**Instructions and Attestation**

*This Check List is designed to facilitate the completion of items required for each data curation submission. Carefully review the Work Plan and Check List and sign the attestation prior to submitting a Data Curation response. Items added in Cycle 6 are highlighted. Submissions which do not conform to these requirements will not be approved.*

**Preparing for Data Curation**

1. Revise the ETL programs as needed to adhere to the current [CDM v4.1 specifications](https://pcornet.imeetcentral.com/p/ZgAAAAAAnnU0) (including the specifications, the parseable file, and the Value Set Reference File) and [PCORnet Data Checks v6.](https://pcornet.imeetcentral.com/p/aQAAAAADvYYZ) 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 (e.g. PATID), set the lengths as short as possible to minimize the space needed, and harmonize the lengths across all tables to facilitate cross-table querying. NOTE: Failure to harmonize variable lengths will cause warnings in SAS logs.
2. If you have made any modifications to your CDM data structure, 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. If your data is stored in SAS datasets, consider using the [Create Indexes](https://pcornet.imeetcentral.com/p/aQAAAAADklz0) program to create indexes.
4. Include the most recent data available, ideally through 3 months prior to the query run date or later. Failure to include recent data will result in an Investigative Data Check exception in the Empirical Data Curation report.
5. Include new data for as many tables as possible and ETL changes as needed.
6. Populate optional fields, especially those examined by the Data Checksa as completely and accurately as possible.
7. 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. The DataMart and Network IDs and Names are listed in the [HARVEST Reference Table](https://pcornet.imeetcentral.com/p/aQAAAAADonhO). Confirm that the REFRESH\_DATES are the most recent date on which data were extracted, transformed and loaded into the CDM.
8. Data are expected to persist across refreshes. Selected data persistence issues will result in an Investigative Data Check exception in the Empirical Data Curation report.
9. Create a static copy of the SAS version of your DataMart.

**Actions to Take after Running the Potential 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 will result in a required Data Check exception in the Empirical Data Curation report.

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

1. Review the log file(s) 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, COND\_L3\_TYPE, MEDADM\_L3\_TYPE, OBSCLIN\_L3\_TYPE, OBSGEN\_L3\_TYPE, PROCM\_L3\_TYPE, PRO\_L3\_PXTYPE
   3. *Highest volume codes*: DIA\_L3\_DX, DISP\_L3\_NDC, LAB\_L3\_LOINC, LAB\_L3\_SNOMED, MEDADM\_L3\_CODE\_TYPE, OBSCLIN\_L3\_CODE\_TYPE, OBSGEN\_L3\_CODE\_TYPE, PRES\_L3\_RXCUI, PRO\_L3\_PX
   4. *Other new tables not included in the EDC report*: All results for OBS\_CLIN and OBS\_GEN except for OBS\_CLIN\_N and OBS\_GEN\_N.

**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 Empirical Data Curation (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 to Data Check 2.06 (lab outliers) must be further explained in the Investigative Data Check 2.06 Exceptions spreadsheet. Exceptions should not be classified as “warrants further investigation” for more than one cycle.
   2. **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.

**Completing the ETL ADD and submitting results**

1. Update all sections of the online ETL ADD with information about the current DataMart refresh. The ETL ADD must include the name and email of the person who assessed the implications of changes between refreshes and Investigative Data Check exceptions and approved the submission.
2. Sign and date this document and return it with the other query documents as instructed in the Work Plan.

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

*Name Date*