**Instructions and Attestation**

*This Check List is designed to facilitate the completion of required items for each data curation submission. Carefully review the Work Plan and Check List and sign the attestation below prior to submitting the Data Curation response. This document formalizes guidance that has previously been communicated to network partners. Items added in Cycle 5 are highlighted. Submissions which do not conform to these requirements will not be approved.*

*I confirm that I have followed the instructions and met the requirements defined in the Work Plan and Check List.*

*Name Date*

**Preparing for Data Curation**

1. Revise the ETL programs as needed to adhere to the [CDM v4.1 Specifications and Implementation Guidance](https://pcornet.imeetcentral.com/p/ZgAAAAAAnnU0) and [PCORnet Data Checks v5.](https://pcornet.imeetcentral.com/p/aQAAAAADmL2b) If you are not adhering to the Implementation Guidance, be prepared to explain why in the ETL Annotated Data Dictionary (ETL ADD). For fields with undefined field lengths, set the lengths as short as possible to minimize the space needed, and harmonize the lengths across all tables to facilitate cross-table querying.
2. Run the [PCORnet CDM Diagnostic Query v4.10](https://pcornet.imeetcentral.com/p/aQAAAAADmLPP) to ensure table-level and field-level conformance to CDM v4.1.
3. Include the most recent data available, ideally through 3 months prior to the query run date or later.
4. Include new data for as many tables as possible and ETL changes as needed.
5. Populate optional fields, especially those examined by the Data Checksa as completely and accurately as possible.
6. Verify the accuracy of **all** fields in the HARVEST table. These data are used extensively by the Coordinating Center to characterize and analyze DataMart-level attributes such as date obfuscation, data latency, and SAS configuration. The DataMart and Network IDs and Names are listed in the [HARVEST Reference Table](https://pcornet.imeetcentral.com/p/aQAAAAAC-f60). Confirm that the REFRESH\_DATES are the most recent date on which data were loaded. Remember to update CDM\_VERSION to v4.1.
7. Data are expected to persist across refreshes. Contact [drnoc@pcornet.org](mailto:drnoc@pcornet.org) to obtain approval prior to removing tables or records.
8. Create a static copy of the SAS version of your DataMart.

**Actions to Take after Running the Code Errors Query**

1. Review the log file for warnings and errors and remediate them as needed b *Note*: statements with the word \_ERROR\_ or the phrase “set the ERROR detection macro variable“ are acceptable.
2. Review the Potential Code Errors report. If 5% or more of the records for any given code type are potentially erroneous, investigate to determine the source of the error. An error rate of 5% or more must be approved by local personnel who have assessed the implications for PCORnet queries. Please note that queries search for codes within code types, so misclassified codes will be omitted from query results. The reason for the error(s) and the email of the person who approved this must be documented in the Other Questions page of the ETL Annotated Data Dictionary (ETL ADD).

**Actions to take after running the Data Curation Query**

1. Review the log file for warnings and errors and remediate them as needed.b
2. Review the data curation query output tables. At a minimum, review the following tables:
   1. *Basic metadata*: DATAMART\_ALL, XTBL\_L3\_METADATA, XTBL\_L3\_DATES
   2. *Code type frequencies:* DIA\_L3\_DXTYPE\_ENCTYPE, COND\_L3\_TYPE, MEDADM\_L3\_CODE\_TYPE, OBSCLIN\_L3\_TYPE, PROCM\_L3\_TYPE, PRO\_L3\_PXTYPE\_ENCTYPE
   3. *Highest volume codes*: DIA\_L3\_DX, DISP\_L3\_NDC, LAB\_L3\_LOINC, LAB\_L3\_SNOMED, MEDADM\_L3\_CODE\_TYPE, OBSCLIN\_L3\_CODE\_TYPE, PRES\_L3\_RXCUI, PRO\_L3\_PX
   4. *Other new tables not included in the EDC report*: DEM\_L3\_PATPREFLANG, DIA\_L3\_DXPOA, DIA\_L3\_PDX\_DETAIL, DISP\_L3\_DOSE\_DIST, DISP\_L3\_DOSEUNIT, DISP\_L3\_ROUTE, ENC\_L3\_PAYERTYPE1, ENC\_L3\_PAYERTYPE2, ENC\_L3\_FACILITYTYPE, MEDADM\_L3\_DOSEADM, MEDADM\_L3\_DOSEADMUNIT, MEDADM\_L3\_ROUTE, MEDADM\_L3\_SOURCE, PRES\_L3\_DISPASWRTN, PRES\_L3\_PRNFLAG, PRES\_L3\_RXDOSEFORM, PRES\_L3\_RXDOSEODR\_DIST, PRES\_L3\_RXDOSEODRUNIT, PRES\_L3\_ROUTE, PRES\_L3\_SOURCE, PRES\_L3\_RAWRXMED, PROCM\_L3\_ITEMFULLNAME, PROCM\_L3\_ITEMNM, PROCM\_L3\_MEASURE\_FULLNAME, PROCM\_L3\_MEASURENM, PRO\_L3\_PPX, PROV\_L3\_NPIFLAG, PROV\_L3\_SPECIALTY, PROV\_L3\_SPECIALTY\_GROUP, and PROV\_L3\_SEX
3. Confirm that the number of records for ICD9 and ICD10 diagnosis codes and ICD9, ICD10, and CPT/HCPCS procedure codes has remained stable or increased since the previous DataMart refresh. You can do this by summarizing the number of records per type from the DIA\_L3\_DXTYPE\_ENCTYPE and PRO\_L3\_PXTYPE\_ENCTYPE tables. If desired, you may use the *refresh\_comparison\_tool.xlsx* to make this comparison. A decrease of 5% or more since the prior refresh must be approved by local personnel who have assessed the implications for PCORnet queries. Please note that queries search for codes within code types, so misclassified codes will be omitted from query results. The reason for the decrease(s) and the email of the person who approved this change must be documented in the Data Persistence page of the ETL Annotated Data Dictionary (ETL ADD).

**Actions to take after running the Empirical Data Curation Query**

1. Review the log file for warnings and errors and remediate them as needed. b
2. Review the entire EDC report. Verify that all tables and charts which are relevant for your DataMart are present and that all information is accurate.
3. Address the following items in the EDC report.
   1. **Data Check exception summary**: Exceptions to required data checks are highlighted in red and must be corrected before returning results. Exceptions to investigative data checks are highlighted in blue and must be investigated and explained in the ETL ADD. Exceptions should not be classified as “warrants further investigation” for more than one cycle.
   2. **Table ID**: Verify the count of patients and records in each table. Confirm that the number of patients and records has remained stable or increased since the previous DataMart refresh. If desired, you may use the *refresh\_comparison\_tool.xlsx* to make this comparison. A decrease of 5% or more since the prior refresh must be approved by local personnel who have assessed the implications for PCORnet queries. The reason for the decrease(s) and the email of the person who approved the change must be documented in the Data Persistence page of the ETL ADD.
   3. **Table IE**: Verify the count of records per encounter type in each table. Confirm that the number of records for AV, ED, and IP encounters has remained stable or increased since the previous DataMart refresh. If desired, you may use the *refresh\_comparison\_tool.xlsx* to make this comparison. A decrease of 5% or more since the prior refresh must be approved by local personnel who have assessed the implications for PCORnet queries. The reasons for the decrease and the email of the person who approved the change must be documented in the Data Persistence page of the ETL ADD.
   4. **All charts**: Investigate significant inflection points and other unexpected patterns and explain them in the Chart Trend Analysis and Feedback section of the ETL ADD.
4. Update all sections of the online ETL ADD with information about the current DataMart refresh. This includes addressing any unanswered comments in the DataMart Specific Feedback section. *Note*: In order to ensure the accuracy of the ETL ADD information, previous answers to the questions on the following pages have been deleted: Data Persistence, Investigative Data Checks, Chart Trend Analysis and Feedback, and Implementation Guidance.
5. Sign and date this document and return it with the other query documents as instructed in the Work Plan.