



# Work Plan: Data Curation Query Package

---

**V6.06**  
**Jan 03, 2023**

## TABLE OF CONTENTS

I.	PURPOSE AND SCOPE.....	3
II.	POTENTIAL CODE ERRORS QUERY .....	5
III.	DATA CURATION QUERY DEFINITIONS .....	8
IV.	DATA CURATION QUERY OUTPUT TABLES .....	10
V.	EMPIRICAL DATA CURATION REPORT.....	24
VI.	PROGRAM PACKAGE FILE STRUCTURE .....	26
VII.	FILES INCLUDED IN QUERY REQUEST .....	27
VIII.	OUTPUT FILES .....	28
IX.	QUERY INPUT AND OUTPUT DIAGRAM.....	31
X.	RESPONDING TO THE QUERY PACKAGE .....	32
XI.	VERSION HISTORY .....	35

## **I. Purpose and Scope**

The purpose of the Data Curation Query Package v6.06 is to characterize the data in PCORnet Common Data Model (CDM) v6.0. This package examines all 23 tables. The package consists of the Potential Code Errors query, a Data Curation query and an Empirical Data Curation Report which summarizes key information from the query output and evaluates the results against PCORnet's Data Check v13. Output tables will be produced by running SAS programs against static local DataMarts in PCORnet CDM v6.0 with SAS data types.

Query results will be used by the PCORnet Coordinating Center's Distributed Research Network Operations Center (DRN OC) to ensure a foundational level of data quality across the networks. Approved results may be used to provide initial feasibility estimates for prep-to-research queries, inform study planning activities, and to create DataMart-level, CRN-level or network-level reports. Data aggregated at the network level may be shared publicly. DataMart-level results may be published within PCORnet and data from this query can be used to inform the PCORnet Coordinating Center of data availability and fitness for use in response to PCORnet queries, to enable the Coordinating Center to share high-level counts of data variables with requestors, to present PCORnet-level Aggregate Data on public-facing websites, (e.g., [PCORnet.org](http://PCORnet.org)) in manuscripts consistent with the guidelines of the International Committee of Medical Journal Editors ("ICMJE"), and for use in PCORnet marketing materials.

To provide the DRN OC with additional insight into the query results, the ETL Annotated Data Dictionary (ETL ADD) must be updated prior to submitting the response to this query.

### **Low Cell Count Threshold**

Users may specify a low cell count threshold that establishes the minimum number of observations required to protect against possible identification of subject data. Query results greater than zero and less than the threshold will be changed to BT (below threshold) and treated as missing. For example, if a DataMart sets a low cell count threshold of 5, cell counts between 1 and 4 will be changed to BT. The low cell count threshold applies to all query results except for descriptive statistics. The low cell count threshold treatment for each query is shown in Section IV.

The default low cell count threshold value is set to zero (0) in accordance with the CRN Scope of Work. If this value is changed it will be highlighted in orange in the Empirical Data Curation report.

### **Potential Code Errors Report**

The query package produces a Potential Code Errors report which (a) identifies exceptions to the expected code length or content for ICD9/ICD10 diagnosis codes; ICD9/ICD10-PCS and CPT/HCPCS procedure codes; LOINC codes; RXNORM\_CUI codes; and NDC codes by applying the heuristics described in Section II and (b) identifies LOINC codes that are in the incorrect CDM table based on the LOINC CLASSTYPE attribute. The expected CLASSTYPE for each CDM table is included in the CDM specification and listed in Section II.

### **Data Curation tables**

The query package produces up to 295 query output tables depending on how many CDM tables are populated and how the program is executed. Information about each output table is provided in Section IV.

## Empirical Data Curation Report

The query package produces an Empirical Data Curation (EDC) report. The EDC Report summarizes key information from the data curation query output tables and identifies exceptions to the PCORnet Data Checks. The table of contents is shown in Section V.

### Lookback Date

The lookback date is the earliest date included in the query results. The lookback date will be calculated by subtracting the lookback period from the Query Response Date which is the date the data curation query (specifically, the cross-table portion) was run. The default lookback period is 10 years (120 months). **The look back value should not be modified unless you are instructed to do so.**

This date restriction does not apply to the DEMOGRAPHIC, DEATH, HARVEST, DEATH\_CAUSE, PROVIDER, LDS\_ADDRESS\_HISTORY, HASH\_TOKEN, or LAB\_HISTORY tables. The dates used to select records in the other tables are ENROLLMENT.ENR\_START\_DATE, ENROLLMENT.ENR\_END\_DATE, ENCOUNTER.ADMIT\_DATE, DIAGNOSIS.ADMIT\_DATE, PROCEDURES.ADMIT\_DATE, VITAL.MEASURE\_DATE, DISPENSING.DISPENSE\_DATE, LAB\_RESULT\_CM.RESULT\_DATE, CONDITION.REPORT\_DATE, PRO\_CM.PRO\_DATE, PRESCRIBING.RX\_ORDER\_DATE, PCORNET\_TRIAL.TRIAL\_ENROLL\_DATE, MED\_ADMIN.MEDADMIN\_START\_DATE, OBS\_CLIN.OBSCLIN\_START\_DATE, OBS\_GEN.OBSGEN\_START\_DATE and IMMUNIZATION.VX\_RECORD\_DATE. Since some of these variables do not have to be populated, records with missing dates will also be included in the data curation query results.

Questions about this query package should be sent to Jennifer Xu ([Jennifer.xu@duke.edu](mailto:Jennifer.xu@duke.edu)) and Laura Qualls ([laura.qualls@duke.edu](mailto:laura.qualls@duke.edu)).

## II. Potential Code Errors Query

The purpose of the Potential Code Errors Program is to help network partners identify exceptions to the expected formats for selected codes. Output tables will be produced by running SAS programs against static local DataMarts in PCORnet CDM 6.0 format with SAS data types.

This program has two components. First, it identifies exceptions to the expected code length or content for 8 coding terminologies used in one or more CDM tables. Heuristics are conservative to allow for all potential implementations (e.g. current LOINC codes are 5+ digits, but the program allows for the shorter deprecated codes; ICD10 procedure codes must be 7 alphanumeric characters). These heuristics will **not** identify all erroneous codes, and will only evaluate codes which are classified as one of the qualifying terminologies (e.g. a ICD9 diagnosis code labeled as a SNOMED-CT code type will not be evaluated). The CDM specifications (available on [pcorner.org](http://pcorner.org)) provide guidance on addressing potential errors (General Implementation Guidance issue #5). The heuristics and CDM tables for each terminology are shown in the table below.

Terminology	Unexpected length (after removing decimals if applicable)	Unexpected string	Unexpected alphabetical character	Unexpected numeric character	CDM tables
CPT/HCPCS	Less than 5	00000x or 99999x	n/a	No numeric characters	IMMUNIZATION; OBS_GEN; PROCEDURES;
CVX	Not 2-3	n/a	Any alphabetical character	n/a <sup>c</sup>	IMMUNIZATION
ICD9 diagnosis	Not 3-5	000.x	Any alphabetical character other than E or V	No numeric characters	CONDITION; DIAGNOSIS; OBS_GEN
ICD10 diagnosis	Not 3-7	000.x or 999.x	First character is not alphabetical	No numeric characters	CONDITION; DIAGNOSIS; OBS_GEN
ICD9 procedures	Not 3-4	00.00	Any alphabetical character	n/a <sup>c</sup>	OBS_GEN; PROCEDURES
ICD10 procedures	Not 3-7 <sup>a</sup>	0000000 or 9999999	n/a	n/a	OBS_GEN; PROCEDURES
LOINC	Not 3-8 <sup>b</sup>	No hyphen in the penultimate position	Any alphabetical character	n/a <sup>c</sup>	LAB_RESULT_CM; LAB_HISTORY; OBS_CLIN; OBS_GEN
NDC	Not 11	00000000000 or 99999999999	Any alphabetical character <sup>2</sup>	n/a <sup>c</sup>	DISPENSING; IMMUNIZATION; MED_ADMIN; OBS_GEN; PROCEDURES
RXNORM_CUI	Not 2-7	n/a	Any alphabetical character	n/a <sup>c</sup>	IMMUNIZATION; MED_ADMIN; OBS_GEN; PRESCRIBING

a. Must be 7 digits for billing but this allows for shorter codes which may appear in electronic health record data.

b. Current LOINC codes are 5+ digits but this allows for shorter deprecated codes.

c. Redundant with the unexpected alphabetical character rule.

The program also identifies LOINC codes that are in the incorrect CDM table based on the LOINC CLASSTYPE attribute. This section contains content from LOINC® (<http://loinc.org>). The LOINC table, LOINC codes, and LOINC panels and forms file are copyright © 1995-2022, Regenstrief Institute, Inc. and the Logical Observation Identifiers Names and Codes (LOINC) Committee and available at no cost under the license at <http://loinc.org/terms-of-use>. This program uses LOINC release 2.73. The expected CLASSTYPE for each CDM table is specified in the CDM specification and listed below:

CDM table	Expected CLASSTYPE(s)
LAB_RESULT_CM	1 (laboratory class)
LAB_HISTORY	1 (laboratory class)
OBS_CLIN	2 (clinical class)
OBS_GEN	2 (clinical class), 3 (Claims attachments) or 4 (Survey)

The data dictionary for the tables produced by the Potential Code Errors query is as follows:

#### code\_summary

*Note: This dataset will include 1 row for each observed combination of table and code type.*

Field	Description
DATAMARTID	
QUERY_DATE	Query run date
TABLE	CONDITION, DIAGNOSIS, DISPENSING, IMMUNIZATION, LAB_RESULT_CM, MED_ADMIN, OBS_CLIN, OBS_GEN, PRESCRIBING, and PROCEDURES
CODE_TYPE	09, DX09, PX09, 10, DX10, PX10, CH, CX, ND, LC, and RX
RECORDS	Count of records
CODES	Count of distinct codes
BAD_RECORDS	Count of potentially bad records
BAD_CODES	Count of potentially bad codes
BAD_RECORD_PCT	The percent of records which are potentially bad records. Calculated as $\text{Bad\_Records} / \text{Records} * 100$ .
MISPLACED_RECORDS	The percent of records which are placed in the wrong CDM table. Only applies to records with CODE_TYPE=LC, and will be 0 for all other code types.
MISPLACED_CODES	The percent of codes which are placed in the wrong CDM table. Only applies to records with CODE_TYPE=LC, and will be 0 for all other code types.
MISPLACED_RECORD_PCT	The percent of records which are placed in the wrong CDM table. Only applies to records with CODE_TYPE=LC, and will be 0 for all other code types. Calculated as $\text{Misplaced\_Records} / \text{Records} * 100$ .

**bad\_condition; bad\_dx; bad\_immunization; bad\_px; bad\_disp; bad\_pres; bad\_lab; bad\_lab\_hist; bad\_medadmin; bad\_obsclin; bad\_obsgen**

*Note: Tables will only be created if the table contains at least 1 record for the relevant code types.*

Field	Description
[Varies]	Pseudoidentifier for the table, e.g. encounterID
CODE_TYPE	09, 10, CH, CX, ND, LC, or RX
CODE	The code. This field is renamed from the table-specific fields, e.g. DX, PX, LAB_LOINC, and MEDADMIN_CODE.
CODE_CLEAN	Uppercase code which discards decimals, dashes, commas, spaces and trailing blanks
CODE_LENGTH	Length of code_clean
ANYALPHA	The position of the first alphabetical character; 0 if there are no alphabetic characters

ANYDIGIT	The position of the first numeric character; 0 if there are no numeric characters
TABLE	DIAGNOSIS, DISPENSING, LAB_RESULT_CM, MED_ADMIN, OBS_GEN, OBS_CLIN, PRESCRIBING, or PROCEDURES
UNEXP_LENGTH	Error indicator. Yes=1; No=0; null=not applicable.
UNEXP_ALPHA	Error indicator. Yes=1; No=0; null=not applicable.
UNEXP_STRING	Error indicator. Yes=1; No=0; null=not applicable.
UNEXP_NUMERIC	Error indicator. Yes=1; No=0; null=not applicable.

#### **misplaced\_loincs**

<b>Field</b>	<b>Description</b>
TABLE	LAB_RESULT_CM, LAB_HISTORY, OBS_CLIN, or OBS_GEN
RECORDID	The identifier for the record. This field is renamed from the table-specific fields (LAB_RESULT_ID, LABHISTORYID, OBSCLINID, or OBSGENID)
CODE	The LOINC code. This field is renamed from the table-specific fields (LAB_LOINC, OBSGEN_CODE, OBSCLIN_CODE)
CODE_DESC	The LONG_COMMON_NAME of the LOINC code.
CLASSTYPE	1, 2, 3 or 4
EXPECTED_CLASSTYPE	1 (for LAB_RESULT_CM and LAB_HISTORY), 2 (for OBS_CLIN), or 2,3, or 4 (for OBS_GEN)

### III. Data Curation Query Definitions

The definitions for variables included in the query output are as follows:

- **ADMIT\_DATE Mismatch:** These fields are replicated from the ENCOUNTER table to the PROCEDURES and DIAGNOSES table. The number of mismatched records is the number of records in PROCEDURES or DIAGNOSIS where these fields do not match the value in the ENCOUNTER table.
- **ALL\_N or RECORD\_N or N:** Count of records with non-missing values for the specified field.
- **DATASET:** CDM table name
- **DIAGNOSISID\_PLUS:** DIAGNOSIS records where (missing<ADMIT\_DATE<=REFRESH\_MAX or missing<DX\_DATE<=REFRESH\_MAX) and ENCOUNTERID is not null and DX\_TYPE is not in ('NI','UN','OT','').
- **DISTINCT\_N:** Count of records with unique values for the specified field.
- **DISTINCT\_N\_PCT:** The percent of records with unique values for the specified field.
- **DISTINCT\_ENC\_ID\_N:** Count of records with unique values for ENCOUNTERID.
- **DISTINCT\_PATID\_N:** Count of records with unique values for PATID.
- **DISTINCT\_PROVIDERID\_N:** Count of records with unique values for PROVIDERID.
- **DISTINCT\_VISIT\_N:** Count of unique visits in the ENCOUNTER table. Visits are a concatenation of PATID + PROVIDER\_ID + ENC\_TYPE + ADMIT\_DT.
- **ELIG\_RECORD\_N:** Count of records in the ENCOUNTER table where PATID, PROVIDER\_ID, ENC\_TYPE, and ADMIT\_DT are all populated.
- **ENC\_TYPE Mismatch:** These fields are replicated from the ENCOUNTER table to the PROCEDURES and DIAGNOSES table. The number of mismatched records is the number of records in PROCEDURES or DIAGNOSIS where these fields do not match the value in the ENCOUNTER table.
- **EXP\_SPECIMEN\_SOURCE:** The expected specimen source based on the values established by LOINC®.
- **KNOWN\_TEST:** Total number of records where LAB\_LOINC is not null.
- **KNOWN\_TEST\_RESULT:** Total number of records where (1) LAB\_LOINC is not null and (2) RESULT\_NUM is not null and RESULT\_MODIFIER is not in (null, NI, UN, OT) or (3) RESULT\_QUAL is not in (null, NI, UN, OT)
- **KNOWN\_TEST\_RESULT\_NUM:** Total number of records where the test and result are known, as follows: (1) LAB\_LOINC is not null and (2) RESULT\_NUM is not null and (3) RESULT\_MODIFIER is not in (null, NI, UN, OT).
- **KNOWN\_TEST\_RESULT\_NUM\_SOURCE:** Total number of records where the test and result are known, as follows: (1) LAB\_LOINC is not null and (2) RESULT\_NUM is not null and (3) RESULT\_MODIFIER is not in (null, NI, UN, OT) and (4) SPECIMEN\_SOURCE is not in (null, NI, UN, OT, UNK\_SUB, SMPLS, SPECIMEN)
- **KNOWN\_TEST\_RESULT\_NUM\_UNIT:** Total number of records where the test and result are known, as follows: (1) LAB\_LOINC is not null and (2) RESULT\_NUM is not null and (3) RESULT\_MODIFIER is not in (null, NI, UN, OT) and (4) RESULT\_UNIT is not in (null, NI, UN, OT)
- **KNOWN\_TEST\_RESULT\_NUM\_SRCE\_UNIT:** Total number of records where the test and result are known, as follows: (1) LAB\_LOINC is not null and (2) RESULT\_NUM is not null and (3) RESULT\_MODIFIER is not in (null, NI, UN, OT) and (4) SPECIMEN\_SOURCE is not in (null, NI, UN, OT, UNK\_SUB, SMPLS, SPECIMEN) and (5) RESULT\_UNIT is not in (null, NI, UN, OT)
- **KNOWN\_TEST\_NUM\_RESULT\_RANGE:** Total number of records where the test, numeric result, and normal range are all known, as follows: (1) LAB\_LOINC is not null and (2) RESULT\_NUM is not null and (3) RESULT\_MODIFIER is not in (null, NI, UN, OT) and (4) one of the following is true: (4a) NORM\_MODIFIER\_LOW='EQ' and NORM\_MODIFIER\_HIGH='EQ' and NORM\_RANGE\_LOW is not null and NORM\_RANGE\_HIGH is not null or (4b) NORM\_MODIFIER\_LOW in ('GT','GE') and



NORM\_MODIFIER\_HIGH='NO' and NORM\_RANGE\_LOW is not null and NORM\_RANGE\_HIGH is null or (4c) NORM\_MODIFIER\_HIGH in ('LE','LT') and NORM\_MODIFIER\_LOW='NO' and NORM\_RANGE\_HIGH is not null and NORM\_RANGE\_LOW is null.

- **KNOWN\_TEST\_RESULT\_PLAUSIBLE:** Total number of records where (1) LAB\_LOINC is not null **and** (2) RESULT\_DATE is not missing and is less than the maximum refresh date **and** either (3) RESULT\_NUM is not null and RESULT\_MODIFIER is not in (null, NI, UN, OT) **or** (4) RESULT\_QUAL is not in (null, NI, UN, OT)
- **MEDADMINID\_PLUS:** MED\_ADMIN records where missing<MEDADMIN\_START\_DATE<=REFRESH\_MAX and MEDADMINID is not null MEDADMIN\_CODE is not null and MEDADMIN\_TYPE is not in ('NI','UN','OT',' ').
- **NMISS or NULL\_N:** Count of records with null or missing values for the specified field.
- **ENCOUNTERID Orphan:** An ENCOUNTERID which is not in the ENCOUNTER table and appears in any other table.
- **LENGTH OF STAY:** Length of stay (LOS) is calculated by the following logic: If DISCHARGE\_DATE is null, then LOS=. Otherwise, LOS=DISCHARGE\_DATE-ADMIT\_DATE.
- **PATID Orphan:** A PATID which is not in the DEMOGRAPHIC table and appears in any other table.
- **PRESCRIBINGID\_PLUS:** PRESCRIBING records where (missing<RX\_START\_DATE <=REFRESH\_MAX or missing<RX\_ORDER\_DATE <= REFRESH\_MAX ) and RXNORM\_CUI is not null.
- **PROCEDURESID\_PLUS:** PROCEDURES records where missing<ADMIT\_DATE<=REFRESH\_MAX or missing<PX\_DATE<=REFRESH\_MAX ) and ENCOUNTERID is not null and PX\_TYPE is not in ('NI','UN','OT',' ').
- **PROVIDERID orphan:** A PROVIDERID which is not in the PROVIDER table and appears in any other table.
- **RECORD\_PCT:** The percent of all records. Will be blank for rows with values of 0 or BT (below threshold).
- **RECORD\_PCT\_CAT:** The percent of records within a category. Will be blank for rows with values of 0 or BT (below threshold).
- **RECORD\_N\_RXCUI:** Count of records with non-missing values for RXNORM\_CUI.
- **RECORD\_N\_LOINC:** Count of records with non-missing values for LOINC.
- **RESPONSE\_DATE:** Date the query package was run (ie, SAS system date).
- **QUERY\_PACKAGE:** Query package name.
- **RXNORM\_CUI\_TTY\_TIER:** The term type (TTY) that the RXNORM\_CUI is mapped to. Tier 1: RXNORM\_CUI\_TTY in ('SCD','SBD','BPCK','GPCK'). Tier 2: RXNORM\_CUI\_TTY in ('SBDG','SCDG','SBDG','SCDG','SBDC','BN','MIN'). Tier 3: RXNORM\_CUI\_TTY in ('SCDC','PIN','IN'). Tier 4: RXNORM\_CUI\_TTY in ('DF','DFG'). NULL or missing=RXNORM\_CUI\_TTY='NULL or missing'.
- **STAT:** Descriptive statistic (e.g. minimum, maximum, median).
- **TAG:** CDM field name
- **VALID\_DISTINCT\_N:** Count of records with unique and valid values for a specific field.
- **VALID\_DISTINCT\_N\_PCT:** The percent of records with unique and valid values for the specified field.
- **VALID\_N:** Number of records in a valid format. Used for fields without a prespecified value set. For the HASH\_TOKEN table, this refers to non-error tokens.
- **VISIT:** As stated in the PCORnet Common Data Model, for the Encounter table, “each record will generally reflect a unique combination of PATID, ADMIT\_DATE, PROVIDERID, and ENC\_TYPE”. Thus, a visit is a concatenation of PATID + ADMIT\_DATE+ PROVIDERID + ENC\_TYPE.

## IV. Data Curation Query Output Tables

For table shells of each dataset, please refer to the Technical Specifications available on the Data Curation page on [iMeet](#).

ID	PCORnet Table(s)	Output table	Output table description
1	CONDITION	cond_l3_condition	CONDITION frequency
2	CONDITION	cond_l3_n	Counts PATID, ENCOUNTERID, and CONDITIONID
3	CONDITION	cond_l3_rdate_y	REPORT_DATE year frequency
4	CONDITION	cond_l3_rdate_ym	REPORT_DATE year month frequency
5	CONDITION	cond_l3_source	CONDITION_SOURCE frequency
6	CONDITION	cond_l3_status	CONDITION_STATUS frequency
7	CONDITION	cond_l3_type	CONDITION_TYPE frequency
8	DEATH	death_l3_date_y	DEATH_DATE year frequency
9	DEATH	death_l3_date_ym	DEATH_DATE year month frequency
10	DEATH	death_l3_impute	DEATH_DATE_IMPUTE frequency
11	DEATH	death_l3_match	DEATH_MATCH_CONFIDENCE frequency
12	DEATH	death_l3_n	Counts non-missing, distinct, and missing PATID and DEATHID
13	DEATH	death_l3_source	DEATH_SOURCE frequency
14	DEATH	death_l3_source_ym	DEATH_SOURCE and DEATH_DATE year month crosstab
15	DEATH_CAUSE	deathc_l3_code	DEATH_CAUSE_CODE frequency
16	DEATH_CAUSE	deathc_l3_conf	DEATH_CAUSE_CONFIDENCE frequency
17	DEATH_CAUSE	deathc_l3_n	Counts PATID, DEATH_CAUSE, and DEATHCID
18	DEATH_CAUSE	deathc_l3_source	DEATH_CAUSE_SOURCE frequency
19	DEATH_CAUSE	deathc_l3_type	DEATH_CAUSE_TYPE frequency
20	DEMOGRAPHIC	dem_l3_ageyrst1	Descriptive statistics for age. Age is calculated as current age or age at death if death date is known. If multiple death records exist, the earlier death date is used.
21	DEMOGRAPHIC	dem_l3_ageyrst2	Age group frequency. Age is calculated as current age or age at death if death date is known. If multiple death records exist, the earlier death date is used.
22	DEMOGRAPHIC	dem_l3_genderdist	GENDER_IDENTITY frequency
23	DEMOGRAPHIC	dem_l3_orientdist	SEXUAL_ORIENTATION FREQUENCY
24	DEMOGRAPHIC	dem_l3_hispdist	HISPANIC frequency
25	DEMOGRAPHIC	dem_l3_n	Counts non-missing, distinct, and missing PATID
26	DEMOGRAPHIC	dem_l3_patpreflang	PAT_PREF_LANGUAGE_SPOKEN frequency

ID	PCORnet Table(s)	Output table	Output table description
27	DEMOGRAPHIC	dem_l3_racedist	RACE frequency
28	DEMOGRAPHIC	dem_l3_sexdist	SEX frequency
29	DEMOGRAPHIC	dem_obs_l3_ageyrsdist1	Descriptive statistics for age for patients with at least 1 ED, EI, IP, OS, TH, or AV encounter within the past 5 years
30	DEMOGRAPHIC	dem_obs_l3_ageyrsdist2	Age group frequency for patients with at least 1 ED, EI, IP, OS, TH, or AV encounter within the past 5 years
31	DEMOGRAPHIC	dem_obs_l3_genderdist	GENDER_IDENTITY frequency for patients with at least 1 ED, EI, IP, OS, TH, or AV encounter within the past 5 years
32	DEMOGRAPHIC	dem_obs_l3_hispdist	HISPANIC frequency for patients with at least 1 ED, EI, IP, OS, TH, or AV encounter within the past 5 years
33	DEMOGRAPHIC	dem_obs_l3_orientdist	SEXUAL_ORIENTATION frequency for patients with at least 1 ED, EI, IP, OS, TH, or AV encounter within the past 5 years
34	DEMOGRAPHIC	dem_obs_l3_patpreflang	PAT_PREF_LANGUAGE_SPOKEN frequency for patients with at least 1 ED, EI, IP, OS, TH, or AV encounter within the past 5 years
35	DEMOGRAPHIC	dem_obs_l3_racedist	RACE frequency for patients with at least 1 ED, EI, IP, OS, TH, or AV encounter within the past 5 years
36	DEMOGRAPHIC	dem_obs_l3_sexdist	SEX frequency for patients with at least 1 ED, EI, IP, OS, TH, or AV encounter within the past 5 years
37	DIAGNOSIS	dia_l3_adate_y	ADMIT_DATE year frequency
38	DIAGNOSIS	dia_l3_adate_ym	ADMIT_DATE year month frequency
39	DIAGNOSIS	dia_l3_dx	DX frequency
40	DIAGNOSIS	dia_l3_dxtype	DX_TYPE frequency
41	DIAGNOSIS	dia_l3_dx_dxtype	DX and DX_TYPE crosstab
42	DIAGNOSIS	dia_l3_dxpoa	DX_POA frequency
43	DIAGNOSIS	dia_l3_dxsource	DX_SOURCE frequency
44	DIAGNOSIS	dia_l3_dxtype_adate_y	DX_TYPE and ADMIT_DATE year crosstab
45	DIAGNOSIS	dia_l3_dxtype_dxsource	DX_TYPE and DX_SOURCE crosstab
46	DIAGNOSIS	dia_l3_dxtype_encype	DX_TYPE and ENC_TYPE crosstab
47	DIAGNOSIS	dia_l3_encype	ENC_TYPE frequency
48	DIAGNOSIS	dia_l3_encype_adate_ym	ENC_TYPE and ADMIT_DATE year month crosstab
49	DIAGNOSIS	dia_l3_n	Counts PATID, ENCOUNTERID, and DIAGNOSISID, and DIAGNOSISID_PLUS
50	DIAGNOSIS	dia_l3_origin	DX_ORIGIN frequency
51	DIAGNOSIS	dia_l3_pdx	PDX frequency
52	DIAGNOSIS	dia_l3_pdx_encype	PDX and ENC_TYPE crosstab
53	DIAGNOSIS	dia_l3_pdxgrp_encype	PDX group and ENC_TYPE crosstab

ID	PCORnet Table(s)	Output table	Output table description
54	DIAGNOSIS	dia_l3_dxdate_y	DX_DATE year frequency
55	DIAGNOSIS	dia_l3_dxdate_ym	DX_DATE year month frequency
56	DIAGNOSIS	dia_l3_dcgroup	Count of PATID by data curation diagnosis group
57	DISPENSING	disp_l3_ndc	NDC frequency
58	DISPENSING	disp_l3_ddate_y	DISPENSE_DATE year frequency
59	DISPENSING	disp_l3_ddate_ym	DISPENSE_DATE year month frequency
60	DISPENSING	disp_l3_dispamt_dist	Descriptive statistics for DISPENSE_AMT
61	DISPENSING	disp_l3_dose_dist	Descriptive statistics for DISPENSE_DOSE_DISP
62	DISPENSING	disp_l3_doseunit	DISPENSE_DOSE_DISP_UNIT frequency
63	DISPENSING	disp_l3_route	DISPENSE_ROUTE frequency
64	DISPENSING	disp_l3_n	Counts non-missing, distinct, and missing PATID, DISPENSINGID, PRESCRIBINGID, NDC, and valid NDCs. Valid NDCs are 11 digits with no dashes, ie. HIPAA format.
65	DISPENSING	disp_l3_supdist2	Record count by category of RX_DAYS_SUPP
66	DISPENSING	disp_l3_source	DISPENSE_SOURCE frequency
67	ENCOUNTER	enc_l3_adate_y	ADMIT_DATE year frequency
68	ENCOUNTER	enc_l3_adate_ym	ADMIT_DATE year month frequency
69	ENCOUNTER	enc_l3_admsrc	ADMITTING_SOURCE frequency
70	ENCOUNTER	enc_l3_dash2	Counts the number of patients with any AV, ED, IP, EI, TH, or OS encounter record with a populated ADMIT_DATE during the designated period prior to the maximum ADMIT_DATE. If the maximum ADMIT_DATE is in the future, the current date is used instead.
71	ENCOUNTER	enc_l3_ddate_y	DISCHARGE_DATE year frequency
72	ENCOUNTER	enc_l3_ddate_ym	DISCHARGE_DATE year month frequency
73	ENCOUNTER	enc_l3_disdisp	DISCHARGE_DISPOSITION frequency
74	ENCOUNTER	enc_l3_disstat	DISCHARGE_STATUS frequency
75	ENCOUNTER	enc_l3_drg_type	DRG_TYPE frequency
76	ENCOUNTER	enc_l3_encype	ENC_TYPE frequency. (Note: Visits are a concatenation of PATID + PROVIDER_ID + ENC_TYPE + ADMIT_DT. ELIG_RECORD_N is a count of records where all fields used to define a visit are populated)
77	ENCOUNTER	enc_l3_encype_adate_y	ENC_TYPE and ADMIT_DATE year month crosstab

ID	PCORnet Table(s)	Output table	Output table description
78	ENCOUNTER	enc_l3_encype_adate_ym	ENC_TYPE and ADMIT_DATE year month crosstab
79	ENCOUNTER	enc_l3_encype_admsrc	ENC_TYPE by ADMITTING_SOURCE crosstab
80	ENCOUNTER	enc_l3_encype_ddate_ym	ENC_TYPE and DISCHARGE_DATE year month crosstab
81	ENCOUNTER	enc_l3_encype_disdisp	ENC_TYPE and DISCHARGE_DISPOSITION crosstab
82	ENCOUNTER	enc_l3_encype_disstat	ENC_TYPE and DISCHARGE_STATUS crosstab
83	ENCOUNTER	enc_l3_encype_drg	ENC_TYPE and DRG_TYPE crosstab
84	ENCOUNTER	enc_l3_n	Counts non-missing, distinct, and missing PATID, ENCOUNTERID, and PROVIDERID, FACILITYID, and FACILITY_LOCATION
85	ENCOUNTER	enc_l3_payertype1	PAYER_TYPE_PRIMARY frequency
86	ENCOUNTER	enc_l3_payertype2	PAYER_TYPE_SECONDARY frequency
87	ENCOUNTER	enc_l3_facilitytype	FACILITY_TYPE frequency
88	ENCOUNTER	enc_l3_facilityloc	FACILITY_LOCATION frequency
89	ENCOUNTER	enc_l3_facilitytype_facilityloc	FACILITY_TYPE and FACILITY_LOCATION (first 5 digits of a zip code) crosstab
90	ENCOUNTER	enc_l3_los_dist	Frequency of length of stay (LOS) category by ENC_TYPE
91	ENROLLMENT	enr_l3_basedist	ENR_BASIS frequency
92	ENROLLMENT	enr_l3_n	Counts non-missing, distinct, and missing PATID, ENR_START_DATE, and ENROLLID (combination of PATID, ENR_START_DATE, and ENR_BASIS)
93	ENROLLMENT	enr_l3_chart	CHART frequency
94	LAB_RESULT_CM	lab_l3_abn	ABN_IND frequency
95	LAB_RESULT_CM	lab_l3_dcgroup	Frequency by DC_LAB_GROUP
96	LAB_RESULT_CM	lab_l3_high	NORM_MODIFIER_HIGH frequency
97	LAB_RESULT_CM	lab_l3_loc	RESULT_LOC frequency
98	LAB_RESULT_CM	lab_l3_loinc	LAB_LOINC frequency
99	LAB_RESULT_CM	lab_l3_loinc_result_num	RESULT_NUM descriptive statistics by LAB_LOINC code
100	LAB_RESULT_CM	lab_l3_loinc_source	LAB_LOINC and SPECIMEN_SOURCE crosstab for a subset of LOINC codes
101	LAB_RESULT_CM	lab_l3_low	NORM_MODIFIER_LOW frequency
102	LAB_RESULT_CM	lab_l3_mod	RESULT_MODIFIER frequency

ID	PCORnet Table(s)	Output table	Output table description
103	LAB_RESULT_CM	lab_l3_n	Counts non-missing, distinct, and missing PATID, LAB_RESULT_CM_ID, and ENCOUNTERID
104	LAB_RESULT_CM	lab_l3_priority	PRIORITY frequency
105	LAB_RESULT_CM	lab_l3_px_pxttype	LAB_PX and LAB_PXTYPE crosstab
106	LAB_RESULT_CM	lab_l3_px_type	LAB_PX_TYPE frequency
107	LAB_RESULT_CM	lab_l3_qual	RESULT_QUAL frequency
108	LAB_RESULT_CM	lab_l3_raw_name	RAW_LAB_NAME frequency
109	LAB_RESULT_CM	lab_l3_rdate_y	RESULT_DATE year frequency
110	LAB_RESULT_CM	lab_l3_rdate_ym	RESULT_DATE year month frequency
111	LAB_RESULT_CM	lab_l3_recordc	Frequency of records with varying levels of completeness across variables
112	LAB_RESULT_CM	lab_l3_snomed	RESULT_SNOMED frequency
113	LAB_RESULT_CM	lab_l3_source	SPECIMEN_SOURCE frequency
114	LAB_RESULT_CM	lab_l3_unit	RESULT_UNIT frequency
115	LAB_RESULT_CM	lab_l3_loinc_result_num_5y	RESULT_NUM descriptive statistics by LAB_LOINC code in the 5 years of the lookback period
116	LAB_RESULT_CM	lab_l3_loinc_source_5y	LAB_LOINC and SPECIMEN_SOURCE crosstab for a subset of LOINC codes in the 5 years of lookback period
117	LAB_RESULT_CM	lab_l3_n_5y	Counts PATID, LAB_RESULT_CM_ID, and ENCOUNTERID in the 5 years of lookback period
118	LAB_RESULT_CM	lab_l3_recordc_5y	Frequency of records with varying levels of completeness across variables in the 5 years of lookback period
119	LAB_RESULT_CM	lab_l3_rsource	LAB_RESULT_SOURCE frequency
120	LAB_RESULT_CM	lab_l3_lsource	LAB_LOINC_SOURCE frequency
121	LAB_RESULT_CM	lab_l3_loinc_unit	LAB_LOINC and RESULT_UNIT crosstab
122	MED_ADMIN	medadm_l3_doseadm	Descriptive statistics for MEDADMIN_DOSE_ADM
123	MED_ADMIN	medadm_l3_doseadmunit	MEDADMIN_DOSE_ADMIN_UNIT frequency

ID	PCORnet Table(s)	Output table	Output table description
124	MED_ADMIN	medadm_l3_n	Counts MEDADMINID, PATID, and MEDADMINID_PLUS
125	MED_ADMIN	medadm_l3_route	MEDADMIN_ROUTE frequency
126	MED_ADMIN	medadm_l3_source	MEDADMIN_SOURCE frequency
127	MED_ADMIN	medadm_l3_type	MEDADMIN_TYPE frequency
128	MED_ADMIN	medadm_l3_sdate_y	MEDADMIN_START_DATE year frequency
129	MED_ADMIN	medadm_l3_sdate_ym	MEDADMIN_START_DATE year month frequency
130	MED_ADMIN	medadm_l3_code_type	MEDADMIN_TYPE and MEDADMIN_CODE crosstab
131	MED_ADMIN	medadm_l3_rxcui_tier	MEDADMIN_CODE frequency by tier of RXCUI term type
132	MED_ADMIN	medadm_l3_rxcui_tier_5y	MEDADMIN_CODE frequency by tier of RXCUI term type in the 5 years of the lookback period
133	MULTIPLE	datamart_all	DataMart metadata including variable names, variable lengths, data types and number of observations. Used to assess conformance to the required SAS structure for the PCORnet Common Data Model (CDM) v3.1.
134	MULTIPLE	elapsed_all	Displays the query start time, query end time, and query run time for each table created by the data_curation_all program, the cumulative run time for the program and the dataset loading time. Will only be present if the 'all' option is used.
135	MULTIPLE	elapsed_main	Displays the query start time, query end time, and query run time for each table created by the data_curation_main program, the cumulative run time for the program and the dataset loading time. Will only be present if the programs are run separately.
136	MULTIPLE	elapsed_lab	Displays the query start time, query end time, and query run time for each table created by the data_curation_lab program, the cumulative run time for the program and the dataset loading time. Will only be present if the programs are run separately.
137	MULTIPLE	elapsed_xtbl	Displays the query start time, query end time, and query run time for each



ID	PCORnet Table(s)	Output table	Output table description
			table created by the data_curation_xtbl program, the cumulative run time for the program and the dataset loading time. The DATAMART_ALL table is not included because it is just a print. Will only be present if the programs are run separately.
138	MULTIPLE	elapsed_obsclin	Displays the query start time, query end time, and query run time for each table created by the data_curation_obsclin program, the cumulative run time for the program and the dataset loading time. Will only be present if the programs are run separately.
139	MULTIPLE	xtbl_l3_dash1	Counts the number of patients with any VITAL record with a populated MEASURE_DATE and a diagnosis record with a populated ADMIT_DATE and DX during the designated period prior to the maximum DIAGNOSIS.ADMIT_DATE. If the maximum ADMIT_DATE is in the future, the current date is substituted.
140	MULTIPLE	xtbl_l3_dash2	Counts the number of patients with any VITAL record with a populated MEASURE_DATE and a DIAGNOSIS record with a populated DX and ADMIT_DATE and either a PRESCRIBING record with a populated RXNORM_CUI and RX_START_DATE or a DISPENSING record with a populated DISPENSE_DATE and NDC during the designated period of time prior to the maximum DIAGNOSIS.ADMIT_DATE. If the maximum ADMIT_DATE is in the future, the current date is substituted.
141	MULTIPLE	xtbl_l3_dash3	Counts the number of patients with any VITAL record with a populated MEASURE_DATE and a DIAGNOSIS record with a populated DX and ADMIT_DATE and either (a PRESCRIBING record with a populated RXNORM_CUI and RX_ORDER_DATE or a DISPENSING record with a populated DISPENSE_DATE and NDC) and a LAB_RESULT_CM record and RESULT_DATE during the designated period of time prior to the maximum DIAGNOSIS.ADMIT_DATE. If the



ID	PCORnet Table(s)	Output table	Output table description
			maximum ADMIT_DATE is in the future, the current date is substituted.
142	MULTIPLE	xtbl_l3_date_logic	Identifies illogical relationships between BIRTH_DATE, DEATH_DATE, and key dates in other tables
143	MULTIPLE	xtbl_l3_dates	Descriptive statistics and counts of records with future dates or dates prior to January 2010 for all date fields.
144	MULTIPLE	xtbl_l3_lab_encype	# of records and patients with lab records by encounter type.
145	MULTIPLE	xtbl_l3_metadata	HARVEST fields; maximum refresh date; query package; response date; low cell count threshold; operating system; SAS version and packages; SAS datastore (data or views); and query run time. There should only be 1 record in this table. The DATAMARTID and REFRESH_MAX fields are used extensively throughout the query package.
146	MULTIPLE	xtbl_l3_mismatch	Counts the number of records where there is a mismatch between a parent and child table. These checks include ENCOUNTERIDs that are not in the ENCOUNTER table; PATIDs that are not in the DEMOGRAPHIC table; PROVIDERIDs that are not in the PROVIDER table; and discordance in the fields that are replicated from the ENCOUNTER table to the PROCEDURES and DIAGNOSIS tables.
147	MULTIPLE	xtbl_l3_non_unique	Identify encounters which are associated with more than 1 patient (PATID) in the same table
148	MULTIPLE	xtbl_l3_pres_encype	# of records and patients with prescribing records by encounter type.
149	MULTIPLE	xtbl_l3_times	Descriptive statistics for all time fields.
150	MULTIPLE	xtbl_l3_race_enc	# of records and patients by RACE among patients with at least 1 encounter after 2009 (from 2010)
151	MULTIPLE	xtbl_l3_medadm_encype	# of records and patients with medication administration records by encounter type
152	MULTIPLE	xtbl_l3_refresh_date	Refresh dates from HARVEST and # of records for all tables from DATAMART_ALL
153	OBS_CLIN	obsclin_l3_n	Counts OBSCLINID and PATID
154	OBS_CLIN	obsclin_l3_code_type	OBSCLIN_TYPE and OBSCLIN_CODE crosstab

ID	PCORnet Table(s)	Output table	Output table description
155	OBS_CLIN	obsclin_l3_mod	OBSCLIN_RESULT_MODIFIER frequency
156	OBS_CLIN	obsclin_l3_qual	OBSCLIN_RESULT_QUAL frequency
157	OBS_CLIN	obsclin_l3_runit	OBSCLIN_RESULT_UNIT frequency
158	OBS_CLIN	obsclin_l3_type	OBSCLIN_TYPE frequency
159	OBS_CLIN	obsclin_l3_source	OBSCLIN_SOURCE frequency
160	OBS_CLIN	obsclin_l3_abn	OBSCLIN_ABN_IND frequency
161	OBS_CLIN	obsclin_l3_sdate_y	OBSCLIN_START_DATEyear frequency
162	OBS_CLIN	obsclin_l3_sdate_ym	OBSCLIN_START_DATEyear month frequency
163	OBS_CLIN	obsclin_l3_ht	HT group frequency
164	OBS_CLIN	obsclin_l3_ht_dist	Descriptive statistics for height
165	OBS_CLIN	obsclin_l3_wt	WT group frequency
166	OBS_CLIN	obsclin_l3_wt_dist	Descriptive statistics for weight
167	OBS_CLIN	obsclin_l3_diastolic	Diastolic group frequency
168	OBS_CLIN	obsclin_l3_systolic	Systolic group frequency
169	OBS_CLIN	obsclin_l3_bmi	BMI group frequency
170	OBS_CLIN	obsclin_l3_code_unit	OBSCLIN_CODE and OBSCLIN_RESULT_UNIT crosstab
171	OBS_CLIN	obsclin_l3_recordc	Frequency of records with varying levels of completeness across variables
172	OBS_GEN	obsngen_l3_mod	OBSGEN_RESULT_MODIFIER frequency
173	OBS_GEN	obsngen_l3_tmod	OBSGEN_TABLE_MODIFIER frequency
174	OBS_GEN	obsngen_l3_n	Counts OBSGENID, PATID, ENCOUNTERID and OBSGEN_PROVIDERID
175	OBS_GEN	obsngen_l3_qual	OBSGEN_RESULT_QUAL frequency
176	OBS_GEN	obsngen_l3_runit	OBSGEN_RESULT_UNIT frequency
177	OBS_GEN	obsngen_l3_type	OBSGEN_TYPE frequency
178	OBS_GEN	obsngen_l3_code_type	OBSGEN_TYPE and OBSGEN_CODE crosstab
179	OBS_GEN	obsngen_l3_source	OBSGEN_SOURCE frequency
180	OBS_GEN	obsngen_l3_abn	OBSGEN_ABN_IND frequency
181	OBS_GEN	obsngen_l3_sdate_y	OBSGEN_START_DATEyear frequency
182	OBS_GEN	obsngen_l3_sdate_ym	OBSGEN_START_DATEyear month frequency
183	OBS_GEN	obsngen_l3_recordc	Frequency of records with varying levels of completeness across variables
184	OBS_GEN	obsngen_l3_code_unit	OBSGEN_CODE and OBSGEN_RESULT_UNIT crosstab
185	PCORNET_TRIAL	trial_l3_n	Counts PATID, TRIALID, PARTICIPANTID, and TRIAL_KEY
186	PRESCRIBING	pres_l3_basis	RX_BASIS frequency
187	PRESCRIBING	pres_l3_dispaswrtn	RX_DISPENSE_AS_WRITTEN frequency

ID	PCORnet Table(s)	Output table	Output table description
188	PRESCRIBING	pres_l3_freq	RX_FREQUENCY frequency
189	PRESCRIBING	pres_l3_n	Counts non-missing, distinct, and missing PATID, PRESCRIBINGID, ENCOUNTERID, RX_PROVIDERID, PRESCRIBINGID_PLUS
190	PRESCRIBING	pres_l3_odate_y	RX_ORDER_DATE year frequency
191	PRESCRIBING	pres_l3_odate_ym	RX_ORDER_DATE year month frequency
192	PRESCRIBING	pres_l3_prnflag	RX_PRN_FLAG frequency
193	PRESCRIBING	pres_l3_rxcui	RXCUI frequency
194	PRESCRIBING	pres_l3_rxcui_rxsup	Descriptive statistics for RX_DAYS_SUPPLY by RXNORM_CUI
195	PRESCRIBING	pres_l3_rxcui_tier	RXNORM_CUI frequency by tier of term type
196	PRESCRIBING	pres_l3_rxdoseform	RX_DOSE_FORM frequency
197	PRESCRIBING	pres_l3_rxdoseodr_dist	Descriptive statistics for RX_DOSE_ORDERED
198	PRESCRIBING	pres_l3_rxdoseodrunit	RX_DOSE_ORDERED_UNIT frequency
199	PRESCRIBING	pres_l3_rxqty_dist	Descriptive statistics for RX_QUANTITY
200	PRESCRIBING	pres_l3_rxrefill_dist	Descriptive statistics for RX_REFILLS
201	PRESCRIBING	pres_l3_route	RX_ROUTE frequency
202	PRESCRIBING	pres_l3_source	RX_SOURCE frequency
203	PRESCRIBING	pres_l3_rawrxmed	RAW_RX_MED_NAME frequency
204	PRESCRIBING	pres_l3_supdist2	Record count by category of RX_DAYS_SUPPLY
205	PRESCRIBING	pres_l3_rxcui_5y	RXNORM_CUI frequency and term type information in the 5 years of lookback period
206	PRESCRIBING	pres_l3_rxcui_tier_5y	RXNORM_CUI frequency by tier of term type in the 5 years of lookback period
207	PRO_CM	procm_l3_cat	PRO_CAT frequency
208	PRO_CM	procm_l3_itemfullname	PRO_ITEM_FULLNAME frequency
209	PRO_CM	procm_l3_loinc	PRO_LOINC frequency
210	PRO_CM	procm_l3_itemnm	PRO_ITEM_NAME frequency

ID	PCORnet Table(s)	Output table	Output table description
211	PRO_CM	procm_l3_measure_fullname	PRO_MEASURE_FULLNAME frequency
212	PRO_CM	procm_l3_measurenm	PRO_MEASURE_NAME frequency
213	PRO_CM	procm_l3_method	PRO_METHOD frequency
214	PRO_CM	procm_l3_mode	PRO_MODE frequency
215	PRO_CM	procm_l3_n	Counts PRO_CM_ID, PATID, and ENCOUNTERID
216	PRO_CM	procm_l3_pdate_y	PRO_DATE year frequency
217	PRO_CM	procm_l3_pdate_ym	PRO_DATE year month frequency
218	PRO_CM	procm_l3_type	PRO_TYPE FREQUENCY
219	PRO_CM	procm_l3_source	PRO_SOURCE frequency
220	PRO_CM	procm_l3_measure_loinc	PRO_MEASURE_LOINC frequency
221	PROCEDURES	pro_l3_adate_y	ADMIT_DATE year frequency
222	PROCEDURES	pro_l3_adate_ym	ADMIT_DATE year month frequency
223	PROCEDURES	pro_l3_encype	ENC_TYPE frequency
224	PROCEDURES	pro_l3_encype_adate_ym	ENC_TYPE and ADMIT_DATE year month crosstab
225	PROCEDURES	pro_l3_n	Counts non-missing, distinct, and missing PATID, ENCOUNTERID, PROCEDURESID, PROVIDERID, and PROCEDURESID_PLUS
226	PROCEDURES	pro_l3_ppx	PPX FREQUENCY
227	PROCEDURES	pro_l3_px	PX frequency
228	PROCEDURES	pro_l3_pstype	PX_TYPE frequency
229	PROCEDURES	pro_l3_px_pstype	PX and PX_TYPE crosstab
230	PROCEDURES	pro_l3_pxdate_y	PX_DATE year frequency
231	PROCEDURES	pro_l3_pxsouce	PX_SOURCE frequency
232	PROCEDURES	pro_l3_pstype_adate_y	PX_TYPE and ADMIT_DATE year crosstab
233	PROCEDURES	pro_l3_pstype_encype	PX_TYPE and ENC_TYPE crosstab

ID	PCORnet Table(s)	Output table	Output table description
234	PROCEDURES	pro_l3_dcgroup	Count of PATID by data curation procedures group
235	PROVIDER	prov_l3_n	Counts PROVIDERID and PROVIDER_NPI
236	PROVIDER	prov_l3_npiflag	PROVIDER_NPI_FLAG frequency
237	PROVIDER	prov_l3_specialty	PROVIDER_SPECIALTY_PRIMARY frequency
238	PROVIDER	prov_l3_specialty_group	PROVIDER_SPECIALTY_PRIMARY group frequency
239	PROVIDER	prov_l3_sex	PROVIDER_SEX frequency
240	PROVIDER	prov_l3_specialty_encype	PROVIDER_SPECIALTY_PRIMARY and ENC_TYPE crosstab by PROVIDER_SPECIALTY_GROUP
241	VITAL	vit_l3_bmi	BMI frequency
242	VITAL	vit_l3_bp_position_type	BP_POSITION_TYPE frequency
243	VITAL	vit_l3_dash1	Counts the number of patients with any vital record with a populated MEASURE_DATE during the designated period prior to the maximum MEASURE_DATE. If the maximum MEASURE_DATE is in the future, the current date is substituted.
244	VITAL	vit_l3_diastolic	DIASTOLIC frequency
245	VITAL	vit_l3_ht	HT frequency
246	VITAL	vit_l3_ht_dist	Descriptive statistics for HT
247	VITAL	vit_l3_mdate_y	MEASURE_DATE year frequency
248	VITAL	vit_l3_mdate_ym	MEASURE_DATE year month frequency
249	VITAL	vit_l3_n	Counts non-missing, distinct, and missing PATID, ENCOUNTERID, and VITALID
250	VITAL	vit_l3_smoking	SMOKING frequency
251	VITAL	vit_l3_systolic	SYSTOLIC frequency
252	VITAL	vit_l3_tobacco	TOBACCO frequency
253	VITAL	vit_l3_tobacco_type	TOBACCO_TYPE frequency
254	VITAL	vit_l3_vital_source	VITAL_SOURCE frequency
255	VITAL	vit_l3_wt	WT frequency
256	VITAL	vit_l3_wt_dist	Descriptive statistics for WT
257	LDS_ADDRESS_HISTORY	ldsadrs_l3_n	Counts for PATID, ADDRESSID, ADDRESS_ZIP5, and ADDRESS_ZIP9

ID	PCORnet Table(s)	Output table	Output table description
258	LDS_ADDRESS_HISTORY	ldsadrs_l3_adrsuse	ADDRESS_USE frequency
259	LDS_ADDRESS_HISTORY	ldsadrs_l3_adrstype	ADDRESS_TYPE frequency
260	LDS_ADDRESS_HISTORY	ldsadrs_l3_adrspref	ADDRESS_PREFERRED frequency
261	LDS_ADDRESS_HISTORY	ldsadrs_l3_adrscity	ADDRESS_CITY frequency
262	LDS_ADDRESS_HISTORY	ldsadrs_l3_adrsstate	ADDRESS_STATE frequency
263	LDS_ADDRESS_HISTORY	ldsadrs_l3_adrszip5	ADDRESS_ZIP5 frequency
264	LDS_ADDRESS_HISTORY	ldsadrs_l3_adrszip9	ADDRESS_ZIP9 frequency
265	IMMUNIZATION	immune_l3_n	Counts for PATID, IMMUNIZATIONID, ENCOUNTERID, PROCEDURESID, VX_PROVIDERID
266	IMMUNIZATION	immune_l3_rdate_y	VX_RECORD_DATE year frequency
267	IMMUNIZATION	immune_l3_rdate_ym	VX_RECORD_DATE year month frequency
268	IMMUNIZATION	immune_l3_adate_y	VX_ADMIN_DATE year frequency
269	IMMUNIZATION	immune_l3_adate_ym	VX_ADMIN_DATE year month frequency
270	IMMUNIZATION	immune_l3_codetype	VX_CODE_TYPE frequency
271	IMMUNIZATION	immune_l3_code_codetype	VX_CODE and VX_CODETYPE crosstab
272	IMMUNIZATION	immune_l3_status	VX_STATUS frequency
273	IMMUNIZATION	immune_l3_statusreason	VX_STATUS_REASON frequency
274	IMMUNIZATION	immune_l3_source	VX_SOURCE frequency
275	IMMUNIZATION	immune_l3_dose_dist	VX_DOSE descriptive statistics
276	IMMUNIZATION	immune_l3_doseunit	VX_DOSE_UNIT frequency
277	IMMUNIZATION	immune_l3_route	VX_ROUTE frequency
278	IMMUNIZATION	immune_l3_bodysite	VX_BODY_SITE frequency
279	IMMUNIZATION	immune_l3_manufacturer	VX_MANUFACTURER frequency
280	IMMUNIZATION	immune_l3_lotnum	VX_LOT_NUM frequency
281	HASH_TOKEN	hash_l3_n	Count for PATIDs and Hash Tokens
282	HASH_TOKEN	hash_l3_token_availability	Count for PATIDs with all possible combination of tokens
283	LAB_HISTORY	labhist_l3_n	Counts LABHISTORYID
284	LAB_HISTORY	labhist_l3_loinc	LAB_LOINC frequency
285	LAB_HISTORY	labhist_l3_sexdist	SEX frequency

<b>ID</b>	<b>PCORnet Table(s)</b>	<b>Output table</b>	<b>Output table description</b>
286	LAB_HISTORY	labhist_l3_racedist	RACE frequency
287	LAB_HISTORY	labhist_l3_min_wks	Descriptive statistics for AGE_MIN_WKS
288	LAB_HISTORY	labhist_l3_max_wks	Descriptive statistics for AGE_MAX_WKS
289	LAB_HISTORY	labhist_l3_unit	RESULT_UNIT frequency
290	LAB_HISTORY	labhist_l3_low	NORM_MODIFIER_LOW frequency
291	LAB_HISTORY	labhist_l3_high	NORM_MODIFIER_HIGH frequency
292	LAB_HISTORY	labhist_l3_pdstart_y	PERIOD_START year frequency
293	LAB_HISTORY	labhist_l3_pdend_y	PERIOD_END year frequency
294	LAB_HISTORY	labhist_l3_rlow_dist	Descriptive statistics for NORM_RANGE_LOW
295	LAB_HISTORY	labhist_l3_rhigh_dist	Descriptive statistics for NORM_RANGE_HIGH

## V. Empirical Data Curation Report

The data from all data curation query output tables except for the *elapsed* and *datamart\_all* datasets is compiled into a normalized dataset. The Empirical Data Curation (EDC) Report is produced from this dataset. The EDC Report summarizes key information from the query output tables and identifies exceptions to the PCORnet Data Checks. The report includes a table of contents, a data check exception summary, and up to 52 tables and charts, depending upon the number of CDM tables which are populated. The table of contents is below.

Section	Table	Table Description	Data Check(s)
n/a	n/a	Data Check Exception Summary	n/a
Section I: Descriptive Information	Table IA	Demographic Summary	n/a
	Table IB	Potential Pools of Patients	3.04, 3.05
	Table IC	Height, Weight, and Body Mass Index (BMI)	n/a
	Table ID	Records, Refresh Dates, Patients, Encounters, and Date Ranges by Table	1.18
	Table IE	Records Per Table by Encounter Type	n/a
	Table IF	Date Obfuscation or Imputation	n/a
	Table IG	Lab Results For Selected Lab Tests	3.13
	Table IH	Patients with Selected Diagnoses	n/a
	Table II	Patients with Selected Procedures	n/a
	Chart IA	Trend in Vital Measures by Measurement Date, Past 5 Years	n/a
	Chart IB	Trend in Encounters by Admit Date and Encounter Type, Past 5 Years	n/a
	Chart IC	Trend in Institutional Encounters by Discharge Date and Encounter Type, Past 5 Years	n/a
	Chart ID	Trend in Laboratory Results by Result Date, Past 5 Years	n/a
	Chart IE	Trend in Prescribed Medications by Rx Order Date, Past 5 Years	n/a
	Chart IF	Trend in Dispensed Medications by Dispense Date, Past 5 Years	n/a
	Chart IG	Trend in Administered Medications by Start Date, Past 5 Years	n/a
	Chart IH	Trend in Condition Records by Report Date, Past 5 Years	n/a
	Chart II	Trend in Death Records by Death Date and Source, Past 5 Years	n/a
	Chart IJ	Trend in Immunization Records by Vx Record Date, Past 5 Years	n/a
	Chart IK	Trend in Clinical Observation Records by Start Date, Past 5 Years	n/a
	Chart IL	Trend in General Observation Records by Start Date, Past 5 Years	n/a
Section II: Data Model Conformance	Table IIA	Primary Key Errors	1.05
	Table IIB	Values Outside of Common Data Model (CDM) Specifications	1.06
	Table IIC	Non-Permissible Missing Values	1.07
	Table IID	Diagnostic Errors	1.01, 1.02, 1.03, 1.04, 1.17
	Table IIE	Orphan Records, Replication Errors, Encounter Duplication and Hash Token Duplication	1.08, 1.09, 1.10, 1.11, 1.12, 1.14, 1.15, 1.19
	Table IIF	Potential Code Errors and Misplaced Codes	1.13, 1.16
Section III: Data Plausibility	Table IIIA	Future Dates	2.01
	Table IIIB	Records with Extreme Values	2.02



Section	Table	Table Description	Data Check(s)
	Table IIIC	Illogical Dates	2.03
	Table IIID	Encounters Per Visit and Per Patient	2.04
	Table IIIE	Laboratory Result Specimen Source Discrepancies	2.05
	Table IIIF	Quantitative Lab Result Outliers, Selected Tests	2.06
	Table IIIG	Monthly Record Volume Outliers, Selected Domains	2.08
	Chart IIIA	Monthly Record Volume Outliers, Encounters	2.08
	Chart IIIB	Monthly Record Volume Outliers, Diagnoses	2.08
	Chart IIIC	Monthly Record Volume Outliers, Procedures	2.08
	Chart IIID	Monthly Record Volume Outliers, Vitals	2.08
	Chart IIIE	Monthly Record Volume Outliers, Prescribing	2.08
	Chart IIIF	Monthly Record Volume Outliers, Labs	2.08
Section IV: Data Completeness and Plausibility	Table IVA	Diagnosis Records Per Encounter, Overall and by Encounter Type	3.01
	Chart IVA	Diagnosis Records Per Encounter by Admit Date and Encounter Type, Past 5 Years	n/a
	Table IVB	Procedure Records Per Encounter, Overall and by Encounter Type	3.02
	Chart IVB	Procedure Records Per Encounter by Admit Date and Encounter Type, Past 5 Years	n/a
	Table IVC	Missing or Unknown Values, Required Tables	3.03
	Table IVD	Missing or Unknown Values, Optional Tables	3.03
	Table IVE	Principal Diagnoses for Institutional Encounters	2.07, 3.06
	Table IVF	Data Latency and Completeness of Encounter, Diagnosis and Procedure Data, Past 2 Years	3.07
	Table IVG	Data Latency and Completeness of Vital, Prescription, and Lab Data, Past 2 Years	3.11
	Table IVH	RXNORM Term Type Mapping, Overall and Past 5 Years	3.08, 3.15
	Table IVI	Laboratory Result Data Completeness, Overall and Past 5 Years	3.09, 3.10, 3.12
	Table IVI_Ref	Laboratory Result Data Completeness Definitions	n/a
	Table IVJ	Data Latency and Completeness of Medication Administration, Dispensing and Clinical Observation Data, Past 2 Years	3.14
Section V: Data Persistence	Table VA	Changes in Tables	4.01
	Table VB	Changes in Selected Encounter Types and Domains	4.02
	Table VC	Changes in Selected Code Types	4.03

## VI. Program Package File Structure

Each request package distributed by PCORnet's DRN OC contains several sub-folders to organize program inputs and outputs. The subfolders must reside within an outer folder labeled with the query name. The subfolders are as follows:

- ***dmlocal***: Contains output generated by the request that should be saved locally but not returned to DRN OC. Output may be used locally or to facilitate follow-up queries.
- ***drnoc***: Contains output generated by the request that should be returned to the DRN OC. These tables consist of aggregate data/output and transfer the minimum required to answer the analytic question.
- ***sasprograms***: Contains the master SAS program that must be edited and then executed locally.
- ***infolder***: Contains all input programs and files needed to execute the request. These are created for each request by the DRN OC Data Curation team; the contents of this folder should not be edited.

## VII. Files Included in Query Request

The following files are included in the Zip file distributed with the query request.

Subfolder	Type	Files
n/a	PDF and Word Document	Cycle 13 Data Curation Query Package Checklist.docx Data Curation Query Package v6.06 Work Plan.pdf Additional Approval Criteria and Exemption Request Form.docx
/infolder	SAS programs	data_curation_query_base.sas data_curation_query_lab.sas data_curation_query_obsclin.sas data_curation_query_main.sas data_curation_print.sas data_curation_query_xtbl.sas edc_prep.sas edc_report.sas edc_template.sas normalization.sas potential_code_errors.sas
	Reference datasets	loinc.cpt: Contains loinc.sas7dat, which is used for the misplaced records portion of the Potential Code Errors query and Data Check 1.16 dc_reference.cpt: See DC_Reference.cpt below edc_reference.cpt: See EDC_Reference.cpt below
/sasprograms	SAS programs	01_run_code_errors.sas 02_run_queries.sas 03_run_edc_prep.sas 04_run_edc_report.sas

### dc\_reference.cpt

Dataset	Description
facility_type pat_pref_language_spoken payer_type provider_specialty_primary_qual _route _dose_form state specimen_source _unit vx_body_site vx_manufacturer	Derived from the Valuesets tab of the current <a href="#">parseable file</a> .
obsclin_codes	Contains the codes used for mapping select concepts to LOINC. Derived from the CODES tab of the current <a href="#">Valueset Reference File</a> .
lab_loinc_ref	Contains the specimens (LOINC “system”) for selected LOINC codes (high volume and/or included in the DC Lab Groups). Used in Table III E and Data Check 2.05.
lab_dcgroup_ref	Codelists for the Data Curation Lab Groups. Used indirectly in Data Checks 2.05, 2.06 and 3.13. This dataset is also available as an Excel file on the Data Curation page on <a href="#">iMeet</a>
rxnorm_cui_ref	RXCUI reference list from UMLS. Used in EDC Table IV H and Data Check 2.08.
dx_dcgroup_ref	Code lists for selected diagnosis concepts. Used in EDC Table I H.
px_dcgroup_ref	Code lists for selected procedure concepts. Used in EDC Table II.

Dataset	Description
loinc_exclude	Contains LOINC codes where the values may contain Personal Identifiable Information. These codes are excluded from the OBS_CLIN, OBS_GEN, LAB_RESULT_CM, and PRO_CM tables.
cdm_version	Contains the allowable values for HARVEST.CDM_VERSION.

#### edc\_reference.cpt

Dataset	Description
dc_summary	Data check descriptions and network results
dc_tables	List of tables produced by 02_run_queries
footers	Text for the EDC report footers
headers	Text for the EDC report headers
lab_volume_ref	Contains the lower boundary for the expected volume of selected data curation lab groups, stratified for pediatric and non-pediatric datamarts. Used in Data Check 3.13.
pediatric_datamarts	List of pediatric DataMarts
required_structure	Derived from the Fields tab of the current parseable file
specimen_source_category	Groups actual specimen sources into categories for Data Check 2.05
Tbl_ivi_ref	Text for Table IVI_Ref
Toc	Text for the Table of Contents
q2_stat_dlg_loinc	Contains clinical and statistical ranges, stratified for pediatric and non-pediatric datamarts for Data Check 2.06

## VIII. Output Files

**Local files (dmlocal folder).** DMID=DataMart ID; DATE=response date

Produced by	File description
pcornet_code_errors.sas	code_summary (SAS dataset and csv file) misplaced_loincs (SAS dataset and csv file) <i>Up to 11 error files, if relevant code types are present:</i> bad_condition (SAS dataset and csv file) bad_dx (SAS dataset and csv file) bad_px (SAS dataset and csv file) bad_pres (SAS dataset and csv file) bad_lab (SAS dataset and csv file) bad_lab_hist (SAS dataset and csv file) bad_disp (SAS dataset and csv file) bad_medadmin (SAS dataset and csv file) bad_obsclin (SAS dataset and csv file) bad_obsgen (SAS dataset and csv file) bad_immunization (SAS dataset and csv file)
data_curation_query_base.sas; data_curation_query_main.sas; data_curation_query_lab.sas; data_curation_query_obsclin.sas; data_curation_query_xtbl.sas	Up to 295 output tables (SAS datasets and csv files; see <a href="#">section IV</a> ) and set.log (contains the output results of the PROC SETINIT procedure. The set.log information is used to populate XTBL_L3_METADATA.
normalization.sas	[DMID]_[DATE]_dc_norm.sas7dat

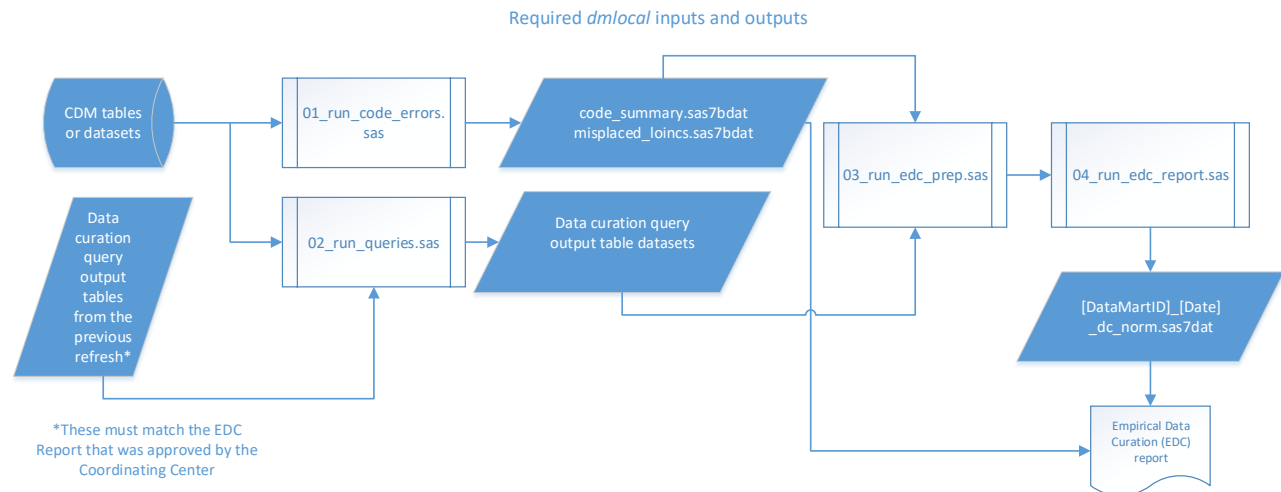
**Files to be returned to the DRN OC (*drnoc* folder).** DMID=DataMart ID; DATE=response date.

File name	Program produced by	File description
[DMID]_[DATE]_potential_code_errors.log	potential_code_errors.sas	The SAS log file for the program. Must be checked for errors and warnings.
[DMID]_[DATE]_Potential_Code_Errors.pdf	potential_code_errors.sas	The report produced by the program.
[DMID]_[DATE]_code_summary.cpt	potential_code_errors.sas	A SAS transport file containing the code error summary dataset produced by the program
<i>If all data curation queries are run at once</i> [DMID]_[DATE]_data_curation_all.cpt  <b>or</b>  <i>If data curation queries are run separately</i> [DMID]_[DATE]_data_curation_main.cpt [DMID]_[DATE]_data_curation_lab.cpt [DMID]_[DATE]_data_curation_obsclin.cpt [DMID]_[DATE]_data_curation_xtbl.cpt	data_curation_query_base.sas; data_curation_query_main.sas; data_curation_query_lab.sas; data_curation_query_obsclin.sas; data_curation_query_xtbl.sas	A SAS transport file containing all the SAS datasets produced by the program(s).
<i>If all data curation queries are run at once</i> [DMID]_[DATE]_data_curation_all.pdf  <b>or</b>  <i>If data curation queries are run separately</i> [DMID]_[DATE]_data_curation_main.pdf [DMID]_[DATE]_data_curation_lab.pdf [DMID]_[DATE]_data_curation_obsclin.pdf [DMID]_[DATE]_data_curation_xtbl.pdf	data_curation_query_base.sas; data_curation_query_main.sas; data_curation_query_lab.sas; data_curation_query_obsclin.sas; data_curation_query_xtbl.sas; data_curation_print.sas	A PDF containing a partial print of the output tables for the benefit of non-programmers. For ease of readability, it excludes the first three columns of the table (DataMartID, Response Date, and Query Package), and large tables are limited to the 100 most frequent observations. Empty tables are not printed.
<i>If all data curation queries are run at once</i> [DMID]_[DATE]_data_curation_query_all.log [DMID]_[DATE]_data_curation_query_base.log  <b>or</b>  <i>If data curation queries are run separately</i> [DMID]_[DATE]_data_curation_query_base.log [DMID]_[DATE]_data_curation_query_main.log [DMID]_[DATE]_data_curation_query_lab.log [DMID]_[DATE]_data_curation_query_obsclin.log [DMID]_[DATE]_data_curation_query_xtbl.log	data_curation_query_base.sas; data_curation_query_main.sas; data_curation_query_lab.sas; data_curation_query_obsclin.sas; data_curation_query_xtbl.sas	The SAS log files for the programs. Must be checked for errors and warnings.
[DMID]_[DATE]_data_curation_progress_report.rtf	data_curation_query_base.sas; data_curation_query_main.sas; data_curation_query_lab.sas; data_curation_query_obsclin.sas; data_curation_query_xtbl.sas	A rtf file containing table names and their processing time
[DMID]_[DATE]_dc_nom.cpt	normalization.sas	A SAS transport file containing a normalized version of all data curation query output tables except the <i>elapsed</i> and <i>datamart_all</i> datasets.
[DMID]_[DATE]_normalization.log	normalization.sas	The SAS log file for the program. Must be checked for errors and warnings.
[DMID]_[DATE]_EDCRPT.log	edc_report.sas	The SAS log file for the program. Must be checked for errors and warnings.

<b>File name</b>	<b>Program produced by</b>	<b>File description</b>
[DMID]_[ DATE]_EDCRPT.pdf	edc_report.sas	The report produced by the program.

## IX. Query Input and Output Diagram

The diagram below illustrates how the query package uses information from the CDM tables and prior data curation results to produce the datasets in the *dmlocal* and *dmoc* folder.



## X. Responding to the Query Package

- 1) Prepare for the query as instructed in the Query Package Checklist.
- 2) Open the query package. Extract the contents, save them locally as described in Sections VI, and create the *drnoc* and *dmlocal* folders.
- 3) If the CDM data is stored in database tables, do the following. Otherwise proceed to Step 4.
  - a) Consider compressing large tables to improve query response time.
  - b) Modify the user inputs to use appropriate SAS/ACCESS options on a LIBNAME statement so that the program knows where to find the database tables. The examples below show connection information for an Oracle database; connecting to other database systems may require different connection information.
    - (1) In the *sasprograms* folder, open **01\_run\_code\_errors.sas**, **02\_run\_queries.sas** and **04\_run\_edc\_report.sas** and edit the `dpath` variable to include the appropriate database connection information. Be sure to use the `%str()` function to mask the embedded equal signs. For example: 

```
%let dpath = %str(oracle user="myuserid"
orapw=mypasswd path=mydbname schema=myschema) ;
```
    - (2) In the *infolder* folder, open the **data\_curation\_query\_base.sas** program edit the `libname pccordata` statement on Line 32 to remove the quotation marks, as: `libname pccordata &dpath;`
- 4) Open all programs in the *sasprograms* folder and modify the directory paths and inputs as instructed below. For reasons of compatibility and standardization, directory paths must meet the following criteria:

- DO use forward slashes (e.g./) which are always compatible on both UNIX and WINDOWS.
  - DO use end of path separators (e.g./xyz/ and not /xyz) which are assumed by many programs.
  - DO use beginning of path separators (e.g./xyz) on UNIX.
  - DO NOT use beginning of path separators on WINDOWS (e.g.P:/xyz not /P:/xyz).
  - DO NOT surround directory paths with quotes (e.g./xyz/ not "/xyz").

  - a) After `%let dpath=`, provide the directory path where your PCORnet CDM SAS data is located.
  - b) After `%let qpath=`, provide the outer folder where the required folders were created.
  - c) In the **02\_run\_queries.sas** program, populate the following user inputs. NOTE: User inputs must be populated correctly for the programs to run correction.
    - i) After `%let threshold=`, change the default value of 0 if required by your institution.
    - ii) After `%let ets_installed=`, change the default value of Yes if needed as instructed in the program.
    - iii) After `%let _grp,` provide one of the query group process option: `all`, `main`, `lab`, `obsclin`, or `xtbl`. DO NOT mix and match query process group options. The default value is `all` and can be changed as described below:
      - (1) *To run the programs all at once:* Select “all” to run the data curation query programs as a batch; this option is recommended if you are not an experienced SAS user and for the final submission to the Coordinating Center.
      - (2) *To run the programs sequentially:* This is recommended for partners who have long run times and want to be able to remediate issues which only affect certain tables more easily. To do so, select one of the 4 options (`main`, `lab`, `obsclin`, or `xtbl`) for the initial run and then repeat with the remaining options as instructed in Step 6.
    - iv) After `%let lookback=`, leave the default value of 10.
  - d) In the **04\_run\_edc\_report.sas** program, populate the following user inputs:
    - i) After `%let ppath=`, provide the outer folder containing the most recently approved query results (i.e. results for the previous DataMart refresh). If this is the DataMart’s first refresh and the DataMart does not have results from a previous DataMart refresh, provide the same value that was provided for `dpath`.



- 5) Open the **01\_run\_code\_errors.sas** program. Run the program and review the log and output as instructed in the Query Package Checklist.
- 6) Open the **02\_run\_queries.sas** program. Run the program, either 1 time if you selected %let \_grp=all, or 4 times in the sequence you desire (e.g. first with lab, next with xtbl, next with obsclin, and finally with main). As it processes each query program, the program will print results to a PDF file, create permanent SAS datasets for each output table, and import all permanent SAS datasets into a SAS transport file. Review the logs and output (see [section IV](#) and [section VIII](#)) as instructed in the Query Package Checklist. You may wish to review the output tables which could contain required data check exceptions before proceeding. If you are working in Windows and executing the queries sequentially, you will need to close all open applications (e.g. PC SAS and Microsoft Word) before running the next program. Otherwise, you will get an error message from SAS like "Fatal ODS error has occurred. Unable to continue processing this output destination" and "File is in use". You can monitor the query progress by checking the [DMID]\_[DATE]\_data\_curation\_progress\_report.rtf document in the *dmoc* folder. Depending on your SAS processing environment, you may also see the same information in the SAS OUTPUT window or RESULTS window.
- 7) Open and run the **03\_run\_edc\_prep.sas** program. All data curation datasets must be present before proceeding with the EDC portion of this package. To ensure that this is the case, review the output in the result window. You should see a statement that says "No datasets are missing". If a dataset is missing, it will be listed in the output. If there is no output, confirm that you entered the correct information after %let qpath=. If necessary, rectify problems by returning to the **02\_run\_queries.sas** program to create the missing datasets.
- 8) Open and run the **04\_run\_edc\_report.sas** program. This program will first call the **normalization.sas** program to create a dataset which combines all the data curation query output tables ([DMID]\_[DATE]\_dc\_norm.sas7bdat). It will then call **edc\_report.sas** to create the Empirical Data Curation (EDC) report from the [DMID]\_[DATE]\_dc\_norm.sas7bdat dataset and the EDC reference datasets and prints results to a PDF file.
- 9) Review the EDC logs and output as instructed in the Query Package Checklist. In investigating Data Check exceptions, you may wish to review the reference datasets used by the data checks (see [Section VII](#)). You may access these by opening the dc\_reference.cpt or edc\_reference.cpt files, as shown in the example below.
 

```
libname outlib 'F:/pcor-net/myproject/';
%let infile= 'F:/pcor-net/myproject/infolder/edc_reference.cpt';
proc cimport infile=&infile library=outlib;
run;
```
- 10) If you need to modify your CDM data after running the queries, follow these guidelines for re-running the **02\_run\_queries.sas** programs, and then rerun the **03\_run\_edc\_prep.sas** and **04\_run\_edc\_report.sas** programs. DO NOT mix and match query process group options. If you do not use the "all" option, follow these guidelines for determining which query group(s) need to be re-run:
  - a) You must re-run the *main* program unless the only change you made is to a field in the HARVEST table that is not a REFRESH date or the DATAMARTID.
  - b) You must re-run the *xtbl* program if records were added or deleted, dates or identifier fields were changed, RACE or ENC\_TYPE were changed, or the HARVEST table was changed.
  - c) You must re-run the *lab* program if you make any changes to the LAB\_RESULT\_CM or LAB\_HISTORY tables.
  - d) You must re-run the *obsclin* program if you make any changes to the OBS\_CLIN table
- 11) Update the online ETL Annotated Dictionary as instructed in the Query Package Checklist.

- 12) If desired, verify the contents of the cpt files by using a `proc cimport` statement, as shown in the example below:

```
libname outlib 'F:/pcornet/myproject/';  
%let infile= 'F:/pcornet/myproject/T1D3_20151101_data_curation.cpt';  
proc cimport infile=&infile library=outlib;  
run;
```

- 13) Return the files in the *drnoc* folder (see [section VIII](#)) and a signed Query Package Checklist. If there is more than one version of any of the files in the *drnoc* folder, archive and/or delete the earlier versions and only return the ones with the most recent date (i.e., those reflecting the final results). Zip the contents of the *drnoc* folder into a file with your datamartid (e.g. [DATAMARTID]\_DRNOC.zip).
- 14) Retain all output from the final run in the *dmlocal* folder for use in subsequent data curation queries as shown in the Query Input and Output Diagram (see [Section IX](#)).

## XI. Version History

Date	Version	Description
Feb 3, 2016	v3.00	Original release.
Mar 17, 2016	v3.01	Corrected truncation of some query results by increasing field lengths. In VITAL_L3_HT, height categories of "<0" and "0-10" were both displaying as "0-10" due to a precision issue with PROC FORMAT/PROC MEANS; this was corrected. In PRO_L3_PXDATE_Y was incorrectly labeled ADMIT_DATE; this was corrected to PX_DATE. Updated all documentation and code to v3.01.
Nov 7, 2016	v3.02	Added queries of DEATH, DISPENSING, LAB_RESULT_CM, and PRESCRIBING (35 queries). Added 7 cross-table queries. Revised 14 queries (retained backwards compatibility). Revised the low cell count threshold logic to conform to PCORnet's new minimum bin size policy. Added the Empirical Data Curation Report.
Nov 18, 2016	v3.03	Eliminated the need for the SAS ACCESS/Interface to PC Files module. Resolves the following warning: "WARNING: In a call to the CATS function, the buffer allocated for the result was not long enough to contain the concatenation of all the arguments."
Mar 21, 2017	v3.04	Modified the programs so that optional variables which are 100% missing will not cause errors or omissions. In ENC_L3_ENCTYPE, corrected the calculations for ELIG_RECORD_N and UNIQUE_VISIT_N. In XTBL_L3_DASH2 and XTBL_L3_DASH3, changed the logic to use PRESCRIBING_RX_ORDER_DATE instead of RX_START_DATE. In XTBL_L3_DASH3, changed the logic to not require LAB_RESULT_CM.LAB_NAME to be populated. In Empirical Data Curation (EDC) Table IIE, corrected the highlighting and added the PRESCRIBING table for orphan ENCOUNTERIDS. In Table IIIB, corrected the percentage calculations. In EDC Table IVD, corrected the "% of encounters without a principal diagnosis" calculation.
Jul 5, 2017	V3.10	Modified queries to conform to CDM v3.1. Added queries of the CONDITION, PCORNET_TRIAL, DEATH_CAUSE, and PRO_CM tables. Added 12 queries pertaining to previously characterized tables. Revised 31 queries. Incorporated PCORnet Data Checks v3.
Sept 18, 2017	V3.11	In the Data Curation query, corrected an omission in the "enc_l3_enctype_disdisp" query. In EDC Table IIB, added RX_QUANTITY_UNIT and corrected calculation for PX_TYPE. In EDC Table IVC, added DX_ORIGIN. In EDC Table IVE, corrected the percentage calculation.
Nov 20, 2017	V3.12	Incorporated the PCORnet Code Errors v3 program. In the Data Curation query, added 13 queries pertaining to previously characterized tables; revised 3 queries; and deprecated 6 queries. In the Empirical Data Curation report, incorporated PCORnet Data Checks v4, added 1 table, and revised 14 tables.
June 8, 2018	V4.10	Modified existing queries to conform to CDM v4.1. Incorporated the PCORnet Code Errors v5 program. In the Data Curation query, added 40 queries (24 pertaining to previously characterized tables; 16 for tables new to CDM v4.1); revised 27 queries; and deprecated 2 queries. In the Empirical Data Curation report, incorporated PCORnet Data Checks v5, added 1 table and 2 charts, and revised 15 tables.
June 29, 2018	V4.11	Corrected the DIAGNOSIS and PROCEDURES information in Table ID. Added additional DATE_MGMT fields to Table IIB.
Oct 8, 2018	V4.12	Corrected minor bugs in v4.11. Split the data curation program into 3 programs. Separated the data curation and code errors "run" programs. In the Data Curation program, modified the lookback logic to remove the date restriction from the DEATH table and include records with non-missing dates. Added an Empirical Data Curation (EDC) preparation program. In the Empirical Data Curation program, updated the reference files to reflect Cycle 5 results and to exclude LOINC codes which have no variation from Data Check 2.06.
Dec 20, 2018	V4.13	Full data curation for OBS_CLIN and OBS_GEN tables. For the data curation query, added query progress check report; updated Value Set Reference File to v1.5; and added dia_l3_dxtype and pro_l3_pxtype. Updated Potential Code Errors to v6 which incorporates OBS_CLIN, OBS_GEN, and

Date	Version	Description
		MED_ADMIN. In the Empirical Data Curation report, incorporated PCORnet Data Checks v6, added 4 tables and 1 chart, revised 11 tables and charts and switched to PDF format.
Mar 19, 2019	V4.14	Corrected minor bugs in v4.13 affecting PRO_L3_PXTYPE and Data Check 2.07. Updated the Data Curation Lab Group reference file to v3.1 and the network-wide results displayed in the EDC report to the most recently available data.
Jun 17, 2019	V4.15	Set the low cell count threshold to 0. Added Section IX: Query Input and Output Diagram. Added information about the new CDM Value Set Conformance Query to Section X: Responding to the Query Package.
Oct 7, 2019	V5.10	Additions and revisions to support CDM v5.1 and PCORnet Data Checks v7. In the Data Curation query, added 42 queries (17 pertaining to previously characterized tables; 25 for tables new to CDM v5.1); revised 30 queries; and updated the reference files for Data Curation Lab Groups, RXCUIs, and lab outliers. In the Empirical Data Curation report, revised 21 tables and added 4 tables and 7 charts.
Jan 6, 2020	V5.11	Corrected minor bugs in V5.10. Updated the Data Check programming so that (a) Data Check 2.08 (monthly outliers) will work for DataMarts that have SAS_ETC, and (b) in Data Check 2.06 (lab outliers) the pediatric lab reference range will be applied to pediatric DataMarts created in October 2019.
Apr 7, 2020	V5.12	In the Data Curation program, added 1 new query and revised 31 queries. In DIA_L3_DASH1, ENC_L3_DASH1, ENC_L3_DASH2, XTBL_L3_DASH1, XTBL_L3_DASH2, XTBL_L3_DASH3, and VIT_L3_DASH1, the logic was changed to use SAS systemdate to calculate a consistent timespan for all DataMarts. Updated Empirical Data Curation programs to incorporate PCORnet Data Checks v8, to suppress printing of Tables IIB and IIC if all fields conform to specifications, and to allow the user to designate if SAS_ETC is installed. Removed Data Curation table shells from the WorkPlan since these are available in the technical specifications posted on iMeet.
July 6, 2020	V5.13	In the Data Curation program, removed 10 query output tables that were no longer needed and incorporated a new parseable file (2020-06-17-PCORnet-Common-Data-Model-v5dot1-parseable.xlsx). In the Empirical Data Curation programs, fixed a defect in Data Check 3.13 that was failing to flag some exceptions for lab tests where the lab volume percent was below threshold but above 0, and modified Data Check 2.05 to use a lookup table.
Dec 21, 2020	V6.00	Updated for the CDM v6.0 specifications and a new parseable file (2020_10_22_PCORnet_Common_Data_Model_v6dot0_parseable). In the Data Curation program, added 20 query output tables and revised 37 query output tables. In the Empirical Data Curation program, updated existing tables and data checks to support CDM v6.0, added Data Checks 1.15 and 1.16, revised the logic for Data Checks 3.01, 3.02, 3.07 and 3.09, and added charts IK and IL.
Apr 5, 2021	V6.01	In the Data Curation program, added 9 query output tables and added a new query group process option, obsclin.
Apr 7, 2021	V6.02	Fixed a bug in the Empirical Data Curation portion of the package that was causing duplicate rows in Table IVI and errors on the Data Check Summary page for Data Checks 3.10 and 3.12.
Jul 6, 2021	V6.03	Updated for a new parseable file (2021_04_12_PCORnet_Common_Data_Model_v6dot0_parseable). In the Data Curation program, added 6 new query output tables and revised 13 query output tables. In the Empirical Data Curation program, added 1 new table; revised the logic for Data Checks 2.07, 2.08, 3.02, 3.03, 3.06, and 4.02; and added Data Checks 1.17, 3.14, and 3.15.
Jan 3, 2022	V6.04	Updated for a new parseable file (2021_11_29_PCORnet_Common_Data_Model_v6dot0_parseable). In the Data Curation program, revised the default lookback period to 10 years, revised the lookback date for ENROLLMENT to consider either ENR_START_DATE or ENR_END_DATE, added 2 new query output tables, and revised 9 query output tables. In the Empirical Data Curation program, fixed the calculations for Data Checks 3.06 and 2.07 in Table IVE; fixed the calculation for PRESCRIBING Tier 1 brand in Table IVH; incorporated PCORnet Data Checks v11 by revising Data Checks 1.09 and 3.03 and adding Data Checks 1.18 and 1.19.
Jul 1, 2022	V6.05	In the Data Curation program, added 9 new query output tables and revised 9 query output tables. In the Empirical Data Curation program, fixed an error in Table IVE that affected Data Check 3.06 and incorporated <a href="#">PCORnet Data Checks v12</a> .

Date	Version	Description
Jan 03, 2023	V6.06	In the Data Curation program, revised 3 query output tables and revised the logic for calculating Length of Stay. In the Empirical Data Curation program, incorporated PCORnet Data Checks v13 and fixed errors in Data Check 1.19 and 2.07.