



Work Plan: CMS Data Transformation Package v2.0

MODIFICATION HISTORY

Date	Version	Description
09/30/2017	V1.0	
01/19/2018	V2.0	The version 2 package will accommodate Medicare source files having either short and long column names.

TABLE OF CONTENTS

I.	PURPOSE AND SCOPE	3
II.	PROGRAM PACKAGE FILE STRUCTURE.....	4
III.	FILES INCLUDED IN THE PACKAGE	4
IV.	OUTPUT FILES.....	5
V.	RUNNING THE PACKAGE	7

I. PURPOSE AND SCOPE

The purpose of the Data Transformation Program package is to transform Medicare Research Identifiable files (RIF) to the PCORnet Common Data Model (CDM) v3.1 to facilitate and standardize the use of Medicare data in conducting PCORnet studies. Documentation about the CDM is available at <http://pcornet.org/pcornet-common-data-model/>.

The focus of the current Data Transformation Programs is to transform the following Medicare RIF files.

- Master Beneficiary Summary File - Base (A/B/D)
- Inpatient Institutional Claims .
- Outpatient Institutional Claims
- Carrier (i.e., Physician/Supplier) Claims
- Part D Prescription Drug Events

Other claim type files exist (e.g., Skilled Nursing Facility, Home Health) and may be added to this package in the future. The package does not include the Medicare Provider Analysis and Review (MedPAR) file transformation. Differences between the RIF and MedPAR files are described [here](#).

The Medicare data are able to populate the following PCORnet CDM tables:

- Demographic
- Enrollment
- Encounter
- Diagnosis
- Procedures
- Dispensing
- Death

The table below shows which Medicare files contribute to which PCORnet CDM v3.1 tables.

Input: Medicare source data	Output: PCORnet CDM v3.1 tables
Master Beneficiary Summary Files - Base (A/B/D)	Demographic Enrollment Death
Inpatient Claims Outpatient Claims Carrier Claims	Encounter Diagnosis Procedures
Part D Event (PDE) Files	Dispensing

The programs allow users to indicate which Medicare files are available for transformation and what type of column names, short or long, are used in the files. The programs are designed to run against CCW Medicare data in either J or K version file record layouts.

The data transformation programs initially generate annual CDM Encounter, Diagnosis, Procedures and Dispensing tables, to parallel how Medicare data are typically received. In a later step, these annual files are combined into single tables or views. The advantage of this approach is to avoid transforming the same year of claims data multiple times.

Questions about this query package should be sent to Yinghong Zhang (yinghong.zhang@duke.edu).

II. PROGRAM PACKAGE FILE STRUCTURE

The Data Transformation Program package contains several subfolders to organize program inputs and outputs. It is recommended (but not mandatory) that these subfolders reside within an outer folder labeled "ETL". The subfolders are as follows:

- **cdm_v31:** Contains all PCORnet CDM v3.1 tables/views as well as crosswalk tables generated by the transformation programs.
- **sasprograms:** Contains the master SAS program that must be edited and then executed locally.
- **infolder:** Contains all input programs needed to execute the master program. The contents of this folder should not be edited.
- **outfolder:** Contains output and log files generated by the programs that may be used to facilitate the debugging and internal checking.

III. FILES INCLUDED IN THE PACKAGE

The following files will be included in the Zip file distributed with the package.

sas_programs

- pcornet_cdm_etl_master_file.sas

infolder

- build_dx_px.sas
- build_enroll.sas
- build_harvest.sas
- build_xwlks.sas
- colist.sas7bdat
- create_CDM_v31_tablesviews.sas
- etl_carr.sas
- etl_denom.sas
- etl_enrollment.sas
- etl_ip.sas
- etl_op.sas
- etl_pde.sas

- pcncdm_formats.sas
- rtm_utility
- varlists.sas

CMS Data Transformation Package Work Plan v2.0.pdf
 2017-12-14 CMS PCORnet CDM 3.1 Mapping.xlsx

IV. OUTPUT FILES

SAS datasets generated in the cdm_v31 folder

File name	File description
death.sas7bdat	SAS dataset for PCORnet CDM Death table
demographic.sas7bdat	SAS dataset for PCORnet CDM Demographic table
diagnosisyyyy.sas7bdat	Annual SAS datasets for PCORnet CDM Diagnosis table
dispensingyyyy.sas7bdat (if available)	Annual SAS datasets for PCORnet CDM Dispensing table
encounteryyyy.sas7bdat	Annual SAS datasets for PCORnet CDM Encounter table
enrollment.sas7bdat	SAS dataset for PCORnet CDM Enrollment table
harvest.sas7bdat	SAS dataset for PCORnet CDM Harvest table
proceduresyyyy.sas7bdat	Annual SAS datasets for PCORnet CDM Procedures table
facility.sas7bdat	The crosswalk file for pseudo identifier to real facility code in Medicare data. It will be used to assign the pseudo identifier to the FACILITYID field.
provider.sas7bdat	The crosswalk file for pseudo identifier to real physician NPI number in Medicare data. It will be used to assign the pseudo identifier to the PROVIDERID field.

Notes: yyyy in file names represents four-digit year

The following SAS datasets will be generated in the cdm_v31 folder if you choose to use a permanent dataset to concatenate annual tables for each CDM table:

File name	File description
diagnosis.sas7bdat	Combined SAS dataset for PCORnet CDM Diagnosis table

File name	File description
dispensing.sas7bdat (if available)	Combined SAS dataset for PCORnet CDM Dispensing table
encounter.sas7bdat	Combined SAS dataset for PCORnet CDM Encounter table
procedures.sas7bdat	Combined SAS dataset for PCORnet CDM Procedures table

The following SAS DATA step views will be generated in the cdm_v31/dsview folder if you choose to use a view to concatenate annual tables for each CDM table:

File name	File description
death.sas7bview	SAS view for PCORnet CDM Death table
demographic.sas7bview	SAS view for PCORnet CDM Demographic table
diagnosis.sas7bview	Combined SAS view for PCORnet CDM Diagnosis table
dispensing.sas7bview (if available)	Combined SAS view for PCORnet CDM Dispensing table
encounter.sas7bview	Combined SAS view for PCORnet CDM Encounter table
enrollment.sas7bview	SAS view for PCORnet CDM Enrollment table
harvest.sas7bview	SAS view for PCORnet CDM Harvest table
procedures.sas7bview	Combined SAS view for PCORnet CDM Procedures table

Output and log files in the outfolder folder

File name	File description
PDF files	These PDF files contain the output summary of data transformation and mapping for programs in the infolder folder. Some output files report summary information by year. The create_CDM_v31_tablesviews.pdf file includes process runtime information for all ETL programs.
Log files	The SAS log file for the programs in the infolder folder.

V. RUNNING THE PACKAGE

- 1) The Medicare RIF files used as source data must be SAS datasets with original variable names. Use the SAS programs sent with the Medicare data distribution to convert the raw text files to the SAS datasets required by this package.
- 2) The source data can reside in the same folder, or in subfolders based on the file type.
- 3) Open the Data Transformation Programs package. Extract the contents, save them locally as described in Sections II and III.
- 4) Open the **sasprograms** folder and edit the SAS program file 'pcornet_cdm_etl_master_file.sas'.
- 5) Modify the directory paths as follows. For reasons of compatibility and standardization, directory paths must meet the following criteria:

- DO use forward slashes (e.g. /) which are always compatible with 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/").

- a) Set EPATH option. This is the root directory for data transformation workplans.
- b) Set SDPATH option. This is the main directory where the source data tables are stored.
- 6) Provide the sub-directory names for each type of source data as follows. *Skip this section if all source data are stored in the main directory.*
 - a) Set SDMBSFAB option. This is the sub-directory where mbsf_ab/mbsf_abcd data are stored.
 - b) Set SDIP option. This is the sub-directory where Inpatient claim data are stored.
 - c) Set SDOP option. This is the sub-directory where Outpatient claim data are stored.
 - d) Set SDCARR option. This is the sub-directory where Carrier claim data are stored.
 - e) Set SDMBSFD option. This is the sub-directory where mbsf_d/mbsf_abcd data are stored.
 - f) Set SDPDE option. This is the sub-directory where Part D drug event data are stored.
- 7) Provide the file name prefixes (e.g. mbsf_ab) for the Medicare SAS datasets, leaving off the four-digit year. The defaults are J version file names. You can replace them with K version file names or your custom names. Leave blank if the source table is not available at your site.
 - a) Set MBSFTBL option. This is the mbsf_ab/mbsf_abcd dataset name.
 - b) Set IPTBL option. This is the inpatient claim base dataset name.
 - c) Set IPREVTBL option. This is the inpatient revenue dataset name.
 - d) Set OPTBL option. This is the outpatient claim base dataset name.

- e) Set OPREVTBL option. This is the outpatient revenue dataset name.
 - f) Set CARTBL option. This is the carrier claim base dataset name.
 - g) Set CARLNTBL option. This is the carrier line dataset name.
 - h) Set PDETBL option. This is the Part D drug event dataset name.
 - i) Set PTDENOMTBL option. This is the mbsf_d/mbsf_abcd dataset name.
- 8) Set the following macro variables.
- a. Set NAMETYPE option. This is the indicator for the type of column names in your source files. The default is S. Set to L if source tables using long column names.
 - b. Set MKXWLK_HRVS option. This option indicates whether or not the harvest table and crosswalk files need to be created in this process. The default is Y. Set to N if harvest table and crosswalk tables have already been created.
 - c. Set STARTYR option. This is the first year of source data to transform.
 - d. Set ENDYR option. This is the last year of source data to transform.
 - e. Set PDSTARTYR option. This is the first year of Part D source data to transform.
 - f. Set PDENDYR option. This is the last year of Part D source data to transform.
 - g. Set DSVIEW option. This option indicates whether or not you want to combine the Annual tables to single views.
- 9) Save and run 'pcornet_cdm_etl_master_file'.
- 10) Check the log file for errors and warnings.
- 11) Review the output files.
- 12) Check the newly generated CDM tables in the cdm_v31 folder.