subset of CDM fields, is also useful tool for implementation. All documentation is available here: <https://pcornet.org/data/>
View useful tools for the CDM, such as the CDM-ERRATA and CDM-GUIDANCE issue trackers, on the PCORnet GitHub CDM Forum: <https://github.com/CDMFORUM>

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

The CDM specifications for version 3.1 and above incorporate the Implementation Guidance that has been developed for PCORnet. The Implementation Guidance is intended to help reduce the variability in how network partners populate their CDM datamarts. It provides recommendations and preferred approaches when there are multiple interpretations of the CDM specification or if there is unexpected complexity in a partner’s source data. ~~The Implementation Guidance is intended to be a living document, and as such, will be updated more frequently than the CDM specification itself.~~

To accommodate the addition of this material, the CDM page size has been increased from US Letter to US Legal (8.50”x14”). For best results when printing, use Legal-size paper.

2. Overview of the PCORnet Common Data Model (CDM)

2.1. License and Use

This work is licensed under the Creative Commons Attribution 4.0 International License, and is provided on an "as is" basis without warranties or conditions of any kind. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The PCORnet Distributed Research Network (DRN) operations center and infrastructure, including the Common Data Model (CDM), is led by the PCORnet Coordinating Center and overseen by the governance established by PCORnet's stakeholders.

The PCORnet CDM was originally based on the Mini-Sentinel Common Data Model v4.0 (MSCDM v4.0; <https://www.sentinelinitiative.org/about/sentinel-common-data-model>) 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. The PCORnet CDM is positioned within healthcare standard terminologies (including ICD, SNOMED, CPT, HCPCS, and LOINC®) to enable interoperability with and responsiveness to evolving data standards.

This material contains content from LOINC® (<http://loinc.org>). The LOINC Table, LOINC Table Core, LOINC Panels and Forms File, LOINC Answer File, LOINC Part File, LOINC Group File, LOINC Document Ontology File, LOINC Hierarchies, LOINC Linguistic Variants File, LOINC/RSNA Radiology Playbook, and LOINC/IEEE Medical Device Code Mapping Table are copyright © 1995-2022, Regenstrief Institute, Inc. and the Logical Observation Identifiers Names and Codes (LOINC) Committee and is available at no cost under the license at <https://loinc.org/kb/license/>. (Updated in v4.0.)

2.2. Introduction

What is the CDM?

*The PCORnet Common Data Model (CDM)
is a **specification** that defines a **standard organization** and **representation** of data for the
PCORnet Distributed Research Network.*

The PCORnet CDM is a key component of the PCORnet Distributed Research Network (DRN) infrastructure. PCORnet developed the PCORnet DRN to be a "...functional distributed research network that facilitates multi-site patient-centered research across the Clinical Research Networks (CRNs) and other interested contributors. The distributed network will enable the conduct of observational research and clinical trials while allowing each participating organization to maintain physical and operational control over its data." [Data Standards, Security, and Network Infrastructure Task Force (DSSNI charter), 2014]

For more details of CDM development, additional references include:

- CDM abstracts presented at scientific conferences: <https://github.com/CDMFORUM/CDM-GUIDANCE/wiki/CDM-related-Abstracts>

2.3. History of Releases and Modifications

<https://pcornt.org/data/>

Note on version conventions: Major releases are denoted in whole number increments (e.g., v1.0, v2.0, v3.0). Minor releases are denoted with decimal increments (e.g., v1.1, v1.2) and will be used for bug fixes and minor adjustments. **Updates to the HARVEST table are typically not listed in the release notes as this metadata table is expected to change to reflect every CDM expansion.**

Reference Table: History of Releases		
<i>Version</i>	<i>Date of Release</i>	<i>Description of Release</i>
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. Special thanks to members of the task force who volunteered to serve on the CDM working group.
v2.0	2015-02-27	The v2.0 release includes: <ul style="list-style-type: none"> • Four new tables (DISPENSING, CONDITION, PRO_CM, LAB_RESULT_CM) • Four new fields in existing tables (VITAL.TOBACCO, VITAL.TOBACCO_TYPE, PROCEDURE.PX_TYPE, PROCEDURE.PX_SOURCE) • Additional guidance and descriptions
v3.0	2015-06-01	The v3.0 release includes: <ul style="list-style-type: none"> • Five new tables (PRESCRIBING, PCORNET_TRIAL, DEATH, DEATH_CAUSE, and HARVEST) • Ten new fields in existing tables (DISPENSING.DISPENSINGID, DISPENSING.PRESCRIBINGID, VITAL.VITALID, VITAL.SMOKING, CONDITION.CONDITIONID, CONDITION.ONSET_DATE, PRO_CM.PRO_CM_ID, DIAGNOSIS.DIAGNOSISD, PROCEDURES.PROCEDURESID, LAB_RESULT_CM.LAB_RESULT_CM_ID) • Modification to relational integrity specifications • Modification to date formatting practices • New specifications specific to SAS data types • Additional guidance, clarifications, and descriptions
v3.0	2015-07-29	Document updated with licensing information and new PCORnet.org URL. No technical specifications have been modified.
v3.1	2016-11-15	Please note: New and modified fields are indicated in green to assist with visually scanning the document (in addition to the descriptive comments). The v3.1 release includes: <ul style="list-style-type: none"> • Four new fields (DEMOGRAPHIC.SEXUAL_ORIENTATION, DEMOGRAPHIC.GENDER_IDENTITY, DIAGNOSIS.DX_ORIGIN, PRESCRIBING.RX_QUANTITY_UNIT) • Encounter types value set expanded to include observation stays and institutional professional consults • Collapsed value set of procedure terminologies so that CPT and HCPCS are grouped into single category • Clarified expected number of digits for RDBMS number formatting • Date of death no longer a required field for DEATH table • Enrollment table basis now includes drug coverage

Reference Table: History of Releases

<i>Version</i>	<i>Date of Release</i>	<i>Description of Release</i>
V4.0	2018-01-02	<p>The v4.0 release includes:</p> <ul style="list-style-type: none"> • Four new tables (PROVIDER, OBS_CLIN, OBS_GEN, MED_ADMIN) • Thirty-two new fields in existing tables (PAT_PREF_LANGUAGE_SPOKEN, PAYER_TYPE_PRIMARY, PAYER_TYPE_SECONDARY, FACILITY_TYPE, DX_POA, PPX, DISPENSE_DOSE_DISP, DISPENSE_DOSE_DISP_UNIT, DISPENSE_ROUTE, RESULT_SNOMED, PRO_TYPE, PRO_ITEM_LOINC, PRO_ITEM_NAME, PRO_RESPONSE_TEXT, PRO_ITEM_VERSION, PRO_MEASURE_NAME, PRO_MEASURE_SEQ, PRO_MEASURE_SCORE, PRO_MEASURE_THETA, PRO_MEASURE_SCALED_TSCORE, PRO_MEASURE_STANDARD_ERROR, PRO_MEASURE_COUNT_SCORED, PRO_MEASURE_VERSION, PRO_ITEM_FULLNAME, PRO_ITEM_TEXT, PRO_MEASURE_FULLNAME, RX_DOSE_ORDERED, RX_DOSE_ORDERED_UNIT, RX_PRN_FLAG, RX_ROUTE, RX_SOURCE, RX_DISPENSE_AS_WRITTEN) • Renamed PRO_RESPONSE field to PRO_RESPONSE_NUM. • PRO_ITEM field deprecated. • Renamed PRO_LOINC field to PRO_ITEM_LOINC. • ADMITTING_SOURCE value set expanded to include intra-hospital admitting source. • DX_ORIGIN and PX_SOURCE value sets expanded to include diagnoses/procedures derived or imputed through analytical procedures (e.g., natural language processing). • SPECIMEN_SOURCE value set expanded to include all values from the LOINC SYSTEM part. • RESULT_QUAL value set expanded. • RESULT_UNIT value set expanded to include common UCUM units. • RX_QUANTITY_UNIT value set modified to align with RxNorm terminology. • RX_FREQUENCY value set expanded to include “every evening” and “once” concepts. • Renamed RX_QUANTITY_UNIT field to RX_DOSE_FORM for clarity and consistency. • LAB_NAME field deprecated. • Date management and refresh date fields for new tables added to HARVEST table. • Required/not null ENCOUNTERID constraint removed from DIAGNOSIS and PROCEDURES tables. • Required/not null PRO_RESPONSE constraint removed from PRO_CM table. • Domain descriptions updated for DISPENSING, LAB_RESULT_CM and PRO_CM tables. • Deprecation of Implementation Guidance Reference Table 3 and CDM Reference Table (PRO Common Measures). • Modifications to the foreign key descriptions for several tables. • Concept of PRIVATE and Supplemental tables introduced. • Various updates to the Implementation Guidance

Reference Table: History of Releases

<i>Version</i>	<i>Date of Release</i>	<i>Description of Release</i>
V4.1	2018-05-15	<p>The v4.1 release includes:</p> <ul style="list-style-type: none"> • Field length updates for PAYER_TYPE_PRIMARY and PAYER_TYPE_SECONDARY. • Foreign key updates for DISPENSING.PATID, PCORNET_TRIAL.PATID and DEATH.PATID. • Typo correction for MEDADMIN.MEDADMIN_PROVIDERID. • Renamed MEDADMIN_END_DATE_MGMT to MEDADMIN_STOP_DATE_MGMT. • Update to description of PRO_MEASURE_FULLNAME. • Various updates to the Implementation Guidance. • Updates to entries of the FACILITY_TYPE, ROUTE and PAYER_TYPE value sets included in the Value Set Appendix. • Removal of the DISPENSE_FORM value set from the Value Set Appendix.
V5.0	2019-07-16	<p>The v5.0 release includes:</p> <ul style="list-style-type: none"> • Three new tables: HASH_TOKEN, LDS_ADDRESS_HISTORY, IMMUNIZATION. • Modifications to field length and definition of ENCOUNTER.FACILITY_LOCATION. • Modifications to value sets for six fields (CONDITION_TYPE, CONDITION_SOURCE, OBSGEN_TYPE, VITAL_SOURCE, DEATH_SOURCE, DEATH_CAUSE_SOURCE) • Seven new fields in existing tables (DX_DATE, DISPENSE_SOURCE, LAB_RESULT_SOURCE, LAB_LOINC_SOURCE, PRO_SOURCE, OBSCLIN_SOURCE, OBSGEN_SOURCE) • Three new PRIVATE tables: PRIVATE_DEMOGRAPHIC, PRIVATE_ADDRESS_HISTORY, PRIVATE_ADDRESS_GEOCODE • Various updates to the Implementation Guidance
V5.1	2019-09-10	<p>The v5.1 release includes:</p> <ul style="list-style-type: none"> • Three new fields in HASH_TOKEN: TOKEN_03, TOKEN_04, TOKEN_16. • Deprecation of 5 fields in HASH_TOKEN: TOKEN_12, TOKEN_17, TOKEN_21, TOKEN_22, TOKEN_23. • Updates to the field descriptions in HASH_TOKEN: TOKEN_01, TOKEN_02, TOKEN_05. • Corrections to the constraints of PRIVATE_DEMOGRAPHIC. • Corrections to the field length and definition of PRIVATE_GEOCODE.GEOCODE_COUNTY. • Updates to the Implementation Guidance in PRESCRIBING, HASH_TOKEN, PRIVATE_DEMOGRAPHIC and PRIVATE_ADDRESS_HISTORY. • Correction to PRO_SOURCE value set to align with the parseable file for this field. • HARVEST table: REFRESH_LDS_ADDRESS_HISTORY_DATE is shortened to REFRESH_LDS_ADDRESS_HX_DATE. • Clarified implementation guidance for future dates and PATID consistency. • Clarified field implementation guidance for ADDRESS_PERIOD_END in the ADDRESS_HISTORY table.

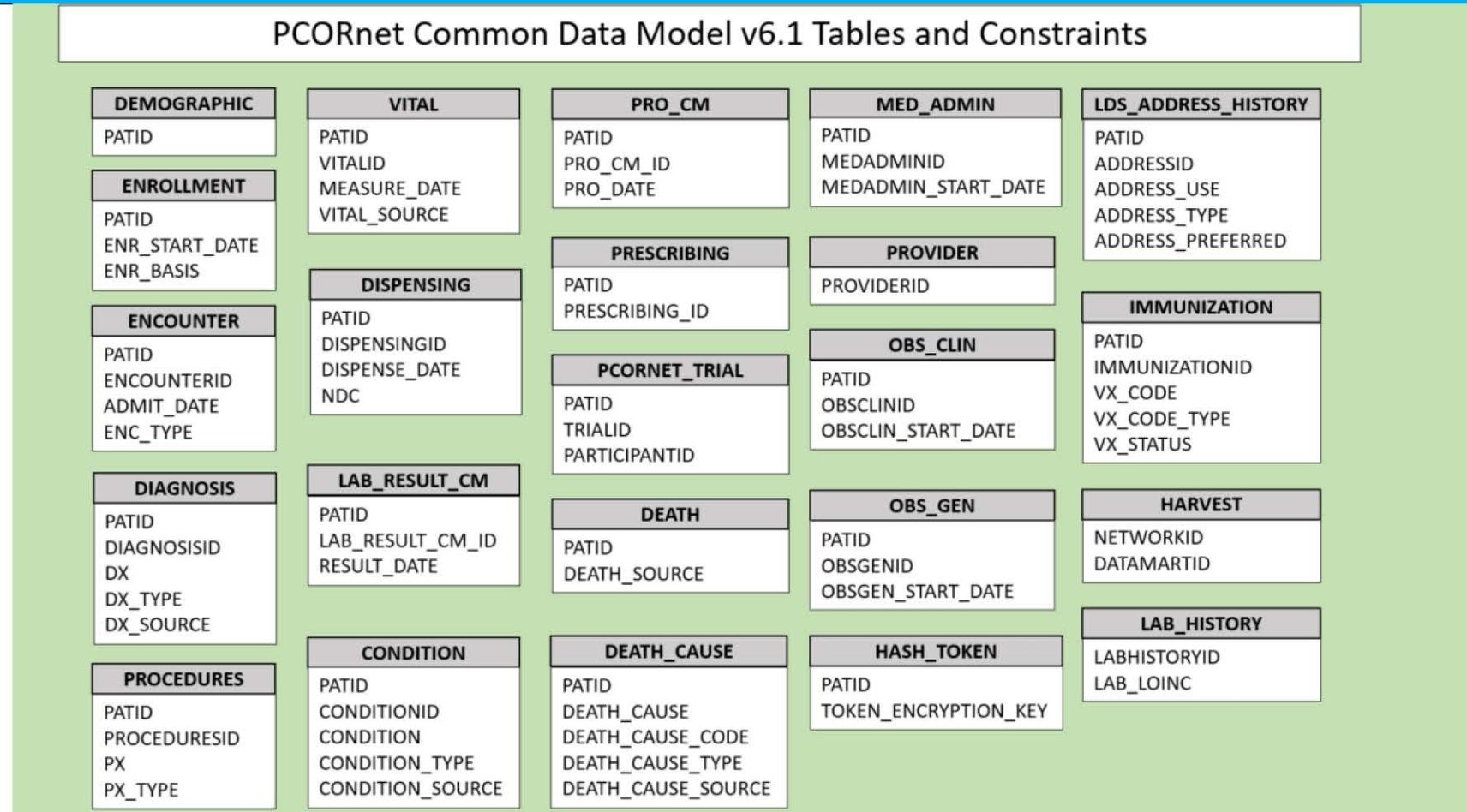
Reference Table: History of Releases

<i>Version</i>	<i>Date of Release</i>	<i>Description of Release</i>
V6.0	2020-10-22	<p>Please note: v6.0 updates are highlighted in green to assist with visually scanning the document (in addition to the descriptive comments).</p> <p>The v6.0 release includes:</p> <ul style="list-style-type: none">• One new table (LAB_HISTORY)• 9 new fields in existing tables (OBSCLIN_STOP_DATE, OBSCLIN_STOP_TIME, OBSCLIN_ABN_IND, OBSGEN_STOP_DATE, OBSGEN_STOP_TIME, OBSGEN_ABN_IND, OBSCLIN_STOP_DATE_MGMT, OBSGEN_STOP_DATE_MGMT, TOKEN_ENCRYPTION_KEY).• Renamed OBSCLIN_DATE field to OBSCLIN_START_DATE.• Renamed OBSCLIN_TIME field to OBSCLIN_START_TIME.• Renamed OBSGEN_DATE field to OBSGEN_START_DATE.• Renamed OBSGEN_TIME field to OBSGEN_START_TIME.• Renamed OBSCLIN_DATE_MGMT field to OBSCLIN_START_DATE_MGMT.• Renamed OBSGEN_DATE_MGMT field to OBSGEN_START_DATE_MGMT.• ENC_TYPE value set expanded to include Telehealth.• ADMITTING_SOURCE value set expanded to include Emergency Medical Service.• CONDITION_SOURCE value set expanded to include Patient Chief Complaint.• OBSGEN_TABLE_MODIFIED value set expanded to include IMMUNIZATION and LDS_ADDRESS_HISTORY.• OBSCLIN_SOURCE and OBSGEN_SOURCE value set modified to include Patient-reported, Patient device direct feed, Healthcare delivery setting, Healthcare device direct feed, as well as deprecate Order/EHR.• Modification to the value set type for CDM_VERSION.• Modification to the value set descriptive text for OBSGEN_TYPE, OBSLIN_SOURCE, and OBSGEN_SOURCE.• Update to the definition for LAB_LOINC, OBSCLIN_START_DATE_MGMT, and OBSGEN_START_DATE_MGMT.• Added implementation guidance for harmonized field lengths.• Update to the implementation guidance for OBS_CLIN and HASH_TOKEN.• Update to field-level implementation guidance for ADMITTING_SOURCE, RAW_RESULT, RAW_CONDITION, OBSLIN_SOURCE, OBSCLIN_START_DATE, OBSCLIN_START_TIME, OBSGEN_SOURCE, OBSCLIN_START_TIME, and OBSGEN_START_TIME.• Update to the Domain Description for OBS_CLIN.• Update to the constraints for OBS_CLIN and OBS_GEN.• Added language about deprecating VITAL.• Various updates to the Value Set Appendix.

Reference Table: History of Releases

<i>Version</i>	<i>Date of Release</i>	<i>Description of Release</i>
V6.1	2023-04-03	<p>Please note: The v6.0 updates remain with green highlighting. The v6.1 updates are highlighted in yellow.</p> <p>The v6.1 release includes:</p> <ul style="list-style-type: none">• Modification of the value set for SEXUAL_ORIENTATION (DEMOGRAPHIC and PRIVATE_DEMOGRAPHIC).• Addition of ADDRESS_COUNTY to LDS_ADDRESS_HISTORY.• Deprecation of TOKEN_ENCRYPTION_KEY from HARVEST.• Addition of TOKEN_ENCRYPTION_KEY, 26 new HASH_TOKEN fields and updates to constraints, relational integrity and table guidance to HASH_TOKEN.• Updates to the table-level guidance for CONDITION, HARVEST, LAB_RESULT_CM, OBS_CLIN, PCORNET_TRIAL, PRO_CM and VITAL as well as the domain descriptions for OBS_CLIN and PRO_CM.• Updated field description for ENC_TYPE and all *_LOINC fields.• Specified a field length for MEDADMIN_CODE, OBSCLIN_CODE, OBSGEN_CODE, OBSCLIN_RESULT_SNOMED, RESULT_SNOMED, and VX_CODE.• Addition of guidance on undefined field lengths and missing time elements to General Implementation Guidance; other updates to the implementation guidance.• Renamed Value Set Appendix to Value Set Reference File. Updated language in CDM to reflect this change.• Various updates to the Value Set Reference File and Parseable File (see INFO tabs in those files for more detail).• Addition of Table Summaries (Core Tables).

2.4. Overview Diagram



2.5. Implementation Expectations

Partners should populate all core CDM tables if data are available in their source system(s). **All core CDM tables must be present in an instantiation of the CDM, even if the table is empty.** This is important because some components of the querying platform need to locate a given table, even if zero records are present in that table. The fields that are required to be populated for all records in a given table are listed in the “constraints” section of each table description. Any table and/or field in the CDM may be required for a partner’s participation in a given study or other PCORnet activity. In assessing foundational data quality, the PCORnet Data Curation query packages may query any CDM table or field (**except for PRIVATE tables**).

3. Design of the CDM

3.1. Special Topics for CDM Modeling

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.**” [emphasis added]

This guiding principle 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 v4.0, including data characterization approaches and the various tools for complex distributed analytics.

Relational Integrity

Database programmers will notice that fields used as primary/foreign keys, especially PATID and ENCOUNTERID, are specified as text instead of numbers. 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.

Please note that all tables must be present in an instantiation of the CDM, even if data are not populated in every table.

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, each RDBMS will be expected to implement its own native date data type for dates, which will be supported by the Entity Framework technology stack¹. The CDM will always separate date fields and time fields for consistency, and employ a naming convention of suffix “_DATE” or “_TIME”.

All times should be recorded within the local time zone. A uniform time stamp or GMT offset is not expected.

¹ <https://msdn.microsoft.com/en-us/data/ef.aspx>
<https://pcornet.org/data/>

A SAS date is different from an RDBMS date. A SAS date is a value that represents the number of days between January 1, 1960 and the specified date. SAS can perform calculations on dates ranging from A.D. 1582 to A.D. 19,900. Dates before January 1, 1960, are negative numbers; dates after are positive numbers. (Guidance added in v3.1.)

Number Formatting

SAS Number fields have a *byte length* of 8 [SAS Numeric(8)]. This corresponds to an 8-byte floating-point number of approximately 16 significant digits. When deciding on the precision/scale for their RDBMS Number fields, partners should ensure that they do not store numbers in a way that would overflow the SAS numeric data type, which would result in a loss of data when generating a SAS dataset from the RDBMS. RDBMS Number can be implemented as any appropriate RDBMS number concept, such as DECIMAL or DOUBLE data types. Although some RDBMS's have a specific data type called "NUMBER" (such as Postgres), the CDM does not imply that this specific data type should be implemented.

When deciding on the parameters to choose for their RDBMS number fields, network partners should choose a combination that does not result in additional, artificial decimal precision. For example:

- The value **1.1** should **never** be modified to become **1.10000000**
- The integer value of **1** should **never** be modified to become **1.0** or **1.00000000**

Instead of specifying a precision (total number of digits) and scale (digits to the right of a decimal point) for RDBMS Number data types, the CDM spec has been revised to just read "RDBMS Number." Partners should specify the parameters that are most appropriate for their RDBMS that that does not cause a loss of data when generating SAS datasets from the RDBMS or nor result in additional, artificial decimal precision.

Missing or Unknown Data Values

The PCORnet CDM will use the HL7 conventions of "Null Flavors" (<https://terminology.hl7.org/1.0.0/CodeSystem-v3-NullFlavor.html> ~~http://hl7.org/implementation/standards/fhir/v3/NullFlavor/~~) as a basis for representing missing or unknown values **in all tables except for the LAB_HISTORY table**. Specifically, for fields where an enumeration is present (i.e., a categorical set of values), we will populate null values as follows:

A data field is not present in the source system. (populate with null)

A data field for an enumeration is present in the source system, **but the source value is null or blank**. (populate with NI=No Information)

A data field for an enumeration is present in the source system, but the source value **explicitly denotes an unknown value**. (populate with UN=Unknown)

A data field for an enumeration is present in the source system, but the source value **cannot be mapped to the CDM**. (populate with OT=Other)

This guidance is only applicable for categorical text fields, not for numbers or dates.

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 analysis or query, allowing for these decisions to be driven by appropriateness for each individual analysis.

PCORnet CDM Guiding Principle #7 states,
“**The CDM will reflect variables and values found in the local data.** If some data are coded in a way that is unique to a site, mapping the data to a standardized format will be necessary. **Values in the source data before mapping will also be included in the CDM.** Derived variables should be avoided.” [emphasis added]

“Raw” Fields

The data model uses a convention for “raw data fields.” These 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.

The data model uses a convention for “raw data fields.” In general, these fields are used to store the originating source value, prior to mapping into PCORnet CDM value set. They may also be used for source-specific ontologies. For tables that may contain narrative or textual results (e.g., LAB_RESULT_CM, CONDITION), the “RAW” field can be used to store result values that would be used in queries or other analyses. Partners should not use RAW fields to store internal record IDs.

The “RAW” fields are intended to support data provenance and may be queried by the DRN OC or others to facilitate quality assurance activities. These fields will have a naming convention of prefix “RAW_”. Partners are encouraged to populate as many RAW fields as possible, but if there are concerns that specific records or fields may contain PHI, those values should be excluded. However, **we discourage the complete exclusion of RAW field values from the CDM**, as that will limit PCORnet’s ability to assess data quality.

Case Sensitivity

All RDBMS implementations are case-sensitive. Schema implementations for Oracle, Microsoft SQL Server, and PostgreSQL should be in uppercase (table name, column names, etc.). Value set codes should reflect the case formatting in the CDM specification and/or Value Set [Reference File](#).

Avoidance of Padding

Numbers should not be “padded” with extra zeroes. Text fields should not be “padded” with spaces before or after the actual textual values.

Additional Fields

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

As stated in PCORnet CDM Guiding Principle #8,
“CRNs may include additional domains and data elements in localized versions of the PCORnet CDM.”

Incomplete Date Guidance

In situations where the exact day or month is unknown or not available, it is still necessary to have a valid date for native RDBMS and SAS date data types. In this situation, please use this specific imputation strategy:

- If the day is missing, use the **first day of the month** to create a valid date value with the existing month and year.
- In the uncommon situation where a month is missing, **use January 1** to create a value date value with the existing year.

The HARVEST table indicators of DATE_ fields are used to indicate the presence of incomplete dates within the data, and the specific details of imputation would be described in the extract-transform-load (ETL) Annotated Data Dictionary (ADD). The convention of the RAW_ fields can also be deployed to indicate the presence and original value of incomplete dates, if desired.

Expanded value sets

Version 4.0 of the PCORnet CDM introduced the concept of expanded value sets for fields with dozens or hundreds of allowable options (e.g., LAB_RESULT_CM.RESULT_UNIT, PRESCRIBING.RX_DOSE_FORM). To reduce the size of the CDM specification document, these value sets are provided in a supplementary Value Set [Reference File](#), which is co-located with the CDM specification on the PCORnet web site (<https://pcornet.org/data/>). It is expected that these value sets will only be updated as part of a CDM version update, but there may be extenuating circumstances where an out-of-sequence update is required. For each value set, we list the raw/expected source value and the corresponding analytics-friendly string to be used when populating the CDM.

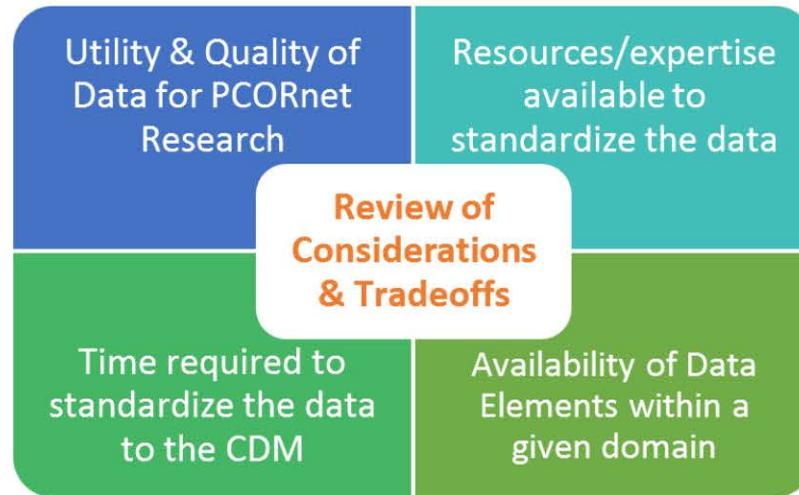
Supplemental tables

Version 4.0 of the PCORnet CDM also formalized the concept of “supplemental” CDM tables. These are tables outside of the core CDM that used to support study-specific activities that involve the participation of many network partners. These table definitions may be managed in separate document(s) outside of the core CDM specification and may be refined out of cycle with the rest of the CDM. Over time, some of these tables may be promoted to the core CDM, at which point they will be governed by the versioning processes of the core. Network partners are not expected to populate these tables unless they are participating in a study that utilizes them.

3.2. Development Notes

PCORnet CDM Guiding Principle #2 states,
“It is expected that not all CRNs will have data needed to populate all parts of the PCORnet CDM. It is the responsibility of the CRNs to communicate availability of each data domain and element.” [emphasis added]

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 assessment of considerations and tradeoffs is an integral part of decision-making based on pragmatism and analytic value.



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.

PCORnet CDM Guiding Principle #6 states,
 “Other common data elements and common data model initiatives exist. PCORnet will draw from the experience of others within and outside of PCORI, leveraging existing successful approached and data model definitions wherever possible.”

The model was initially 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 have been a basis for strategy and decisions. 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 #4,
 “The PCORnet CDM will be developed in a **modular, incremental, and extensible fashion**. New types of data will be needed, or newly available, during the life of PCORnet. Data domains and data elements will be added, revised, and deprecated throughout an iterative CDM lifecycle.” [emphasis added]

3.3. 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 (e.g., 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.

PCORnet Distributed Research Network Guiding Principle #2 states,
“CRNs will control how their data are used as allowed by internal governance policies. Data resources developed for PCORnet will stay within the CRNs and under their control.” [emphasis added]

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 should include (but are not limited to):

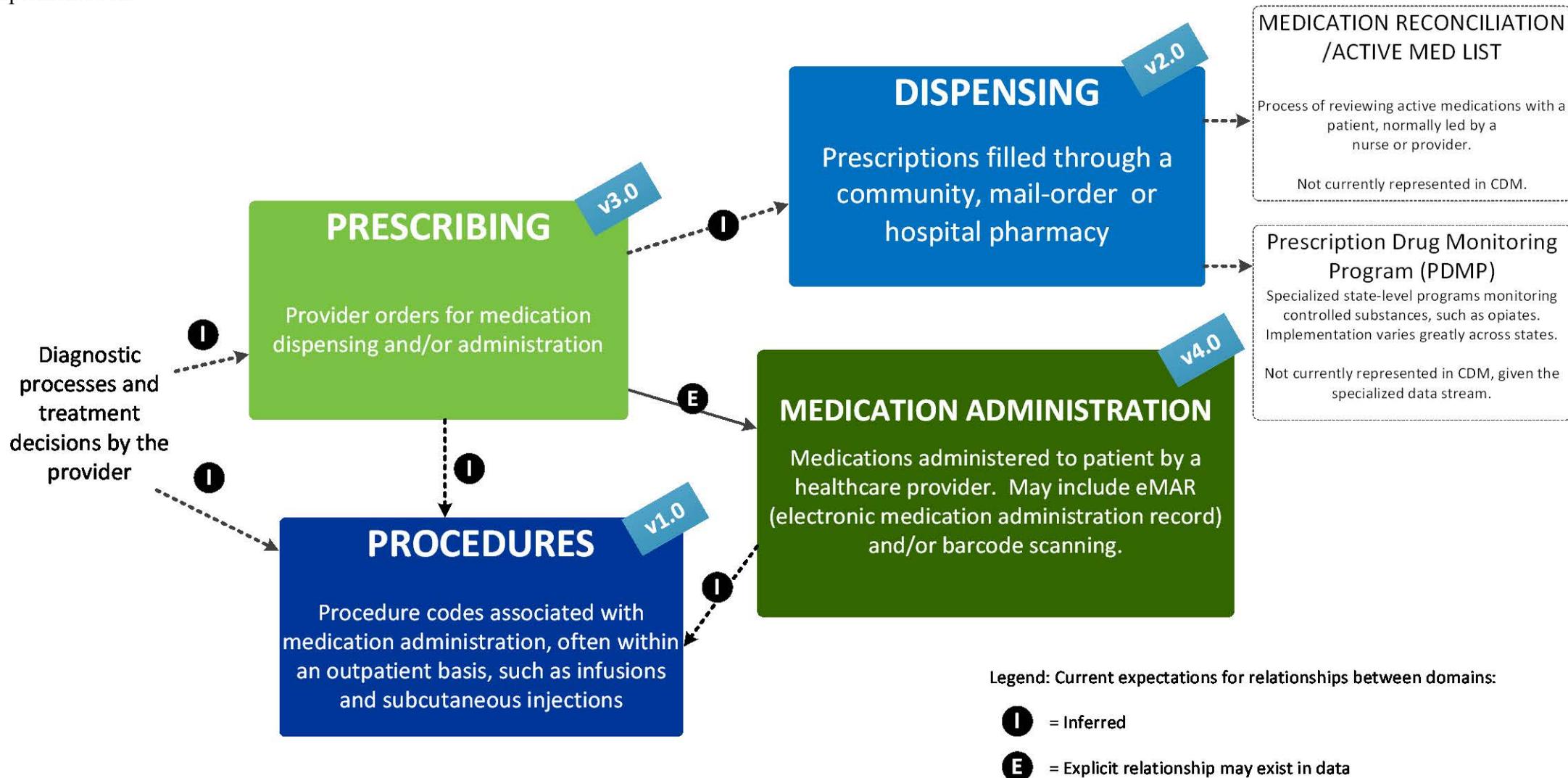
- PATID crosswalk
- PROVIDER crosswalk

PRIVATE tables

Version 4.0 of the PCORnet CDM introduced the concept of “PRIVATE” tables, which are intended to provide standardized representations for the commonly-used PHI elements that are necessary for certain analytic activities (e.g., patient linkage, geocoding). These tables will not be directly queried by the DRN OC and can remain physically and logically separate from the rest of the CDM. These tables will initially be developed as Supplemental tables through the process described in Section 3.1.

3.4. The Continuum of Medication-related Data Domains

This diagram represents our expectations for the **current state** of medication-related data stores in clinical systems, and is meant to assist in the assessment of data availability for PCORnet CDM implementation.



4. Implementation Guidance

Implementation Guidance is intended to help reduce the variability in how partners populate their CDM datamarts. It provides recommendations and describes preferred approaches when there are multiple interpretations of the CDM specifications or if there is unexpected complexity in a partner’s source data. The Implementation Guidance is divided into three sections, plus a set of reference tables:

1. *General guidance*: The guidance applies to more than 1 CDM table/domain. These items are included in this section of the CDM Specification.
2. *Table-level guidance*: The guidance applies to the table/domain in general or applies to more than one field in the table. Table-level guidance is provided in each table’s description before the table specification.
3. *Field-level guidance*: The guidance applies to implementation decisions that are specific to a given field in a table. Field-level guidance is provided as an extra column in the CDM table specification.
4. *Reference tables*: When applicable, reference tables have been created to provide additional guidance to network partners.

Guidance updated as part of a major CDM revision is highlighted in **green**. Guidance updated between major CDM versions is highlighted in **yellow**.

GENERAL Implementation Guidance (spans more than 1 table)	
<i>Topic</i>	<i>Guidance</i>
1 – Population of RAW fields	<p>If a given record/observation can be directly mapped into the PCORnet CDM, there is no need to populate the RAW values. RAW values may be used when partners need to employ a mapping in order to populate a given table (e.g., converting local diagnosis codes to ICD-9 or ICD-10, mapping internal procedure codes to one of the specified PCORnet procedure terminologies or converting several dozen race values to the values specified in the PCORnet CDM). Populating the RAW fields is optional, but if a partner chooses to do so, they should use the following strategy:</p> <ul style="list-style-type: none"> • If there is a 1:1 relationship between the source value and the CDM, the RAW value can be populated on the same record. <p>In cases where there is a 1:many relationship for multiple records that are part of encounter or have the same timestamp (e.g., 3 local diagnosis codes recorded in the same encounter that map to a single ICD9 code), all of the local values should be concatenated into a single RAW observation, with the values separated by a pipe delimiter (“ ”). This prevents the creation of duplicate records in the PCORnet CDM.</p>
2 – Date Obfuscation and/or Truncation	<p>Do not include shifted or truncated dates within the CDM, with the exception of birth date for patients >89, if required by local regulations. Aggregate or de-identified queries that return age for patients >89 will appropriately bucket the results before they are returned.</p>
3 – Patients with age > 89	<p>PCORnet queries issued by the DRN OC will bucket all patients age 90+ into one age group to limit the risk of re-identification. Some partners’ local regulatory restrictions may require additional protection of date-related data on patients age > 89. For example, some partners might be asked to remove the birth date while others might require that it be shifted to mask the patient’s true date of birth. If the birth date must be shifted, these regulatory restrictions would subsequently prevent the same shift from being applied to the rest of the patient’s dates of service, resulting in inconsistencies in the data (see General Guidance #2). In this case, partners should consider shifting the value in the DEMOGRAPHIC.BIRTH_DATE field for these patients to a dummy date of January 1, 1900. If this approach is taken, partners should structure their ETL to shift birth dates from patients currently over 89, as well as those that will “age out” and turn 89 before the next expected datamart refresh.</p>

GENERAL Implementation Guidance (spans more than 1 table)	
<i>Topic</i>	<i>Guidance</i>
4 – Mapping source data to standard terminologies when multiple options exist	<p>Some partners with Epic EHRs are able to access diagnoses/problem list entries coded via vocabulary mapping software/middleware (e.g., IMO). Such software/middleware provides mappings to multiple terminologies, so partners have several options from which to choose when populating their datamart.</p> <p>Partners should utilize the vocabulary or terminology that most closely reflects the standard typically used to encode data in that domain at the time the observation was recorded. For instance, diagnosis codes would use ICD-9 before October 2015, and ICD-10 afterwards; procedure codes would be encoded in CPT/HCPCS or in ICD-9 before October 2015 and in CPT/HCPCS or ICD-10 afterwards; problem list values would be encoded in SNOMED. This will allow partners to avoid the potential duplication of records (e.g., if combining facility billing diagnoses coded in ICD with professional diagnoses coded in SNOMED or mapping the same EHR record to ICD-9 and ICD-10) and more readily allow the application of existing validated algorithms. In these situations, the original codes can be retained within the RAW fields to allow for additional studies that seek to execute exploratory analyses using alternative mappings. This guidance may be revised in the future to incorporate the published findings of subsequent validation studies that compare the downstream analytical results that arise from various mapping strategies.</p>
5 – Approach when there are known errors or missingness in TYPE fields (e.g., PROCEDURES.PX_TYPE, LAB_RESULT_CM.LAB_RESULT_PX_TYPE, ENCOUNTER.DRG_TYPE, or DIAGNOSIS.DX_TYPE)	<p>The PCORnet CDM Guiding Principles ask that data from source systems be populated in the CDM “as is.” However, some fields in the CDM that are used to identify attributes about the data, such as the TYPE fields, are critical to the operation of the PCORnet analytic queries. If the TYPE field is incorrect, the query will end up using a mismatched code set and return invalid/empty results. If the errors in the source for the TYPE fields are systemic (e.g., an interface reports all procedure billing codes as ICD-10, when they are really a mix of ICD-10 and CPT), and the partner is able to isolate and rectify the issue, then the preference is that these data be cleaned before populating the CDM. The same guidance applies if TYPE fields are not populated or are not transmitted from the source system. The heuristic or algorithm used by the partner should be documented in their ETL ADD. <i>If there are known systemic issues in other fields or domains, these data should not be cleaned, but partners should make every attempt to resolve or rectify the issue within their source system(s).</i> Any known systemic issues should also be documented in the ETL ADD.</p>
6 – Corrected/updated source values	<p>If a measure/laboratory result has been corrected in the source system, and both the original and corrected values are present in the source system, partners should only include the corrected (e.g., most recent) value in their CDM datamart. If a measure/laboratory result is included in their CDM datamart that has later been found to be canceled, partners should remove that value from their CDM datamart in subsequent refreshes.</p>
7 – Precision/scale for RDBMS NUMBER fields	<p>This guidance has been deprecated as of CDM v3.1.</p>
8 – Study-specific datamart restrictions	<p>The DRN OC will work with networks that operate using a centralized model to ensure that site-specific results are available for all queries in the same fashion as distributed data marts.</p>
9 – PATIDs & other IDs	<p>To the extent possible, Partners should maintain PATIDs and all other ID fields (ENCOUNTERID, DIAGNOSISID, PROVIDERID, etc.) across refreshes. In addition, the lengths of these fields should be harmonized across all tables to facilitate cross-table querying (see General Guidance #18).</p>

GENERAL Implementation Guidance (spans more than 1 table)

Topic	Guidance
10 – Units of Measure	<p>Starting with PCORnet CDM v4.0, the value set for the units of measure fields (e.g., RESULT_UNIT in LAB_RESULT_CM, OBSGEN and CLINOBS OBSCLIN, and *_STRENGTH*_UNIT in PRESCRIBING, DISPENSING and MED_ADMIN) has been expanded to include values from the Unified Code for Units of Measure (UCUM). The value set in the Value Set Reference File reflects the list of values curated by LOINC that are most often used in clinical reporting (https://loinc.org/usage/units/). This list likely includes more values than a partner would have in their source system, but we are providing a more comprehensive list to err on the side of completeness.</p> <p>When mapping from source data to the UCUM values, partners should choose the code that is the closest match to their source data. It is expected that most units will have a 1:1 match with the UCUM list, though in some cases, there will be different spelling variations or representations that need to be harmonized or made singular (e.g., mapping PERCENT to % or liter(s) to liter). For laboratory test results and clinical observations, partners should ensure that there is no discordance between the selected unit of measure and the corresponding test code (e.g., choosing % as a unit of measure when the LOINC code indicates the result is reported as mg/dL).</p> <p>It is possible that a partner may have unit values that are not part of the LOINC-curated list. This is because the universe of UCUM codes is essentially unbounded (http://unitsofmeasure.org/ucum.html#datyp2apdxatblxmp). We have standardized on the LOINC-curated list to allow for testing of model conformance. Units that fall outside of this value set should be mapped to “OT,” with the source value stored in the appropriate RAW field, even if they are technically “valid” UCUM codes.</p> <p>Please note that there are several instances where a single unit code maps to multiple entries (e.g., nmol/mg{prot} maps to both “nanomole of ½ cystine per milligram of protein” and “nanomole per milligram of protein”). This is a known issue and reflective of how the curated list of codes were created by LOINC. The values were distinct at one point but collapsed due to field length restrictions. In addition, there are instances where units differ solely by letter case (e.g., “A” – ampere, “a” – year; “Ms” – megasecond, “ms” – millisecond). As with all valuesets, we will retain case sensitivity when making comparisons on this field.</p>
11 – Use of “OT” when mapping to expanded value sets	<p>As partners map their source data to the expanded value sets, the value of “OT” (Other) may be used in cases where a source value is present, but does not match any of the responses in the value set. Additionally, “OT” may be used in those cases when a data partner has not completely mapped all of their source data to the CDM value set for a given variable (e.g., UNIT, ROUTE, FACILITY_TYPE, etc.). In this case, partners can include the code from the Value Set Reference File for those entries that have been mapped and “OT” for the remainder until they have been assigned a code from the value set. Unmapped values should be specified in the corresponding RAW field.</p>
12 – Required vs. Optional fields	<p>The fields that are required to be populated for all records in a given table are listed in the “constraints” section of each table description. Any table and/or field in the CDM may be required for a partner’s participation in a given study or other PCORnet activity. In assessing foundational data quality, the PCORnet Data Curation query packages may query any CDM table or field (Language replicated from Section 2.5 – Implementation Expectations).</p>
13 – Storing PROs or other observations with date-type responses	<p>When loading PROs with a date-type response, store the character/string version of the date in PRO_RESPONSE_TEXT field, and the SAS Date format (Numeric) of the response in PRO_RESPONSE_NUM. Analytically, the PRO_RESPONSE_NUM value will be used in queries. The same logic applies for observations with date-type responses that might be stored within OBS_CLIN or OBS_GEN.</p>

GENERAL Implementation Guidance (spans more than 1 table)	
<i>Topic</i>	<i>Guidance</i>
14 – Future Dates	If the source system assigns all unspecified end dates with a far-off future date instead of null (e.g., year 4700), these values should be set to null when loading the CDM. Only do this for true “dummy” dates that should truly be blank/null, and only for dates that are systematically assigned at the source.
15 – Correspondence between LOINC classes and CDM tables	The LOINC Class Type variable has 4 values (as of May 2018): 1 = Lab Class, 2 = Clinical Class, 3 = Claims Attachment, 4 = Survey. This information can be found on search.loinc.org or in the LOINC Core files. In general, observations with a class type of “1” will be stored in LAB_RESULT_CM. Observations with a class type of “2” will be loaded into OBS_CLIN, and observations with a class type of “4” will generally be stored in PRO_CM, but not definitively.
16 – Immunizations	Vaccines captured as part of an immunization workflow should be stored in the IMMUNIZATION table. If immunizations are present in other data streams (e.g., medication administrations, procedures, etc.), however, they should not be removed from those tables.
17 – Lab LOINC mappings	If partners discover that laboratory results have been assigned an incorrect LOINC code within their source system(s), they should make every effort to have it corrected by the personnel who maintain those mappings. In the meantime, it is acceptable to assign the proper LOINC code before loading those results into the CDM.
18 – Harmonized field lengths	Fields with undefined lengths that are present in more than one CDM table must have consistent field lengths in all tables. This constraint applies to the following fields: PATID, ENCOUNTERID, PRESCRIBINGID, PROCEDURESID, and all PROVIDERID fields (PROVIDERID, MEDADMIN_PROVIDERID, OBSGEN_PROVIDERID, OBSCLIN_PROVIDERID, RX_PROVIDERID, and VX_PROVIDERID).
19 – Undefined field lengths	For character fields with undefined field lengths, sites should use the minimum length needed to store the data in order to save storage space and reduce query processing time. Field lengths for non-RAW fields should not exceed 256 characters unless it is absolutely necessary.
20 – Missing time elements	For tables with start / stop times (e.g., MED_ADMIN), if there is an incomplete start or stop time within the source system (e.g., minutes or hours are missing), leave the field blank.

5. Table Summaries (Core Tables)

Table	Domain Description	Table Implementation Guidance
DEMOGRAPHIC	Demographics record the direct attributes of individual patients	<ul style="list-style-type: none"> 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, use the most recent value.
ENROLLMENT	Enrollment is a concept that defines a period of time during which a person is expected to have complete data capture. This concept is often insurance-based, but other methods of defining enrollment are possible.	<ul style="list-style-type: none"> For partners that do not have insurance-based enrollment information for some of their patients, other approaches can be used to identify periods during which complete medical capture is expected. Members with medical insurance coverage, with or without drug coverage, or should be included. If a patient has both medical and drug coverage, create the appropriate enrollment records for each. A break in insurance coverage 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 (i.e., if care is not observed, it likely did not happen). 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.
ENCOUNTER	Encounters are interactions between patients and providers within the context of healthcare delivery.	<ul style="list-style-type: none"> Each ENCOUNTERID will generally reflect a unique combination of PATID, ADMIT_DATE, PROVIDERID and ENC_TYPE. Every diagnosis and procedure recorded during the encounter should have a separate record in the DIAGNOSIS or PROCEDURES 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 for different encounter types on the same day, however, such as a physician appointment that leads to a hospitalization, would generally correspond to multiple encounters within the ENCOUNTER table. Rollback or voided transactions and other adjustments should be processed before populating this table. 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 with variation in how each field is populated. Do not include scheduled encounters. Partners should ensure that “administrative” encounters (e.g., e-mail, phone, documentation-only), are coded to the appropriate encounter type, which is typically “OA” for outpatient visits.
DIAGNOSIS	Diagnosis codes indicate the results of diagnostic processes	<ul style="list-style-type: none"> This table should capture all uniquely recorded diagnoses for all encounters, with the exception of problem list entries. If partners have access to multiple versions of each diagnosis within a given encounter (e.g., admitting,

Table	Domain Description	Table Implementation Guidance
	<p>and medical coding within healthcare delivery. Data in this table are expected to be from healthcare-mediated processes and reimbursement drivers.</p>	<p>interim, final), the preference is to prioritize final or discharge diagnoses. A value should be specified in DX_SOURCE to indicate the classification of the diagnosis.</p> <ul style="list-style-type: none"> • Diagnoses from problem lists will be captured in the CONDITION table. • If a patient has multiple diagnoses associated with one encounter, then there would be one record in this table for each diagnosis. • ENCOUNTERID should be populated for DIAGNOSIS and PROCEDURES. The definitions of the DIAGNOSIS and PROCEDURES tables are dependent upon a healthcare context; therefore, the encounter basis is necessary and the ENCOUNTERID, PROVIDERID, ENCOUNTER_TYPE, and ADMIT_DATE from the associated ENCOUNTER record should be included. While not desirable, a low percentage of orphan records is permissible to accommodate instances in which the associated ENCOUNTER details are missing from the source data. • Data in this table are expected to be from healthcare-mediated processes and reimbursement drivers, including technical/facility billing, professional billing and other data streams. Do not omit billing data unless it is unavailable from the source system or the partner is certain that the diagnoses loaded from the non-billing system (e.g., the EHR) represents completely the diagnosis data available from the billing system. Data from these different streams have different analytical utility so there is a benefit to including both if available. • Diagnoses are often only related to the treatment of the patient during the specific encounter. Chronic conditions that are not be pertinent to the treatment of a specific encounter, for example, would not be expected to be present. • If a local vocabulary is used, but <i>cannot</i> be mapped to a standard vocabulary such as ICD-9-CM, DX_TYPE should be populated as “Other” and the local value stored in DX. If the local value can be mapped to a standard vocabulary, follow the guidance around the population of Raw fields (General Guidance #1). • Partners should continue to populate ADMIT_DATE, even if they are populating DX_DATE. Analyses may leverage either date, or both. DX_DATE can be particularly useful for identifying diagnoses or conditions that might have developed over the course of a long inpatient encounter.
PROCEDURES	<p>Procedure codes indicate the discreet medical interventions and diagnostic testing, such as surgical procedures and lab orders, delivered within a healthcare context.</p>	<ul style="list-style-type: none"> • This table should capture all uniquely recorded procedures for all encounters, including office or evaluation and management visits, diagnostic testing, laboratory test orders, medication administrations, or other services rendered by a clinician. • If a patient has multiple procedures ordered during one encounter, then there would be one record in this table for each procedure. • ENCOUNTERID should be populated for DIAGNOSIS and PROCEDURES. The definitions of the DIAGNOSIS and PROCEDURES tables are dependent upon a healthcare context; therefore, the encounter basis is necessary and the ENCOUNTERID, PROVIDERID, ENCOUNTER_TYPE, and ADMIT_DATE from the associated ENCOUNTER record should be included. While not desirable, a low percentage of orphan records

Table	Domain Description	Table Implementation Guidance
		<p>is permissible to accommodate instances in which the associated ENCOUNTER details are missing from the source data.</p> <ul style="list-style-type: none"> • Data in this table are expected to be from healthcare-mediated processes and reimbursement drivers, including technical/facility billing, professional billing and other data streams. Do not omit billing data unless it is unavailable from the source system or the partner is certain that the procedures loaded from the non-billing system (e.g., the EHR) represents completely the procedure data available from the billing system • If a local vocabulary is used, but cannot be mapped to a standard vocabulary such as ICD-9-CM, PX_TYPE should be populated as “Other” and the local value stored in PX. If the local value can be mapped to a standard vocabulary, follow the guidance around the population of Raw fields (General Guidance #1). • Evidence of medications administered in outpatient settings should be present in the PROCEDURES table if that information is included with other billed/ordered PROCEDURES. • Evidence of inpatient administrations should be present in the PROCEDURES table if that information is included with other billed/ordered PROCEDURES. • DO NOT include records from medication administration sources (e.g., electronic medication administration records) in this table. • If possible to determine from the source data, only include procedures that have actually occurred. • Inclusion of laboratory orders – If possible, partners should include laboratory orders within the PROCEDURES table to support potential studies of appropriate laboratory monitoring. This includes those orders without a corresponding result in the LAB_RESULT_CM table. Do not include canceled orders.
VITAL	Vital signs (such as height, weight, and blood pressure) directly measure an individual’s current state of attributes.	<ul style="list-style-type: none"> • The deprecation of the VITAL table has been postponed. Partners should continue to populate VITAL with the relevant observations. Vital measures can also be stored in OBS_CLIN. • This table includes measurements recorded in both healthcare and non-healthcare settings. • The VITAL table contains one record per result/entry. Multiple measurements may exist in source data (for example, 3 blood pressure readings on the same day); in this case, <u>each</u> measurement would be a <u>separate</u> record. If multiple vitals are collected at the same time (e.g., height, weight and blood pressure recorded at the start of an encounter), it is permissible to store these values in a single record. This table should be populated with all available measures, with the possible exception(s) noted below. • If a partner has access to vital signs that are sourced from a device feed, they should make an assessment about data volume before including these measures, particularly if multiple readings per day are present for a large percentage of their population. Measures should not be averaged or aggregated. <ul style="list-style-type: none"> ○ For healthcare device data sources: If multiple readings are available and the volume of data is judged by the data partner to be too burdensome for inclusion, using the set of values that were recorded directly in the medical record is preferred over any algorithmic selection process.

Table	Domain Description	Table Implementation Guidance
		<ul style="list-style-type: none"> For personal device data sources: If multiple readings are available and the volume of data is judged by the data partner to be too high for inclusion, the project/study leadership should define a method for selecting individual measurements. in the ETL-ADD.
DISPENSING	Prescriptions filled through a community, mail-order or hospital pharmacy. Outpatient dispensing may not be directly captured within healthcare systems.	<ul style="list-style-type: none"> Each record represents an outpatient pharmacy dispensing. This domain is commonly available in claims data, but may not be available in many EHR data sources. Dispensing records are different from medication orders or prescribing records, data from medication administration activities, as well as the medication reconciliation of the active medication list. Administered medications should NOT be stored in this table. They should be stored in the MED_ADMIN table. Evidence of medications administered in outpatient settings, such as infusions given in medical practices, or those administered in an inpatient setting may be present in the PROCEDURES table if that level of detail is available in the source procedure data. Rollback transactions and other adjustments that are indicative of a dispensing being canceled or not picked up by the member should be processed (removed) before populating this table. This may be handled differently by Data Partners and may be affected by billing cycles. In the uncommon situation where one NDC is dispensed more than once for a given patient on a given day, it is acceptable to combine the values from the multiple dispensings for days supply and number of units
LAB_RESULT_CM	This table is used to store quantitative and qualitative measurements from blood and other body specimens.	<ul style="list-style-type: none"> Only records with actual lab results should be included in this table. If the result suggests that the test was run (e.g., result is "borderline" or "inconclusive") include it. But if the test is not resulted for any reason (specimen not sufficient, patient did not show), then do not include it. If lab results are stored using local or custom codes, partners should ensure that the assigned LOINC code has been validated by a subject matter expert or similar process If a LOINC code is available for a given result, the LAB_LOINC field should be populated. If a LOINC code is available for the <i>order</i>, that value can be used to populate the LAB_PX field. Note that one order can correspond to many different results. Each result should have its own record in the LAB_RESULT_CM table. If the same LOINC code is used to populate both the order and the result, partners should ensure that the LAB_LOINC field is populated. Inclusion of additional lab results - Partners should include <i>all</i> available laboratory results within their LAB_RESULT_CM table. If the result has a <i>validated</i> LOINC code, the LAB_LOINC field should be populated. Otherwise, the LAB_LOINC field should be blank. The RAW_LAB_NAME field can be used to keep track of the various lab results until the appropriate LOINC code is assigned. Lab results beyond the 11 originally included in the PCORnet CDM are being requested in order to establish a denominator of potentially available lab results. Over time, the number of unmapped results is expected to decrease. Results for labs performed as a service for outside institutions do not need to be included. Results from external vendors (e.g., LabCorp, Quest) should be included when available.

Table	Domain Description	Table Implementation Guidance
		<ul style="list-style-type: none"> • Clinical LOINC Concepts – Only include Laboratory LOINC concepts in this table. Do not include clinical LOINC concepts (e.g., EKG results). These records may be stored in the OBS_CLIN table. • Standing orders - Partners should populate the date fields to the best of their ability. For results that are tied to standing laboratory orders, even if LAB_ORDER_DATE reflects the date of the original standing order, SPECIMEN_DATE and/or RESULT_DATE would be expected to correspond to the time when the sample was collected/resulted. Analyses will take both dates into consideration. • Units of measure – A given LOINC code may have many acceptable units of measure. If the RESULT_UNIT field is not populated, it may not be possible to use a result analytically. • Verifying LOINC mappings – At most health systems, laboratory results are typically not associated with a LOINC code at the time they are generated but are assigned a code after the fact. In order to verify that the LOINC code has been appropriately assigned, the PCORnet Coordinating Center will verify that the metadata associated with the result, such as SPECIMEN_SOURCE and RESULT_UNIT, are valid options. Partners should ensure that these fields are populated. Do not derive these values based on metadata associated with the selected LOINC code. • Titers – titer results that are returned as ratios (e.g., 1:4) should be stored in the RESULT_QUAL field even if these ratios are not explicitly enumerated in the QUAL valueset. Beginning in CDMv6.1, data model conformance checks for RESULT_QUAL will include the enumerated list and text strings of [X]:[Y].
CONDITION	A condition represents a patient’s diagnosed and self-reported health conditions and diseases. The patient’s medical history and current state may both be represented.	<ul style="list-style-type: none"> • This table includes both healthcare and non-healthcare settings. • Rollback or voided transactions and other adjustments should be processed (removed) before populating this table. • These records should NOT be duplicated in the DIAGNOSIS table.
PRO_CM	This table is used to store responses to patient-reported outcome measures (PROs), surveys, and questionnaires. This table can be used to store item-level responses as well as the overall score for each measure.	<ul style="list-style-type: none"> • This version of the PRO_CM table is not optimized for representational efficiency. Certain values will be duplicated across records, and many fields will be blank for certain records. Over time, the structure of this table is expected to evolve as PCORnet better defines the analytical use of PROs across the network. Until then, this table has been defined to support the broadest range of possible use cases at the expense of representational efficiency. • The PRO_CM table can be used to store both individual item-level responses, as well as the overall score for the measure/instrument. Each item response will be stored in an individual record. Measure/instrument scores can be stored along with the item-level responses that are associated with that measure (where available). See the figure below for an example of how to populate this table. • If partners are populating PRO item responses or measure scores and are unsure of the PRO_ITEM_NAME, PRO_ITEM_LOINC, PRO_MEASURE_NAME and/or PRO_MEASURE_LOINC, they should populate

Table	Domain Description	Table Implementation Guidance
		<p>PRO_ITEM_FULLNAME and PRO_MEASURE_FULLNAME instead. These fields can be considered analogous to RAW fields.</p> <ul style="list-style-type: none"> • For the PRO_CM fields with variable field lengths, partners should choose an appropriate field length based on the characteristics of the data are loading into the table. As we use these tables analytically as part of PCORnet studies, we will determine whether it is more efficient to define specific field lengths. • If a patient completes a survey, but skips a question, create a record in the PRO_CM table as you would for other items in the survey (i.e., include the appropriate date/time fields and other relevant metadata). Then leave PRO_RESPONSE_TEXT and PRO_RESPONSE_NUM blank, as these fields are not required. Do not create empty records if the patient did not actually see the question. • The PRO_CM table can be used to store the results from questionnaires where the provider or caregiver are providing their interpretation or assessment of the patient’s status. Despite the name, it is not restricted solely to patient-reported outcomes. The table is designed to represent survey-type responses. General observations about patients, however, like pain scores recorded in an inpatient or surgical setting, should be stored in the OBS_CLIN table. See General Guidance #15 for additional details or contact the DRN OC with questions. • In general, patient-reported social determinants of health (e.g., food instability) would be expected to be found in this table. However, if partners who are participating in initiatives like the National COVID Cohort Collaborative have already loaded those records into OBS_GEN, it is acceptable for them to remain there. • For responses that can be represented as either a number or a text (e.g., a numeric value of 2 corresponds to “rarely” in a value set), the expectation is to store whatever is recorded in the source system. Both should be populated if present, but partners are not expected to derive the other if not. If values are represented based on a sequence number, store the actual value, not the sequence number.
PRESCRIBING	<p>Provider orders for medication dispensing and/or administration. These orders may take place in any setting, including the inpatient or outpatient basis.</p>	<ul style="list-style-type: none"> • If a medication cannot be mapped to RxNorm, it should still be present and RAW_RX_MED_NAME should be populated. • This table can be used to store all medication orders, regardless of encounter type (e.g., inpatient, outpatient, ED) and can include orders for medications that are to be dispensed as well as for those that are to be administered. • If including orders derived through natural language processing (NLP), make sure that RX_SOURCE has been populated for all records. • See Reference Table 4 for the ordering strategy for RxNorm Term Types. • Do not populate CDM fields with information derived from the RXCUI (e.g., RX_DOSE_ORDERED, RX_DOSE_FORM). Populate fields only if data are captured in the source system as a discrete value. • Populate records with the RXCUI as it existed at the time the order was entered, even if the RXCUI is no longer active. Do not attempt to update inactive RXCUIs with a more recent value. • Medications with approved formulations should have an RXCUI that can adequately represent all ingredients with a single code (e.g., SBD, SCD, MIN). For medication mixtures that lack RXCUIs that can represent all of

Table	Domain Description	Table Implementation Guidance
		<p>the component ingredients (e.g., IV mixtures prepared at an inpatient or compounding pharmacy), each individual medication from the order set should be included as a separate record with a unique PRESCRIBINGID. If partners wish to preserve the fact that the records belong to the same order, they do so by creating and populating a new <i>optional</i> ORDERID field. Medications with a 1:1 correspondence between the order and RXCUI could have the PRESCRIBINGID stored in the ORDERID. Orders with a 1:many RXCUI relationship would have different PRESCRIBINGIDs but the same ORDERID. Future versions of the CDM may formalize this guidance.</p>
PCORNET_TRIAL	Patients who are enrolled in PCORnet clinical trials and PCORnet studies.	<ul style="list-style-type: none"> Partners may use the PCORNET_TRIAL table to maintain mappings between PCORnet CDM PATIDs and external trial or study IDs. Partners wishing to use this table will need to register their TRIALID with the DRN OC. Please contact the DRN OC if you plan to utilize this table. TRIALIDs that start with “PT_” and “PS_” are reserved for PCORnet Trials and PCORnet Studies. Partners should refrain from using TRIALIDs that start with these characters. The TRIALID “ADPT” is also reserved. One patient participating in multiple trials or studies will have multiple records Each PCORnet trial or study will define its parameters for study or trial participation, and will provide study-specific instructions on how to populate the fields in this table. Patients who decline to participate in a trial or study or do not meet eligibility criteria should <u>not</u> be included in this table Patients who enroll in a trial or study but later withdraw <u>should</u> be included in this table, with a date in the TRIAL_WITHDRAW_DATE field. so that their withdrawal status and date are recognized and used to appropriately manage reporting of CDM data back to the coordinating center In most cases, trials will be expected to have a trial database that is separate from the CDM Randomization assignment is not included in this table due to the potential for unblinding. PATID is not generally appropriate for use as a PARTICIPANTID because it is not disambiguated across networks.
DEATH	Reported mortality information for patients.	<ul style="list-style-type: none"> One patient may potentially have multiple records in this table because their death may be reported by different sources. Deaths represented in the ENCOUNTER.DISCHARGE_DISPOSITION and ENCOUNTER.DISCHARGE_STATUS would generally be expected to be present in this table (see field-level guidance for DEATH.DEATH_SOURCE).
DEATH_CAUSE	The individual causes associated with a reported death.	<ul style="list-style-type: none"> When legacy data have conflicting reports, please make a local determination as to which to use. There is typically a 1-2 year lag in death registry data.
MED_ADMIN	Records of medications administered to patients by healthcare providers. These	<ul style="list-style-type: none"> If a medication cannot be mapped to RxNorm or NDC, it should still be present and RAW_MEDADMIN_NAME should be populated.

Table	Domain Description	Table Implementation Guidance
	<p>administrations may take place in any setting, including inpatient, outpatient or home health encounters.</p>	<ul style="list-style-type: none"> • Only include administrations that were actually delivered to the patient, if that level of specificity is available in the source system. • Patient-reported medication administrations are not within the scope of this table. • See Reference Table 4 for the ordering strategy for RxNorm Term Types. • Do not populate CDM fields with information derived from the RXCUI (e.g., MEDADMIN_DOSE_ADMIN). Populate fields only if data are captured in the source system as a discrete value. • Populate records with the RXCUI as it existed at the time the order was entered, even if the RXCUI is no longer active. Do not attempt to update inactive RXCUIs with a more recent value. • If a medication mixture contains multiple RXCUIs (e.g., inpatient mixture), each individual medication from the order set should be included as an individual record with a unique MEDADMINID. Each individual medication is expected to have a unique dose. • ENCOUNTERID is expected to be present for records in the MED_ADMIN table. • For administrations where the amount ordered is listed as a rate (e.g., infusions), if the DOSE/DOSE_UNIT values are specified as a rate, and those fields are stored discretely in your EHR, populate the relevant CDM fields. Assuming that START_DATE/START_TIME and STOP_DATE/STOP_TIME are also populated, it will be possible to compute the rate analytically. Otherwise, leave blank.
PROVIDER	<p>Data about the providers who are involved in the care processes documented in the CDM.</p>	<ul style="list-style-type: none"> • Include one record per provider. • When populating provider specialty, if multiple values are available, use the specialty believed to be primary.
OBS_CLIN	<p>Standardized qualitative and quantitative clinical observations about a patient, including vital signs. Some observations may also be represented in the VITAL table.</p>	<ul style="list-style-type: none"> • The OBS_CLIN table is intended to store standardized clinical observations that have been recorded about a patient. • Examples of the types of observations that can be stored in this table include pulmonary function test results (e.g., FEV1, FVC, FEV1/FVC), echocardiogram results (e.g., left ventricle ejection fraction) and vital signs (e.g., temperature). • Vital signs that can be stored in the VITAL table may also be stored in this table (e.g., systolic blood pressure), though VITAL should continue to be populated. The DRNOC maintains a list of LOINC codes that can be used to represent these observations. The Value Set Appendix contains a list of codes that can be used to map records that were previously stored in the VITAL table. For completeness purposes, it contains both preferred and discouraged codes. Partners should select the code(s) that best represent the data in their source systems but should not utilize a code that encodes more information than is available in their data (e.g., do not use a code that specifies a method if the method is unknown). In general, discouraged codes should not be used. • Decisions on what to include in this table and how to prioritize the population of those records are expected to be driven primarily by potential funding opportunities.

Table	Domain Description	Table Implementation Guidance
		<ul style="list-style-type: none"> • This table provides a generalized structure for storing observations and is not optimized for analytical efficiency. As elements from this table are used in studies and/or distributed queries, additional representations of those data elements (i.e., new table structures) may be required to better support those activities. • If partners are populating pain scores (not pain-related PRO surveys) captured in an inpatient or surgical setting, these values would be expected to be present in this table, not PRO_CM. • If an observation has a value set that includes an option of “not documented or assessed” (or similar), these values should be included in the CDM if they are present in the source system. Do not derive if an observation is missing. • If a partner has access to vital signs that are sourced from a device feed, they should make an assessment about data volume before including these measures, particularly if multiple readings per day are present for a large percentage of their population. Measures should not be averaged or aggregated. • For healthcare device data sources: If multiple readings are available and the volume of data is judged by the data partner to be too burdensome for inclusion, using the set of values that were recorded directly in the medical record is preferred over any algorithmic selection process. • For personal device data sources: If multiple readings are available and the volume of data is judged by the data partner to be too high for inclusion, the project/study leadership should define a method for selecting individual measurements and this logic should be documented in the ETL ADD. • Observations recorded together (e.g., diastolic and systolic blood pressure) should have the same date(s) and time(s).
OBS_GEN	Table to store everything else.	<ul style="list-style-type: none"> • Partners may use this table to store network- or study-specific data elements. • Records in this table are not expected to be used in queries distributed by the DRN OC. • This table provides a generalized structure for storing observations and is not optimized for analytical efficiency. As elements from this table are used in studies and/or distributed queries, additional representations of those data elements (i.e., new table structures) may be required to better support those activities.
HASH_TOKEN	Encrypted hash tokens that are used to match patient records across DataMarts using privacy-preserving record linkage methods.	<ul style="list-style-type: none"> • Every patient in the DEMOGRAPHIC table is expected to have one record in the HASH_TOKEN table for each TOKEN_ENCRYPTION_KEY. • Tokens are generated from personally-identifiable information (PII). This information can be stored in each partner’s PRIVATE_DEMOGRAPHIC table and PRIVATE_ADDRESS_HISTORY table. The PII is used as input to the Datavant DeId/tokenization module. Tokens should not be placed into the CDM until they have been transformed into Site-PCORnet transit tokens using the Datavant Link/transform-tokens module. • If PII fields are populated with dummy values by default (e.g., 999-99-9999 for phone number or SSN), these values should be removed before running the Datavant DeID/tokenization module. If PII is not available within the source system or if local restrictions prevent its use, leave the input field blank. • Tokens are generated based on data availability. If input data is not present for a given token strategy (e.g., combination of PII elements), no token will be generated and an error code will be produced instead. These

Table	Domain Description	Table Implementation Guidance
		<p>token error codes should be loaded into the HASH_TOKEN table (i.e., there should not be any null values). Do not suppress the error codes in the output of the Datavant software.</p> <ul style="list-style-type: none"> • Each successfully generated token has a fixed length of 44 characters. Do not enforce a 44-character constraint, however, to accommodate the error codes generated in the case of tokenization failure. • Tokens should be generated as part of every refresh. Partners can choose to generate tokens for all patients, or only for those patients who were added between refreshes or had updates to their PII. • Select tokens generated using the Datavant DeID/tokenization module are certified as de-identified data via the HIPAA Expert Determination method in accordance with the HIPAA Privacy Rule (45 CFR parts 160 and 164). All tokens in the HASH_TOKEN table satisfy this criteria and are controlled via the Datavant DeID PCORnet configuration settings within the Datavant software. • Additional token strategies are available and can be implemented as needed on a per-study basis based on the study-specific data dictionary. • See the Supplemental Guide on Privacy-Preserving Record Linkage for additional implementation details and guidance (separate document). • Partners that wish to include tokens encrypted using multiple keys for internal purposes should add a TOKEN_ENCRYPTION_KEY field and store the key name there. When responding to PCORnet queries/data curation, the table should be filtered to only include tokens used in PCORnet linkages. The name of this key should also be included in the TOKEN_ENCRYPTION_KEY field in HARVEST. • Partners should include the TOKEN_ENCRYPTION_KEY string that is output from the Datavant Link software. Please ensure that this table includes tokens encrypted using the PCORnet key (includes “pcornet” or similar in the output string). If partners are encrypting their tokens using multiple keys, it is acceptable to include those additional records in this table.
LDS_ADDRESS_HISTORY	Longitudinal record of a patient’s address that adheres to the requirements of a Limited Data Set.	<ul style="list-style-type: none"> • Expect multiple records per individual • This table is currently limited to addresses in the United States. • Partners can limit records in this table to validated addresses if known. • If partners have difficulty constructing a longitudinal address history for patients within their DataMart, they should prioritize populating the current address for each patient.
IMMUNIZATION	Records of immunizations that have been delivered within the health system as well as reports of those administered elsewhere.	<ul style="list-style-type: none"> • Do not include study vaccines.
HARVEST	Attributes associated with the specific PCORnet datamart	<ul style="list-style-type: none"> • If partners need to impute date values, whether for a portion of the date (e.g., month) or the entire string, a value of “02” should be chosen for the relevant DATE_MGMT field(s).

Table	Domain Description	Table Implementation Guidance
	implementation, including data refreshes.	<ul style="list-style-type: none"> • If partners must impute the entire date for a field, this should only be done for those dates that are required. Optional fields should be left blank in these situations. • All date obfuscation and imputation strategies must be thoroughly documented in the Extract, Transform and Load (ETL) Annotated Data Dictionary (ADD). • Partners should refrain from obfuscating dates within the CDM, with the possible exception of BIRTH_DATE (see General Guidance #2). Future versions of the CDM may remove options “03” and “04” from the value set of the DATE_MGMT fields.
LAB_HISTORY	Table for storing historical information about units of measure and reference ranges for laboratory test results.	<ul style="list-style-type: none"> • This table is intended to serve as a resource for partners as they develop their extract-transform-load (ETL) procedures to populate LAB_RESULT_CM. It is designed to store details related to units of measure and normal ranges for laboratory results that do not include such values at the record level. Partners can use this table as a reference during their ETL to look up and populate the relevant fields in LAB_RESULT_CM for those records. • It is expected that partners will be able to find information on units of measure and normal range via reference material maintained by their clinical labs (electronic documents or online catalogs). It is not necessary to derive this metadata from individual lab results. DO NOT DERIVE THESE DATA FROM LOINC RESOURCES. • While values for this table may need to be entered manually, a relatively small number of tests (~150) typically cover the vast majority of testing volume (>85%), which should minimize any data entry burden. • Partners may not need to populate this table, particularly if they are already able to meet the network lab data quality metrics. • Partners do not need to create records in this table for results that come from external reference labs (e.g., Quest, LabCorp). The DRN OC can provide this information on request. • The DRN OC may reference this table if labs are needed as part of a particular study or analysis and the units and/or reference range is missing from the result(s). Partners may be asked to populate records in this table for those corresponding tests. • Every record in this table should be unique.

6. Core CDM Table Specifications

6.1. Table: DEMOGRAPHIC

DEMOGRAPHIC Domain Description:
Demographics record the direct attributes of individual patients.

Relational Integrity:

The DEMOGRAPHIC table contains one record per patient.

Primary Key: PATID

Constraints:

PATID (unique; required, not null)

DEMOGRAPHIC Table Implementation Guidance

Guidance

- 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, use the most recent value.

DEMOGRAPHIC Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-Level Implementation Guidance</i>
PATID	RDBMS Text(x)	SAS Char(x)	.	<p>Arbitrary person-level identifier used to link across tables.</p> <p>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.</p> <p>The PATID must be unique within each PCORnet data mart. Creating a unique identifier within a network would be beneficial and acceptable. The PATID is not the basis for linkages across data partners.</p>	MSCDM v4.0	
BIRTH_DATE	RDBMS Date	SAS Date (Numeric)	.	Date of birth.	MSCDM v4.0	
BIRTH_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)	.	Time of birth.	PCORnet Source of time format: ISO 8601	
SEX	RDBMS Text(2)	SAS Char(2)	A=Ambiguous F=Female M=Male NI=No information UN=Unknown OT=Other	Sex assigned at birth.	<p>MSCDM v4.0 with modified field size and value set</p> <p>Source: Administrative Sex (HL7)</p> <p>http://phinvads.cdc.gov/vads/ViewValueSet.action?id=06D34BBC-617F-DD11-B38D-00188B398520</p>	<p>The “Ambiguous” category may be used for individuals who are physically undifferentiated from birth. The “Other” category may be used for individuals who are undergoing gender re-assignment.</p> <p>If a value of “X” is recorded in the source system, map to “OT”.</p>

DEMOGRAPHIC Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-Level Implementation Guidance</i>
SEXUAL_ORIENTATI ON	RDBMS Text(2)	SAS Char(2)	AS=Asexual BI=Bisexual GA=Gay LE=Lesbian QU=Queer HO=Lesbian, Gay or Homosexual QS=Questioning ST=Straight SE=Something else MU=Multiple sexual orientations DC=Decline to answer NI=No information UN=Unknown OT=Other	Sexual orientation.	Source: Health IT Certification Criteria, 2015 Base Edition, modified with expert advisory within PCORnet https://www.federalregister.gov/documents/2015/10/16/2015-25597/2015-edition-health-information-technology-health-it-certification-criteria-2015-edition-base	<ul style="list-style-type: none"> Sites should map to the most granular value supported by their data. The entry for “Lesbian, Gay, or Homosexual” was added for those partners who capture sexual orientation based on the ONC Meaningful Use value set.
GENDER_IDENTITY	RDBMS Text(2)	SAS Char(2)	M=Man F=Woman TM=Transgender male/Trans man/Female-to-male TF=Transgender female/Trans woman/Male-to- female GQ=Genderqueer/N on-binary SE=Something else MU=Multiple gender categories DC=Decline to answer NI=No information UN=Unknown OT=Other	Current gender identity.	Source: Health IT Certification Criteria, 2015 Base Edition, modified with expert advisory within PCORnet https://www.federalregister.gov/documents/2015/10/16/2015-25597/2015-edition-health-information-technology-health-it-certification-criteria-2015-edition-base	Use Genderqueer (“GQ”) for patients who report a non-binary gender identify.

DEMOGRAPHIC Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-Level Implementation Guidance</i>
HISPANIC	RDBMS Text(2)	SAS Char(2)	Y=Yes N=No R=Refuse to answer 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 v4.0 with modified field size and value set Compatible with “OMB Hispanic Ethnicity” (Hispanic or Latino, Not Hispanic or Latino)	Populating RACE and HISPANIC if race and ethnicity are not captured separately within the source system (e.g., “Hispanic or Latino” is included as a selection under Race) - for patients with a known race (e.g., Race is something other than “Hispanic or Latino”, partners should set HISPANIC to "OT" and RACE to the appropriate race code. For patients who are listed as having a race of “Hispanic,” partners should set HISPANIC to "Y" and RACE to "OT". In this situation, the combined race/ethnicity field is treated as known field capturing values for both race and ethnicity, which is why the preference is to use “OT” instead of “NI”.
RACE	RDBMS Text(2)	SAS Char(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. 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.	MSCDM v4.0 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	

DEMOGRAPHIC Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-Level Implementation Guidance</i>
BIOBANK_FLAG	RDBMS Text(1)	SAS Char(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 blood, urine, or tissue (eg, skin cells, organ tissues). 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”.	PCORnet	This field is a derived attribute and is not expected to be an explicit data field within a source system
PAT_PREF_LANGUAG E_SPOKEN	RDBMS Text(3)	SAS Char(3)	See Value Set Reference File for a list of acceptable values.	Preferred spoken language of communication as expressed by the patient.	ISO639-2	<ul style="list-style-type: none"> This information may be documented in the EHR or an external registry. Do not impute or derive if not expressly defined in the source system. Analytically, will assume that “NI” corresponds to a preferred language of English. Use the value of “ZHO” (Chinese) for both Mandarin and Cantonese, and place the specific value in the RAW field. Within the ISO639-2 value set, there is no distinction between the two. https://www.loc.gov/standards/iso639-2/faq.html#24 Use “OT” for American Sign Language
RAW_SEX	RDBMS Text(x)	SAS Char(x)	.	Field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_SEXUAL_ORIENTATI ON	RDBMS Text(x)	SAS Char(x)	.	Field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet	

DEMOGRAPHIC Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-Level Implementation Guidance</i>
RAW_GENDER_IDENTITY	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_HISPANIC	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_RACE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_PAT_PREF_LANGUAGE_SPOKEN	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	

6.2. Table: ENROLLMENT

ENROLLMENT Domain Description:

Enrollment is a concept that defines a period of time during which a person is expected to have complete data capture. This concept is often insurance-based, but other methods of defining enrollment are possible.

Relational Integrity:

The ENROLLMENT table contains one record per unique combination of PATID, ENR_START_DATE, and ENR_BASIS.

Please note: Each form of coverage (the ENR_BASIS) would have a separate record; for example, if a patient has both medical coverage and drug coverage, these would be 2 separate records, potentially with different enrollment dates for each record.

Composite Primary Key: PATID, ENR_START_DATE, ENR_BASIS

Foreign Key:

ENROLLMENT.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

Constraints:

PATID + ENR_START_DATE + ENR_BASIS (unique)

PATID (required, not null)

ENR_START_DATE (required, not null)

ENR_BASIS (required, not null)

ENROLLMENT Table Implementation Guidance

Guidance

- For partners that do not have insurance-based enrollment information for some of their patients, other approaches can be used to identify periods during which complete medical capture is expected.
- Members with medical insurance coverage, with or without drug coverage, or should be included. If a patient has both medical and drug coverage, create the appropriate enrollment records for each.
- A break in insurance coverage 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 (i.e., if care is not observed, it likely did not happen). 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.

ENROLLMENT Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-Level Implementation Guidance</i>
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier used to link across tables.	MSCDM v4.0	
ENR_START_DATE	RDBMS Date	SAS Date (Numeric)	.	Date of the beginning of the enrollment period. If the exact date is unknown, use the first day of the month.	MSCDM v4.0 with modified field name	<ul style="list-style-type: none"> For implementation of the CDM, a long span of longitudinal data is desirable. However, an enrollment record is intended to represent the dates for which there is complete capture of all medically-attended events, so the data partner's knowledge of the validity and usability of the data should inform their choice of enrollment start date, especially for historical data more than a decade old. If ENR_BASIS is encounter-based ("E"), the enrollment start date should not be earlier than the earliest encounter date in the datamart for that patient. If ENR_BASIS is based on insurance coverage ("I"), then the enrollment start date <i>may</i> occur before the earliest encounter date in the datamart for that patient.
ENR_END_DATE	RDBMS Date	SAS Date (Numeric)	.	Date of the end of the enrollment period. If the exact date is unknown, use the last day of the month.	MSCDM v4.0 with modified field name	

ENROLLMENT Table Specification

Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-Level Implementation Guidance
CHART	RDBMS Text(1)	SAS Char(1)	Y=Yes N=No	<p>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.</p> <p>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.</p>	MSCDM v4.0	<ul style="list-style-type: none"> 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 patient. This field is a derived attribute and is not expected to be an explicit data field within a source system
ENR_BASIS	RDBMS Text(1)	SAS Char(1)	I=Medical insurance coverage D=Outpatient prescription drug coverage G=Geography A=Algorithmic E=Encounter-based	<p>ENR_BASIS is a property of the time period defined. A patient can have multiple entries in the table.</p> <p>Details of categorical definitions:</p> <p>Medical insurance coverage: The start and stop dates are based upon enrollment where the health plan has any responsibility for covering medical care for the member during this enrollment period (i.e., if you expect to observe medical care provided to this member during the enrollment period).</p> <p>Outpatient prescription drug coverage: The start and stop dates are based on enrollment where the health plan has any responsibility for covering outpatient prescription drugs for the member during this enrollment period (i.e., if you expect to observe outpatient pharmacy dispensings for this member during this enrollment period). (New value set item added in v3.1.)</p> <p>Geography: An assertion of complete data capture between the start and end dates based upon geographic characteristics, such as regional isolation.</p> <p>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.</p> <p>Encounter-based: The start and stop dates are populated from the earliest-observed encounter and latest-observed encounter.</p> <p>Field definition and value sets modified in v3.1 to include drug coverage.</p>	PCORnet Based upon the HMORN VDW and Sentinel CDM v6.0	<ul style="list-style-type: none"> When an algorithmic method is used to determine the ENR_BASIS, the exact details should be described in the ETL ADD. This field is a derived attribute and is not expected to be an explicit data field within a source system

6.3. Table: ENCOUNTER

ENCOUNTER Domain Description:

Encounters are interactions between patients and providers within the context of healthcare delivery.

Relational Integrity:

The ENCOUNTER table contains one record per unique encounter.

Primary Key: ENCOUNTERID

Foreign Key:

ENCOUNTER.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

ENCOUNTER.PROVIDERID is a foreign key to PROVIDER.PROVIDERID (many-to-one relationship)

Constraints:

ENCOUNTERID (unique; required, not null)

PATID (required, not null)

ADMIT_DATE (required, not null)

ENC_TYPE (required, not null)

ENCOUNTER Table Implementation Guidance

Guidance

- Each ENCOUNTERID will generally reflect a unique combination of PATID, ADMIT_DATE, PROVIDERID and ENC_TYPE.
- Every diagnosis and procedure recorded during the encounter should have a separate record in the DIAGNOSIS or PROCEDURES 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 for different encounter types on the same day, however, such as a physician appointment that leads to a hospitalization, would generally correspond to multiple encounters within the ENCOUNTER table.
- Rollback or voided transactions and other adjustments should be processed before populating this table.
- 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 with variation in how each field is populated.
- Do not include scheduled encounters.
- Partners should ensure that “administrative” encounters (e.g., e-mail, phone, documentation-only), are coded to the appropriate encounter type, which is typically “OA” for outpatient visits.

ENCOUNTER Table Specification

Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
ENCOUNTERID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary encounter-level identifier. Used to link across tables, including the ENCOUNTER, DIAGNOSIS, and PROCEDURES tables.	MSCDM v4.0	
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier used to link across tables.	MSCDM v4.0	All PATIDs in this table must be present in the DEMOGRAPHIC table.
ADMIT_DATE	RDBMS Date	SAS Date (Numeric)	.	Encounter or admission date.	MSCDM v4.0 with modified field name	
ADMIT_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)	.	Encounter or admission time.	PCORnet Source of time format: ISO 8601	

ENCOUNTER Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
DISCHARGE_DATE	RDBMS Date	SAS Date (Numeric)	.	Discharge date.	MSCDM v4.0 with modified field name	Should be populated for all Inpatient Hospital Stay (IP), Non-Acute Institutional Stay (IS) encounter types and ED-to-Inpatient (EI) encounter types. May be populated for Emergency Department (ED) encounter types. Should be missing for ambulatory visit (AV or OA) encounter types, though it may be present for Observation Stays.
DISCHARGE_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)	.	Discharge time.	PCORnet Source of time format: ISO 8601	
PROVIDERID	RDBMS Text(x)	SAS Char(x)	.	Code for the provider who is most responsible for this encounter. As with the PATID, the provider code is a pseudoidentifier with a consistent crosswalk to the real identifier.	MSCDM v4.0	<ul style="list-style-type: none"> PROVIDERID generally refers to the person most responsible for providing care during the encounter, though it may also correspond to a device (e.g., MRI) for certain procedure-only encounters. If populated, PROVIDERID must be present in the PROVIDER table.
FACILITY_LOCATION	RDBMS Text(5)	SAS Char(5)	.	Geographic location (5-digit zip code).	MSCDM v4.0 with modified field type	Updated in CDM v5.0 to support 5-digit zip codes.

ENC_TYPE	RDBMS Text(2)	SAS Char(2)	AV=Ambulatory Visit ED=Emergency Department EI=Emergency Department Admit to Inpatient Hospital Stay (permissible substitution) IP=Inpatient Hospital Stay IS=Non-Acute Institutional Stay OS=Observation Stay IC=Institutional Professional Consult (permissible substitution) TH=Telehealth 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 facility 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. Does not include observation stays, where known. Emergency Department Admit to Inpatient Hospital Stay: Permissible substitution for preferred state of separate ED and IP records. Only for use with data sources where the individual records for ED and IP cannot be distinguished. 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. Does not include observation stays, where known. Observation Stay: "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 . Institutional Professional Consult: Permissible substitution when services provided by a medical professional cannot be combined with the given encounter record, such as a specialist consult in an inpatient setting; this situation can be common with claims data sources. This includes physician consults for patients during inpatient encounters that are not directly related to the cause of the admission (e.g. a ophthalmologist consult for a patient with diabetic ketoacidosis) (guidance updated in v4.0). Non-Acute Institutional Stay: Includes inpatient hospice, skilled nursing facility (SNF), inpatient rehab center, nursing home, residential, overnight non-hospital dialysis, and other non-hospital overnight stays. Telehealth: Includes telemedicine or virtual visits, which can be conducted via video, phone or other means. Other Ambulatory Visit: Includes other non-overnight AV encounters such as hospice visits, rehab visits home health visits, skilled nursing 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.	MSCDM v4.0 with modified value set	<ul style="list-style-type: none"> • Observation stays– If partners are able to identify observation stays within their data, these encounters should be labeled “OS.” Typical observation stays last 24-48 hours. If partners find that they have observation stays that last significantly longer (e.g., weeks), they should notify the DRN OC. This should also be documented in the ETL_ADD. • Same-day surgery, OT/PT, and provider office visits for treatment/testing should be labeled as “AV.” • For the situation where an Emergency Department (ED) encounter leads to a hospital admission <ul style="list-style-type: none"> ○ The optimal, preferred state is to have one record for the ED (ENC_TYPE=ED), and a separate record for the hospital admission (ENC_TYPE=IP) ○ However, this separation does not always exist in source data records. If the source system combines the ED and IP basis into one concept, a permissible substitution is to use ENC_TYPE=EI ○ Never merge separate ED and IP records together. • Services rendered in an inpatient setting that cannot be combined with the facility encounter – Inpatient (IP) and ED to Inpatient (EI) encounter types should be limited to encounters which include the facility component of the admission since these data are required to fully populate the expected fields (e.g. Discharge Date, Admitting Source, Discharge Disposition, Discharge Status). If a partner has data for professional services that occur in an inpatient care setting that cannot be combined with the associated facility encounter, the partner should map
----------	------------------	-------------	--	---	------------------------------------	---

ENCOUNTER Table Specification

Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
						<p>these services to "OT." and document this in their ETL ADD.</p> <ul style="list-style-type: none"> Generally, a reimbursement basis will determine the source system classification, instead of physical location. For example, a patient may occupy a hospital bed during an observation that is not classified as an inpatient hospital stay. Please note that stand-alone urgent care facilities are usually not established as Emergency Departments. Routine telephone and e-mail consultations / contacts should be classified as Other Ambulatory encounters, not Telehealth. Telehealth visits are expected to be associated with diagnosis and/or procedure codes.
FACILITYID	RDBMS Text(x)	SAS Char(x)	.	<p>Arbitrary local facility code that identifies the hospital or clinic. Used for chart abstraction and validation.</p> <p>FACILITYID can be a true identifier, or a pseudoidentifier with a consistent crosswalk to the true identifier retained by the source data partner.</p>	MSCDM v4.0 with modified field name	<ul style="list-style-type: none"> If populating FACILITY_TYPE, ensure that multiple FACILITY_TYPE values are not used to describe the same FACILITYID. If a facility exists that operates both inpatient and outpatient units and is described in the source system by the same facility id, a potential solution is to append the source id with a suffix to create "sub facilities" that can be used to distinguish locations with different levels of service.

ENCOUNTER Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
DISCHARGE_DISPOSITION	RDBMS Text(2)	SAS Char(2)	A=Discharged alive E=Expired NI=No information UN=Unknown OT=Other	Vital status at discharge.	MSCDM v4.0 with modified field size and value set	Should be populated for Inpatient Hospital Stay (IP), Non-Acute Institutional Stay (IS) and ED-to-Inpatient (EI) encounter types. May be populated for Emergency Department (ED) encounter types. Should be missing for ambulatory visit (AV or OA) encounter types, though it may be present for Observation Stays.
DISCHARGE_STATUS	RDBMS Text(2)	SAS Char(2)	AF=Adult Foster Home AL=Assisted Living Facility AM=Against Medical Advice 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	Discharge status.	MSCDM v4.0 with modified value set	Should be populated for Inpatient Hospital Stay (IP), Non-Acute Institutional Stay (IS) and ED-to-Inpatient (EI) encounter types. May be populated for Emergency Department (ED) encounter types. Should be missing for ambulatory visit (AV or OA) encounter types, though it may be present for Observation Stays.

ENCOUNTER Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
DRG	RDBMS Text(3)	SAS Char(3)	.	3-digit Diagnosis Related Group (DRG). 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 v4.0	<ul style="list-style-type: none"> Should be populated for Inpatient Hospital Stay (IP), Non-Acute Institutional Stay (IS) and ED-to-Inpatient (EI) encounter types. May be populated for Emergency Department (ED) encounter types. Should be missing for ambulatory visit (AV or OA) encounter types, though it may be present for Observation Stays. Use leading zeroes for codes less than 100. For records with multiple DRGs assigned, choose the most appropriate one based on the source data.
DRG_TYPE	RDBMS Text(2)	SAS Char(2)	01=CMS-DRG (old system) 02=MS-DRG (current system) NI=No information UN=Unknown OT=Other	DRG code version.	MSCDM v4.0 with modified field size and value set	<ul style="list-style-type: none"> MS-DRG (current system) began on 10/1/2007. Should be populated for Inpatient Hospital Stay (IP), Non-Acute Institutional Stay (IS) and ED-to-Inpatient (EI) encounter types. May be populated for Emergency Department (ED) encounter types. Should be missing for ambulatory visit (AV or OA) encounter types, though it may be present for Observation Stays. This field is a derived attribute and is not expected to be an explicit data field within a source system

ENCOUNTER Table Specification

Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
ADMITTING_SOURCE	RDBMS Text(2)	SAS Char(2)	AF=Adult Foster Home AL=Assisted Living Facility AV=Ambulatory Visit ED=Emergency Department ES=Emergency Medical Service 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 IH=Intra-hospital NI=No information UN=Unknown OT=Other	Admitting source.	MSCDM v4.0 with modified value set	<ul style="list-style-type: none"> Should be populated for Inpatient Hospital Stay (IP), Non-Acute Institutional Stay (IS) and ED-to-Inpatient (EI) encounter types. May be populated for Emergency Department (ED) encounter types. Should be missing for ambulatory visit (AV or OA) encounter types, though it may be present for Observation Stays. When a patient is discharged from one distinct entity in a hospital and admitted to another, <i>resulting in a separate claim</i>, use "IH" Use "ES" for patient arrival by ambulance or helicopter (emergency medical services) who are not transferred from AL, HS, IP, NH, SN or IH sources. Arrival by non-emergency chair-car, van or hired service should not be coded as ES. In general, these patients will be admitted to the ED for assessment and stabilization.

ENCOUNTER Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PAYER_TYPE_PRIMARY	RDBMS Text(5)	SAS Char(5)	See Value Set Reference File for a list of acceptable values.	Categorization of payer type for primary payer associated with the encounter	PHDSC Source of Payment Typology	<ul style="list-style-type: none"> Do not derive if not already assigned through a validated process (e.g., by hospital billing department) Map to the most granular value the source data will support. Additional information can be found in the Source of Payment Typology User Guide, located here: https://www.nahdo.org/sopt Even if payer type is not specified, populating RAW_PAYER_TYPE_PRIMARY and RAW_PAYER_NAME_PRIMARY will allow a value to be determined through PCORnet-wide or study-specific process

ENCOUNTER Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PAYER_TYPE_SECONDARY	RDBMS Text(5)	SAS Char(5)	See Value Set Reference File for a list of acceptable values.	Categorization of payer type for secondary payer associated with the encounter	PHDSC Source of Payment Typology	<ul style="list-style-type: none"> Do not derive if not already assigned through a validated process (e.g., by hospital billing department) Map to the most granular value the source data will support. Additional information can be found in the Source of Payment Typology User Guide, located here: https://www.nahdo.org/sopt Even if payer type is not specified, populating RAW_PAYER_TYPE_SECONDARY and RAW_PAYER_NAME_SECONDARY will allow a value to be determined through PCORnet-wide or study-specific process

ENCOUNTER Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
FACILITY_TYPE	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Description of the facility where the encounter occurred.	PCORnet	<ul style="list-style-type: none"> Do not assign more than one FACILITY_TYPE to a single FACILITYID If a facility exists that operates both inpatient and outpatient units and is described in the source system by the same facility id, a potential solution is to append the source id with a suffix to create “sub facilities” that can be used to distinguish locations with different levels of service. For office visits at an academic medical center, select the most appropriate hospital outpatient clinic value. Pediatric specialty clinics should map the relevant specialty clinic, if available. The PEDIATRIC facility types would best be applied to a General Pediatrics clinic. A draft mapping between the CMS Place of Service value set and FACILITY_TYPE can be found here: https://github.com/CDMFORUM/CDM-GUIDANCE/issues/67
RAW_SITEID	RDBMS Text(x)	SAS Char(x)	.	Field for locally-defined identifier intended for local use; for example, where a network may have multiple sites contributing to a central data repository. This attribute may be sensitive in certain contexts; the intent is for internal network use only, and not to enable site quality comparisons.	PCORnet	
RAW_ENC_TYPE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	

ENCOUNTER Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RAW_DISCHARGE_DISPOSITION	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_DISCHARGE_STATUS	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_DRG_TYPE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_ADMITTING_SOURCE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_FACILITY_TYPE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_PAYER_TYPE_PRIMARY	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_PAYER_NAME_PRIMARY	RDBMS Text(x)	SAS Char(x)	.	Primary payer name as denoted in the source system. Used to derive PAYER_TYPE_PRIMARY if validated process does not exist.	PCORnet	<ul style="list-style-type: none"> Name of secondary payer associated with the encounter. Partners should only populate if local governance allows it.
RAW_PAYER_ID_PRIMARY	RDBMS Text(x)	SAS Char(x)	.	Primary PAYER identifier as denoted in the source system. Used to derive PAYER_TYPE_PRIMARY if validated process does not exist.	PCORnet	<ul style="list-style-type: none"> Identifier associated with the primary payer. Partners should only populate if local governance allows it.
RAW_PAYER_TYPE_SECONDARY	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_PAYER_NAME_SECONDARY	RDBMS Text(x)	SAS Char(x)	.	Secondary payer name as denoted in the source system. Used to derive PAYER_TYPE_SECONDARY if validated process does not exist.	PCORnet	<ul style="list-style-type: none"> Name of secondary payer associated with the encounter. Partners should only populate if local governance allows it.
RAW_PAYER_ID_SECONDARY	RDBMS Text(x)	SAS Char(x)	.	Secondary PAYER identifier as denoted in the source system. Used to derive PAYER_TYPE_SECONDARY if validated process does not exist.	PCORnet	<ul style="list-style-type: none"> Identifier associated with the secondary payer. Partners should only populate if local governance allows it.

6.4. Table: DIAGNOSIS

DIAGNOSIS Domain Description:

Diagnosis codes indicate the results of diagnostic processes and medical coding within healthcare delivery.

Data in this table are expected to be from healthcare-mediated processes and reimbursement drivers.

Relational Integrity:

The DIAGNOSIS table contains one record per DIAGNOISID.

Primary Key: DIAGNOSISID

Foreign Keys:

DIAGNOSIS.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

DIAGNOSIS.ENCOUNTERID is a foreign key to ENCOUNTER.ENCOUNTERID (many-to-one relationship)

DIAGNOSIS.PROVIDERID is a foreign key to PROVIDER.PROVIDERID (many-to-one relationship)

Constraints:

DIAGNOSISID (unique; required, not null)

PATID (required, not null)

DX (required, not null)

DX_TYPE (required, not null)

DX_SOURCE (required, not null)

DIAGNOSIS Table Implementation Guidance

Guidance

- This table should capture all uniquely recorded diagnoses for all encounters, with the exception of problem list entries. If partners have access to multiple versions of each diagnosis within a given encounter (e.g., admitting, interim, final), the preference is to prioritize final or discharge diagnoses. A value should be specified in DX_SOURCE to indicate the classification of the diagnosis.
- Diagnoses from problem lists will be captured in the CONDITION table.
- If a patient has multiple diagnoses associated with one encounter, then there would be one record in this table for each diagnosis.
- ENCOUNTERID should be populated for DIAGNOSIS and PROCEDURES. The definitions of the DIAGNOSIS and PROCEDURES tables are dependent upon a healthcare context; therefore, the encounter basis is necessary and the ENCOUNTERID, PROVIDERID, ENCOUNTER_TYPE, and ADMIT_DATE from the associated ENCOUNTER record should be included. While not desirable, a low percentage of orphan records is permissible to accommodate instances in which the associated ENCOUNTER details are missing from the source data.
- Data in this table are expected to be from healthcare-mediated processes and reimbursement drivers, including technical/facility billing, professional billing and other data streams. **Do not omit** billing data unless it is unavailable from the source system or the partner is certain that the diagnoses loaded from the non-billing system (e.g., the EHR) represents completely the diagnosis data available from the billing system. Data from these different streams have different analytical utility so there is a benefit to including both if available.
- Diagnoses are often only related to the **treatment** of the patient during the specific encounter. Chronic conditions that are not be pertinent to the treatment of a specific encounter, for example, would not be expected to be present.
- If a local vocabulary is used, but *cannot* be mapped to a standard vocabulary such as ICD-9-CM, DX_TYPE should be populated as “Other” and the local value stored in DX. If the local value can be mapped to a standard vocabulary, follow the guidance around the population of Raw fields (General Guidance #1).
- Partners should continue to populate ADMIT_DATE, even if they are populating DX_DATE. Analyses may leverage either date, or both. DX_DATE can be particularly useful for identifying diagnoses or conditions that might have developed over the course of a long inpatient encounter.

DIAGNOSIS Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
DIAGNOSISID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique record.	PCORnet	
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier. Used to link across tables.	MSCDM v4.0	All PATIDs in this table must be present in the DEMOGRAPHIC table.
ENCOUNTERID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary encounter-level identifier. Used to link across tables.	MSCDM v4.0	All ENCOUNTERIDs in this table must be present in the ENCOUNTER table.
ENC_TYPE	RDBMS Text(2)	SAS Char(2)	AV=Ambulatory Visit ED=Emergency Department EI=Emergency Department Admit to Inpatient Hospital Stay (permissible substitution) IP=Inpatient Hospital Stay IS=Non-Acute Institutional Stay OS=Observation Stay IC=Institutional Professional Consult (permissible substitution) TH=Telehealth 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 v4.0 with modified value set	Should be non-null for all records replicated from ENCOUNTER table (guidance added in v4.0).
ADMIT_DATE	RDBMS Date	SAS Date (Numeric)	.	Please note: This is a field replicated from the ENCOUNTER table. See the ENCOUNTER table for definitions.	MSCDM v4.0 with modified field name	

DIAGNOSIS Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PROVIDERID	RDBMS Text(x)	SAS Char(x)	.	Identifier associated with the provider most responsible for the diagnosis.	MSCDM v4.0	<ul style="list-style-type: none"> The PROVIDERID from the ENCOUNTER can be used if provider assigning the diagnosis is unknown. Use the ID of the attending provider if the diagnosis is assigned by a resident/intern. All PROVIDERIDs must be present in the PROVIDER table.
DX	RDBMS Text(18)	SAS Char(18)	.	<p>Diagnosis code.</p> <p>Some codes will contain leading zeroes, and different levels of decimal precision may also be present. This field is a character field, not numeric, to accommodate these coding conventions.</p>	MSCDM v4.0	<ul style="list-style-type: none"> 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. Do not include leading zeros beyond those that are part of the code itself (i.e., represent ICD-9 diagnosis 001.9 as “001.9”, not “000001.9” or some other variation). Decimal points may or may not be present for ICD-9/ICD-10 codes. If the decimal point is missing from source data, do not add. If it is present in source data, do not remove.

DIAGNOSIS Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
DX_TYPE	RDBMS Text(2)	SAS Char(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 v4.0 with modified field name	This field is a derived attribute and is not expected to be an explicit data field within a source system
DX_DATE	RDBMS Date	SAS Date (Numeric)	.	Date diagnosis was recorded, if known.	PCORnet	<ul style="list-style-type: none"> This field is meant to represent when a diagnosis was first recorded in the source system during a given encounter, not necessarily when a patient was first diagnosed with a given condition. If the source system records a date for each diagnosis for every day of the encounter, only populate this field if it actually represents the date the diagnosis was first recorded. Otherwise, leave blank.
DX_SOURCE	RDBMS Text(2)	SAS Char(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.	PCORnet	<ul style="list-style-type: none"> It is not necessary to populate interim diagnoses unless readily available. Ambulatory encounters would generally be expected to have a source of “Final.”

DIAGNOSIS Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
DX_ORIGIN	RDBMS Text(2)	SAS Char(2)	OD=Order/EHR BI=Billing CL=Claim DR=Derived NI=No information UN=Unknown OT=Other	Source of the diagnosis information. Billing pertains to internal healthcare processes and data sources. Claim pertains to data from the bill fulfillment, generally data sources held by insurers and other health plans.	PCORnet	<ul style="list-style-type: none"> • Use “OD” for diagnoses entered into the EHR that are associated with an Order. • Use “OD” for any diagnosis associated with an encounter that is entered into the EHR by a provider. • Use “BI” for all diagnoses that are generated through the physician and hospital billing process. • Use “DR” for all diagnoses that are derived or imputed through analytical procedures (e.g., natural language processing). This does not apply to diagnoses that have been mapped from vocabulary mapping software/middleware (e.g., IMO) See General Guidance #4. In those instances, use “OD” or “BI”, depending on the provenance of the diagnosis.

DIAGNOSIS Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PDX	RDBMS Text(2)	SAS Char(2)	P=Principal S=Secondary X=Unable to Classify NI=No information UN=Unknown OT=Other	Principal discharge diagnosis flag.	MSCDM v4.0 with modified field size and value set	<ul style="list-style-type: none"> Value expected for IP, IS, EI, and OS encounters. May also be present for other encounter types. One principal diagnosis per encounter is expected, although in some instances more than one diagnosis may be flagged as principal. Professional vs. Facility – If a partner has access to both professional and facility diagnosis data for a given encounter, the facility diagnoses should be used to populate this field. Partners should document their approach in their ETL ADD.
DX_POA	RDBMS Text(2)	SAS Char(2)	Y = Diagnosis present N = Diagnosis not present U = Insufficient documentation W = Clinically undetermined 1 = Unreported / not used NI=No information UN=Unknown OT=Other	Flag to denote whether diagnosis was present on inpatient admission. Y = Diagnosis present at time of inpatient admission N = Diagnosis not present at time of inpatient admission U = Documentation insufficient to determine if the condition was present at the time of inpatient admission W = Clinically undetermined. Provider unable to clinically determine whether the condition was present at the time of inpatient admission 1 = Unreported / not used. Exempt from present-on-admission reporting.	CMS (https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Coding)	<ul style="list-style-type: none"> Include for EI, IP visits only If data are populated for some inpatient diagnoses, but not all, use “UN” for diagnoses with blank/null value Only assign a value of “1” to a diagnosis if it is reported that way in the source system.
RAW_DX	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	

DIAGNOSIS Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RAW_DX_TYPE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_DX_SOURCE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_PDX	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_DX_POA	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	

6.5. Table: PROCEDURES

PROCEDURES Domain Description:

Procedure codes indicate the discrete medical interventions and diagnostic testing, such as surgical procedures and lab orders, delivered within a healthcare context.

Relational Integrity:

The PROCEDURES table contains one record per PROCEDURESID.

Primary Key: PROCEDURESID

Foreign Keys:

PROCEDURES.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

PROCEDURES.ENCOUNTERID is a foreign key to ENCOUNTER.ENCOUNTERID (many-to-one relationship)

PROCEDURES.PROVIDERID is a foreign key to PROVIDER.PROVIDERID (many-to-one relationship)

Constraints:

PROCEDURESID (unique; required, not null)

PATID (required, not null)

PX (required, not null)

PX_TYPE (required, not null)

Note: This table uses the plural form of “procedures” because “procedure” (singular) is often a reserved word in RDBMS’s.

PROCEDURES Table Implementation Guidance

Guidance

- This table should capture all uniquely recorded procedures for all encounters, including office or evaluation and management visits, diagnostic testing, laboratory test orders, medication administrations, or other services rendered by a clinician.
- If a patient has multiple procedures ordered during one encounter, then there would be one record in this table for each procedure.
- ENCOUNTERID should be populated for DIAGNOSIS and PROCEDURES. The definitions of the DIAGNOSIS and PROCEDURES tables are dependent upon a healthcare context; therefore, the encounter basis is necessary and the ENCOUNTERID, PROVIDERID, ENCOUNTER_TYPE, and ADMIT_DATE from the associated ENCOUNTER record should be included. While not desirable, a low percentage of orphan records is permissible to accommodate instances in which the associated ENCOUNTER details are missing from the source data.
- Data in this table are expected to be from healthcare-mediated processes and reimbursement drivers, including technical/facility billing, professional billing and other data streams. **Do not omit** billing data unless it is unavailable from the source system or the partner is certain that the procedures loaded from the non-billing system (e.g., the EHR) represents completely the procedure data available from the billing system
- If a local vocabulary is used, but cannot be mapped to a standard vocabulary such as ICD-9-CM, PX_TYPE should be populated as “Other” and the local value stored in PX. If the local value can be mapped to a standard vocabulary, follow the guidance around the population of Raw fields (General Guidance #1).
- Evidence of medications administered in outpatient settings should be present in the PROCEDURES table if that information is included with other billed/ordered PROCEDURES.
- Evidence of inpatient administrations should be present in the PROCEDURES table if that information is included with other billed/ordered PROCEDURES.
- **DO NOT** include records from medication administration sources (e.g., electronic medication administration records) in this table.
- If possible to determine from the source data, only include procedures that have actually occurred.
- **Inclusion of laboratory orders** – If possible, partners should include laboratory orders within the PROCEDURES table to support potential studies of appropriate laboratory monitoring. This includes those orders without a corresponding result in the LAB_RESULT_CM table. Do not include canceled orders.

PROCEDURES Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PROCEDURESID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique record.	PCORnet	
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier. Used to link across tables.	MSCDM v4.0	All PATIDs in this table must be present in the DEMOGRAPHIC table.
ENCOUNTERID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary encounter-level identifier. Used to link across tables.	MSCDM v4.0	All ENCOUNTERIDs in this table must be present in the ENCOUNTER table.

PROCEDURES Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ENC_TYPE	RDBMS Text(2)	SAS Char(2)	AV=Ambulatory Visit ED=Emergency Department EI=Emergency Department Admit to Inpatient Hospital Stay (permissible substitution) IP=Inpatient Hospital Stay IS=Non-Acute Institutional Stay OS=Observation Stay IC=Institutional Professional Consult (permissible substitution) TH=Telehealth 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 v4.0 with modified field name and value set	Should be non-null for all records replicated from ENCOUNTER table (guidance added in v4.0).
ADMIT_DATE	RDBMS Date	SAS Date (Numeric)	.	Please note: This is a field replicated from the ENCOUNTER table. See ENCOUNTER table for definitions.	MSCDM v4.0 with modified field name	
PROVIDERID	RDBMS Text(x)	SAS Char(x)	.	Identifier of the PROVIDER most associated with the procedure order.	MSCDM v4.0	All PROVIDERIDs must be present in the PROVIDER table.
PX_DATE	RDBMS Date	SAS Date (Numeric)	.	Date the procedure was performed.	PCORnet	PX_DATE should reflect the date the procedure was performed, if known. Leave blank if you cannot determine if or when the procedure was performed.
PX	RDBMS Text(11)	SAS Char(11)	.	Procedure code.	MSCDM v4.0	Decimal points may or may not be present for ICD-9/ICD-10 procedure codes. If the decimal point is missing, do not add. If it is present, do not remove.

PROCEDURES Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PX_TYPE	RDBMS Text(2)	SAS Char(2)	09=ICD-9-CM 10=ICD-10-PCS 11=ICD-11-PCS CH = CPT or HCPCS LC=LOINC ND=NDC RE=Revenue NI=No information UN=Unknown OT=Other	<p>Procedure code type.</p> <p>We include a number of code types for flexibility, but the basic requirement that the code refer to a medical procedure remains.</p> <p>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.</p> <p>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.</p> <p>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.</p> <p>Please note: The “Other” category is meant to identify internal use ontologies and codes.</p>	MSCDM v4.0 with modified field name and value set	<ul style="list-style-type: none"> • Expected/known length of codes for some terminologies in PX_TYPE: <ul style="list-style-type: none"> ○ ICD-9-CM (09): 3-4 numbers, including a decimal ○ ICD-10-PCS (10): 7 alphanumeric characters ○ CPT/HCPCS: 5 alphanumeric characters; may be longer if modifiers are included • CPT and HCPCS codes should be assigned a value of “CH.” • This field may be a derived attribute. In these cases, it would not be expected to be an explicit data field within a source system

PROCEDURES Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PX_SOURCE	RDBMS Text(2)	SAS Char(2)	OD=Order/EHR BI=Billing CL=Claim DR=Derived NI=No information UN=Unknown OT=Other	Source of the procedure information. Order and billing pertain to internal healthcare processes and data sources. Claim pertains to data from the bill fulfillment, generally data sources held by insurers and other health plans.	PCORnet	<ul style="list-style-type: none"> • If evaluation and management (E/M) or level of service (LOS) codes are available, they should be included • This field is a derived attribute and is not expected to be an explicit data field within a source system • Use “OD” for procedures entered into the EHR that are associated with an Order. • Use “OD” for any procedures associated with an encounter that is entered into the EHR by a provider. • Use “BI” for all procedures that are generated through the physician and hospital billing process. • Use “DR” for all procedure records that are derived or imputed through analytical procedures (e.g., natural language processing). This does not apply to procedures mapped from a superset terminology (General Guidance #4).
PPX	RDBMS Text(2)	SAS Char(2)	P=Principal S=Secondary NI=No information UN=Unknown OT=Other	Principal procedure flag.	PCORnet	<ul style="list-style-type: none"> • Value may be present for IP, IS, EI, and OS encounters. • One principal procedure per encounter is expected, although in some instances more than one procedure may be flagged as principal.
RAW_PX	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_PX_TYPE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	

PROCEDURES Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RAW_PPX	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	

6.6. Table: VITAL

VITAL Domain Description:

Vital signs (such as height, weight, and blood pressure) directly measure an individual's current state of attributes.

Relational Integrity:

The VITAL table contains one record per VITALID.

Primary Key: VITALID

Foreign Keys:

VITAL.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

VITAL.ENCOUNTERID is a foreign key to ENCOUNTER.ENCOUNTERID (zero/many-to-one relationship)

Constraints:

VITALID (unique; required, not null)

PATID (required, not null)

MEASURE_DATE (required, not null)

VITAL_SOURCE (required, not null)

VITAL Table Implementation Guidance

Guidance

- The deprecation of the VITAL table has been postponed. Partners should continue to populate VITAL with the relevant observations. Vital measures can also be stored in OBS_CLIN.
- This table includes measurements recorded in both healthcare and non-healthcare settings.
- The VITAL table contains one record per result/entry. Multiple measurements may exist in source data (for example, 3 blood pressure readings on the same day); in this case, each measurement would be a separate record. If multiple vitals are collected at the same time (e.g., height, weight and blood pressure recorded at the start of an encounter), it is permissible to store these values in a single record. This table should be populated with all available measures, with the possible exception(s) noted below.
- If a partner has access to vital signs that are sourced from a device feed, they should make an assessment about data volume before including these measures, particularly if multiple readings per day are present for a large percentage of their population. Measures should not be averaged or aggregated.
 - For healthcare device data sources: If multiple readings are available and the volume of data is judged by the data partner to be too burdensome for inclusion, using the set of values that were recorded directly in the medical record is preferred over any algorithmic selection process.
- For personal device data sources: If multiple readings are available and the volume of data is judged by the data partner to be too high for inclusion, the project/study leadership should define a method for selecting individual measurements. ~~in the ETL-ADD.~~

Figure 1. Example of populated VITAL table.

VITALID	PATID	ENCOUNTERID	MEASURE_DATE	MEASURE_TIME	VITAL_SOURCE	HT	WT	DIASTOLIC	SYSTOLIC	ORIGINAL_BMI	BP_POSITION	TOBACCO	TOBACCO_TYPE
f5a9a07a-f910-11e4-a322-1697f925ec7b	123	98765	1/5/2014	13:51	HC	67							
f5a9a2be-f910-11e4-a322-1697f925ec7b	123	98765	1/5/2014	13:52	HC		150						
f5a9a3fe-f910-11e4-a322-1697f925ec7b	123	98765	1/5/2014	13:55	HC			120	80		01		
f5a9a52a-f910-11e4-a322-1697f925ec7b	123	98765	1/5/2014		HC							01	NI
f5a9a822-f910-11e4-a322-1697f925ec7b	123	98765	1/5/2014	14:02	HC			122	86		01		
f5a9a94e-f910-11e4-a322-1697f925ec7b	123		3/22/2014		PR		145.6						
f5a9aa7a-f910-11e4-a322-1697f925ec7b	123	65432	11/30/2014		HC	67							
f5a9ab9c-f910-11e4-a322-1697f925ec7b	123	65432	11/30/2014		HC		149.3						

The encounter basis is optional.

Measurements on the same date are recorded in different records; however, it is permissible to consolidate into one record if none of the measures were repeated.

In this example, no time was recorded for several of the measures. Although preferable to capture time, we recognize that some source data may not include time precision.

More than one blood pressure reading was collected during this encounter on January 5.

Note: Completely fake data example created de novo, not from existing data.

VITAL Table Specification						
Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
VITALID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique VITAL record.	PCORnet	
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier. Used to link across tables.	MSCDM v4.0	All PATIDs must be present in the DEMOGRAPHIC table.
ENCOUNTERID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary encounter-level identifier. Not all vital sign measures will be associated with a healthcare encounter.	PCORnet	<ul style="list-style-type: none"> ENCOUNTERID should generally be present if the vitals were measured as part of the healthcare delivery captured by this datamart. All ENCOUNTERIDs in this table must be present in the ENCOUNTER table.
MEASURE_DATE	RDBMS Date	SAS Date (Numeric)	.	Date of vitals measure.	MSCDM v4.0	

VITAL Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
MEASURE_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)	.	Time of vitals measure.	MSCDM v4.0 with modified data type Source of time format: ISO 8601	
VITAL_SOURCE	RDBMS Text(2)	SAS Char(2)	PR=Patient-reported PD=Patient device direct feed HC=Healthcare delivery setting HD=Healthcare device direct feed DR=Derived NI=No information UN=Unknown OT=Other	Please note: The “Patient-reported” category can include reporting by patient’s family or guardian.	PCORnet	<ul style="list-style-type: none"> • This field is a derived attribute and is not expected to be an explicit data field within a source system. • If the source of a given vital sign is not explicitly present in the source system, partners should infer a value for VITAL_SOURCE based on the data and workflow used to collect them. If there is uncertainty as to whether the values come directly from a device, partners should use the more general value/context (patient-reported or healthcare delivery setting). If it is not possible to infer whether the value is patient-reported or was collected in a healthcare delivery setting, partners should choose NI (“no information”). • Use “DR” for all vital signs that are derived or imputed through analytical procedures (e.g., natural language processing).

VITAL Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
HT	RDBMS Number(x)	SAS Numeric(length 8)	.	Height (in inches) measured by standing. Only populated if measure was taken on this date. If missing, this value should be null. Decimal precision is permissible.	MSCDM v4.0	
WT	RDBMS Number(x)	SAS Numeric(length 8)	.	Weight (in pounds). Only populated if measure was taken on this date. If missing, this value should be null. Decimal precision is permissible.	MSCDM v4.0	
DIASTOLIC	RDBMS Number(x)	SAS Numeric(length 8)	.	Diastolic blood pressure (in mmHg). Only populated if measure was taken on this date. If missing, this value should be null.	MSCDM v4.0	
SYSTOLIC	RDBMS Number(x)	SAS Numeric(length 8)	.	Systolic blood pressure (in mmHg). Only populated if measure was taken on this date. If missing, this value should be null.	MSCDM v4.0	
ORIGINAL_BMI	RDBMS Number(x)	SAS Numeric(length 8)	.	BMI if calculated in the source system. Decimal precision is permissible. 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	RDBMS Text(2)	SAS Char(2)	01=Sitting 02=Standing 03=Supine NI=No information UN=Unknown OT=Other	Position for orthostatic blood pressure. This value should be null if blood pressure was not measured.	MSCDM v4.0 with modified field name, field size, and value set	

VITAL Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
SMOKING	RDBMS Text(2)	SAS Char(2)	01=Current every day smoker 02=Current some day smoker 03=Former smoker 04=Never smoker 05=Smoker, current status unknown 06=Unknown if ever smoked 07=Heavy tobacco smoker 08=Light tobacco smoker NI=No information UN=Unknown OT=Other	<p>Indicator for any form of tobacco that is smoked.</p> <p>Per Meaningful Use guidance, "...smoking status includes any form of tobacco that is smoked, but not all tobacco use."</p> <p>"'Light smoker' is interpreted to mean less than 10 cigarettes per day, or an equivalent (but less concretely defined) quantity of cigar or pipe smoke. 'Heavy smoker' is interpreted to mean greater than 10 cigarettes per day or an equivalent (but less concretely defined) quantity of cigar or pipe smoke."</p> <p>"...we understand that a "current every day smoker" or "current some day smoker" is an individual who has smoked at least 100 cigarettes during his/her lifetime and still regularly smokes every day or periodically, yet consistently; a "former smoker" would be an individual who has smoked at least 100 cigarettes during his/her lifetime but does not currently smoke; and a "never smoker" would be an individual who has not smoked 100 or more cigarettes during his/her lifetime."</p> <p>http://www.healthit.gov/sites/default/files/standards-certification/2014-edition-draft-test-procedures/170-314-a-11-smoking-status-2014-test-procedure-draft-v1.0.pdf [retrieved May 11, 2015]</p>	PCORnet Meaningful Use Core Measures 9 of 13, Stage 1 (2014 definition) http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/downloads/9_Record_Smoking_Status.pdf [retrieved January 11, 2015]	

VITAL Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
TOBACCO	RDBMS Text(2)	SAS Char(2)	01=Current user 02=Never 03=Quit/former user 04=Passive or environmental exposure 06=Not asked NI=No information UN=Unknown OT=Other	Indicator for any form of tobacco.	MSCDM v4.0 with modified field name, field size, and value set	
TOBACCO_TYPE	RDBMS Text(2)	SAS Char(2)	01=Smoked tobacco only 02=Non-smoked tobacco only 03=Use of both smoked and non-smoked tobacco products 04=None 05=Use of smoked tobacco but no information about non-smoked tobacco use NI=No information UN=Unknown OT=Other	Type(s) of tobacco used.	MSCDM v4.0 with modified field size and value set	
RAW_DIASTOLIC	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_SYSTOLIC	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_BP_POSITION	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	

VITAL Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RAW_SMOKING	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_TOBACCO	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_TOBACCO_TYPE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	

6.7. Table: DISPENSING

DISPENSING Domain Description:

Prescriptions filled through a community, mail-order or hospital pharmacy. Outpatient dispensing may not be directly captured within healthcare systems.

(Domain description updated in v4.0)

Relational Integrity:

The DISPENSING table contains one record per DISPENSINGID.

Primary Key: DISPENSINGID

Foreign Keys:

DISPENSING.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

DISPENSING.PRESCRIBINGID is a foreign key to PRESCRIBING.PRESCRIBINGID (zero-to-many relationship)

Constraints:

DISPENSINGID (unique; required, not null)

PATID (required, not null)

DISPENSE_DATE (required, not null)

NDC (required, not null)

DISPENSING Table Implementation Guidance

Guidance

- Each record represents an outpatient pharmacy dispensing.
- This domain is commonly available in claims data, but may not be available in many EHR data sources.
- Dispensing records are different from medication orders or prescribing records, data from medication administration activities, as well as the medication reconciliation of the active medication list.
- Administered medications should NOT be stored in this table. They should be stored in the MED_ADMIN table. Evidence of medications administered in outpatient settings, such as infusions given in medical practices, or those administered in an inpatient setting may be present in the PROCEDURES table if that level of detail is available in the source procedure data.
- Rollback transactions and other adjustments that are indicative of a dispensing being canceled or not picked up by the member should be processed (removed) before populating this table. This may be handled differently by Data Partners and may be affected by billing cycles.
- In the uncommon situation where one NDC is dispensed more than once for a given patient on a given day, it is acceptable to combine the values from the multiple dispensings for days supply and number of units.

DISPENSING Table Specification

Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
DISPENSINGID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique record.	PCORnet	
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier. Used to link across tables.	MSCDM v4.0	All PATIDs must be present in the DEMOGRAPHIC table.
PRESCRIBINGID	RDBMS Text(x)	SAS Char(x)	.	This is an <u>optional</u> relationship to the PRESCRIBING table, and may not be generally available. One prescribing order may generate multiple dispensing records.	PCORnet	
DISPENSE_DATE	RDBMS Date	SAS Date (Numeric)	.	Dispensing date (as close as possible to date the person received the dispensing).	MSCDM v4.0 with modified field name	
NDC	RDBMS Text(11)	SAS Char(11)	.	National Drug Code in the 11-digit, no-dash, HIPAA format. Please expunge any place holders (such as dashes or extra digits).	MSCDM v4.0 with additional guidance	NDC must be in HIPAA format. Guidance on normalization for other forms of NDC can be found here: http://www.nlm.nih.gov/research/umls/rxnorm/docs/2012/rxnorm_doco_full_2012-1.html (see section 6)

DISPENSING Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
DISPENSE_SOURCE	RDBMS Text(2)	SAS Char(2)	OD=Order/EHR BI=Billing CL=Claim PM=Pharmacy Benefit Manager DR=Derived NI=No information UN=Unknown OT=Other	Source of the dispensing information.	PCORnet	<ul style="list-style-type: none"> • This field is a derived attribute and is not expected to be an explicit data field within a source system • Use “OD” for dispensing records that are sourced from the EHR or internal pharmacy systems. • Use “BI” for all dispensing records that are generated through the physician and hospital billing process. • Use “CL” for dispensing records that are sourced from pharmacy or medical claims. • Use “DR” for all dispensing records that are derived or imputed through analytical procedures (e.g., natural language processing). • Use “PM” for dispensing records sourced from a pharmacy benefit manager (e.g., Surescripts, Express Scripts).
DISPENSE_SUP	RDBMS Number(x)	SAS Numeric(length 8)	.	Days supply. Number of days that the medication supports based on the number of doses as reported by the pharmacist. This amount is typically found on the dispensing record. Integer values are expected. Important: Please do not calculate during CDM implementation. This field should only reflect originating source system calculations.	MSCDM v4.0 with modified field name	

DISPENSING Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
DISPENSE_AMT	RDBMS Number(x)	SAS Numeric(length 8)	.	Number of units (pills, tablets, vials) dispensed. Net amount per NDC per dispensing. This amount is typically found on the dispensing record. Positive values are expected. Important: Please do not calculate during CDM implementation. This field should only reflect originating source system calculations.	MSCDM v4.0 with modified field name	
DISPENSE_DOSE_DISP	RDBMS Number(x)	SAS Numeric(length 8)	.	Dose of a given medication, as dispensed	PCORnet	Do not impute or derive. Populate only if captured in the source system as a discrete value.
DISPENSE_DOSE_DISP_UNIT	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Units of measure associated with the dose of the medication as dispensed	UCUM	<ul style="list-style-type: none"> Do not impute or derive. Populate only if captured in the source system as a discrete value. Choose the standardized unit of measure that is most reflective of the source data. This is a mixed case value set and entries should be handled accordingly.
DISPENSE_ROUTE	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Route of delivery		Do not impute or derive. Populate only if captured in the source system as a discrete value.
RAW_NDC	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.		
RAW_DISPENSE_DOSE_DISP	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.		
RAW_DISPENSE_DOSE_DISP_UNIT	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.		
RAW_DISPENSE_ROUTE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.		

6.8. Table: LAB_RESULT_CM

LAB_RESULT_CM Domain Description:

This table is used to store quantitative and qualitative measurements from blood and other body specimens.

(Domain description updated in v4.0)

Relational Integrity:

The LAB_RESULT_CM table contains one record per LAB_RESULT_CM_ID

Primary Key: LAB_RESULT_CM_ID

Foreign Keys:

LAB_RESULT_CM.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

LAB_RESULT_CM.ENCOUNTERID is a foreign key to ENCOUNTER.ENCOUNTERID (zero/many-to-one relationship)

Constraints:

LAB_RESULT_CM_ID (unique; required, not null)

PATID (required, not null)

RESULT_DATE (required, not null)

LAB_RESULT_CM Table Implementation Guidance

- Only records with actual lab results should be included in this table. If the result suggests that the test was run (e.g., result is "borderline" or "inconclusive") include it. But if the test is not resulted for any reason (specimen not sufficient, patient did not show), then do not include it.
- If lab results are stored using local or custom codes, partners should ensure that the assigned LOINC code has been validated by a subject matter expert or similar process
- If a LOINC code is available for a given result, the LAB_LOINC field should be populated. If a LOINC code is available for the *order*, that value can be used to populate the LAB_PX field. Note that one order can correspond to many different results. Each result should have its own record in the LAB_RESULT_CM table. If the same LOINC code is used to populate both the order and the result, partners should ensure that the LAB_LOINC field is populated.
- **Inclusion of additional lab results** - Partners should include *all* available laboratory results within their LAB_RESULT_CM table. If the result has a *validated* LOINC code, the LAB_LOINC field should be populated. Otherwise, the LAB_LOINC field should be blank. The RAW_LAB_NAME field can be used to keep track of the various lab results until the appropriate LOINC code is assigned. Lab results beyond the 11 originally included in the PCORnet CDM are being requested in order to establish a denominator of potentially available lab results. Over time, the number of unmapped results is expected to decrease. Results for labs performed as a service for outside institutions do not need to be included. Results from external vendors (e.g., LabCorp, Quest) should be included when available.
- **Clinical LOINC Concepts** – Only include Laboratory LOINC concepts in this table. Do not include clinical LOINC concepts (e.g., EKG results). These records may be stored in the OBS_CLIN table.
- **Standing orders** - Partners should populate the date fields to the best of their ability. For results that are tied to standing laboratory orders, even if LAB_ORDER_DATE reflects the date of the original standing order, SPECIMEN_DATE and/or RESULT_DATE would be expected to correspond to the time when the sample was collected/resulted. Analyses will take both dates into consideration.
- **Units of measure** – A given LOINC code may have many acceptable units of measure. If the RESULT_UNIT field is not populated, it may not be possible to use a result analytically.
- **Verifying LOINC mappings** – At most health systems, laboratory results are typically not associated with a LOINC code at the time they are generated but are assigned a code after the fact. In order to verify that the LOINC code has been appropriately assigned, the PCORnet Coordinating Center will verify that the metadata associated with the result, such as SPECIMEN_SOURCE and RESULT_UNIT, are valid options. Partners should ensure that these fields are populated. **Do not derive these values based on metadata associated with the selected LOINC code.**
- **Titers** – titer results that are returned as ratios (e.g., 1:4) should be stored in the RESULT_QUAL field even if these ratios are not explicitly enumerated in the QUAL valueset. Beginning in CDMv6.1, data model conformance checks for RESULT_QUAL will include the enumerated list and text strings of [X]:[Y] as acceptable values.

LAB_RESULT_CM Table Specification

Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
LAB_RESULT_CM_ID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique LAB_RESULT_CM record. Does not need to be persistent across refreshes, and may be created by methods such as sequence or GUID.	PCORnet	

LAB RESULT CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier. Used to link across tables.	MSCDM v4.0	All PATIDs are expected to be in the DEMOGRAPHIC table.
ENCOUNTERID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary encounter-level identifier. Not all lab results will be associated with a healthcare encounter.	PCORnet (modeled upon VITAL table)	<ul style="list-style-type: none"> Populate with the ENCOUNTERID where the lab specimen was collected (i.e., encounter when the lab test was administered). All ENCOUNTERIDs in this table must also be present in the ENCOUNTER table.
LAB_NAME	RDBMS Text(10)	SAS Char(10)	A1C=Hemoglobin A1c CK=Creatine kinase total CK_MB=Creatine kinase MB CK_MBI=Creatine kinase MB/creatinine total CREATININE=Creatinine HGB=Hemoglobin LDL=Low density lipoprotein INR=International normalized ratio TROP_I=Troponin I cardiac TROP_T_QL=Troponin T cardiac (qualitative) TROP_T_QN=Troponin T cardiac (quantitative) NI=No information UN=Unknown OT=Other	Laboratory result common measure, a categorical identification for the type of test, which is harmonized across all contributing data partners. Please note that it is possible for more than one LOINC® code, CPT code, and/or local code to be associated with one LAB_NAME. Value set modified in v3.1 to add “null value” options.	MSCDM v4.0 with modified field name and subset of categorical values	This field is deprecated effective v4.0. Partners should prioritize mapping their labs to LOINC. If the LOINC code for a given result is unknown, partners should populate the name of the lab in RAW_LAB_NAME.

LAB RESULT_CM Table Specification						
Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
SPECIMEN_SOURCE	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Specimen source. All records will have a specimen source; some tests have several possible values for SPECIMEN_SOURCE.	LOINC	Do not derive a specimen source based on the LOINC code. Please map the specimen source present in the source system to the appropriate code in the Value Set Reference File.
LAB_LOINC	RDBMS Text(10)	SAS Char(10)	.	Logical Observation Identifiers, Names, and Codes (LOINC®) from the Regenstrief Institute. Results with local versions of LOINC codes (e.g., LOINC candidate codes) should be included in the RAW_table field, but the LOINC variable should be set to missing. Current LOINC codes are from 3-8 characters long but Regenstrief suggests a length of 10 for future growth. The last digit of the LOINC code is a check digit and is always preceded by a hyphen. All parts of the LOINC code, including the hyphen, must be included. Do not pad the LOINC code with leading zeros.	MSCDM v4.0	<ul style="list-style-type: none"> Use this field to store the LOINC code of the laboratory result. Expected format of LOINC codes: Length of 3-8, hyphen in the penultimate position, no alphabetical characters. Do not populate the LOINC field with dummy codes. If the LOINC code for a result is unknown, leave blank. Pre-release LOINC codes may be included in this field. Proposed codes that are not yet accepted by LOINC should not be populated. Do not use “discouraged” LOINC codes.

LAB_RESULT_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
LAB_RESULT_SOURCE	RDBMS Text(2)	SAS Char(2)	OD=Order/EHR BI=Billing CL=Claim DR=Derived NI=No information UN=Unknown OT=Other	Source of the information for the lab result.	PCORnet	<ul style="list-style-type: none"> • This field is a derived attribute and is not expected to be an explicit data field within a source system • Use “OD” for lab results that are sourced from the EHR or laboratory information management system (LIMS). • Use “BI” for all lab results that are generated through the physician and hospital billing process (it is unlikely that this value will be used). • Use “CL” for laboratory results that are sourced from pharmacy or medical claims. • Use “DR” for all lab results that are derived or imputed through analytical procedures (e.g., natural language processing).

LAB RESULT CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
LAB_LOINC_SOURCE	RDBMS Text(2)	SAS Char(2)	IN=Instrument LM=LIMS (Standalone or EHR) HL=HL7 feed or other interface DW=Data warehouse PC=PCORnet ETL DM=Other CDM NI=No information UN=Unknown OT=Other	<p>Source/provenance of the LOINC code for this result.</p> <p>Details of categorical definitions: Instrument: Assigned by the instrument used to process the sample and generate the result.</p> <p>Laboratory Information Management System (LIMS): Code is assigned by the system used to manage laboratory results, either stand-alone or as part of the EHR.</p> <p>HL7 Feed: Code is assigned from an external HL7 feed or other interface (e.g., by a Health Information Exchange).</p> <p>Data warehouse: LOINC codes are generated during the extract-transform-load (ETL) procedures used to populate a standalone data warehouse or reporting database.</p> <p>PCORnet ETL: LOINC codes are assigned as part of the ETL process used to populate the PCORnet CDM.</p> <p>Other CDM: Codes are assigned through the ETL used to populate a CDM that is upstream from the PCORnet ETL.</p>	PCORnet	<ul style="list-style-type: none"> This field is a derived attribute and is not expected to be an explicit data field within a source system. If a LOINC code is present in laboratory results transmitted from a LIMS to the EHR via an HL7 feed, use “LM” or “IN” as appropriate, unless the LOINC code is assigned via the HL7 interface itself. Then use “HL” If a LOINC code is assigned by an external entity like a public health laboratory, use “HL”
PRIORITY	RDBMS Text(2)	SAS Char(2)	E=Expedite R=Routine S=Stat NI=No information UN=Unknown OT=Other	<p>Immediacy of test. The intent of this variable is to determine whether the test was obtained as part of routine care or as an emergent/urgent diagnostic test (designated as Stat or Expedite).</p>	MSCDM v4.0 with modified value set and modified field name	

LAB RESULT_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RESULT_LOC	RDBMS Text(2)	SAS Char(2)	L=Lab P=Point of Care NI=No information UN=Unknown OT=Other	Location of the test result. Point of Care locations may include anticoagulation clinic, newborn nursery, finger stick in provider office, or home. The default value is 'L' unless the result is Point of Care. There should not be any missing values.	MSCDM v4.0 with modified value set	
LAB_PX	RDBMS Text(11)	SAS Char(11)	.	Variable for local and standard procedure codes, used to identify the originating order for the lab test.	MSCDM v4.0 with modified field name	Can be used to store the procedure code of the laboratory <i>order</i> . If the same LOINC procedure code is used to identify both the order and the result, make sure LAB_LOINC is populated.
LAB_PX_TYPE	RDBMS Text(2)	SAS Char(2)	09=ICD-9-CM 10=ICD-10-PCS 11=ICD-11-PCS CH = CPT or HCPCS LC=LOINC ND=NDC RE=Revenue NI=No information UN=Unknown OT=Other	Procedure code type, if applicable.	MSCDM v4.0 with modified field name and value set	<ul style="list-style-type: none"> CPT and HCPCS codes should be assigned a value of "CH." This field may be a derived attribute. In these situations, it is not expected to be an explicit data field within a source system
LAB_ORDER_DATE	RDBMS Date	SAS Date (Numeric)	.	Date test was ordered.	MSCDM v4.0 with modified field name	
SPECIMEN_DATE	RDBMS Date	SAS Date (Numeric)	.	Date specimen was collected.	MSCDM v4.0 with modified field name	
SPECIMEN_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)	.	Time specimen was collected.	MSCDM v4.0 with modified field name	

LAB RESULT_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RESULT_DATE	RDBMS Date	SAS Date (Numeric)		Result date.	MSCDM v4.0 with modified field name	If RESULT_DATE is unavailable, partners should use the date that is the closest match in their source data. Partners are permitted to populate RESULT_DATE with the value from SPECIMEN_DATE if necessary and note this in their ETL-ADD.
RESULT_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)		Result time.	MSCDM v4.0 with modified field name	
RESULT_QUAL	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Standardized result for qualitative results. This variable should be NI for quantitative results.	LOINC	If qualitative result cannot be harmonized to a value in RESULT_QUAL value set, please ensure that RAW_RESULT is populated with result value. See table-level guidance on how to represent the results of titers.
RESULT_SNOMED	RDBMS Text(18)	SAS Char(18)	.	If the qualitative result has been mapped to SNOMED CT, the corresponding SNOMED code can be placed here.	PCORnet	Do not impute or derive. Populate only if captured in the source system as a discrete value.
RESULT_NUM	RDBMS Number(x)	SAS Numeric(length 8)	.	Standardized/converted result for quantitative results.	MSCDM v4.0 with modified field name	Used to store quantitative results, including the numeric component of numeric results that contain operators (e.g., "<200", ">= 0.5"). See guidance for RESULT_MODIFIER for further details.

LAB RESULT_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RESULT_MODIFIER	RDBMS Text(2)	SAS Char(2)	EQ=Equal GE=Greater than or equal to GT=Greater than LE=Less than or equal to LT=Less than TX=Text NI=No information UN=Unknown OT=Other	Modifier for result values.	MSCDM v4.0 with modified field name and value set	<ul style="list-style-type: none"> For quantitative results, a non-null RESULT_MODIFIER must be present if they data are to be used analytically. Any symbols in the RAW_RESULT value should be reflected in the RESULT_MODIFIER variable. <p>For example, if the original source data value is "<=200" then RAW_RESULT is "<=200" and RESULT_MODIFIER is LE. RESULT_NUM would also be set to "200". If the original source data value is text, then RESULT_MODIFIER=TX If the original source data value is a numeric value, then RESULT_MODIFIER=EQ</p>

LAB RESULT_CM Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RESULT_UNIT	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Converted/standardized units for the quantitative result.	UCUM	<ul style="list-style-type: none"> Chose the standardized unit of measure that is most reflective of the source data. Do not derive based on the LOINC code. A given LOINC code may have many acceptable units of measure (e.g., tests reporting results as mass concentration can utilize mg/dL, g/dL, or similar). It is important that this field be populated in order to use the data analytically. This is a mixed case value set and entries should be handled accordingly.
NORM_RANGE_LOW	RDBMS Text(10)	SAS Char(10)	.	Lower bound of the normal range assigned by the laboratory. Value should only contain the value of the lower bound. The symbols >, <, >=, <= should be removed. For example, if the normal range for a test is >100 and <300, then "100" should be entered.	MSCDM v4.0	

LAB RESULT_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
NORM_MODIFIER_LOW	RDBMS Text(2)	SAS Char(2)	EQ=Equal GE=Greater than or equal to GT=Greater than NO=No lower limit NI=No information UN=Unknown OT=Other	Modifier for NORM_RANGE_LOW values. For numeric results one of the following needs to be true: 1) Both MODIFIER_LOW and MODIFIER_HIGH contain EQ (e.g. normal values fall in the range 3-10) 2) MODIFIER_LOW contains GT or GE and MODIFIER_HIGH contains NO (e.g. normal values are >3 with no upper boundary) 3) MODIFIER_HIGH contains LT or LE and MODIFIER_LOW contains NO (e.g. normal values are <=10 with no lower boundary)	MSCDM v4.0 with modified value set and field name	
NORM_RANGE_HIGH	RDBMS Text(10)	SAS Char(10)	.	Upper bound of the normal range assigned by the laboratory. Value should only contain the value of the upper bound. The symbols >, <, >=, <= should be removed. For example, if the normal range for a test is >100 and <300, then "300" should be entered.	MSCDM v4.0 with modified field length	
NORM_MODIFIER_HIGH	RDBMS Text(2)	SAS Char(2)	EQ=Equal LE=Less than or equal to LT=Less than NO=No higher limit NI=No information UN=Unknown OT=Other	Modifier for NORM_RANGE_HIGH values. For numeric results one of the following needs to be true: 1) Both MODIFIER_LOW and MODIFIER_HIGH contain EQ (e.g. normal values fall in the range 3-10) 2) MODIFIER_LOW contains GT or GE and MODIFIER_HIGH contains NO (e.g. normal values are >3 with no upper boundary) 3) MODIFIER_HIGH contains LT or LE and MODIFIER_LOW contains NO (e.g. normal values are <=10 with no lower boundary)	MSCDM v4.0 with modified value set and field name	

LAB RESULT_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ABN_IND	RDBMS Text(2)	SAS Char(2)	AB=Abnormal AH=Abnormally high AL=Abnormally low CH=Critically high CL=Critically low CR=Critical IN=Inconclusive NL=Normal NI=No information UN=Unknown OT=Other	Abnormal result indicator. This value comes from the source data; do not apply logic to create it. If field is blank in source data, map to the appropriate flavor of null (guidance added in v4.0).	MSCDM v4.0 with modified value set	<ul style="list-style-type: none"> The “NL” flag is only expected for results that are explicitly tagged as normal within the source system. Results that are flagged as “very abnormal” should be mapped to one of the abnormal codes.
RAW_LAB_NAME	RDBMS Text(x)	SAS Char(x)	.	Local name related to an individual lab test.	PCORnet	
RAW_LAB_CODE	RDBMS Text(x)	SAS Char(x)	.	Local code related to an individual lab test.	PCORnet	
RAW_PANEL	RDBMS Text(x)	SAS Char(x)	.	Local code related to a battery or panel of lab tests.	PCORnet	
RAW_RESULT	RDBMS Text(x)	SAS Char(x)	.	The original test result value as seen in your source data. Values may include a decimal point, a sign or text (e.g., POSITIVE, NEGATIVE, DETECTED).	PCORnet	This field should also be used to store narrative results that may be used in future analyses.
RAW_UNIT	RDBMS Text(x)	SAS Char(x)	.	Original units for the result in your source data.	PCORnet	
RAW_ORDER_DEPT	RDBMS Text(x)	SAS Char(x)	.	Local code for ordering provider department.	PCORnet	
RAW_FACILITY_CODE	RDBMS Text(x)	SAS Char(x)	.	Local facility code that identifies the hospital or clinic. Taken from facility claims.	PCORnet	

Implementation Guidance Reference Table 1: Laboratory Results & LOINC Codes

This table has been deprecated as of version 2.0 of the PCORnet Implementation Guidance.

Implementation Guidance Reference Table 2: Laboratory Results and CPT Codes

This table has been deprecated as of version 2.0 of the PCORnet Implementation Guidance

Implementation Guidance Reference Table 3: Laboratory Standard Abbreviations

This table has been deprecated as of version 4.0 of the PCORnet CDM.

6.9. Table: CONDITION

CONDITION Domain Description:

A condition represents a patient's diagnosed and self-reported health conditions and diseases. The patient's medical history and current state may both be represented.

Relational Integrity:

The CONDITION table contains one record per CONDITIONID.

Primary Key: CONDITIONID

Foreign Keys:

CONDITION.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

CONDITION.ENCOUNTERID is a foreign key to ENCOUNTER.ENCOUNTERID (zero/many-to-one relationship)

Constraints:

CONDITIONID (unique; required, not null)

PATID (required, not null)

CONDITION (required, not null)

CONDITION_TYPE (required, not null)

CONDITION_SOURCE (required, not null)

CONDITION Table Implementation Guidance

Guidance

- This table includes both healthcare and non-healthcare settings.
- Rollback or voided transactions and other adjustments should be processed (removed) before populating this table.
- **These records should NOT be duplicated in the DIAGNOSIS table.**

CONDITION Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
CONDITIONID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique record.	PCORnet	
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier. Used to link across tables.	MSCDM v4.0	All PATIDs in this table must be present in the DEMOGRAPHIC table.
ENCOUNTERID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary encounter-level identifier used to link across tables. This should only be populated if the item was collected as part of a healthcare encounter. If more than one encounter association is present, this field should be populated with the ID of the encounter when the condition was first entered into the system. However, please note that many conditions may be recorded outside of an encounter context.	PCORnet (modeled upon VITAL table)	<ul style="list-style-type: none"> If more than one encounter association is present, this field should be populated with the ID of the encounter when the condition was first entered into the system. However, please note that many conditions may be recorded outside of an encounter context. All ENCOUNTERIDs in this table must also be present in the ENCOUNTER table.
REPORT_DATE	RDBMS Date	SAS Date (Numeric)	.	Date condition was noted, which may be the date when it was recorded by a provider or nurse, or the date on which the patient reported it. Please note that this date may not correspond to onset date.	PCORnet (informed by ESP model)	Date condition was noted, which may be the date when it was recorded by a provider or nurse, or the date on which the patient reported it. Please note that this date may not correspond to onset date.
RESOLVE_DATE	RDBMS Date	SAS Date (Numeric)	.	Date condition was resolved, if resolution of a transient condition has been achieved. A resolution date is not generally expected for chronic conditions, even if the condition is managed.	PCORnet	

CONDITION Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ONSET_DATE	RDBMS Date	SAS Date (Numeric)	.	<p>The onset date concept here refers to "the date and time when problem (illness, disorder, or symptom) started" (ONC:MU Clinical Data Set, caDSR 4973971).</p> <p>This is a different concept than report date, which is the date on which the medical status was collected. An onset date should generally be considered independently of the observer or provider. However, the judgment of when a condition "started" depends on the disease, the frequency of visits, and many other factors. It is not clear that any facility or physician employs this field in a manner which can be trusted without validation during analysis.</p>	PCORnet	A value should only be provided where it exists in the source data. It is not calculated.
CONDITION_STATUS	RDBMS Text(2)	SAS Char(2)	AC=Active RS=Resolved IN=Inactive NI=No information UN=Unknown OT=Other	Condition status corresponding with REPORT_DATE.	PCORnet (informed by ESP model)	The value of IN=Inactive may be used in situations where a condition is not resolved, but is not currently active (for example, psoriasis).
CONDITION	RDBMS Text(18)	SAS Char(18)	.	<p>Condition code.</p> <p>Some codes will contain leading zeroes, and different levels of decimal precision may also be present. This field is a character field, not numeric, to accommodate these coding conventions.</p> <p>Please populate the exact value of this diagnosis code, but remove any source-specific suffixes and prefixes.</p>	PCORnet (modeled upon DIAGNOSIS table)	

CONDITION Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
CONDITION_TYPE	RDBMS Text(2)	SAS Char(2)	09=ICD-9-CM 10=ICD-10-CM/ PCS 11=ICD-11-CM/ PCS SM=SNOMED CT HP=Human Phenotype Ontology AG=Algorithmic NI=No information UN=Unknown OT=Other	Condition code type. Please note: The “Other” category is meant to identify internal use ontologies and codes.	PCORnet (modeled upon DIAGNOSIS table)	<ul style="list-style-type: none"> This field is a derived attribute and is not expected to be an explicit data field within a source system. Do not include ICD procedure codes.
CONDITION_SOURCE	RDBMS Text(2)	SAS Char(2)	PR=Patient-reported medical history HC=Healthcare problem list RG=Registry cohort CC=Patient Chief Complaint PC=PCORnet-defined condition algorithm DR=Derived NI=No information UN=Unknown OT=Other	Please note: The “Patient-reported” category can include reporting by a proxy, such as patient’s family or guardian.	PCORnet (modeled upon VITAL table)	<ul style="list-style-type: none"> “Registry cohort” generally refers to cohorts of patients flagged with a certain set of characteristics for management within a health system. “Patient-reported” can include self-reported medical history and/or current medical conditions, not captured via healthcare problem lists or registry cohorts. Use “DR” for all conditions that are derived or imputed through analytical procedures (e.g., natural language processing). This field is a derived attribute and is not expected to be an explicit data field within a source system

CONDITION Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RAW_CONDITION_STATUS	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_CONDITION	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	This field should be used to store text-string versions of a Condition that may be derived later using methods like natural language processing.
RAW_CONDITION_TYPE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_CONDITION_SOURCE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	

6.10. Table: PRO_CM

PRO_CM Domain Description:

This table is used to store responses to patient-reported outcome measures (PROs), surveys, and questionnaires. This table can be used to store item-level responses as well as the overall score for each measure.

(Domain description updated in v4.0)

Relational Integrity:

The PRO_CM table contains one record per PRO_CM_ID.

Primary Key: PRO_CM_ID

Foreign Keys:

PRO_CM.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

PRO_CM.ENCOUNTERID is a foreign key to ENCOUNTER.ENCOUNTERID (zero/many-to-one relationship)

Constraints:

PRO_CM_ID (unique; required, not null)

PATID (required, not null)

PRO_DATE (required, not null)

PRO_CM Table Implementation Guidance

Guidance

- This version of the PRO_CM table is not optimized for representational efficiency. Certain values will be duplicated across records, and many fields will be blank for certain records. Over time, the structure of this table is expected to evolve as PCORnet better defines the analytical use of PROs across the network. Until then, this table has been defined to support the broadest range of possible use cases at the expense of representational efficiency.
- The PRO_CM table can be used to store both individual item-level responses, as well as the overall score for the measure/instrument. Each item response will be stored in an individual record. Measure/instrument scores can be stored along with the item-level responses that are associated with that measure (where available). See the figure below for an example of how to populate this table.
- If partners are populating PRO item responses or measure scores and are unsure of the PRO_ITEM_NAME, PRO_ITEM_LOINC, PRO_MEASURE_NAME and/or PRO_MEASURE_LOINC, they should populate PRO_ITEM_FULLNAME and PRO_MEASURE_FULLNAME instead. These fields can be considered analogous to RAW fields.
- For the PRO_CM fields with variable field lengths, partners should choose an appropriate field length based on the characteristics of the data are loading into the table. As we use these tables analytically as part of PCORnet studies, we will determine whether it is more efficient to define specific field lengths.
- If a patient completes a survey, but skips a question, create a record in the PRO_CM table as you would for other items in the survey (i.e., include the appropriate date/time fields and other relevant metadata). Then leave PRO_RESPONSE_TEXT and PRO_RESPONSE_NUM blank, as these fields are not required. Do not create empty records if the patient did not actually see the question.
- The PRO_CM table can be used to store the results from questionnaires where the provider or caregiver are providing their interpretation or assessment of the patient's status. Despite the name, it is not restricted solely to patient-reported outcomes. The table is designed to represent survey-type responses. General observations about patients, however, like pain scores recorded in an inpatient or surgical setting, should be stored in the OBS_CLIN table. See General Guidance #15 for additional details or contact the DRN OC with questions.
- **In general, patient-reported social determinants of health (e.g., food instability) would be expected to be found in this table. However, if partners who are participating in initiatives like the National COVID Cohort Collaborative have already loaded those records into OBS_GEN, it is acceptable for them to remain there.**
- For responses that can be represented as either a number or a text (e.g., a numeric value of 2 corresponds to “rarely” in a value set), the expectation is to store whatever is recorded in the source system. Both should be populated if present, but partners are not expected to derive the other if not. If values are represented based on a sequence number, store the actual value, not the sequence number.

Note: Not all fields are shown
Note: Scores not correctly calculated

PRO_CM_ID	PATID	ENCOUNTERID	PRO_TYPE	PRO_ITEM_NAME	PRO_ITEM_LOINC	PRO_RESPONSE_TEXT	PRO_RESPONSE_NUM	PRO_MEASURE_NAME	PRO_MEASURE_SEQ	PRO_MEASURE_SCORE	PRO_MEASURE_THETA	PRO_MEASURE_SCALED_TSCORE	PRO_MEASURE_STANDARD_ERR	PRO_MEASURE_COUNT_SCORED	PRO_MEASURE_LOINC	PRO_MEASURE_VERSION	PRO_ITEM_FULLNAME	PRO_ITEM_TEXT	PRO_MEASURE_FULLNAME
Example - PROMIS measure with individual item responses (as text and numeric), as well as the overall measure score (repeated for each field)																			
0001	001	100000	PM	Sleep108	61996-5	Very much	5	PROMIS Sleep Disturbance 8a	123456	20	0.11	51.1	3.7	8	62197-9	1.0		In the past 7 days, my sleep was restless	
0002	001	100000	PM	Sleep115	61985-8	Not at all	5	PROMIS Sleep Disturbance 8a	123456	20	0.11	51.1	3.7	8	62197-9	1.0		In the past 7 days, I was satisfied with my sleep	
0003	001	100000	PM	Sleep116	61986-6	Very much	1	PROMIS Sleep Disturbance 8a	123456	20	0.11	51.1	3.7	8	62197-9	1.0		In the past 7 days, my sleep was refreshing	
0004	001	100000	PM	Sleep44	61999-9	Not at all	1	PROMIS Sleep Disturbance 8a	123456	20	0.11	51.1	3.7	8	62197-9	1.0		In the past 7 days, I had difficulty falling asleep	
0005	001	100000	PM	Sleep87	61992-4	Never	1	PROMIS Sleep Disturbance 8a	123456	20	0.11	51.1	3.7	8	62197-9	1.0		In the past 7 days, I had trouble staying asleep	
0006	001	100000	PM	Sleep90	61993-2	Never	1	PROMIS Sleep Disturbance 8a	123456	20	0.11	51.1	3.7	8	62197-9	1.0		In the past 7 days, I had trouble sleeping	
0007	001	100000	PM	Sleep110	61988-2	Always	1	PROMIS Sleep Disturbance 8a	123456	20	0.11	51.1	3.7	8	62197-9	1.0		In the past 7 days, I got enough sleep	
0008	001	100000	PM	Sleep109	61987-4	Very poor	5	PROMIS Sleep Disturbance 8a	123456	20	0.11	51.1	3.7	8	62197-9	1.0		In the past 7 days, my sleep quality was	
Example - PROMIS measure with a score, but no individual item responses																			
0009	001	100010	PM					PROMIS Sleep Disturbance 8a		17	0.51	55.1	2.9	6		1.0			
Example - "operational" PRO (PHQ9) that has a score, but no item responses																			
0010	001	100001	LC					PHQ9		2					44261-6				Patient Health Questionnaire 9 item
Example - "operational" PRO (disability survey) that has individual responses (text and numeric) that are not combined into a summary score																			
0012	002	200000	LC		69854-8	Ethnicity	1		11111						69919-9		Are you Hispanic, Latino/a, or of Spanish origin?	Are you Hispanic, Latino/a, or of Spanish origin?	Race, ethnicity, sex, primary language, disability - Health and Human Services (HHS) panel
0013	002	200000	LC		69855-5	White	1		11111						69919-9		Race	What is your race?	Race, ethnicity, sex, primary language, disability - Health and Human Services (HHS) panel
0014	002	200000	LC		46098-0	Male	1		11111						69919-9		Sex		Race, ethnicity, sex, primary language, disability - Health and Human Services (HHS) panel
0015	002	200000	LC		68503-2	Very well	1		11111						69919-9		How well do you speak english	How would you rate your ability to speak and understand English OR How well do you speak English?	Race, ethnicity, sex, primary language, disability - Health and Human Services (HHS) panel
0016	002	200000	LC		69856-3	No	2		11111						69919-9		Are you deaf, or do you have serious difficulty hearing?	Are you deaf, or do you have serious difficulty hearing?	Race, ethnicity, sex, primary language, disability - Health and Human Services (HHS) panel
0017	002	200000	LC		69857-1	No	2		11111						69919-9		Are you blind, or do you have serious difficulty seeing, even when wearing glasses?	Are you blind, or do you have serious difficulty seeing, even when wearing glasses?	Race, ethnicity, sex, primary language, disability - Health and Human Services (HHS) panel
0018	002	200000	LC		69858-9	No	2		11111						69919-9		Because of a physical, mental, or emotional condition, do you have serious difficulty concentrating, remembering, or making decisions?	Because of a physical, mental, or emotional condition, do you have serious difficulty concentrating, remembering, or making decisions?	Race, ethnicity, sex, primary language, disability - Health and Human Services (HHS) panel
0019	002	200000	LC		69859-7	No	2		11111						69919-9		Do you have serious difficulty walking or climbing stairs?	Do you have serious difficulty walking or climbing stairs?	Race, ethnicity, sex, primary language, disability - Health and Human Services (HHS) panel
0020	002	200000	LC		69860-5	No	2		11111						69919-9		Do you have difficulty dressing or bathing?	Do you have difficulty dressing or bathing?	Race, ethnicity, sex, primary language, disability - Health and Human Services (HHS) panel
0021	002	200000	LC		69861-3	No	2		11111						69919-9		Because of a physical, mental, or emotional condition, do you have difficulty doing errands alone such as visiting a physician's office or shopping?	Because of a physical, mental, or emotional condition, do you have difficulty doing errands alone such as visiting a physician's office or shopping?	Race, ethnicity, sex, primary language, disability - Health and Human Services (HHS) panel
Example - PRO-CTCAE PRO with individual item responses (text only) that are not combined into a summary score																			
0022	003	300000	PC	PROCTCAE_23A_SCL		Almost Constantly			2222						1.0		PRO-CTCAE Version 1.0 23a. Past Seven Days Frequency of Feeling Pounding or Racing Heartbeat 5 Point Scale	In the last 7 days, how OFTEN did you feel a POUNDING OR RACING HEARTBEAT (PALPITATIONS)?	
0023	003	300000	PC	PROCTCAE_23B_SCL		Mild			2222						1.0		PRO-CTCAE Version 1.0 23b. Past Seven Days Worst Severity Pounding or Racing Heartbeat 5 Point Scale	In the last 7 days, what was the SEVERITY of your POUNDING OR RACING HEARTBEAT (PALPITATIONS) at its WORST?	
0024	003	300000	PC	PROCTCAE_52A_SCL		Moderate			2222						1.0		PRO-CTCAE Version 1.0 52a. Past Seven Days Worst Severity Insomnia 5 Point Scale	In the last 7 days, what was the SEVERITY of your INSOMNIA (INCLUDING DIFFICULTY FALLING ASLEEP, STAYING ASLEEP, OR WAKING UP EARLY) at its WORST?	
0025	003	300000	PC	PROCTCAE_52B_SCL		Very much			2222						1.0		PRO-CTCAE Version 1.0 52b. Past Seven Days How Much Insomnia Interfered with Usual or Daily Activities 5 Point Scale	In the last 7 days, how much did INSOMNIA (INCLUDING DIFFICULTY FALLING ASLEEP, STAYING ASLEEP, OR WAKING UP EARLY) INTERFERE with your usual or daily activities?	

Figure: Example of a populated PRO_CM table (note: not all required fields are shown).

PRO_CM Table Specification						
Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
PRO_CM_ID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique record.	PCORnet	

PRO_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier for the patient for whom the PRO response was captured. Used to link across tables.	MSCDM v4.0	All PATIDs in this table must be present in the DEMOGRAPHIC table.
ENCOUNTERID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary encounter-level identifier used to link across tables. This should only be populated if the item was collected as part of a healthcare encounter.	PCORnet (modeled upon VITAL table)	All ENCOUNTERIDs in this table must be present in the ENCOUNTER table.
PRO_DATE	RDBMS Date	SAS Date (Numeric)	.	The date of the response submission.	PCORnet	
PRO_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)	.	The time of the response submission.	PCORnet Source of time format: ISO 8601	

PRO_CM Table Specification

Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
PRO_ITEM	RDBMS Text(20)	SAS Char(20)	PN_0001=GLOBAL01 PN_0002=GLOBAL02 PN_0003=GLOBAL06 PN_0004=PFA53 PN_0005=EDDEP29 PN_0006=HI7 PN_0007=SLEEP20 PN_0008=SRPPER11_C APS PN_0009=PAININ9 PN_0010=3793R1 PN_0011=28676R1 PN_0012=EOS_P_011 PN_0013=PEDSGLOBA L2 PN_0014=PEDSGLOBA L5 PN_0015=PEDSGLOBA L6 PN_0016=GLOBAL03 PN_0017=GLOBAL04 PN_0018=EDANX53 PN_0019=SAMHSA PN_0020=CAHPS 4.0 PN_0021=PA070 NI=No information UN=Unknown OT=Other	PCORnet identifier for the specific Common Measure item. Please see the Common Measures Reference Table for more details.	PCORnet	<p>Non-PCORnet Common Measure PROs may also be stored in this table. These measures should be labeled with a value of "OT".</p> <p>This field has been deprecated as of CDM v4.0.</p>

PRO_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PRO_TYPE	RDBMS Text(2)	SAS Char(2)	PM=PROMIS NQ=Neuro-QoL AM=ASQC-Me NT=NIH Toolbox PC=PRO_CTCAE LC=LOINC HC=HCAHPS NI=No information UN=Unknown OT=Other	<p>Terminology / vocabulary used to describe the PRO item.</p> <p>More information on PROMIS, Neuro-QoL and ASQC-Me and the NIH Toolbox can be found on the HealthMeasures website. (www.healthmeasures.net)</p> <p>The Patient-Reported Outcome version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE™) is maintained by the National Cancer Institute. (https://healthcaredelivery.cancer.gov/pro-ctcae/)</p> <p>Information on the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) is located here: http://www.hcahpsonline.org</p>	PCORnet	<ul style="list-style-type: none"> For items/measures that belong to one of the listed terminologies and can also be found in LOINC (e.g., PROMIS), list the native terminology for PRO_TYPE. A value of "LOINC" should be used for those items/measures that do not belong to one of the other specified terminologies but can be found in LOINC (e.g., PHQ-9, WHO-5). Information on PRO-CTCAE can also be found in the NCI Common Data Element browser (https://cdebrowser.nci.nih.gov)
PRO_ITEM_NAME	RDBMS Text(x)	SAS Char(x)	.	Short name or code of the PRO item in the vocabulary/terminology specified in PRO_TYPE.	PCORnet	If a short name or code for the PRO item does not exist within the specified terminology, do not create one. Populate PRO_ITEM_FULLNAME instead.

PRO_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PRO_ITEM_LOINC	RDBMS Text(10)	SAS Char(10)	.	<p>LOINC® code for the PRO item, if available.</p> <p>Logical Observation Identifiers, Names, and Codes (LOINC) from the Regenstrief Institute. Current LOINC codes are from 3-8 characters long but Regenstrief suggests a length of 10 for future growth. The last digit of the LOINC code is a check digit and is always preceded by a hyphen. All parts of the LOINC code, including the hyphen, must be included. Do not pad the LOINC code with leading zeros.</p>	PCORnet (modeled on LAB_RESULT_CM table)	
PRO_RESPONSE_TEXT	RDBMS Text(x)	SAS Char(x)	.	Text version of the response recorded for the item, if available/applicable.	PCORnet	
PRO_RESPONSE_NUM	RDBMS Number(x)	SAS Numeric(length 8)	.	The numeric response recorded for the item, if available/applicable.	PCORnet	
PRO_METHOD	RDBMS Text(2)	SAS Char(2)	PA=Paper EC=Electronic PH=Telephonic IV=Telephonic with interactive voice response (IVR) technology NI=No information UN=Unknown OT=Other	Method of administration. Electronic includes responses captured via a personal or tablet computer, at web kiosks, or via a smartphone.	PCORnet	

PRO_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PRO_MODE	RDBMS Text(2)	SAS Char(2)	SF=Self without assistance SA= Self with assistance PR=Proxy without assistance PA=Proxy with assistance NI=No information UN=Unknown OT=Other	The person who responded on behalf of the patient for whom the response was captured. A proxy report is a measurement based on a report by someone other than the patient reporting as if he or she is the patient, such as a parent responding for a child, or a caregiver responding for an individual unable to report for themselves. Assistance excludes providing interpretation of the patient's response.	PCORnet	
PRO_CAT	RDBMS Text(2)	SAS Char(2)	Y=Yes N=No NI=No information UN=Unknown OT=Other	Indicates whether Computer Adaptive Testing (CAT) was used to administer the survey or instrument that the item was part of. May apply to electronic (EC) and telephonic (PH or IV) modes.	PCORnet	

PRO_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PRO_SOURCE	RDBMS Text(2)	SAS Char(2)	OD=Order/EHR BI=Billing CL=Claim SR=Survey system/mobile app DR=Derived NI=No information UN=Unknown OT=Other	Source of the information for the PRO result.	PCORnet	<ul style="list-style-type: none"> • This field is a derived attribute and is not expected to be an explicit data field within a source system • Use “OD” for PRO records that are sourced from the EHR, its patient portal, or any other system used to capture PROs as part of the care process. • Use “BI” for all PRO records that are generated through the physician and hospital billing process (it is unlikely that this value will be used). • Use “SR” for PRO records that are generated from external survey systems or mobile apps as part of a non-care process. • Use “CL” for PRO records that are sourced from pharmacy or medical claims. • Use “DR” for all PRO records that are derived or imputed through analytical procedures (e.g., natural language processing).
PRO_ITEM_VERSION	RDBMS Text(x)	SAS Char(x)	.	Version of the item/question.	PCORnet	

PRO_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PRO_MEASURE_NAME	RDBMS Text(x)	SAS Char(x)	.	Short name or code of the PRO measure/form that item belongs to, if item is being administered as part of a measure	PCORnet	<ul style="list-style-type: none"> • Will be blank if item is not being administered as part of a measure/form • If measure does not have a short name or code within the specified PRO terminology, do not create one. Populate PRO_MEASURE_FULLNAME instead. • If item is part of a PRO measure, the value for this field will be replicated for all items in the measure.
PRO_MEASURE_SEQ	RDBMS Text(x)	SAS Char(x)	.	Arbitrary ID/sequence number used to link PRO item responses that are associated with the same measure/form.	PCORnet	<ul style="list-style-type: none"> • All PRO item responses associated with the same form/measure should have the same value for PRO_MEASURE_SEQ. • Will be blank if item is not part of a PRO measure/form.
PRO_MEASURE_SCORE	RDBMS Number(x)	SAS Numeric(length 8)	.	Overall raw score for the PRO measure.	PCORnet	<ul style="list-style-type: none"> • Will be blank if item is not part of a PRO measure/form. • If item is part of a PRO measure, the value for this field will be replicated for all items in the measure.

PRO_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PRO_MEASURE_THETA	RDBMS Number(x)	SAS Numeric(length 8)	.	The value of theta reported from the CAT PROMIS results. Only applies to items that are administered as part of a measure.	PCORnet	<ul style="list-style-type: none"> Expected when scoring any measure/instrument that uses computer-adaptive testing (CAT) that has been calibrated with item-response theory (IRT), but may not be present for PRO measures/instruments that are not using CAT. Will be blank if item is not part of a PRO measure/form. If item is part of a PRO measure, the value for this field will be replicated for all items in the measure.
PRO_MEASURE_SCALED_TSCORE	RDBMS Number(x)	SAS Numeric(length 8)	.	Standardized score based on the total raw score for the instrument. Only applies to items that are administered as part of a measure.	PCORnet	<ul style="list-style-type: none"> Expected when scoring any measure/instrument that uses CAT that has been calibrated with IRT, but may not be present for PRO measures/instruments that are not using CAT. Will be blank if item is not part of a PRO measure/form. If item is part of a PRO measure, the value for this field will be replicated for all items in the measure.

PRO_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PRO_MEASURE_STAND ARD_ERROR	RDBMS Number(x)	SAS Numeric(length 8)	.	Possible range of the actual final score based on the scaled T-score. Only applies to items that are administered as part of a measure.	PCORnet	<ul style="list-style-type: none"> • Expected when scoring any measure/instrument that uses CAT that has been calibrated with IRT, but may not be present for PRO measures/instruments that are not using CAT. • Will be blank if item is not part of a PRO measure/form. • If item is part of a PRO measure, the value for this field will be replicated for all items in the measure.
PRO_MEASURE_COUNT _SCORED	RDBMS Number(x)	SAS Numeric (length 8)	.	Number of PRO item responses that were involved in the scoring of the measure.	PCORnet	<ul style="list-style-type: none"> • Expected when scoring any measure/instrument that uses CAT that has been calibrated with IRT, but may not be present for PRO measures/instruments that are not using CAT. • Will be blank if item is not part of a PRO measure/form. • If item is part of a PRO measure, the value for this field will be replicated for all items in the measure.

PRO_CM Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PRO_MEASURE_LOINC	RDBMS Text(10)	SAS Char(10)	.	LOINC® code for the PRO item, if available. Logical Observation Identifiers, Names, and Codes (LOINC) from the Regenstrief Institute. Current LOINC codes are from 3-8 characters long but Regenstrief suggests a length of 10 for future growth. The last digit of the LOINC code is a check digit and is always preceded by a hyphen. All parts of the LOINC code, including the hyphen, must be included. Do not pad the LOINC code with leading zeros.	PCORnet	<ul style="list-style-type: none"> Will be blank if item is not part of a PRO measure/form. If item is part of a PRO measure, the value for this field will be replicated for all items in the measure.
PRO_MEASURE_VERSION	RDBMS Text(x)	SAS Char(x)	.	Version of the measure.	PCORnet	<ul style="list-style-type: none"> Will be blank if item is not part of a PRO measure/form. If item is part of a PRO measure, the value for this field will be replicated for all items in the measure.
PRO_ITEM_FULLNAME	RDBMS Text(x)	SAS Char(x)	.	Full name of the PRO item.	PCORnet	
PRO_ITEM_TEXT	RDBMS Text(x)	SAS Char(x)	.	Text of the PRO item question.	PCORnet	
PRO_MEASURE_FULLNAME	RDBMS Text(x)	SAS Char(x)	.	Full name of the PRO measure.	PCORnet	<ul style="list-style-type: none"> Will be blank if item is not part of a PRO measure/form. If item is part of a PRO measure, the value for this field will be replicated for all items in the measure.
RAW_PRO_CODE	RDBMS Text(x)	SAS Char(x)	:	Optional field for originating code, such as LOINC candidate codes that have not yet been adopted	PCORnet	This field is deprecated as of CDM v4.0
RAW_PRO_RESPONSE	RDBMS Text(x)	SAS Char(x)	:	Optional field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet	This field is deprecated as of CDM v4.0

CDM Reference Table: PRO Common Measures

This table has been deprecated as of version 4.0 of the PCORnet CDM.

6.11. Table: PRESCRIBING

PRESCRIBING Domain Description:

Provider orders for medication dispensing and/or administration. These orders may take place in any setting, including the inpatient or outpatient basis.

Relational Integrity:

The PRESCRIBING table contains one record per PRESCRIBINGID.

Primary Key: PRESCRIBINGID

Foreign Keys:

PRESCRIBING.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

PRESCRIBING.ENCOUNTERID is a foreign key to ENCOUNTER.ENCOUNTERID (zero/many-to-one relationship)

PRESCRIBING.RX_PROVIDERID is a foreign key to PROVIDER.PROVIDERID (many-to-one relationship)

Constraints:

PRESCRIBINGID (unique; required, not null)

PATID (required, not null)

PRESCRIBING Table Implementation Guidance

Guidance

- If a medication cannot be mapped to RxNorm, it should still be present and RAW_RX_MED_NAME should be populated.
- This table can be used to store all medication orders, regardless of encounter type (e.g., inpatient, outpatient, ED) and can include orders for medications that are to be dispensed as well as for those that are to be administered.
- If including orders derived through natural language processing (NLP), make sure that RX_SOURCE has been populated for *all* records.
- See Reference Table 4 for the ordering strategy for RxNorm Term Types.
- Do not populate CDM fields with information derived from the RXCUI (e.g., RX_DOSE_ORDERED, RX_DOSE_FORM). Populate fields only if data are captured in the source system as a discrete value.
- Populate records with the RXCUI as it existed at the time the order was entered, even if the RXCUI is no longer active. Do not attempt to update inactive RXCUIs with a more recent value.
- Medications with approved formulations should have an RXCUI that can adequately represent all ingredients with a single code (e.g., SBD, SCD, MIN). For medication mixtures that lack RXCUIs that can represent all of the component ingredients (e.g., IV mixtures prepared at an inpatient or compounding pharmacy), each individual medication from the order set should be included as a separate record with a unique PRESCRIBINGID. If partners wish to preserve the fact that the records belong to the same order, they do so by creating and populating a new *optional* ORDERID field. Medications with a 1:1 correspondence between the order and RXCUI could have the PRESCRIBINGID stored in the ORDERID. Orders with a 1:many RXCUI relationship would have different PRESCRIBINGIDs but the same ORDERID. Future versions of the CDM may formalize this guidance.

PRESCRIBING Table Specification

Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
PRESCRIBINGID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique PRESCRIBING record.	PCORnet	
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier used to link across tables.	MSCDM v4.0	All PATIDs must be present in the DEMOGRAPHIC table.
ENCOUNTERID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary encounter-level identifier. This should be present if the prescribing activity is directly associated with an encounter.	MSCDM v4.0	All ENCOUNTERIDs in this table must be present in the ENCOUNTER table.
RX_PROVIDERID	RDBMS Text(x)	SAS Char(x)	.	Provider code for the provider who prescribed the medication. The provider code is a pseudoidentifier with a consistent crosswalk to the real identifier.	PCORnet	All PROVIDERIDs in this table must be present in the PROVIDER table.

PRESCRIBING Table Specification

Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
RX_ORDER_DATE	RDBMS Date	SAS Date (Numeric)	.	Order date of the prescription by the provider.	MSCDM v4.0	If RX_ORDER_DATE is not known, a permissible substitution is the RX_START_DATE (if known) or the date the order was signed by the provider.
RX_ORDER_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)	.	Order time of the prescription by the provider.	PCORnet	
RX_START_DATE	RDBMS Date	SAS Date (Numeric)	.	Start date of order. This attribute may not be consistent with the date on which the patient actually begin taking the medication.	Based on ESP	
RX_END_DATE	RDBMS Date	SAS Date (Numeric)	.	End date of order (if available).	Based on ESP	
RX_DOSE_ORDERED	RDBMS Number(x)	SAS Numeric(length 8)	.	Dose of a given medication, as ordered by the provider	PCORnet	Do not impute or derive. Populate only if captured in the source system as a discrete value.
RX_DOSE_ORDERED_UNIT	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Units of measure associated with the dose of the medication as ordered by the provider	UCUM	<ul style="list-style-type: none"> Do not impute or derive. Populate only if captured in the source system as a discrete value. Choose the standardized unit of measure that is most reflective of the source data. This is a mixed case value set and entries should be handled accordingly.
RX_QUANTITY	RDBMS Number(x)	SAS Numeric(length 8)	.	Quantity ordered.	Based on OMOP and ESP	Do not impute or derive. Populate only if captured in the source system as a discrete value.

PRESCRIBING Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RX_DOSE_FORM	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	The unit associated with the quantity prescribed. This is equivalent to RxNorm Dose Form.	PCORnet, based on RxNorm attributes	Do not impute or derive. Populate only if captured in the source system as a discrete value.
RX_REFILLS	RDBMS Number(x)	SAS Numeric(length 8)	.	Number of refills ordered (not including the original prescription). If no refills are ordered, the value should be zero.	Based on OMOP and ESP	<ul style="list-style-type: none"> If value is non-numeric, leave field blank and populate RAW_RX_REFILLS with originating source value. Do not impute or derive. Populate only if captured in the source system as a discrete value.
RX_DAYS_SUPPLY	RDBMS Number(x)	SAS Numeric(length 8)	.	Number of days supply ordered, as specified by the prescription.	Based on OMOP	Do not impute or derive. Populate only if captured in the source system as a discrete value.
RX_FREQUENCY	RDBMS Text(2)	SAS Char(2)	01=Every day 02=Two times a day (BID) 03=Three times a day (TID) 04=Four times a day (QID) 05=Every morning 06=Every afternoon 07=Before meals 08=After meals 10=Every evening 11=Once NI=No information UN=Unknown OT=Other	Specified frequency of medication.	PCORnet	Do not impute or derive. Populate only if captured in the source system as a discrete value.

PRESCRIBING Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RX_PRN_FLAG	RDBMS Text(1)	SAS Char(1)	Y=Yes N=No	Flag to indicate that all or part of medication frequency instructions includes “as needed.”	PCORnet	<ul style="list-style-type: none"> Select Y if medication order includes instructions to take medication “as needed” or with any other frequency “as needed” (e.g., Two times a day, as needed). This field is a derived attribute and is not expected to be an explicit data field within a source system
RX_ROUTE	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Route of medication delivery.	RxNorm (SNOMED)	<ul style="list-style-type: none"> The value set for Route is derived from SNOMED and may include values that are more granular than what is present in the source system. If a direct mapping is available, use the appropriate SNOMED code. If there is any possible ambiguity, use “OT” and then store the source value in RAW_RX_ROUTE. For example, an Injection could map to Subcutaneous or Intramuscular or Intraocular, depending on the drug, so that would best be mapped to “OT.” Do not impute or derive. Populate only if captured in the source system as a discrete value.
RX_BASIS	RDBMS Text(2)	SAS Char(2)	01=Order to Dispense 02=Order to administer NI=No information UN=Unknown OT=Other	Basis of the medication order. The PRESCRIBING table can contain orders for many different activities, and this field is intended to connect the provider’s prescribing order with how the order was fulfilled (such as outpatient dispensing or administration by a healthcare professional). (Value set items updated and field definition expanded in v3.1.)	PCORnet	This field is a derived attribute and is not expected to be an explicit data field within a source system

PRESCRIBING Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RXNORM_CUI	RDBMS Text (8)	SAS Char(8)	.	Where an RxNorm mapping exists for the source medication, this field contains the RxNorm concept identifier (CUI) at the highest possible specificity.	PCORnet and NLM UMLS	<ul style="list-style-type: none"> • Ordering Strategy– The ordering strategy for RxNorm Term Types has been updated to indicate a preference of brand name CUIs over generics, when available. It has also been expanded to include addition RxNorm Term Types. Please see Reference Table 4 for more information. • Do not assign a RxNorm Term Type that is not supported by the source data (i.e., if the medication order in source system is missing strength information, do not use any of the RxNorm Term Types that incorporate the strength component). • This field may be a derived attribute and is not necessarily expected to be an explicit data field within a source system • Expected format of RXNORM_CUI codes: Length of 2-7, no alphabetical characters. • Do not assign more than one RxCUI per order. Ensure that a single prescribing record is not assigned multiple RxCUIs.

PRESCRIBING Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RX_SOURCE	RDBMS Text(2)	SAS Char(2)	OD=Order/EHR DR=Derived NI=No information UN=Unknown OT=Other	Source of the prescribing information.	PCORnet	<ul style="list-style-type: none"> This field is a derived attribute and is not expected to be an explicit data field within a source system Use “OD” for medication orders entered into the EHR or for electronic prescriptions. Use “DR” for all medication orders that are derived or imputed through analytical procedures (e.g., natural language processing). This does not apply to medication orders mapped from a superset terminology or drug database (e.g., MediSpan, FDB). For those records, use “OD” (General Guidance #4).
RX_DISPENSE_AS_WRITTEN	RDBMS Text(2)	SAS Char(2)	Y=Yes N=No NI=No information UN=Unknown OT=Other	Flag to indicate whether the provider indicated that the medication order was to be dispensed as written.	PCORnet	<ul style="list-style-type: none"> This information is typically captured within EHRs or e-prescribing as part of the ordering process. Do not impute or derive. Populate only if captured in the source system as a discrete value.
RAW_RX_MED_NAME	RDBMS Text(x)	SAS Char(x)	.	Field for originating, full textual medication name from the source.	PCORnet	
RAW_RX_FREQUENCY	RDBMS Text(x)	SAS Char(x)		Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_RXNORM_CUI	RDBMS Text(x)	SAS Char(x)		Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	

PRESCRIBING Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RAW_RX_QUANTITY	RDBMS Text(x)	SAS Char(x)		Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_RX_NDC	RDBMS Text(x)	SAS Char(x)		Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_RX_DOSE_ORDERED	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_RX_DOSE_ORDERED_UNIT	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_RX_ROUTE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_RX_REFILLS	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.		

Implementation Guidance Reference Table 4: Ordering of RxNorm Term Types							
(Content from the UMLS [https://www.nlm.nih.gov/research/umls/rxnorm/docs/appendix5.html]- Accessed October 2016)							
	RxNorm Term Type		Information incorporated				
	Code	Description	Ingredient(s)	Strength	Dose Form	Brand Name	Notes
<i>Most Preferred</i>	SBD	Semantic Branded Drug	X	X	X	X	
	SCD	Semantic Clinical Drug	X	X	X		
	BPCK	Brand Name Pack	X	X	X	X	
	GPCK	Generic Pack	X	X	X		
	SBDF	Semantic Branded Drug Form	X		X	X	
	SCDF	Semantic Clinical Drug Form	X		X		

↓	SBDG	Semantic Branded Dose Form Group			X	X	
	SCDG	Semantic Clinical Dose Form Group	X		X		
	SBDC	Semantic Branded Drug Component	X	X		X	
	BN	Brand Name				X	
	MIN	Multiple Ingredients	X				
	SCDC	Semantic Clinical Drug Component	X	X			May not be enough to distinguish medication for analysis purposes. If medication contains multiple ingredients, include a record in the PRESCRIBING table for each one.
	PIN	Precise Ingredient	X				
<i>Least Preferred</i>	IN	Ingredient	X				May not be enough to distinguish medication for analysis purposes. If medication contains multiple ingredients, include a record in the PRESCRIBING table for each one.
<i>Do not use</i>	DF	Dose Form			X		Non-specific
<i>Do not use</i>	DFG	Dose Form Group			X		Non-specific
<i>Do not use</i>	PSN	Prescribable Name					Synonym of another TTY; Use original TTY
<i>Do not use</i>	SY	Synonym					Synonym of another TTY; Use original TTY
<i>Do not use</i>	TMSY	Tall Man Lettering Synonym					Synonym of another TTY; Use original TTY

Implementation Guidance Reference Table 4a: RxNorm Term Types with examples

(obtained from RxNav [<https://mor.nlm.nih.gov/RxNav/>] – Accessed October 2016)

RxNorm Term Type						
	Code	Description	Example (Augmentin XR 12 HR 1000 MG Extended Oral Release Tablet)	RxCUI(s)	Example (Z-PAK)	RxCUI(s)
<i>Most Preferred</i>	SBD	Semantic Branded Drug	Augmentin XR 12 HR 1000 MG Extended Release Oral Tablet	861689	Zithromax 250 MG Oral Tablet	212446
	SCD	Semantic Clinical Drug	12 HR Amoxicillin 1000 MG / Clavulanate 62.5 MG Extended Release Oral Tablet	617995	Azithromycin 250 MG Oral Tablet	308460
	BPCK	Brand Name Pack	N/A	N/A	Z-PAK	750149
	GPCK	Generic Pack	N/A	N/A	{6 (Azithromycin 250 MG Oral Tablet) } Pack	749783
	SBDF	Semantic Branded Drug Form	Amoxicillin / Clavulanate Extended Release Oral Tablet [Augmentin]	618038	Azithromycin Oral Tablet [Zithromax]	367697
	SCDF	Semantic Clinical Drug Form	Amoxicillin / Clavulanate Extended Release Oral Tablet	617988	Azithromycin Oral Tablet	370976
	SBDG	Semantic Branded Dose Form Group	Augmentin Oral Product; Augmentin Pill	1174397; 1174308	Zithromax Oral Product; Zithromax Pill	1187674; 1187675
↓	SCDG	Semantic Clinical Dose Form Group	Amoxicillin / Clavulanate Oral Product; Amoxicillin / Clavulanate Pill	1152874; 1152875	Azithromycin Oral Product; Azithromycin Pill	1155011; 1155012
	SBDC	Semantic Branded Drug Component	Amoxicillin 1000 MG / Clavulanate 62.5 MG [Augmentin]	618037	Azithromycin 250 MG [Zithromax]	564001
	BN	Brand Name	Augmentin	151392	Zithromax	169474
	MIN	Multiple Ingredients	Amoxicillin / Clavulanate	19711	N/A	N/A
	SCDC	Semantic Clinical Drug Component	Amoxicillin 1000 MG; Clavulanate 62.5 MG	331055; 617303	Azithromycin 250 MG	315449
	PIN	Precise Ingredient	N/A	N/A	N/A	N/A
<i>Least Preferred</i>	IN	Ingredient	Amoxicillin; Clavulanate	723; 48203	Azithromycin	18631
<i>Do not use</i>	DF	Dose Form	Extended Release Oral Tablet	316945	Oral Capsule; Oral Tablet	316965; 317541
<i>Do not use</i>	DFG	Dose Form Group	Oral Product; Pill	1151131; 1151133	Oral Product; Pill	1151131; 1151133
<i>Do not use</i>	PSN	Prescribable Name				
<i>Do not use</i>	SY	Synonym				
<i>Do not use</i>	TMSY	Tall Man Lettering Synonym				

6.12. Table: PCORNET_TRIAL

PCORNET_TRIAL Domain Description:

Patients who are enrolled in PCORnet clinical trials and PCORnet studies.

Relational Integrity:

The PCORNET_TRIAL table contains one record per unique combination of PATID, TRIALID, and PARTICIPANTID.

Composite Primary Key: PATID, TRIALID, PARTICIPANTID

Foreign Key:

PCORNET_TRIAL.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

Constraints:

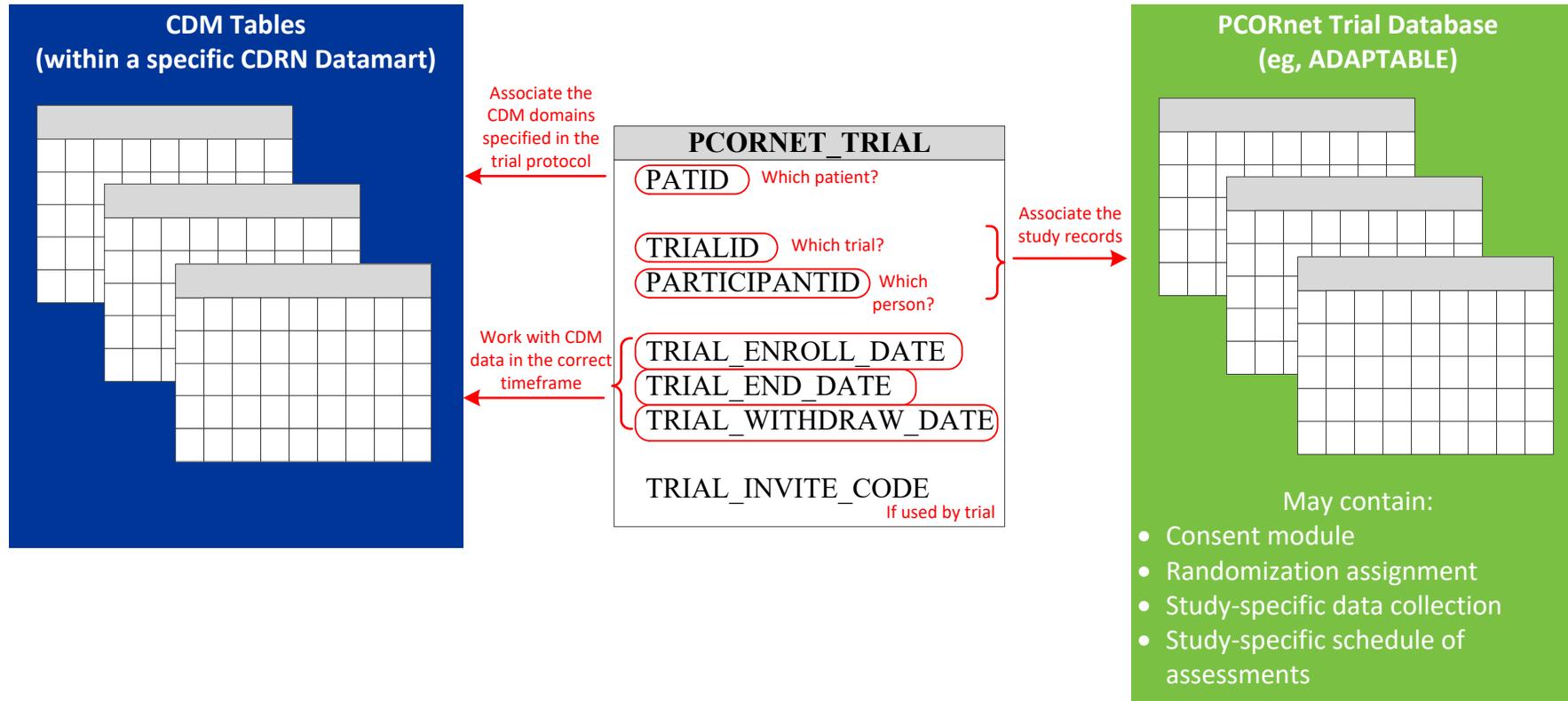
PATID + TRIALID + PARTICIPANTID (unique)

PATID (required, not null)

TRIALID (required, not null)

PARTICIPANTID (required, not null)

The PCORNET_TRIAL table serves as a connector and filter for CDM data within the parameters of a given trial protocol:



PCORNET_TRIAL Table Implementation Guidance

Guidance

- Partners may use the PCORNET_TRIAL table to maintain mappings between PCORnet CDM PATIDs and external trial or study IDs.
- Partners wishing to use this table will need to register their TRIALID with the DRN OC. Please contact the DRN OC if you plan to utilize this table.
- TRIALIDs that start with “PT_” and “PS_” are reserved for PCORnet Trials and PCORnet Studies. Partners should refrain from using TRIALIDs that start with these characters. **The TRIALID “ADPT” is also reserved.**
- One patient participating in multiple trials or studies will have multiple records
- Each PCORnet trial or study will define its parameters **for study or trial participation, and will provide study-specific instructions on how to populate the fields in this table.**
- Patients who decline to participate in a trial or study or do not meet eligibility criteria should not be included in this table
- **Patients who enroll in a trial or study but later withdraw should be included in this table, with a date in the TRIAL_WITHDRAW_DATE field. ~~so that their withdrawal status and date are recognized and used to appropriately manage reporting of CDM data back to the coordinating center~~**
- In most cases, trials will be expected to have a trial database that is separate from the CDM
- Randomization assignment is not included in this table due to the potential for unblinding.
- PATID is not generally appropriate for use as a PARTICIPANTID because it is not disambiguated across networks.

PCORNET_TRIAL Table Specification

<i>Field Name</i>	<i>RDBM Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier used to link across tables.	MSCDM v4.0	All PATIDs must be present in the DEMOGRAPHIC table.
TRIALID	RDBMS Text(20)	SAS Char(20)	.	Each TRIALID is assigned by the PCORnet trial or study’s coordinating center.	PCORnet	
PARTICIPANTID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier used to uniquely identify a participant in a PCORnet trial or study. PARTICIPANTID is never repeated or reused for a specific clinical trial or study, and is generally assigned by trial or study-specific processes. It may be the same as a randomization ID.	PCORnet	

PCORNET TRIAL Table Specification						
<i>Field Name</i>	<i>RDBM Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
TRIAL_SITEID	RDBMS Text(x)	SAS Char(x)	.	Each TRIAL_SITEID is assigned by the PCORnet trial or study coordinating center.	PCORnet	<ul style="list-style-type: none"> This field is a derived attribute and is not expected to be an explicit data field within a source system This field will typically represent the patient's recruiting site for clinical trials. A given patient may have data in more than one PCORnet DataMart.
TRIAL_ENROLL_DATE	RDBMS Date	SAS Date (Numeric)	.	Date on which the participant started participating enrolled in the trial or study (generally coincides with trial or study consent, randomization and/or enrollment process).	PCORnet	
TRIAL_END_DATE	RDBMS Date	SAS Date (Numeric)		Date on which the participant completes participation in the trial or study.		This field may be a person-specific date or it may be study-specific (a stated date for all participants or the date the query is run)
TRIAL_WITHDRAW_DATE	RDBMS Date	SAS Date (Numeric)	.	If applicable, date on which the participant withdraws consent for using their CDM data in the trial or study (generally coincides with trial or study withdrawal processes). from the trial or study.	PCORnet	

PCORNET_TRIAL Table Specification						
<i>Field Name</i>	<i>RDBM Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
TRIAL_INVITE_CODE	RDBMS Text(20)	SAS Char(20)	.	<p>Textual strings used to uniquely identify invitations sent to potential participants, and allows acceptances to be associated back to the originating source.</p> <p>Where used, there should generally be a unique combination of PATID, TRIAL_NAME, and INVITE_CODE within each DataMart.</p> <p>For example, this might include “co-enrollment ID strings” for e-mail invites or “verification codes” for letter invites.</p>	PCORnet	

6.13. Table: DEATH

DEATH Domain Description:
Reported mortality information for patients.

Relational Integrity:

The DEATH table contains one record per unique combination of PATID and DEATH_SOURCE.

Composite Primary Key: PATID, DEATH_SOURCE

Foreign Key:

DEATH.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

Constraints (modified in v3.1)

PATID + DEATH_SOURCE (unique)

PATID (required, not null)

DEATH_SOURCE (required, not null)

DEATH Table Implementation Guidance

Guidance

- One patient may potentially have multiple records in this table because their death may be reported by different sources.
- Deaths represented in the ENCOUNTER.DISCHARGE_DISPOSITION and ENCOUNTER.DISCHARGE_STATUS would generally be expected to be present in this table (see field-level guidance for DEATH.DEATH_SOURCE).

DEATH Table Specification

Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier used to link across tables.	MSCDM v4.0	All PATIDs in this table must be present in the DEMOGRAPHIC table.

DEATH Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
DEATH_DATE	RDBMS Date	SAS Date (Numeric)		Date of death.	MSCDM v4.0 with modified field name and data type	If the death date is completely unknown (e.g., fully imputed), partners should leave it blank.
DEATH_DATE_IMPUTE	RDBMS Text(2)	SAS Char(2)	B=Both month and day imputed D=Day imputed M=Month imputed N=Not imputed NI=No information UN=Unknown OT=Other	When date of death is imputed, this field indicates which parts of the date were imputed.	MSCDM v4.0 with modified field name and valueset	This field is a derived attribute and is not expected to be an explicit data field within a source system
DEATH_SOURCE	RDBMS Text(2)	SAS Char(2)	L=Other, locally defined N=National Death Index D=Social Security S=State Death files T=Tumor data DR=Derived NI=No information UN=Unknown OT=Other		MSCDM v4.0 with modified field name and additional guidance	<ul style="list-style-type: none"> “Other, locally defined” may be used to indicate presence of deaths reported from EHR systems, such as in-patient hospital deaths or dead on arrival. This field is a derived attribute and is not expected to be an explicit data field within a source system Use “DR” for all death records that are derived or imputed through analytical procedures (e.g., natural language processing).
DEATH_MATCH_CONFIDENCE	RDBMS Text(2)	SAS Char(2)	E=Excellent F=Fair P=Poor NI=No information UN=Unknown OT=Other	For situations where a probabilistic patient matching strategy is used, this field indicates the confidence that the patient drawn from external source data represents the actual patient.	MSCDM v4.0 with modified field name and additional guidance	<ul style="list-style-type: none"> Should not be present where DEATH_SOURCE is L (locally-defined). May not be applicable for DEATH_SOURCE=T (tumor registry data). This field is a derived attribute and is not expected to be an explicit data field within a source system

6.14. Table: DEATH_CAUSE

DEATH_CAUSE Domain Description:
The individual causes associated with a reported death.

Relational Integrity:

The DEATH_CAUSE table contains one record per unique combination of PATID, DEATH_CAUSE, DEATH_CAUSE_CODE, DEATH_CAUSE_TYPE, and DEATH_CAUSE_SOURCE.

Composite Primary Key: PATID, DEATH_CAUSE, DEATH_CAUSE_CODE, DEATH_CAUSE_TYPE, DEATH_CAUSE_SOURCE

Foreign Key:

DEATH_CAUSE.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

Constraints:

PATID + DEATH_CAUSE + DEATH_CAUSE_CODE + DEATH_CAUSE_TYPE + DEATH_CAUSE_SOURCE (unique)

PATID (required, not null)

DEATH_CAUSE (required, not null)

DEATH_CAUSE_CODE (required, not null)

DEATH_CAUSE_TYPE (required, not null)

DEATH_CAUSE_SOURCE (required, not null)

DEATH_CAUSE Table Implementation Guidance

Guidance

- When legacy data have conflicting reports, please make a local determination as to which to use. There is typically a 1-2 year lag in death registry data.

DEATH_CAUSE Table Specification

Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier used to link across tables.	MSCDM v4.0	All PATIDs in this table must be present in the DEMOGRAPHIC table.

DEATH_CAUSE Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
DEATH_CAUSE	RDBMS Text(8)	SAS Char(8)	.	Cause of death code. Please include the decimal point in ICD codes (if any).	MSCDM v4.0 with modified field name	
DEATH_CAUSE_CODE	RDBMS Text(2)	SAS Char(2)	09=ICD-9 10=ICD-10 NI=No information UN=Unknown OT=Other	Cause of death code type.	MSCDM v4.0 with modified field name	
DEATH_CAUSE_TYPE	RDBMS Text(2)	SAS Char(2)	C=Contributory I=Immediate/Primary O=Other U=Underlying NI=No information UN=Unknown OT=Other	Cause of death type. There should be only one underlying cause of death.	MSCDM v4.0 with modified field name	This field is a derived attribute and is not expected to be an explicit data field within a source system
DEATH_CAUSE_SOURCE	RDBMS Text(2)	SAS Char(2)	L=Other, locally defined N=National Death Index D=Social Security S=State Death files T=Tumor data DR=Derived NI=No information UN=Unknown OT=Other	Source of cause of death information.	MSCDM v4.0 with modified field name	<ul style="list-style-type: none"> • “Other, locally defined” may be used to indicate presence of deaths reported from EHR systems, such as in-patient hospital deaths or dead on arrival. • This field is a derived attribute and is not expected to be an explicit data field within a source system • Use “DR” for all death cause records that are derived or imputed through analytical procedures (e.g., natural language processing).
DEATH_CAUSE_CONFIDENCE	RDBMS Text(2)	SAS Char(2)	E=Excellent F=Fair P=Poor NI=No information UN=Unknown OT=Other	Confidence in the accuracy of the cause of death based on source, match, number of reporting sources, discrepancies, etc.	MSCDM v4.0 with modified field name	This field is a derived attribute and is not expected to be an explicit data field within a source system

6.15. Table: MED_ADMIN

MED_ADMIN Domain Description:

Records of medications administered to patients by healthcare providers. These administrations may take place in any setting, including inpatient, outpatient or home health encounters.

Relational Integrity:

The MED_ADMIN table contains one record per MEDADMINID.

Primary Key: MEDADMINID

Foreign Keys:

MEDADMIN.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

MEDADMIN.ENCOUNTERID is a foreign key to ENCOUNTER.ENCOUNTERID (many-to-one relationship)

MEDADMIN.MEDADMIN_PROVIDERID is a foreign key to PROVIDER.PROVIDERID (many-to-one relationship)

Constraints:

MEDADMINID (unique; required, not null)

PATID (required, not null)

MEDADMIN_START_DATE (required, not null)

MED_ADMIN Table Implementation Guidance

Guidance

- If a medication cannot be mapped to RxNorm or NDC, it should still be present and RAW_MEDADMIN_NAME should be populated.
- Only include administrations that were actually delivered to the patient, if that level of specificity is available in the source system.
- Patient-reported medication administrations are not within the scope of this table.
- See Reference Table 4 for the ordering strategy for RxNorm Term Types.
- Do not populate CDM fields with information derived from the RXCUI (e.g., MEDADMIN_DOSE_ADMIN). Populate fields only if data are captured in the source system as a discrete value.
- Populate records with the RXCUI as it existed at the time the order was entered, even if the RXCUI is no longer active. Do not attempt to update inactive RXCUIs with a more recent value.
- If a medication mixture contains multiple RXCUIs (e.g., inpatient mixture), each individual medication from the order set should be included as an individual record with a unique MEDADMINID. Each individual medication is expected to have a unique dose.
- ENCOUNTERID is expected to be present for records in the MED_ADMIN table.
- For administrations where the amount ordered is listed as a rate (e.g., infusions), if the DOSE/DOSE_UNIT values are specified as a rate, and those fields are stored discretely in your EHR, populate the relevant CDM fields. Assuming that START_DATE/START_TIME and STOP_DATE/STOP_TIME are also populated, it will be possible to compute the rate analytically. Otherwise, leave blank.

MED_ADMIN Table Specification

Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
MEDADMINID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique MED_ADMIN record.	PCORnet	
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier used to link across tables.	MSCDM v4.0	All PATIDs must be present in the DEMOGRAPHIC table.
ENCOUNTERID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary encounter-level identifier. The ENCOUNTERID should be present.	MSCDM v4.0	All ENCOUNTERIDs must be present in the ENCOUNTER table.
PRESCRIBINGID	RDBMS Text(x)	SAS Char(x)	.	This is an <u>optional</u> relationship to the PRESCRIBING table, and may not be generally available. One prescribing order may generate multiple administration records.	PCORnet	
MEDADMIN_PROVID ERID	RDBMS Text(x)	SAS Char(x)	.	Provider code for the provider who prescribed the medication. The provider code is a pseudoidentifier with a consistent crosswalk to the real identifier.	PCORnet	All PROVIDERIDs must be present in the PROVIDER table.

MED_ADMIN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
MEDADMIN_START_DATE	RDBMS Date	SAS Date (Numeric)	.	Date medication administration started/occurred	PCORnet	Populate for single point-in-time administrations, as well as continuous time administrations, such as infusions.
MEDADMIN_START_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)	.	Time medication administration started/occurred	PCORnet	Populate for single point-in-time administrations, as well as continuous time administrations, such as infusions.
MEDADMIN_STOP_DATE	RDBMS Date	SAS Date (Numeric)	.	Date medication administration ended	PCORnet	Populate for continuous time administrations, such as infusions. Leave blank if the administration is a single point-in-time event.
MEDADMIN_STOP_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)	.	Time medication administration ended	PCORnet	Populate for continuous time administrations, such as infusions. Leave blank if the administration is a single point-in-time event.
MEDADMIN_TYPE	RDBMS Text(2)	SAS Char(2)	ND=NDC RX=RXNORM NI=No information UN=Unknown OT=Other	Medication code type.	PCORnet	<ul style="list-style-type: none"> This field is a derived attribute and is not expected to be an explicit data field within a source system If mapping from medication database (e.g., MediSpan, FDB), and it is possible to map to RxNorm and NDC, RxNorm is the preferred term type. If medication administration records are stored natively as NDC, do not convert to RxNorm.
MEDADMIN_CODE	RDBMS Text(11)	SAS Char(11)	.	Medication code	PCORnet	
MEDADMIN_DOSE_ADMIN	RDBMS Number(x)	SAS Numeric(length 8)	.	Dose of a given medication, as administered by the provider	PCORnet	Do not impute or derive. Populate only if captured in the source system as a discrete value.

MED_ADMIN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
MEDADMIN_DOSE_ADMIN_UNIT	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Units of measure associated with the dose of the medication as administered by the provider	UCUM	<ul style="list-style-type: none"> Do not impute or derive. Populate only if captured in the source system as a discrete value. Choose the standardized unit of measure that is most reflective of the source data. This is a mixed case value set and entries should be handled accordingly.
MEDADMIN_ROUTE	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Route of medication delivery.	RxNorm (SNOMED)	<ul style="list-style-type: none"> The value set for Route is derived from SNOMED and may include values that are more granular than what is present in the source system. If a direct mapping is available, use the appropriate SNOMED code. If there is any possible ambiguity, use "OT" and then store the source value in RAW_MEDADMIN_ROUTE. For example, an Injection could map to Subcutaneous or Intramuscular or Intraocular, depending on the drug, so that would best be mapped to "OT." Do not impute or derive. Populate only if defined in the source system as a discrete value.

MED_ADMIN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
MEDADMIN_SOURCE	RDBMS Text(2)	SAS Char(2)	OD=Order/EHR DR=Derived NI=No information UN=Unknown OT=Other	Source of the medication administration record.	PCORnet	<ul style="list-style-type: none"> This field is a derived attribute and is not expected to be an explicit data field within a source system Use “OD” for medication orders entered into the EHR. Use “DR” for all medication orders that are derived or imputed through analytical procedures (e.g., natural language processing). This does not apply to administrations mapped from a superset terminology or drug database (e.g., MediSpan, FDB). For those records, use “OD” (General Guidance #4).
RAW_MEDADMIN_MED_NAME	RDBMS Text(x)	SAS Char(x)	.	Field for originating, full textual medication name from the source.	PCORnet	
RAW_MEDADMIN_CODE	RDBMS Text(x)	SAS Char(x)		Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_MEDADMIN_DOSE_ADMIN	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_MEDADMIN_DOSE_ADMIN_UNIT	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_MEDADMIN_ROUTE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	

6.16. Table: PROVIDER

PROVIDER Domain Description:

Data about the providers who are involved in the care processes documented in the CDM.

Relational Integrity:

The PROVIDER table contains one record per PROVIDERID.

Primary Key: PROVIDERID

Foreign Keys:

PROVIDER.PROVIDERID is a foreign key to ENCOUNTER.PROVIDERID (one-to-many relationship)

PROVIDER.PROVIDERID is a foreign key to DIAGNOSIS.PROVIDERID (one-to-many relationship)

PROVIDER.PROVIDERID is a foreign key to PROCEDURES.PROVIDERID (one-to-many relationship)

PROVIDER.PROVIDERID is a foreign key to PRESCRIBING.RX_PROVIDERID (one-to-many relationship)

PROVIDER.PROVIDERID is a foreign key to MEDADMIN.MEDADMIN_PROVIDERID (one-to-many relationship)

Constraints:

PROVIDERID (unique; required, not null)

PROVIDER Table Implementation Guidance

Guidance

- Include one record per provider.
- When populating provider specialty, if multiple values are available, use the specialty believed to be primary.

PROVIDER Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PROVIDERID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique PROVIDER record. Does not need to be persistent across refreshes, and may be created by methods such as sequence or GUID.	PCORnet	
PROVIDER_SEX	RDBMS Text(2)	SAS Char(2)	A=Ambiguous F=Female M=Male NI=No information UN=Unknown OT=Other	Sex assigned at birth.	MSCDM v4.0 with modified field size and value set Source: Administrative Sex (HL7) http://phinvads.cdc.gov/vads/ViewValueSet.action?id=06D34BBC-617F-DD11-B38D-00188B398520	The “Ambiguous” category may be used for individuals who are physically undifferentiated from birth. The “Other” category may be used for individuals who are undergoing gender re-assignment. If a value of “X” is recorded in the source system, map to “OT”.
PROVIDER_SPECIALTY_PRIMARY	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Primary specialty of the provider	Healthcare Provider Taxonomy Code Set	
PROVIDER_NPI	RDBMS Number(x)	SAS Numeric(length 8)	.	National Provider Identifier (NPI) of the provider.	PCORnet	<ul style="list-style-type: none"> Partners should only consider populating this field if their local governance allows it. The expectation is that this field will primarily be used to support study-specific activities, though partners may also populate it to support their internal work.
PROVIDER_NPI_FLAG	RDBMS Text(1)	SAS Char(1)	Y=Yes N=No	Flag to indicate whether partner has access to the National Provider Identifier (NPI) of the provider.	PCORnet	This field is a derived attribute and is not expected to be an explicit data field within a source system
RAW_PROVIDER_SPECIALTY_PRIMARY	RDBMS Text(x)	SAS Char(x)	.	Field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet	

6.17. Table: OBS_CLIN

OBS_CLIN Domain Description:

Standardized qualitative and quantitative clinical observations about a patient, including vital signs. Some observations may also be represented in the VITAL table.

Relational Integrity:

The OBS_CLIN table contains one record OBSCLINID

Primary Key: OBSCLINID

Foreign Keys:

OBSCLIN.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

OBSCLIN.ENCOUNTERID is a foreign key to ENCOUNTER.ENCOUNTERID (many-to-one relationship)

OBSCLIN.PROVIDERID is a foreign key to PROVIDER.PROVIDERID (many-to-one relationship)

Constraints:

OBSCLINID (unique; required, not null)

PATID (required, not null)

OBSCLIN.START_DATE (required, not null)

OBS_CLIN Table Implementation Guidance

Guidance

- The OBS_CLIN table is intended to store standardized clinical observations that have been recorded about a patient.
- Examples of the types of observations that can be stored in this table include pulmonary function test results (e.g., FEV1, FVC, FEV1/FVC), echocardiogram results (e.g., left ventricle ejection fraction) and vital signs (e.g., temperature).
- Vital signs that can be stored in the VITAL table may also be stored in this table (e.g., systolic blood pressure), though VITAL should continue to be populated. The DRNOC maintains a list of LOINC codes that can be used to represent these observations. The Value Set Appendix contains a list of codes that can be used to map records that were previously stored in the VITAL table. For completeness purposes, it contains both preferred and discouraged codes. Partners should select the code(s) that best represent the data in their source systems but should not utilize a code that encodes more information than is available in their data (e.g., do not use a code that specifies a method if the method is unknown). In general, discouraged codes should not be used.
- Decisions on what to include in this table and how to prioritize the population of those records are expected to be driven primarily by potential funding opportunities.
- This table provides a generalized structure for storing observations and is not optimized for analytical efficiency. As elements from this table are used in studies and/or distributed queries, additional representations of those data elements (i.e., new table structures) may be required to better support those activities.
- If partners are populating pain scores (not pain-related PRO surveys) captured in an inpatient or surgical setting, these values would be expected to be present in this table, not PRO_CM.
- If an observation has a value set that includes an option of “not documented or assessed” (or similar), these values should be included in the CDM if they are present in the source system. Do not derive if an observation is missing.
- If a partner has access to vital signs that are sourced from a device feed, they should make an assessment about data volume before including these measures, particularly if multiple readings per day are present for a large percentage of their population. Measures should not be averaged or aggregated.
- For healthcare device data sources: If multiple readings are available and the volume of data is judged by the data partner to be too burdensome for inclusion, using the set of values that were recorded directly in the medical record is preferred over any algorithmic selection process.
- For personal device data sources: If multiple readings are available and the volume of data is judged by the data partner to be too high for inclusion, the project/study leadership should define a method for selecting individual measurements and this logic should be documented in the ETL ADD.
- Observations recorded together (e.g., diastolic and systolic blood pressure) should have the same date(s) and time(s).

OBS_CLIN Table Specification

Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
OBSCLINID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique OBS_CLIN record.	PCORnet	
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier. Used to link across tables.	MSCDM v4.0	All PATIDs must be present in the DEMOGRAPHIC table.

OBS_CLIN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ENCOUNTERID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary encounter-level identifier used to link across tables.	MSCDM v4.0	<ul style="list-style-type: none"> Populate with the ENCOUNTERID where the observation was obtained. All ENCOUNTERIDs must be present in the ENCOUNTER table.
OBSCLIN_PROVIDERID	RDBMS Text(x)	SAS Char(x)	.	Provider code for the provider who ordered the observation. The provider code is a pseudoidentifier with a consistent crosswalk to the real identifier.	PCORnet	All PROVIDERIDs must be present in the PROVIDER table.
OBSCLIN_START_DATE	RDBMS Date	SAS Date (Numeric)	.	Date of observation/measurement.	MSCDM v4.0 with modified field name	Populate for both point-in-time and continuous observations/measurements
OBSCLIN_START_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)	.	Time of observation/measurement.	MSCDM v4.0 with modified field name	Populate for both point-in-time and continuous observations/measurements if the timestamp is recorded in the source system. Do not impute if not present.
OBSCLIN_STOP_DATE	RDBMS Date	SAS Date (Numeric)	.	Date observation/measurement ended.	PCORnet	<p>Populate for continuous time observations / measurements.</p> <p>Leave blank if point-in-time observations / measurements.</p>
OBSCLIN_STOP_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)	.	Time observation/measurement ended.	PCORnet	<p>Populate for continuous time observations / measurements if the timestamp is recorded in the source system. Do not impute if not present.</p> <p>Leave blank if single point-in-time observations / measurements.</p>

OBS_CLIN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
OBSCLIN_TYPE	RDBMS Text(2)	SAS Char(2)	LC=LOINC SM=SNOMED CT (observable entity) NI=No information UN=Unknown OT=Other	Terminology / vocabulary used to describe the clinical observation.	PCORnet	
OBSCLIN_CODE	RDBMS Text(18)	SAS Char(18)	.	Code of the clinical observation in the vocabulary/terminology specified in OBSCLIN_TYPE.	PCORnet	<ul style="list-style-type: none"> Results with local versions of LOINC codes (e.g., LOINC candidate codes) should be included in the RAW_ table field. The last digit of the LOINC code is a check digit and is always preceded by a hyphen. All parts of the LOINC code, including the hyphen, must be included. Do not pad codes with leading zeros.
OBSCLIN_RESULT_QUAL	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Standardized result for qualitative results. This variable should be NI for quantitative results.	LOINC	If qualitative result cannot be harmonized to a value in OBSCLIN_RESULT_QUAL value set, please ensure that RAW_OBSCLIN_RESULT is populated with result value.
OBSCLIN_RESULT_TEXT	RDBMS Text(x)	SAS Char(x)	.	Narrative/textual clinical observations	PCORnet	
OBSCLIN_RESULT_SNOMED	RDBMS Text(18)	SAS Char(18)	.	If the qualitative result has been mapped to SNOMED CT, the corresponding SNOMED code can be placed here.	PCORnet	Partners are not expected to derive or impute if not present in the source system.

OBS_CLIN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
OBSCLIN_RESULT_NUM	RDBMS Number(x)	SAS Numeric(length 8)	.	Standardized/converted result for quantitative results.	MSCDM v4.0 with modified field name	Used to store quantitative results, including the numeric component of numeric results that contain operators (e.g., "<200", ">= 0.5"). See guidance for RESULT_MODIFIER for further details.
OBSCLIN_RESULT_MODIFIER	RDBMS Text(2)	SAS Char(2)	EQ=Equal GE=Greater than or equal to GT=Greater than LE=Less than or equal to LT=Less than TX=Text NI=No information UN=Unknown OT=Other	Modifier for result values.	MSCDM v4.0 with modified field name and value set	Any symbols in the RAW_RESULT value should be reflected in the RESULT_MODIFIER variable. For example, if the original source data value is "<=200" then RAW_RESULT=200 and RESULT_MODIFIER=LE. RESULT_NUM would also be set to "200". If the original source data value is text, then RESULT_MODIFIER=TX. If the original source data value is a numeric value, then RESULT_MODIFIER=EQ
OBSCLIN_RESULT_UNIT	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Converted/standardized units for the result.	UCUM	<ul style="list-style-type: none"> Chose the standardized unit of measure that is most reflective of the source data (if applicable). This is a mixed case value set and entries should be handled accordingly.

OBS_CLIN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
OBSCLIN_SOURCE	RDBMS Text(2)	SAS Char(2)	OD=Order/EHR BI=Billing CL=Claim RG=Registry/ ancillary system PR=Patient-reported PD=Patient device direct feed HC=Healthcare delivery setting HD=Healthcare device direct feed DR=Derived NI=No information UN=Unknown OT=Other	Source of the information for the lab result. observation or measurement.	PCORnet	<ul style="list-style-type: none"> This field is a derived attribute and is not expected to be an explicit data field within a source system Use "OD" for clinical observations that are sourced from the EHR. Use "BI" for all clinical observations that are generated through the physician and hospital billing process (it is unlikely that this value will be used). Use "CL" for all clinical observations that are sourced from pharmacy or medical claims. Use "HC" for clinical observations that are sourced from the EHR or other ancillary clinical systems. Use "RG" for observations that are sourced from a separate registry. The "Patient-reported" category can include reporting by patient's family or guardian. Use "DR" for all clinical observations that are derived or imputed through analytical procedures (e.g., natural language processing).

OBS_CLIN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
OBSCLIN_ABN_IND	RDBMS Text(2)	SAS Char(2)	AB=Abnormal AH=Abnormally high AL=Abnormally low CH=Critically high CL=Critically low CR=Critical IN=Inconclusive NL=Normal NI=No information UN=Unknown OT=Other	Abnormal result indicator. This value comes from the source data; do not apply logic to create it. If field is blank in source data, map to the appropriate flavor of null.	MSCDM v4.0 with modified value set	<ul style="list-style-type: none"> The “NL” flag is only expected for results that are explicitly tagged as normal within the source system. Results that are flagged as “very abnormal” should be mapped to one of the abnormal codes.
RAW_OBSCLIN_NAME	RDBMS Text(x)	SAS Char(x)	.	Local name related to an individual clinical observation/measurement.	PCORnet	
RAW_OBSCLIN_CODE	RDBMS Text(x)	SAS Char(x)	.	Local code related to an individual clinical observation/measurement.	PCORnet	
RAW_OBSCLIN_TYPE	RDBMS Text(x)	SAS Char(x)	.	Terminology related to the code in RAW_OBSCLIN_CODE .	PCORnet	
RAW_OBSCLIN_RESULT	RDBMS Text(x)	SAS Char(x)	.	The original test result value as seen in your source data. Values may include a decimal point, a sign or text (e.g., POSITIVE, NEGATIVE, DETECTED). The symbols >, <, >=, <= should be removed from the value and stored in the Modifier variable instead.	PCORnet	
RAW_OBSCLIN_MODIFIER	RDBMS Text(x)	SAS Char(x)	.	The original modifier text as represented in your source data.	PCORnet	
RAW_OBSCLIN_UNIT	RDBMS Text(x)	SAS Char(x)	.	Original units for the result in your source data.	PCORnet	

6.18. Table: OBS_GEN

OBS_GEN Domain Description:
Table to store everything else.

Relational Integrity:

The OBS_GEN table contains one record OBSGENID

Primary Key: OBSGENID

Foreign Keys:

OBSGEN.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

OBSGEN.ENCOUNTERID is a foreign key to ENCOUNTER.ENCOUNTERID (zero/many-to-one relationship)

OBSGEN.PROVIDERID is a foreign key to PROVIDER.PROVIDERID (many-to-one relationship)

Constraints:

OBSGENID (unique; required, not null)

PATID (required, not null)

OBSGEN **START** DATE (required, not null)

OBS_GEN Table Implementation Guidance

Guidance

- Partners may use this table to store network- or study-specific data elements.
- Records in this table are not expected to be used in queries distributed by the DRN OC.
- This table provides a generalized structure for storing observations and is not optimized for analytical efficiency. As elements from this table are used in studies and/or distributed queries, additional representations of those data elements (i.e., new table structures) may be required to better support those activities.

OBS_GEN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
OBSGENID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique OBS_GEN record.	PCORnet	
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier. Used to link across tables.	MSCDM v4.0	All PATIDs must be present in the DEMOGRAPHIC table.
ENCOUNTERID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary encounter-level identifier used to link across tables. This field should be populated if the observation was recorded as part of a healthcare encounter.	PCORnet	<ul style="list-style-type: none"> Populate with the ENCOUNTERID where the observation was recorded. All ENCOUNTERIDs in this table must also be present in the ENCOUNTER table.
OBSGEN_PROVIDERID	RDBMS Text(x)	SAS Char(x)	.	Provider code for the provider who recorded the observation. The provider code is a pseudoidentifier with a consistent crosswalk to the real identifier.	PCORnet	All PROVIDERIDs must be present in the PROVIDER table.
OBSGEN_START_DATE	RDBMS Date	SAS Date (Numeric)		Date of observation/measurement	MSCDM v4.0 with modified field name	Populate for both point-in-time and continuous observations/measurements
OBSGEN_START_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)		Time of observation/measurement.	MSCDM v4.0 with modified field name	Populate for both point-in-time and continuous observations/measurements if the timestamp is recorded in the source system. Do not impute if not present.
OBSGEN_STOP_DATE	RDBMS Date	SAS Date (Numeric)	.	Date observation/measurement ended.	PCORnet	<p>Populate for continuous time observations / measurements.</p> <p>Leave blank if point-in-time observations / measurements.</p>

OBS_GEN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
OBSGEN_STOP_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)	.	Time observation/measurement ended.	PCORnet	Populate for continuous time observations / measurements if the timestamp is recorded in the source system. Do not impute if not present. Leave blank if single point-in-time observations / measurements.
OBSGEN_TYPE	RDBMS Text(30)	SAS Char(30)	09DX=ICD-9-CM-DX 09PX=ICD-10-PCS 09PX=ICD-9-CM-PX 10DX=ICD-10-CM/ PCS 10PX=ICD-10-PCS 11DX=ICD-11-CM/ PCS 11PX=ICD-11-PCS ON=ICD-O (Oncology) SM=SNOMED HP=Human Phenotype Ontology HG=Human Genome Organization LC=LOINC RX=RNXNORM ND=NDC CH=CPT or HCPCS GM=Global Medical Device Nomenclature UD_*=User-defined PC_*=PCORnet reserved NI=No information UN=Unknown OT=Other	Terminology/vocabulary used to describe the observation. Networks/partners can define their own terminologies with strings starting with “UD_”. Strings that start with “PC_” are reserved for network-wide activities and will be assigned by the Coordinating Center.	PCORnet	<ul style="list-style-type: none"> For user-defined values, use the listed convention. If a study involves this table and spans multiple networks, participants should ensure that they chose a UD string that is not in use in those locations. The ICD CM and PCS terminologies are listed as separate options, since codes may be present without decimals. This can lead to collisions between values if they are co-mingled.
OBSGEN_CODE	RDBMS Text(50)	SAS Char(50)	.	Standardized code denoting the observations based on the terminology/vocabulary specified in OBSGEN_TYPE	PCORnet	

OBS_GEN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
OBSGEN_RESULT_QUAL	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Standardized result for qualitative results. This variable should be NI for quantitative results.	LOINC	Use RAW_OBSGEN_RESULT to store qualitative results that cannot be harmonized to the defined value set.
OBSGEN_RESULT_TEXT	RDBMS Text(x)	SAS Char(x)	.	Narrative/textual observations.	PCORnet	
OBSGEN_RESULT_NUM	RDBMS Number(x)	SAS Numeric(length 8)	.	Standardized/converted result for quantitative results.	MSCDM v4.0 with modified field name	Used to store quantitative results, including the numeric component of numeric results that contain operators (e.g., "<200", ">= 0.5"). See guidance for RESULT_MODIFIER for further details.
OBSGEN_RESULT_MODIFIER	RDBMS Text(2)	SAS Char(2)	EQ=Equal GE=Greater than or equal to GT=Greater than LE=Less than or equal to LT=Less than TX=Text NI=No information UN=Unknown OT=Other	Modifier for result values.	MSCDM v4.0 with modified field name and value set	Any symbols in the RAW_RESULT value should be reflected in the RESULT_MODIFIER variable. For example, if the original source data value is "<=200" then RAW_RESULT=200 and RESULT_MODIFIER=LE. RESULT_NUM would also be set to "200". If the original source data value is text, then RESULT_MODIFIER=TX If the original source data value is a numeric value, then RESULT_MODIFIER=EQ

OBS_GEN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
OBSGEN_RESULT_UNIT	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Converted/standardized units for the result.	UCUM	<ul style="list-style-type: none"> Chose the standardized unit of measure that is most reflective of the source data (if applicable). This is a mixed case value set and entries should be handled accordingly.
OBSGEN_TABLE_MODIFIED	RDBMS Text(3)	SAS Char(3)	ENR=ENROLLMENT ENC=ENCOUNTER DX=DIAGNOSIS PX=PROCEDURES VT=VITAL DSP=DISPENSING LAB=LAB_RESULT_CM CON=CONDITION PRO=PRO_CM RX=PRESCRIBING PT=PCORNET_TRIAL DTH=DEATH DC=DEATH_CAUSE MA=MED_ADMIN OC=OBS_CLIN OB=OBS_GEN IM=IMMUNIZATION LDS=LDS_ADDRESS_HISTORY OT=Other	Table name when observation describes attributes of an existing record in the CDM.	PCORnet	<ul style="list-style-type: none"> If observation record modifies something other than the patient (i.e., attribute about an encounter), a link to that table can be included here. If a value is listed in OBSGEN_TABLE_MODIFIED, then a corresponding ID should be listed in OBSGEN_ID_MODIFIED.

OBS_GEN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
OBSGEN_ID_MODIFIED	RDBMS Text(x)	SAS Char(x)	.	Identifier when observation describes attributes of an existing record in the CDM.	PCORnet	<ul style="list-style-type: none"> • If observation record modifies something other than the patient (i.e., attribute about an encounter), a link to that record can be included here. • If a value is listed in OBSGEN_TABLE_MODIFIED, then a corresponding ID should be listed in OBSGEN_ID_MODIFIED. • If modifying a record in OBS_GEN, the value of OBSGEN_ID_MODIFIED must be different than the value of OBSGENID for that record.

OBS_GEN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
OBSGEN_SOURCE	RDBMS Text(2)	SAS Char(2)	OD=Order/EHR BI=Billing CL=Claim RG=Registry / ancillary system SR=Survey system/mobile app PR=Patient-reported PD=Patient device direct feed HC=Healthcare delivery setting HD=Healthcare device direct feed DR=Derived NI=No information UN=Unknown OT=Other	Source of the information for the observation or measurement. lab result.	PCORnet	<ul style="list-style-type: none"> This field is a derived attribute and is not expected to be an explicit data field within a source system Use “OD” for clinical observations that are sourced from the EHR. Use “BI” for all clinical observations that are generated through the physician and hospital billing process (it is unlikely that this value will be used). Use “CL” for all clinical observations that are sourced from pharmacy or medical claims. Use “HC” for clinical observations that are sourced from the EHR or other ancillary clinical systems. Use “RG” for observations that are sourced from a separate registry or ancillary clinical system. The “Patient-reported” category can include reporting by patient’s family or guardian. Use “DR” for all clinical observations that are derived or imputed through analytical procedures (e.g., natural language processing).

OBS_GEN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
OBSGEN_ABN_IND	RDBMS Text(2)	SAS Char(2)	AB=Abnormal AH=Abnormally high AL=Abnormally low CH=Critically high CL=Critically low CR=Critical IN=Inconclusive NL=Normal NI=No information UN=Unknown OT=Other	Abnormal result indicator. This value comes from the source data; do not apply logic to create it. If field is blank in source data, map to the appropriate flavor of null.	MSCDM v4.0 with modified value set	<ul style="list-style-type: none"> The "NL" flag is only expected for results that are explicitly tagged as normal within the source system. Results that are flagged as "very abnormal" should be mapped to one of the abnormal codes.
RAW_OBSGEN_NAME	RDBMS Text(x)	SAS Char(x)	.	Local name related to an individual clinical observation/measurement.	PCORnet	
RAW_OBSGEN_CODE	RDBMS Text(x)	SAS Char(x)	.	Local code related to an individual clinical observation/measurement.	PCORnet	
RAW_OBSGEN_TYPE	RDBMS Text(x)	SAS Char(x)	.	Terminology related to the code in RAW_OBSGEN_CODE.	PCORnet	
RAW_OBSGEN_RESULT	RDBMS Text(x)	SAS Char(x)	.	The original test result value as seen in your source data.	PCORnet	
RAW_OBSGEN_UNIT	RDBMS Text(x)	SAS Char(x)	.	Original units for the result in your source data.	PCORnet	

6.19. Table: HASH_TOKEN

HASH_TOKEN Domain Description:

Encrypted, keyed secure hash tokens that are used to match patient records across DataMarts using privacy-preserving record linkage methods.

Relational Integrity:

The HASH_TOKEN table contains one record per combination of patient and token encryption key.

Composite Primary Key: PATID, TOKEN_ENCRYPTION_KEY

Foreign Key:

HASH_TOKEN.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

Constraints:

PATID (unique; required, not null)

TOKEN_ENCRYPTION_KEY (required, not null)

HASH_TOKEN Table Implementation Guidance

- Guidance*
- Every patient in the DEMOGRAPHIC table is expected to have one record in the HASH_TOKEN table for each TOKEN_ENCRYPTION_KEY.
 - Tokens are generated from personally-identifiable information (PII). This information can be stored in each partner's PRIVATE_DEMOGRAPHIC table and PRIVATE_ADDRESS_HISTORY table. The PII is used as input to the Datavant DeID/tokenization module. Tokens should not be placed into the CDM until they have been transformed into Site-PCORnet transit tokens using the Datavant Link/transform-tokens module.
 - If PII fields are populated with dummy values by default (e.g., 999-99-9999 for phone number or SSN), these values should be removed before running the Datavant DeID/tokenization module. If PII is not available within the source system or if local restrictions prevent its use, leave the input field blank.
 - Tokens are generated based on data availability. If input data is not present for a given token strategy (e.g., combination of PII elements), no token will be generated and an error code will be produced instead. These token error codes should be loaded into the HASH_TOKEN table (i.e., there should not be any null values). Do not suppress the error codes in the output of the Datavant software.
 - Each successfully generated token has a fixed length of 44 characters. Do not enforce a 44-character constraint, however, to accommodate the error codes generated in the case of tokenization failure.
 - Tokens should be generated as part of every refresh. Partners can choose to generate tokens for all patients, or only for those patients who were added between refreshes or had updates to their PII.
 - Select tokens generated using the Datavant DeID/tokenization module are certified as de-identified data via the HIPAA Expert Determination method in accordance with the HIPAA Privacy Rule (45 CFR parts 160 and 164). All tokens in the HASH_TOKEN table satisfy this criteria and are controlled via the Datavant DeID PCORnet configuration settings within the Datavant software.
 - Additional token strategies are available and can be implemented as needed on a per-study basis based on the study-specific data dictionary.
 - See the Supplemental Guide on Privacy-Preserving Record Linkage for additional implementation details and guidance (separate document).
 - Partners that wish to include tokens encrypted using multiple keys for internal purposes should add a TOKEN_ENCRYPTION_KEY field and store the key name there. When responding to PCORnet queries/data curation, the table should be filtered to only include tokens used in PCORnet linkages. The name of this key should also be included in the TOKEN_ENCRYPTION_KEY field in HARVEST.
 - Partners should include the TOKEN_ENCRYPTION_KEY string that is output from the Datavant Link software. Please ensure that this table includes tokens encrypted using the PCORnet key (includes "pcornet" or similar in the output string). If partners are encrypting their tokens using multiple keys, it is acceptable to include those additional records in this table.

HASH_TOKEN Table Specification

Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier. Used to link across tables. PATID is passed through the Datavant software in order to be associated with the generated encrypted keyed hashes.	MSCDM v4.0	All PATIDs should be present in the DEMOGRAPHIC table.
TOKEN_01	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 01. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_02	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 02. Enforced through PCORnet configuration setting.	PCORnet	

HASH_TOKEN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
TOKEN_03	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 03. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_04	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 04. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_05	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 05. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_06	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 06. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_07	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 07. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_08	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 08. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_09	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 09. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_12	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 12. Enforced through PCORnet configuration setting.	PCORnet	This field was deprecated in v5.1 and added back in v6.1
TOKEN_14	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 14. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_15	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 15. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_16	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using token strategy 16 in Datavant DeID. Enforced through PCORnet configuration setting.	PCORnet	

HASH_TOKEN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
TOKEN_17	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 17. Enforced through PCORnet configuration setting.	PCORnet	This field was deprecated in v5.1 and added back in v6.1
TOKEN_18	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 18. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_23	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 23. Enforced through PCORnet configuration setting.	PCORnet	This field was deprecated in v5.1 and added back in v6.1
TOKEN_24	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 24. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_25	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 25. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_26	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 26. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_29	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 29. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_30	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 30. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_101	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 101. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_102	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 102. Enforced through PCORnet configuration setting.	PCORnet	

HASH_TOKEN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
TOKEN_103	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 103. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_104	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 104. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_105	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 105. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_106	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 106. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_107	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 107. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_108	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 108. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_109	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 109. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_110	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 110. Enforced through PCORnet configuration setting.	PCORnet	

HASH_TOKEN Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
TOKEN_111	RDBMS Text(x)	SAS Char(x)	.	Encrypted keyed hash generated from PII using Datavant token strategy 111. Enforced through PCORnet configuration setting.	PCORnet	
TOKEN_ENCRYPTION_KEY	RDBMS Text(x)	SAS Char(x)	.	Name of the key used to generate the encrypted hash tokens.	PCORnet	Descriptive text string output as part of the tokenization process (e.g., "SITENAME-pcornet_TOKEN_ENCRYPTION_KEY")
TOKEN_21	RDBMS Text(x)	SAS Char(x)	:	Encrypted hash generated from PII using token strategy 21.	PCORnet	This field is deprecated as of CDM v5.1
TOKEN_22	RDBMS Text(x)	SAS Char(x)	:	Encrypted hash generated from PII using token strategy 22.	PCORnet	This field is deprecated as of CDM v5.1

6.20. Table: LDS_ADDRESS_HISTORY

LDS_ADDRESS_HISTORY Domain Description:
Longitudinal record of a patient's address that adheres to the requirements of a Limited Data Set.

Relational Integrity:

The LDS_ADDRESS_HISTORY table can contain many records per patient.

Primary Key: ADDRESSID

Foreign Key:

ADDRESS_HISTORY.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

Constraints:

ADDRESSID (unique; required, not null)

PATID (required, not null)

ADDRESS_USE (required, not null)

ADDRESS_TYPE (required, not null)

ADDRESS_PREFERRED (required, not null)

LDS_ADDRESS_HISTORY Table Implementation Guidance

Guidance

- Expect multiple records per individual
- This table is currently limited to addresses in the United States.
- Partners can limit records in this table to validated addresses if known.
- If partners have difficulty constructing a longitudinal address history for patients within their DataMart, they should prioritize populating the current address for each patient.

LDS_ADDRESS_HISTORY Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ADDRESSID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique address record.	PCORnet	
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier. Used to link across tables.	MSCDM v4.0	All PATIDs should be present in the DEMOGRAPHIC table.
ADDRESS_USE	RDBMS Text(2)	SAS Char(2)	HO=Home WO=Work TP=Temp OL=Old/Incorrect NI=No information UN=Unknown OT=Other	Purpose of the address. Details of categorical definitions: Home: A communication address at home. Work: An office address. First choice for business-related contacts during business hours. Temp: A temporary address. Old/Incorrect: This address is no longer in use (or was never correct but retained for records).	FHIR - ADDRESS	<ul style="list-style-type: none"> • This field may be a derived attribute that is not an explicit data field within a source system. • Use the period start/end fields to indicate if an address is no longer valid. Do not change values of HO/WO/TP to OL if a new address is available. • The old/incorrect value is included in case partners are doing a bulk load and it is present in their source system. It is acceptable to exclude these records, however. • If addresses within the source system are reasonably expected to represent the patient's home address, it is acceptable to mark these as "HO."

LDS_ADDRESS_HISTORY Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ADDRESS_TYPE	RDBMS Text(2)	SAS Char(2)	PO=Postal PH=Physical BO=Both NI=No information UN=Unknown OT=Other	Type of address. Details of categorical definitions: Postal: mailing address – PO Boxes and care-of addresses. Physical: A physical address that can be visited. Both: An address that is both physical and postal.	FHIR - ADDRESS	<ul style="list-style-type: none"> • This field may be a derived attribute that is not an explicit data field within a source system. • Addresses that are clearly not postal-only addresses can be marked as “Both”

ADDRESS_PREFERRED	RDBMS Text(2)	SAS Char(2)	Y=Yes N=No	Indicates whether this address is the preferred one for a given patient, address use and address type within a given address period.	PCORnet	<ul style="list-style-type: none"> • This field may be a derived attribute that is not an explicit data field within a source system. • It is possible to have an address marked as preferred for every address type/use within each defined address period. • If a new address period is defined, it is not necessary to set the flag(s) back to “N” for any of the existing address periods. • If only one address is present for a given period, that address should be marked as preferred. • If multiple addresses are present for a period, one should be denoted as preferred. The rest should be labeled as “N”. • If multiple addresses are present for a period without clear institutional guidance as to which is preferred, partners can use a heuristic to make a determination (i.e., address associated with most recent encounter, address used for billing, etc.)
-------------------	------------------	-------------	---------------	--	---------	---

LDS_ADDRESS_HISTORY Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ADDRESS_CITY	RDBMS Text(x)	SAS Char(x)	.	The name of the city, town, village or other community.	FHIR - ADDRESS	
ADDRESS_STATE	RDBMS Text(2)	SAS Char(2)	See Value Set Reference File for a list of acceptable values.	State, as represented by 2-digit postal abbreviation.	PCORnet	
ADDRESS_ZIP5	RDBMS Text(5)	SAS Char(5)	.	5-digit postal code for the address.	PCORnet	
ADDRESS_ZIP9	RDBMS Text(9)	SAS Char(9)	.	9-digit postal code for the address.	PCORnet	Do not include hyphens.
ADDRESS_COUNTY	RDBMS Text(x)	SAS Char(x)	.	The name of the county associated with the address.	PCORnet	<ul style="list-style-type: none"> Partners are not expected to derive county for every address in their CDM. Rather, it is anticipated that this field will be used by individual projects, which will have specific guidance on how to derive it.
ADDRESS_PERIOD_START	RDBMS Date	SAS Date (Numeric)	.	Initial date when the address is known to be in use.	FHIR - ADDRESS	<ul style="list-style-type: none"> If the date the address is known to be in use is unknown, it is acceptable to use the date the record was created in the source system or the date the record was first created in the CDM.

LDS_ADDRESS_HISTORY Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ADDRESS_PERIOD_END	RDBMS Date	SAS Date (Numeric)	.	Date when address was no longer in use.	FHIR - ADDRESS	<ul style="list-style-type: none"> Only the current address period is expected to have a null value. All previous periods are expected to have a defined end date.

6.21. Table: IMMUNIZATION

IMMUNIZATION Domain Description:

Records of immunizations that have been delivered within the health system as well as reports of those administered elsewhere.

Relational Integrity:

The IMMUNIZATION table contains one record per IMMUNIZATIONID.

Primary Key: IMMUNIZATIONID

Foreign Keys:

IMMUNIZATION.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

IMMUNIZATION.ENCOUNTERID is a foreign key to ENCOUNTER.ENCOUNTERID (zero/many-to-one relationship)

IMMUNIZATION.VX_PROVIDERID is a foreign key to PROVIDER.PROVIDERID (zero/many-to-one relationship)

IMMUNIZATION.PROCEDURESID is a foreign key to PROCEDURES.PROCEDURESID (zero/many-to-one relationship)

Constraints:

IMMUNIZATIONID (unique; required, not null)

PATID (required, not null)

VX_CODE (required, not null)

VX_CODE_TYPE (required, not null)

VX_STATUS (required, not null)

IMMUNIZATION Table Implementation Guidance

Guidance

- Do not include study vaccines.

IMMUNIZATION Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
IMMUNIZATIONID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique IMMUNIZATION record.	PCORnet	
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier used to link across tables.	MSCDM v4.0	All PATIDs must be present in the DEMOGRAPHIC table.
ENCOUNTERID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary encounter-level identifier. This should be present if the immunization activity is directly associated with an encounter.	MSCDM v4.0	All ENCOUNTERIDs in this table must be present in the ENCOUNTER table.
PROCEDURESID	RDBMS Text(x)	SAS Char(x)	.	This is an optional relationship to the PROCEDURES table and is not expected to be available for all immunizations. One procedure may generate multiple immunization records.	PCORnet	All PROCEDURESIDs in this table must be present in the PROCEDURES table.
VX_PROVIDERID	RDBMS Text(x)	SAS Char(x)	.	Provider code for the provider who delivered the immunization. The provider code is a pseudoidentifier with a consistent crosswalk to the real identifier.	PCORnet	All PROVIDERIDs in this table must be present in the PROVIDER table.
VX_RECORD_DATE	RDBMS Date	SAS Date (Numeric)	.	Date immunization was recorded (i.e., date record was created).	FHIR – Immunization	
VX_ADMIN_DATE	RDBMS Date	SAS Date (Numeric)	.	Date immunization was administered, if known.	FHIR - Immunization	<ul style="list-style-type: none"> • Leave blank if unknown. • Partial dates are allowed. Follow the CDM date imputation guidance and ensure the HARVEST table is properly updated.

IMMUNIZATION Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
VX_CODE_TYPE	RDBMS Text(2)	SAS Char(2)	CX=CVX ND=NDC CH = CPT or HCPCS RX=RXNORM NI=No information UN=Unknown OT=Other	Immunization code type.	PCORnet	<ul style="list-style-type: none"> This field is a derived attribute and is not expected to be an explicit data field within a source system If immunizations are represented within the source system by multiple terminologies, choose the one with the highest level of granularity.
VX_CODE	RDBMS Text(11)	SAS Char(11)	.	Immunization code	PCORnet	
VX_STATUS	RDBMS Text(2)	SAS Char(2)	CP=Completed ER=Entered in error ND=Not Done IC=Incomplete NI=No information UN=Unknown OT=Other	Status of the immunization.	FHIR – Immunization; PCORnet	It is not necessary to load immunizations that were entered in error if partners can easily distinguish these data in the source system(s).
VX_STATUS_REASON	RDBMS Text(2)	SAS Char(2)	IM=Immunity MP=Medical precaution OS=Out of stock PO=Patient objection NI=No information UN=Unknown OT=Other	Reason immunization is incomplete or not done.	FHIR – Immunization	This field is not expected to be populated for immunizations with a status of “CP” or “ER”

IMMUNIZATION Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
VX_SOURCE	RDBMS Text(2)	SAS Char(2)	OD=Internal administration EF=External feed IS=Immunization Information Systems PR=Patient-reported DR=Derived NI=No information UN=Unknown OT=Other	Source of the prescribing information.	PCORnet	<ul style="list-style-type: none"> This field is a derived attribute, and is not expected to be an explicit data field within a source system Use “DR” for all immunizations that are derived or imputed through analytical procedures (e.g., natural language processing). Use “IS” for immunizations sourced from the CDC Immunization Information Systems
VX_DOSE	RDBMS Number(x)	SAS Numeric(length 8)	.	Dose of a given immunization	PCORnet	Do not impute or derive. Populate only if captured in the source system as a discrete value.
VX_DOSE_UNIT	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Units of measure associated with the dose of the immunization as delivered by the provider	UCUM	<ul style="list-style-type: none"> Do not impute or derive. Populate only if captured in the source system as a discrete value. Choose the standardized unit of measure that is most reflective of the source data. This is a mixed case value set and entries should be handled accordingly.

IMMUNIZATION Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
VX_ROUTE	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Route of immunization delivery.	SNOMED	<ul style="list-style-type: none"> The value set for Route is derived from SNOMED and may include values that are more granular than what is present in the source system. If a direct mapping is available, use the appropriate SNOMED code. If there is any possible ambiguity, use “OT” and then store the source value in RAW_VX_ROUTE. For example, an Injection could map to Subcutaneous or Intramuscular or Intraocular, depending on the drug, so that would best be mapped to “OT.” Do not impute or derive. Populate only if captured in the source system as a discrete value.
VX_BODY_SITE	RDBMS Text (x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Body site where the immunization was delivered.	FHIR – Immunization (SNOMED)	<ul style="list-style-type: none"> Most immunizations are expected in the right arm (“RA”) or left arm (“LA”)
VX_MANUFACTURER	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Manufacturer of the immunization, coded using MVX terminology.	PCORnet	
VX_LOT_NUM	RDBMS Text(x)	SAS Char(x)	.	Lot number of the immunization.	FHIR - Immunization	<ul style="list-style-type: none"> This information is typically captured within EHRs or e-prescribing as part of the ordering process. Do not impute or derive. Populate only if captured in the source system as a discrete value.

IMMUNIZATION Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
VX_EXP_DATE	RDBMS Date	SAS Date (Numeric)	.	Expiration date of the immunization.	FHIR – Immunization	
RAW_VX_NAME	RDBMS Text(x)	SAS Char(x)	.	Field for originating, full textual immunization name from the source.	PCORnet	
RAW_VX_CODE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_VX_CODE_TYPE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_VX_DOSE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_VX_DOSE_UNIT	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_VX_ROUTE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_VX_BODY_SITE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_VX_STATUS	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_VX_STATUS_REASON	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_VX_MANUFACTURER	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	

6.22. Table: HARVEST

HARVEST Domain Description:

Attributes associated with the specific PCORnet datamart implementation, including data refreshes.

Relational Integrity:

The HARVEST table contains one record per unique combination of NETWORKID and DATAMARTID.

Composite Primary Key: NETWORKID, DATAMARTID

Constraints:

NETWORKID + DATAMARTID (unique)

NETWORKID (required, not null)

DATAMARTID (required, not null)

Imputation and Obfuscation definitions:

- “No imputation or obfuscation”: For any and every date value that is present, no methods of imputation and/or obfuscation have been applied. This does not imply that every record has a date value.
- “Imputation for incomplete dates”: Some or all date values were imputed from incomplete dates, but no obfuscation was performed.
- “Date obfuscation”: Some or all date values were obfuscated, but no imputation of incomplete dates was performed. Obfuscation can also be called “shifting” or “masking.”
- “Both imputation and obfuscation”: Some or all date values were imputed, and some or all date values were obfuscated (does not necessarily need to be on the same record).

Imputation refers to the practice of adding day or month precision for incomplete dates (ie, where a specific day or specific month is not present).

Obfuscation, also known as date shifting, is a technique not recommended within PCORnet. However, where this practice exists, this table allows the situation to be recognized for analytic consideration.

HARVEST Table Implementation Guidance

Guidance

- If partners need to impute date values, whether for a portion of the date (e.g., month) or the entire string, a value of “02” should be chosen for the relevant DATE_MGMT field(s).
- If partners must impute the entire date for a field, this should only be done for those dates that are required. Optional fields should be left blank in these situations.
- All date obfuscation and imputation strategies must be thoroughly documented in the Extract, Transform and Load (ETL) Annotated Data Dictionary (ADD).
- Partners should refrain from obfuscating dates within the CDM, with the possible exception of BIRTH_DATE (see General Guidance #2). Future versions of the CDM may remove options “03” and “04” from the value set of the DATE_MGMT fields.

HARVEST Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
NETWORKID	RDBMS Text(10)	SAS Char(10)	.	This identifier is assigned by the PCORnet Distributed Research Network Operations Center (DRN OC)	PCORnet	
NETWORK_NAME	RDBMS Text(20)	SAS Char(20)	.	Descriptive name of the network. This identifier is assigned by the PCORnet Distributed Research Network Operations Center (DRN OC)	PCORnet	
DATAMARTID	RDBMS Text(10)	SAS Char(10)	.	This identifier is assigned by the PCORnet Distributed Research Network Operations Center (DRN OC)	PCORnet	
DATAMART_NAME	RDBMS Text(20)	SAS Char(20)	.	Descriptive name of the datamart. This identifier is assigned by the PCORnet Distributed Research Network Operations Center (DRN OC)	PCORnet	

HARVEST Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
DATAMART_PLATFORM	RDBMS Text(2)	SAS Char(2)	01=SQL Server 02=Oracle 03=PostgreSQL 04=MySQL 05=SAS NI=No information UN=Unknown OT=Other			
CDM_VERSION	RDBMS Number(x)	SAS Numeric(length 8)	-	Version currently implemented within this datamart (for example, 1.0, 2.0, 3.0).	PCORnet	Replicated row for ease of reading.
CDM_VERSION	RDBMS Text(3)	SAS Char(3)	010=1.0 020=2.0 030=3.0 031=3.1 040=4.0 041=4.1 050=5.0 051=5.1 060=6.0 061=6.1	Version currently implemented within this datamart.	PCORnet	
DATAMART_CLAIMS	RDBMS Text(2)	SAS Char(2)	01=Not present 02=Present NI=No information UN=Unknown OT=Other	Datamart includes claims data source(s).	PCORnet	

HARVEST Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
DATAMART_EHR	RDBMS Text(2)	SAS Char(2)	01=Not present 02=Present NI=No information UN=Unknown OT=Other	Datamart includes EHR data source(s).	PCORnet	
TOKEN_ENCRYPTION_KEY	RDBMS Text(x)	SAS Char(x)	-	Name of the key used to generate the encrypted hash tokens.	PCORnet	• This field has been deprecated as of CDM v6.1
BIRTH_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the BIRTH_DATE field in the DEMOGRAPHIC table.	PCORnet	
ENR_START_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the ENR_START_DATE field in the ENROLLMENT table.	PCORnet	

HARVEST Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ENR_END_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the ENR_END_DATE field in the ENROLLMENT table.	PCORnet	
ADMIT_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the ADMIT_DATE field in the ENCOUNTER table.	PCORnet	
DISCHARGE_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the DISCHARGE_DATE field in the ENCOUNTER table.	PCORnet	

HARVEST Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
DX_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the DX_DATE field in the ENCOUNTER table.	PCORnet	
PX_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the PX_DATE field in the PROCEDURES table.	PCORnet	
RX_ORDER_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the RX_ORDER_DATE field in the PRESCRIBING table.	PCORnet	

HARVEST Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RX_START_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the RX_START_DATE field in the PRESCRIBING table.	PCORnet	
RX_END_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the RX_END_DATE field in the PRESCRIBING table.	PCORnet	
DISPENSE_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the DISPENSE_DATE field in the DISPENSING table.	PCORnet	

HARVEST Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
LAB_ORDER_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the LAB_ORDER_DATE field in the LAB_RESULT_CM table.	PCORnet	
SPECIMEN_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the SPECIMEN_DATE field in the LAB_RESULT_CM table.	PCORnet	
RESULT_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the RESULT_DATE field in the LAB_RESULT_CM table.	PCORnet	

HARVEST Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
MEASURE_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the MEASURE_DATE field in the VITAL table.	PCORnet	This field will not be deprecated as part of has been deprecated as of CDM v6.1
ONSET_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the ONSET_DATE field in the CONDITION table.	PCORnet	
REPORT_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the REPORT_DATE field in the CONDITION table.	PCORnet	

HARVEST Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RESOLVE_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the RESOLVE_DATE field in the CONDITION table.	PCORnet	
PRO_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the PRO_DATE field in the PRO_CM table.	PCORnet	
DEATH_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the DEATH_DATE field in the DEATH table.	PCORnet	

HARVEST Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
MEDADMIN_START_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the MEDADMIN_START_DATE field in the MED_ADMIN table.	PCORnet	
MEDADMIN_STOP_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the MEDADMIN_STOP_DATE field in the MED_ADMIN table.	PCORnet	
OBSCLIN_START_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the OBSCLIN_START_DATE field in the OBS_CLIN table.	PCORnet	

HARVEST Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
OBSCLIN_STOP_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the OBSCLIN_STOP_DATE field in the OBS_CLIN table.	PCORnet	
OBSGEN_START_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the OBSGEN_START_DATE field in the OBS_GEN table.	PCORnet	
OBSGEN_STOP_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the OBSGEN_STOP_DATE field in the OBS_GEN table.	PCORnet	

HARVEST Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ADDRESS_PERIOD_START_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the ADDRESS_PERIOD_START field in the LDS_ADDRESS_HISTORY table.	PCORnet	
ADDRESS_PERIOD_END_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the ADDRESS_PERIOD_END field in the LDS_ADDRESS_HISTORY table.	PCORnet	
VX_RECORD_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the VX_RECORD_DATE field in the IMMUNIZATION table.	PCORnet	

HARVEST Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
VX_ADMIN_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the VX_ADMIN_DATE field in the IMMUNIZATION table.	PCORnet	
VX_EXP_DATE_MGMT	RDBMS Text(2)	SAS Char(2)	01=No imputation or obfuscation 02=Imputation for incomplete dates 03=Date obfuscation 04=Both imputation and obfuscation NI=No information UN=Unknown OT=Other	Data management strategy employed for the VX_EXP_DATE field in the IMMUNIZATION table.	PCORnet	
REFRESH_DEMOGRAPHIC_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the DEMOGRAPHIC table. This date should be null if the table does not have records.	PCORnet	
REFRESH_ENROLLMENT_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the ENROLLMENT table. This date should be null if the table does not have records.	PCORnet	
REFRESH_ENCOUNTER_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the ENCOUNTER table. This date should be null if the table does not have records.	PCORnet	
REFRESH_DIAGNOSIS_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the DIAGNOSIS table. This date should be null if the table does not have records.	PCORnet	
REFRESH_PROCEDURES_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the PROCEDURES table. This date should be null if the table does not have records.	PCORnet	

HARVEST Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
REFRESH_VITAL_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the VITAL table. This date should be null if the table does not have records.	PCORnet	This field will not be deprecated as part of has been deprecated as of CDM v6.1
REFRESH_DISPENSING_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the DISPENSING table. This date should be null if the table does not have records.	PCORnet	
REFRESH_LAB_RESULT_CM_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the LAB_RESULT_CM table. This date should be null if the table does not have records.	PCORnet	
REFRESH_CONDITION_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the CONDITION table. This date should be null if the table does not have records.	PCORnet	
REFRESH_PRO_CM_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the PRO_CM table. This date should be null if the table does not have records.	PCORnet	
REFRESH_PRESCRIBING_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the PRESCRIBING table. This date should be null if the table does not have records.	PCORnet	
REFRESH_PCORNET_TRIAL_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the PCORNET_TRIAL table. This date should be null if the table does not have records.	PCORnet	
REFRESH_DEATH_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the DEATH table. This date should be null if the table does not have records.	PCORnet	
REFRESH_DEATH_CAUSE_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the DEATH_CAUSE table. This date should be null if the table does not have records.	PCORnet	
REFRESH_MED_ADMIN_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the MED_ADMIN table. This date should be null if the table does not have records.	PCORnet	
REFRESH_OBS_CLIN_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the OBS_CLIN table. This date should be null if the table does not have records.	PCORnet	

HARVEST Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
REFRESH_PROVIDER_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the PROVIDER table. This date should be null if the table does not have records.	PCORnet	
REFRESH_OBS_GEN_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the OBS_GEN table. This date should be null if the table does not have records.	PCORnet	
REFRESH_HASH_TOKEN_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the HASH_TOKEN table. This date should be null if the table does not have records.	PCORnet	
REFRESH_LDS_ADDRESS_HX_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the LDS_ADDRESS_HISTORY table. This date should be null if the table does not have records.	PCORnet	
REFRESH_IMMUNIZATION_DATE	RDBMS Date	SAS Date (Numeric)	.	Most recent date on which the present data were loaded into the IMMUNIZATION table. This date should be null if the table does not have records.	PCORnet	

6.23. Table: LAB_HISTORY

LAB_HISTORY Domain Description:

Table for storing historical information about units of measure and reference ranges for laboratory test results.

Relational Integrity:

The LAB_HISTORY table contains one record per LABHISTORYID.

Primary Key: LABHISTORYID

Constraints:

LABHISTORYID (unique; required, not null)

LAB_LOINC (required, not null)

LAB_HISTORY Table Implementation Guidance

Guidance

- This table is intended to serve as a resource for partners as they develop their extract-transform-load (ETL) procedures to populate LAB_RESULT_CM. It is designed to store details related to units of measure and normal ranges for laboratory results that do not include such values at the record level. Partners can use this table as a reference during their ETL to look up and populate the relevant fields in LAB_RESULT_CM for those records.
- It is expected that partners will be able to find information on units of measure and normal range via reference material maintained by their clinical labs (electronic documents or online catalogs). It is not necessary to derive this metadata from individual lab results. **DO NOT DERIVE THESE DATA FROM LOINC RESOURCES.**
- While values for this table may need to be entered manually, a relatively small number of tests (~150) typically cover the vast majority of testing volume (>85%), which should minimize any data entry burden.
- Partners may not need to populate this table, particularly if they are already able to meet the network lab data quality metrics.
- Partners do not need to create records in this table for results that come from external reference labs (e.g., Quest, LabCorp). The DRN OC can provide this information on request.
- The DRN OC may reference this table if labs are needed as part of a particular study or analysis and the units and/or reference range is missing from the result(s). Partners may be asked to populate records in this table for those corresponding tests.
- Every record in this table should be unique.

LABHISTORYID	LAB_LOINC	LAB_FACILITYID	SEX	RACE	AGE_MIN_WKS	AGE_MAX_WKS	RESULT_UNIT	NORM_RANGE_LOW	NORM_MODIFIER_LOW	NORM_RANGE_HIGH	NORM_MODIFIER_HIGH	PERIOD_START	PERIOD_END	RAW_LAB_NAME	RAW_UNIT	RAW_RANGE
1	51733-4		1	NS	NS	0	4 %	85	EQ	97	EQ	1/1/15		O2 Sat Arterial	%	85.00 - 97.00
2	51733-4		1	NS	NS	4	7826 %	90	EQ	100	EQ	1/1/15		O2 Sat Arterial	%	90.00 - 100.00
3	2093-3		2	F	NS	0	104.356 mg/dL	44	EQ	181	EQ	1/1/13	12/31/18	Chol	mg/dL	44.00-181.00
4	2093-3		2	F	NS	104.346	7826 mg/dL		NO	199	LE	1/1/13	12/31/18	Chol	mg/dL	<=199.00
5	2093-3		2	F	NS	0	104.356 mg/dL	40	EQ	186	EQ	1/1/19		Chol	mg/dL	40.00-186.00
6	2093-3		2	F	NS	104.346	7826 mg/dL		NO	205	LE	1/1/19		Chol	mg/dL	<=205.00

Figure 1: Example of populated records for LAB_HISTORY. The first two records detail metadata for an Arterial O2 Saturation test, with different ranges for patients between 0-4 weeks and then >4 weeks. These ranges went into effect on 1/1/2015 and are still valid. The remaining 4 rows illustrate ranges and units for a serum cholesterol test in females. There is a range for patients age 0 – 2 years and then >2 years. The first set of ranges were valid from 1/1/2013 to 12/31/2018. They were replaced with new references ranges on 1/1/2019, as shown in the final two rows.

LAB HISTORY Table Specification						
Field Name	RDBMS Data Type	SAS Data Type	Predefined Value Sets and Descriptive Text for Categorical Fields	Definition / Comments	Data Element Provenance	Field-level Implementation Guidance
LABHISTORYID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique LAB HISTORY record.	PCORnet	
LAB_LOINC	RDBMS Text(10)	SAS Char(10)	.	Logical Observation Identifiers, Names, and Codes (LOINC®) from the Regenstrief Institute. Current LOINC codes are from 3-8 characters long but Regenstrief suggests a length of 10 for future growth. The last digit of the LOINC code is a check digit and is always preceded by a hyphen. All parts of the LOINC code, including the hyphen, must be included. Do not pad the LOINC code with leading zeros.	MSCDM v4.0	<ul style="list-style-type: none"> Use this field to store the LOINC code of the laboratory test. Expected format of LOINC codes: Length of 3-8, hyphen in the penultimate position, no alphabetical characters. Pre-release LOINC codes may be included in this field. Proposed codes that are not yet accepted by LOINC should not be populated. Do not use “discouraged” LOINC codes.
LAB_FACILITYID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary local code that identifies the facility providing lab results. Used in instances where partners are aggregating laboratory results from a number of different organizations. LAB_FACILITYID can be a true identifier, or a pseudoidentifier with a consistent crosswalk to the true identifier retained by the source data partner.	MSCDM v4.0 with modified field name	<ul style="list-style-type: none"> Does not have to correlate with FACILITYID in ENCOUNTER.

LAB HISTORY Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
SEX	RDBMS Text(2)	SAS Char(2)	A=Ambiguous F=Female M=Male NS = Not specified	Sex to which this unit and range applies.	MSCDM v4.0 with modified field size and value set Source: Administrative Sex (HL7) http://phinvads.cdc.gov/vads/ViewValueSet.action?id=06D34BBC-617F-DD11-B38D-00188B398520	<ul style="list-style-type: none"> Use “NS” for instances where the range and unit do not apply to a specific sex.
RACE	RDBMS Text(2)	SAS Char(2)	01=American Indian or Alaska Native 02=Asian 03=Black or African American 04=Native Hawaiian or Other Pacific Islander 05=White OT=Other NS = Not specified	Race to which this unit and range applies. 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. 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.	MSCDM v4.0 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	<ul style="list-style-type: none"> There is a push to move away from reference ranges based on race, but this field is included to capture historical data. If unit/range applies to multiple specified races, create multiple entries, one per race. Use “NS” for instances where the range and unit do not apply to a specific race.

LAB HISTORY Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
AGE_MIN_WKS	RDBMS Number(x)	SAS Numeric(length 8)	.	Minimum age to which this unit and range applies, in weeks.		<ul style="list-style-type: none"> For age cut-offs that are not listed in weeks, use the following conversions: <ul style="list-style-type: none"> Months: 4.348 weeks per month Years: 52.178 weeks per year Days: 7 days per week For unit / range records with no minimum age, use 0.
AGE_MAX_WKS	RDBMS Number(x)	SAS Numeric(length 8)	.	Maximum age to which this unit and range applies, in weeks.		<ul style="list-style-type: none"> For age cut-offs that are not listed in weeks, use the following conversions: <ul style="list-style-type: none"> Months: 4.348 weeks per month Years: 52.178 weeks per year Days: 7 days per week For unit / range records with no maximum age, use 7826 (~150 years).
RESULT_UNIT	RDBMS Text(x)	SAS Char(x)	See Value Set Reference File for a list of acceptable values.	Converted/standardized units for the specified LOINC.	UCUM	<ul style="list-style-type: none"> Chose the standardized unit of measure that is most reflective of the source data. Do not derive based on the LOINC code. This is a mixed case value set and entries should be handled accordingly.
NORM_RANGE_LOW	RDBMS Text(10)	SAS Char(10)	.	Lower bound of the normal range assigned by the laboratory. Value should only contain the value of the lower bound. The symbols >, <, >=, <= should be removed. For example, if the normal range for a test is >100 and <300, then "100" should be entered.	MSCDM v4.0	

LAB HISTORY Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
NORM_MODIFIER_LOW	RDBMS Text(2)	SAS Char(2)	EQ=Equal GE=Greater than or equal to GT=Greater than NO=No lower limit NI=No information UN=Unknown OT=Other	Modifier for NORM_RANGE_LOW values. For numeric results one of the following needs to be true: 1) Both MODIFIER_LOW and MODIFIER_HIGH contain EQ (e.g. normal values fall in the range 3-10) 2) MODIFIER_LOW contains GT or GE and MODIFIER_HIGH contains NO (e.g. normal values are >3 with no upper boundary) 3) MODIFIER_HIGH contains LT or LE and MODIFIER_LOW contains NO (e.g. normal values are <=10 with no lower boundary)	MSCDM v4.0 with modified value set and field name	
NORM_RANGE_HIGH	RDBMS Text(10)	SAS Char(10)	.	Upper bound of the normal range assigned by the laboratory. Value should only contain the value of the upper bound. The symbols >, <, >=, <= should be removed. For example, if the normal range for a test is >100 and <300, then "300" should be entered.	MSCDM v4.0 with modified field length	
NORM_MODIFIER_HIGH	RDBMS Text(2)	SAS Char(2)	EQ=Equal LE=Less than or equal to LT=Less than NO=No higher limit NI=No information UN=Unknown OT=Other	Modifier for NORM_RANGE_HIGH values. For numeric results one of the following needs to be true: 1) Both MODIFIER_LOW and MODIFIER_HIGH contain EQ (e.g. normal values fall in the range 3-10) 2) MODIFIER_LOW contains GT or GE and MODIFIER_HIGH contains NO (e.g. normal values are >3 with no upper boundary) 3) MODIFIER_HIGH contains LT or LE and MODIFIER_LOW contains NO (e.g. normal values are <=10 with no lower boundary)	MSCDM v4.0 with modified value set and field name	

LAB_HISTORY Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PERIOD_START	RDBMS Date	SAS Date (Numeric)	.	Initial date when the unit and range are known to be in use for the specified age range, race, sex and facility.		<ul style="list-style-type: none"> If the start date of the unit and range is unknown, leave blank. There should only be one blank PERIOD_START for a given LOINC, Facility, Sex, Race, Age range, Unit and normal range.
PERIOD_END	RDBMS Date	SAS Date (Numeric)	.	Date when the unit and range are no longer in use for the specified age range, race, sex and facility.		<ul style="list-style-type: none"> Only the current period is expected to have a null value for end date. All previous periods are expected to have a defined end date.
RAW_LAB_NAME	RDBMS Text(x)	SAS Char(x)	.	Local name related to an individual lab test.	PCORnet	
RAW_UNIT	RDBMS Text(x)	SAS Char(x)	.	Original units in your source data.	PCORnet	
RAW_RANGE	RDBMS Text(x)	SAS Char(x)	.	Original reference range in your source data.	PCORnet	

7. Supplemental Table Specifications

7.1. Supplemental Table: PRIVATE_DEMOGRAPHIC

PRIVATE_DEMOGRAPHIC Domain Description:

Protected table that is intended to provide a standardized representation of the personally-identifiable information (PII) that is needed to support local activities related to record linkage.

Relational Integrity:

The PRIVATE_DEMOGRAPHIC table contains one record per PATID.

Primary Key: PATID

Foreign Key:

PRIVATE_DEMOGRAPHIC.PATID is a foreign key to DEMOGRAPHIC.PATID (one-to-one relationship)

Constraints:

PATID (unique, required, not null)

PAT_FIRSTNAME (required, not null)

PAT_LASTNAME (required, not null)

PRIVATE_DEMOGRAPHIC Table Implementation Guidance

Guidance

- This table is intended to provide a standardized representation for PII used in PCORnet privacy-preserving record linkage. It is not necessary to create this table if existing processes can generate the relevant PII from local source system(s).
- For DataMarts that incorporate data from multiple organizations/institutions, PRIVATE_DEMOGRAPHIC can be used to assist de-duplication efforts. See field-level guidance for more information.
- Include most recent values for each field.
- This table can be kept logically and physically separate from the rest of the CDM.
- This table replicates the fields of the DEMOGRAPHIC table. They are included in case partners wish to create a master table that can be used to populate DEMOGRAPHIC within the CDM. SAS data types are included for the replicated fields to help with ETL programming, but partners are not expected to create a SAS-based version of this PRIVATE table. It is not necessary to populate the replicated fields in this table if DEMOGRAPHIC is loaded through a separate ETL process.

PRIVATE_DEMOGRAPHIC Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier. Used to link across tables. Corresponds to PATID in the DEMOGRAPHIC table.	MSCDM v4.0	All PATIDs should be present in the DEMOGRAPHIC table.
ORG_PATID	RDBMS Text(x)		.	Corresponds to PATID in the originating CDM if receiving data from multiple managing organizations.	PCORnet	Can leave null if not combining records for multiple organizations.
MANAGING_ORG	RDBMS Text(x)		.	Organization where record originated	PCORnet	
MRN	RDBMS Text(x)			Local Medical Record Number for the patient	PCORnet	
PAT_FIRSTNAME	RDBMS Text(x)		.	Given name of the patient.	PCORnet	
PAT_MIDDLENAME	RDBMS Text(x)			Middle name of the patient.	PCORnet	
PAT_LASTNAME	RDBMS Text(x)		.	Surname or family name.	PCORnet	
PAT_MAIDENNAME	RDBMS Text(x)		.	Surname or family name prior to marriage.	PCORnet	

PRIVATE DEMOGRAPHIC Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PAT_SSN	RDBMS Text(x)		.	Social Security Number (SSN; whole or part). If only part of the SSN is available, replace digits with the letter "X" (e.g. XXX-XX-8453).	FHIR-managing Organization	<ul style="list-style-type: none"> Hyphens can be included.
BIRTH_DATE	RDBMS Date	SAS Date (Numeric)	.	Date of birth. Corresponds to BIRTH_DATE in the DEMOGRAPHIC table.	MSCDM v4.0	
BIRTH_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)	.	Time of birth. Corresponds to BIRTH_TIME in the DEMOGRAPHIC table.	PCORnet Source of time format: ISO 8601	
PRIMARY_EMAIL	RDBMS Text(x)		.	Primary e-mail address for the patient.	PCORnet	
PRIMARY_PHONE	RDBMS Text(10)		.	Primary phone number for the patient (if known). 10-digit US phone number.	PCORnet.	<ul style="list-style-type: none"> Remove punctuation (parentheses, hyphens). Cell phone preferred over home phone.
SEX	RDBMS Text(2)	SAS Char(2)	A=Ambiguous F=Female M=Male NI=No information UN=Unknown OT=Other	Sex assigned at birth. Corresponds to SEX in the DEMOGRAPHIC table.	MSCDM v4.0 with modified field size and value set Source: Administrative Sex (HL7) http://phinvads.edc.gov/vads/ViewValueSet.action?id=06D34BBC-617F-DD11-B38D-00188B398520	The "Ambiguous" category may be used for individuals who are physically undifferentiated from birth. The "Other" category may be used for individuals who are undergoing gender re-assignment. If a value of "X" is recorded in the source system, map to "OT".

PRIVATE DEMOGRAPHIC Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
SEXUAL_ORIENTATION	RDBMS Text(2)	SAS Char(2)	AS=Asexual BI=Bisexual GA=Gay LE=Lesbian QU=Queer HO=Lesbian, Gay or Homosexual QS=Questioning ST=Straight SE=Something else MU=Multiple sexual orientations DC=Decline to answer NI=No information UN=Unknown OT=Other	Sexual orientation. Corresponds to SEXUAL_ORIENTATION in the DEMOGRAPHIC table.	Source: Health IT Certification Criteria, 2015 Base Edition, modified with expert advisory within PCORnet https://www.federalregister.gov/documents/2015/10/16/2015-25597/2015-edition-health-information-technology-health-it-certification-criteria-2015-edition-base	<ul style="list-style-type: none"> Sites should map to the most granular value supported by their data. The entry for “Lesbian, Gay, or Homosexual” was added for those partners who capture sexual orientation based on the ONC Meaningful Use value set.
GENDER_IDENTITY	RDBMS Text(2)	SAS Char(2)	M=Man F=Woman TM=Transgender male/Trans man/Female-to-male TF=Transgender female/Trans woman/Male-to-female GQ=Genderqueer/Non-binary SE=Something else MU=Multiple gender categories DC=Decline to answer NI=No information UN=Unknown OT=Other	Current gender identity. Corresponds to GENDER_IDENTITY in the DEMOGRAPHIC table.	Source: Health IT Certification Criteria, 2015 Base Edition, modified with expert advisory within PCORnet https://www.federalregister.gov/documents/2015/10/16/2015-25597/2015-edition-health-information-technology-health-it-certification-criteria-2015-edition-base	Use Genderqueer (“GQ”) for patients who report a non-binary gender identify.

PRIVATE DEMOGRAPHIC Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
HISPANIC	RDBMS Text(2)	SAS Char(2)	Y=Yes N=No R=Refuse to answer 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. Corresponds to HISPANIC in the DEMOGRAPHIC table.	MSCDM v4.0 with modified field size and value set Compatible with “OMB Hispanic Ethnicity” (Hispanic or Latino, Not Hispanic or Latino)	Populating RACE and HISPANIC if race and ethnicity are not captured separately within the source system (e.g., “Hispanic or Latino” is included as a selection under Race) - for patients with a known race (e.g., Race is something other than “Hispanic or Latino”, partners should set HISPANIC to "OT" and RACE to the appropriate race code. For patients who are listed as having a race of “Hispanic,” partners should set HISPANIC to "Y" and RACE to "OT". In this situation, the combined race/ethnicity field is treated as known field capturing values for both race and ethnicity, which is why the preference is to use “OT” instead of “NI”.

PRIVATE DEMOGRAPHIC Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RACE	RDBMS Text(2)	SAS Char(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	<p>Please use only one race value per patient. Corresponds to RACE in the DEMOGRAPHIC table.</p> <p>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. 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.</p>	<p>MSCDM v4.0 with modified field size and value set</p> <p>Original value set is based upon U.S. Office of Management and Budget (OMB) standard, and is compatible with the 2010 U.S. Census</p>	
BIOBANK_FLAG	RDBMS Text(1)	SAS Char(1)	Y=Yes N=No	<p>Flag to indicate that one or more biobanked specimens are stored and available for research use. Examples of biospecimens could include blood, urine, or tissue (eg, skin cells, organ tissues). If biospecimens are available, locally maintained “mapping tables” would be necessary to map between the DEMOGRAPHIC record and the originating biobanking system(s).</p> <p>If no known biobanked specimens are available, this field should be marked “No”.</p>	PCORnet	This field is a derived attribute and is not expected to be an explicit data field within a source system

PRIVATE_DEMOGRAPHIC Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
PAT_PREF_LANGUAGE_SPOKEN	RDBMS Text(3)	SAS Char(3)	See Value Set Reference File for a list of acceptable values.	Preferred spoken language of communication as expressed by the patient. Corresponds to PAT_PREF_LANGUAGE_SPOKEN in the DEMOGRAPHIC table.	ISO639-2	<ul style="list-style-type: none"> This information may be documented in the EHR or an external registry. Do not impute or derive if not expressly defined in the source system. Analytically, will assume that "NI" corresponds to a preferred language of English. Use the value of "ZHO" (Chinese) for both Mandarin and Cantonese, and place the specific value in the RAW field. Within the ISO639-2 value set, there is no distinction between the two. https://www.loc.gov/standards/iso639-2/faq.html#24
RAW_PAT_NAME	RDBMS Text(x)		.	Full name of the patient, as represented in the source.	PCORnet	<ul style="list-style-type: none"> If components of patient name are not stored as separate fields, can be used to store full name before parsing into first/middle/last. If components of name are stored separately, they can be combined into a single string in this field (first middle last).

PRIVATE_DEMOGRAPHIC Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
RAW_SEX	RDBMS Text(x)	SAS Char(x)	.	Field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_SEXUAL_ORIENTATION	RDBMS Text(x)	SAS Char(x)	.	Field for originating value of field, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_GENDER_IDENTITY	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_HISPANIC	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_RACE	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	
RAW_PAT_PREF_LANGUAGE_SP OKEN	RDBMS Text(x)	SAS Char(x)	.	Field for originating value, prior to mapping into the PCORnet CDM value set.	PCORnet	

7.2. Supplemental Table: PRIVATE_ADDRESS_HISTORY

PRIVATE_ADDRESS_HISTORY Domain Description:

Protected table that can be used to store elements of a patient's address that are considered personal health information (PHI). These data can be used for geocoding or other linkage projects.

Relational Integrity:

The PRIVATE_ADDRESS_HISTORY table can contain many records per patient.

Primary Key: ADDRESSID

Foreign Key:

PRIVATE_ADDRESS_HISTORY.PATID is a foreign key to DEMOGRAPHIC.PATID (many-to-one relationship)

Constraints:

ADDRESSID (unique; required, not null)

PATID (required, not null)

ADDRESS_USE (required, not null)

ADDRESS_TYPE (required, not null)

ADDRESS_PREFERRED (required, not null)

PRIVATE_ADDRESS_HISTORY Table Implementation Guidance

Guidance

- The only address-related information used by the current configuration of the PCORnet privacy-preserving record linkage solution is a patient's current zip code. It is not necessary to create and populate this table solely for that field. However, partners who are participating in the PCORnet geographic surveillance pilots or those who wish to have their patient populations included in any future geographic characterizations of PCORnet will need to populate LDS_ADDRESS_HISTORY. In those situations, there may be a benefit in using PRIVATE_ADDRESS_HISTORY as a local staging table.
- Expect multiple records per individual.
- This table is currently limited to addresses in the United States.
- Partners can limit records in this table to validated addresses if known.
- If partners have difficulty constructing a longitudinal address history for patients within their DataMart, they should prioritize populating the current address for each patient.
- Geocoded information about the address should be stored in PRIVATE_ADDRESS_GEOCODE.
- This table can be kept logically and physically separate from the rest of the CDM. The LDS_ADDRESS_HISTORY table represents a subset of this table. It removes the fields that are considered personal health information under HIPAA (ADDRESS_STREET, ADDRESS_DETAIL, RAW_ADDRESS_TEXT).
- This table replicates the fields of the LDS_ADDRESS_HISTORY table. They are included in case partners wish to create a master table that can be used to populate LDS_ADDRESS_HISTORY within the CDM. SAS data types are included for the replicated fields to help with ETL programming, but partners are not expected to create a SAS-based version of this PRIVATE table. It is not necessary to populate the replicated fields in this table if LDS_ADDRESS_HISTORY is loaded through a separate ETL process.

PRIVATE_ADDRESS_HISTORY Table Specification

<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ADDRESSID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary identifier for each unique address record.	PCORnet	
PATID	RDBMS Text(x)	SAS Char(x)	.	Arbitrary person-level identifier. Used to link across tables.	MSCDM v4.0	All PATIDs should be present in the DEMOGRAPHIC table.

PRIVATE_ADDRESS_HISTORY Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ADDRESS_USE	RDBMS Text(2)	SAS Char(2)	HO=Home WO=Work TP=Temp OL=Old/Incorrect NI=No information UN=Unknown OT=Other	<p>Purpose of the address.</p> <p>Details of categorical definitions: Home: A communication address at home. Work: An office address. First choice for business-related contacts during business hours. Temp: A temporary address. Old/Incorrect: This address is no longer in use (or was never correct but retained for records).</p>	FHIR - ADDRESS	<ul style="list-style-type: none"> • This field may be a derived attribute that is not an explicit data field within a source system. • Use the period start/end fields to indicate if an address is no longer valid. Do not change values of HO/WO/TP to OL if a new address is available. • The old/incorrect value is included in case partners are doing a bulk load and it is present in their source system. It is acceptable to exclude these records, however. • If addresses within the source system are reasonably expected to represent the patient's home address, it is acceptable to mark these as "HO."

PRIVATE_ADDRESS_HISTORY Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ADDRESS_TYPE	RDBMS Text(2)	SAS Char(2)	PO=Postal PH=Physical BO=Both NI=No information UN=Unknown OT=Other	Type of address. Details of categorical definitions: Postal: mailing address – PO Boxes and care-of addresses. Physical: A physical address that can be visited. Both: An address that is both physical and postal.	FHIR - ADDRESS	<ul style="list-style-type: none"> • This field may be a derived attribute that is not an explicit data field within a source system. • Addresses that are clearly not postal-only addresses can be marked as “Both”

PRIVATE_ADDRESS_HISTORY Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ADDRESS_PREFERRED	RDBMS Text(2)	SAS Char(2)	Y=Yes N=No	Indicates whether this address is the preferred one for a given patient, address use and address type within a given address period.	PCORnet	<ul style="list-style-type: none"> • This field is a derived attribute and is not expected to be an explicit data field within a source system. • It is possible to have an address marked as preferred for every address type/use within each defined address period. Partners do not need to set the flag back to “N” every time the address is updated. • If only one address is present for a given period, that address can be marked as preferred. • If multiple addresses are present for a period, one should be denoted as preferred. • If multiple addresses are present for a period without clear institutional guidance as to which is preferred, partners can use a heuristic to make a determination (i.e., address associated with most recent encounter, address used for billing, etc.)

PRIVATE_ADDRESS_HISTORY Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ADDRESS_STREET	RDBMS Text(x)		.	Primary address line (e.g., street name and number)	PCORnet	If possible, standardize according to US Post Office conventions.
ADDRESS_DETAIL	RDBMS Text(x)		.	Remaining address details (e.g., suite, post office box, other details)		<ul style="list-style-type: none"> Information in this field may be used for deduplication/linkage of people who live in the same building. Contents are unlikely to add additional information when geocoding.
ADDRESS_COUNTY	RDBMS Text(x)		.	The name of the county associated with the address.	PCORnet	
ADDRESS_CITY	RDBMS Text(x)	SAS Char(x)	.	The name of the city, town, village or other community.	FHIR - ADDRESS	
ADDRESS_STATE	RDBMS Text(2)	SAS Char(2)	See Value Set Reference File for a list of acceptable values.	State, as represented by 2-digit postal abbreviation.	PCORnet	
ADDRESS_ZIP5	RDBMS Text(5)	SAS Char(5)	.	5-digit postal code for the address.	FHIR - ADDRESS	
ADDRESS_ZIP9	RDBMS Text(9)	SAS Char(9)	.	9-digit postal code for the address.		Do not include hyphens.
ADDRESS_PERIOD_START	RDBMS Date	SAS Date (Numeric)	.	Initial date when the address known to be in use.	FHIR - ADDRESS	If the date the address is known to be in use is unknown, it is acceptable to use the date the record was created in the source system or the date the record was first created in the CDM.

PRIVATE_ADDRESS_HISTORY Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
ADDRESS_PERIOD_END	RDBMS Date	SAS Date (Numeric)	.	Date when address was no longer in use.	FHIR - ADDRESS	Only the current address period should have a null value. All previous periods are expected to have a defined end date.
RAW_ADDRESS_TEXT	RDBMS Text(x)		.	Text representation of the address	FHIR - ADDRESS	The full address can be represented in this field and may contain information beyond the street and the PO Box.

7.3. Supplemental Table: PRIVATE_ADDRESS_GEOCODE

PRIVATE_ADDRESS_GEOCODE Domain Description:
Protected table that can be used to store the results of geocoding algorithms.

Relational Integrity:

The PRIVATE_ADDRESS_GEOCODE table can contain many records per ADDRESSID.

Primary Key: GEOCODEID

Foreign Key:

PRIVATE_ADDRESS_GEOCODE.ADDRESSID is a foreign key to PRIVATE_ADDRESS.ADDRESSID (many-to-one relationship)

Constraints:

GEOCODEID (unique; required, not null)

ADDRESSID (required, not null)

PRIVATE_ADDRESS_GEOCODE Table Implementation Guidance

Guidance

- This table can contain more than one record per address.
- This table is designed to accommodate documentation of algorithmic linkage.
- This table can be kept logically and physically separate from the rest of the CDM.
- There are no implementation expectations related to this table at this time. Future PCORnet projects may reference it in some fashion, but in the meantime, partners may utilize it if they require geocoding for any of their local analyses.

PRIVATE_ADDRESS_GEOCODE Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
GEOCODEID	RDBMS Text(x)		.	Arbitrary identifier for each unique geocode record.	PCORnet	
ADDRESSID	RDBMS Text(x)		.	Arbitrary identifier for each unique address record.	PCORnet	All ADDRESSIDs should be in the PRIVATE_ADDRESS_HISTORY table.
GEOCODE_STATE	RDBMS Text(2)		See Value Set Reference File for a list of acceptable values.	State associated with the address.	PCORnet	
GEOCODE_COUNTY	RDBMS Text(x)		.	County associated with the address.	PCORnet	
GEOCODE_LONGITUDE	RDBMS Text(x)		.	Longitude assigned to the address by the geocoding algorithm. WGS84 format.	FHIR – Location.position.longitude	
GEOCODE_LATITUDE	RDBMS Text(x)		.	Latitude assigned to the address by the geocoding algorithm. WGS84 format.	FHIR – Location.position.latitude	
GEOCODE_BLOCK	RDBMS Text(x)		.	Census Block associated with the address.	PCORnet	
GEOCODE_TRACT	RDBMS Text(x)		.	Census Tract associated with the address.	PCORnet	
GEOCODE_GROUP	RDBMS Text(x)		.	Census Block Group associated with the address.	PCORnet	
GEOCODE_ZCTA	RDBMS Text(x)		.	Zip code tabulation area associated with the address.	PCORnet	
GEOCODE_CUSTOM	RDBMS Text(x)		.	Field that can be used to store a custom/local geocode for a given address.	PCORnet	
GEOCODE_CUSTOM_TEXT	RDBMS Text(x)		.	Text description of the terminology/approach used to assign the custom geocode.	PCORnet	

PRIVATE_ADDRESS_GEOCODE Table Specification						
<i>Field Name</i>	<i>RDBMS Data Type</i>	<i>SAS Data Type</i>	<i>Predefined Value Sets and Descriptive Text for Categorical Fields</i>	<i>Definition / Comments</i>	<i>Data Element Provenance</i>	<i>Field-level Implementation Guidance</i>
SHAPEFILE	RDBMS Text(x)		.	Name of shapefile used in the geocoding process.	PCORnet	<ul style="list-style-type: none"> • Include year along with file name.
GEO_ACCURACY	RDBMS Text(2)		Z9=ZIP9 Z5=ZIP5 CN=County CY=City ST=State TR=Tract BL=Block SR=Street NI=No information UN=Unknown OT=Other	Level of accuracy of the geocoded address based on the method used. ZIP9: 9-digit zip code ZIP5: 5-digit zip code County: County level City: City, village or community State: State Tract: Census tract Block: Census block Street: Street	PCORnet	<ul style="list-style-type: none"> • The geocoding software should provide this information after addresses are geocoded. • Important when merging data collected/aggregated at different levels.
GEO_PROV_REF	RDBMS Text(x)		.	Reference to the methodology/software and parameters used to assign the geocode.	PCORnet	
ASSIGNMENT_DATE	RDBMS Date		.	Date that the geocoding was completed.	PCORnet	