



## Data Curation Check List, Query Package v3.11

### Preparing for Data Curation

1. Include the most recent data available, ideally data through July 31, 2017 or later.
2. If you populate the LAB\_RESULT\_CM table, at a minimum include all common measures which pertain to your population and map them to both LAB\_LOINC and LAB\_NAME.
3. Ensure that ICD diagnosis and/or procedure codes after the ICD10 implementation date of October 2015 are captured and assigned to the correct code type.
4. Populate the following data provenance variables as precisely as possible: PX\_SOURCE, DX\_SOURCE, DX\_ORIGIN, VITAL\_SOURCE, and DEATH\_SOURCE.
5. Run the [Code Errors v2](#) program to identify 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.
6. Verify the accuracy of **all** fields in the HARVEST table. The DataMart and Network IDs and Names are listed in the [HARVEST Reference Table](#). These data are used extensively by the Coordinating Center to characterize and analyze DataMart-level attributes (e.g. date obfuscation, data recency, debugging SAS programs based on SAS configuration, etc.).
7. Create a static copy of the SAS version of your DataMart.

### After running the Data Curation Query

Review the data curation query output tables (PDF and/or SAS datasets). At a minimum, review the following: DATAMART\_ALL; XTBL\_L3\_METADATA; XTBL\_L3\_DATES, DIA\_L3\_DX, PRO\_L3\_PX, LAB\_L3\_NAME, LAB\_L3\_LOINC, DISP\_L3\_NDC, PRES\_L3\_RXCUI, VIT\_L3\_HT, VIT\_L3\_WT, VIT\_L3\_SYSTOLIC, VIT\_L3\_DIASTOLIC.

### After running the Empirical Data Curation Query

At a minimum, review the following:

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.
2. **Table ID:** Verify the count of patients and records in each table. Confirm that the number of patients has remained stable or increased since the previous DataMart refresh. Significant changes between refreshes must be explained in the ETL ADD.
3. **Table IE:** Confirm that the distribution is similar between the Encounter, Procedures, and Diagnosis tables. Significant differences should be explained in the ETL ADD.
4. **All charts:** Investigate significant inflection points and other unexpected patterns and explain them in the ETL ADD.
5. **Verify that the ETL ADD is complete and includes explanations and comments as instructed above.** DataMarts will not be approved until a complete ETL ADD is submitted.