Workplan: CDM Diagnostic Query V3.00

December 21, 2015

Table of Contents

I. Purpose and Scope

The purpose of this document is to describe PCORnet's CDM Diagnostic Query V3.00. This query examines a DataMart's metadata and evaluates the DataMart's conformance to the PCORnet Common Data Model (CDM) V3.0 SAS table structure, variable names, variable lengths, and data types. As noted on page 7 of the PCORnet CDM v3.0, "all tables must be present in an instantiation of the CDM, even if data are not populated in every table" (http://www.pcornet.org/pcornet-common-data-model/). This query examines all 15 tables and all fields except the RAW fields.

II. Files Included in Query Request

1. Workplan CDM Diagnostic Query V3.00.doc
2. Diagnostic.sas
3. Required_structure.cpt

III. Output Files

The program creates three output files: a diagnostic report (.PDF), a metadata file (.CPT), and a program log (.LOG). All file names will contain the user's DATAMARTID from the HARVEST table and the query response date. Descriptions of the diagnostic report and SAS table are provided below.

DIAGNOSTIC RESULTS.PDF

The SAS program produces a report which contains information about each of the 15 PCORnet CDM v3.0 tables, formatted as follows and containing one of the following Condition statements in each row.

Table Name	Field Name	Condition

(1) Expected Condition statements
   a. Table is present, meets requirements, and is populated. (Expected for the DEMOGRAPHIC, ENROLLMENT, ENCOUNTER, DIAGNOSIS, PROCEDURES, VITAL, and HARVEST tables).
   b. Table is present, meets requirement, and is not populated. (Permissible for the DISPENSING, LAB_RESULT_CM, CONDITION PRO_CM, PRESCRIBING, PCORNET_TRIAL, DEATH, and DEATH_CAUSE tables).

(2) Unexpected Condition statements
   a. Table is not present
   b. Required numeric field is not present
   c. Required character field is not present
   d. Required character field is numeric
   e. Required numeric field is character
   f. Required character field is present but of unexpected length

METADATA.CPT

The program produces a temporary SAS dataset called datamart_all.sas7bdat which contains the following SAS metadata for each table or variable. The only counts included in this table are the total number of observations in each table. To facilitate secure and stable transmission to the DRN OC, this dataset is saved as a SAS transport data set file (metadata.cpt). SAS transport data set files are machine-independent files that let users move SAS data sets between computers running different operating systems. Instructions for opening these files are included in Section IV. Running the SAS Program.

Library Name
Library Member Name
Data Set Label
Special Data Set Type
Variable Name
Variable Type
Variable Length
Variable Number
Variable Label
Variable Format
Variable Format Length
Number of Format Decimals
Variable Informat
Variable Informat Length
Number of Informat Decimals
Justification
Position in Buffer
Observations in Data Set
Engine Name
Create Date
Last Modified Date
Deleted Observations in Data Set
Use of Variable in Indexes
Library Member Type
Number of Indexes for Data Set
Password Protection
Update Flags
Compression Routine
Reuse Space
Sorted and/or Validated
Position of Variable in Sorted By Clause
Host Character Set
Collating Sequence
Sort Option; No Duplicate Keys
Sort Option; No Duplicate Records
Encryption Routine
Point to Observation
Maximum Number of Generations
Generation Number
Next Generation Number
Character Variables Transcoded
Number of observations

IV. Running the SAS program

1) Open CDM_DIAGNOSTIC_V300.zip and extract the contents. Save the DIAGNOSTIC.SAS and REQUIRED_STRUCTURE.CPT files in the same directory on your network. The directory should be reserved for query processing and should not be the same directory that contains the PCORnet CDM V3.0 data.
2) Using the SAS Editor, open DIAGNOSTIC.SAS. The program should not be altered in any way except to modify the directory paths specified in the program, following the instructions below.
   a) 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/")
   b) The directory paths as written expect the CDM data to be in SAS datasets, and will need to be modified slightly to run the queries on data stored in database tables. You will need to use the 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.
      - 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);`
      - Edit the libname pcordata statement to remove the quotation marks, as: `libname pcordata &dpath;`
3) Save and submit the program.
4) Review the output files and make any necessary updates to the DataMart’s tables so that only expected statements appear in the PDF. To unpack the .cpt transport file into a SAS dataset, use a proc cimport statement, as shown in the example below:

```
libname outlib 'F:/pcornet/myproject/'
%let infile= ‘F:/pcornet/myproject/DATAMARTID_YYYYMMDD_METADATA.CPT’;
proc cimport infile=&infile library=outlib;
run;
```

5) Zip the three output files (PDF, LOG and CPT files) into a file with the name of the file distribution query (e.g PROD_P02_DIAG_FDPRO_DIAQ_C1TEST_r001_v01.zip). Return the zip file to the PCORnet DRN OC via the DRN Query Tool. Instructions on how to respond to file distribution query requests are found here: https://popmednet.atlassian.net/wiki/display/DOC/File+Distribution.
6) Questions about this query package should be sent to Laura Qualls (laura.qualls@duke.edu).