

Cycle 8 Data Curation Query Package Check List

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. 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 v5.1 specifications](#) (including the specifications and the parseable file) and [PCORnet Data Checks v8](#). 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 [CDM Diagnostic Query v5.10](#) to ensure table-level and field-level conformance to CDM v5.1.
3. If your data is stored in SAS datasets, consider using the [Create Indexes v2](#) program to improve run times.
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 checks, 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](#). Confirm that the REFRESH_DATES are the most recent date on which data were extracted, transformed and loaded into the CDM. DataMarts with an incorrect CDM_VERSION field will not be approved.
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. Consider running the self-service CDM Value Set Conformance Query to identify and resolve value set conformance errors prior to running the data curation query. The [CDM Value Set Conformance Query v5.11](#) examines the values populated in the 130 fields (20 tables) which have defined value sets, and allow users to designate which table(s) to check. The query produces a PDF file summarizing results and SAS datasets and CSV files of the non-conforming records if any are present.
10. Create a static copy of the SAS version of your DataMart.

Cycle 8 Data Curation Query Package Check List

Actions to take after running the Potential Code Errors Query

1. Review the log file for warnings and errors and remediate any errors. *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 any errors.
2. Review the data curation query output tables. At a minimum, review the following tables:
 - **Basic metadata:** DATAMART_ALL, XTBL_L3_METADATA, XTBL_L3_DATES
 - **Code type frequencies:** DIA_L3_DXTYPE, COND_L3_TYPE, MEDADM_L3_TYPE, OBSCLIN_L3_TYPE, OBSGEN_L3_TYPE, PROCM_L3_TYPE, PRO_L3_PXTYPE, IMMUNE_L3_CODETYPE
 - **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, IMMUNE_CODE_CODETYPE

Actions to take after running the Empirical Data Curation Query

1. Review the log file for warnings and errors and remediate any errors.
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.
 - 3.1. **Low cell counts and date obfuscation:** Exceptions to the CRN Statement of Work expectations regarding low cell count suppression and date obfuscation are highlighted in orange on Page 1 and in Table 1F.
 - 3.2. **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.
 - 3.3. **Data Persistence Checks:** For Data Check 4.01, 4.02 and 4.03, confirm that the data in the Previous Refresh column matches the ‘Current Refresh’ in the previous refresh’s EDC report Table VA, Table VB and Table VC.

V4.15 EDC Report Table VA

Table	Records	
	Previous Refresh	Current Refresh
DEMOGRAPHIC		4,976
ENROLLMENT		6,120
ENCOUNTER		120,765
DIAGNOSIS		219,482
PROCEDURES		281,109
VITAL		7,022
DEATH		1,485
PRESCRIBING		15,840
DISPENSING		14,738
LAB_RESULT_CM		14,708
CONDITION		7,997
DEATH_CAUSE		1,469
PRO_CM		104,496
PROVIDER		11,019
MED_ADMIN		4,976
OBS_CLIN		4,976
OBS_GEN		4,976

V5.10 EDC Report Table VA

Table	Previous Refresh
DEMOGRAPHIC	4,976
ENROLLMENT	6,120
ENCOUNTER	120,765
DIAGNOSIS	219,482
PROCEDURES	281,109
VITAL	7,022
DEATH	1,485
PRESCRIBING	15,840
DISPENSING	14,738
LAB_RESULT_CM	14,708
CONDITION	7,997
DEATH_CAUSE	1,469
PRO_CM	104,496
PROVIDER	11,019
MED_ADMIN	4,976
OBS_CLIN	4,976
OBS_GEN	4,976



Cycle 8 Data Curation Query Package Check List

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 _____