



Work Plan: CMS Data Transformation Package V4.0

MODIFICATION HISTORY

Date	Version	Description
09/30/2017	V1.0	Original release
01/19/2018	V2.0	The version 2 package will accommodate Medicare source files having either short and long column names.
01/08/2019	V3.0	The version 3 package will transform CMS data to PCORnet CDM v4.1
5/31/2019	V3.1	The version 3.1 package will have updated SAS transport file
5/22/2020	V4.0	The version 4 package will transform CMS data to PCORnet CDMv5.1

TABLE OF CONTENTS

I.	PURPOSE AND SCOPE.....	3
II.	PROGRAM PACKAGE FILE STRUCTURE	4
III.	FILES INCLUDED IN THE PACKAGE	5
IV.	OUTPUT FILES	5
V.	RUNNING THE PACKAGE.....	8
VI.	APPENDIX A.	10

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) v5.1 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
- Skilled Nursing Facility Claims
- Part D Prescription Drug Events
- Medicare Provider Analysis and Review (MedPAR)

The MedPAR file is an alternate source of information about inpatient hospital and skilled nursing facility (SNF) care, users can transform either the MedPAR file or Inpatient Claims and Skilled Nursing Facility Claims, but not both, to the PCORnet Common Data Model.

Other claim type files exist (e.g., DME, Home Health) and may be added to this package in the future.

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

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

For all CDM tables where no Medicare data are populated, empty tables will be generated. The table below shows which Medicare files contribute to which PCORnet CDM tables.

Input: Medicare source data	Output: PCORnet CDM tables
Master Beneficiary Summary Files - Base (A/B/D)	Demographic Enrollment Death Lds_address_history
Inpatient Claims Outpatient Claims Carrier Claims Skilled Nursing Facility (SNF) Claims Medpar File	Encounter Diagnosis Procedures Provider
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 Medicare files received from the CMS data distributor 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_v51:** Contains all PCORnet CDM v5.1 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_provider.sas
- build_xwls.sas
- create_CDM_v51_tablesviews.sas
- etl_carr.sas
- etl_denom.sas
- etl_enrollment.sas
- etl_ip.sas
- etl_op.sas
- etl_medpar.sas
- etl_snf.sas
- etl_pde.sas
- pcncdm_formats.sas
- rtm_utility
- varlists.sas
- reference_file.cpt
 - colist.sas7bdat
 - provider_specialty_mapping.sas7bdat
 - state_code_mapping

CMS Data Transformation Package Work Plan v4.0.pdf

2020-04-23 CMS PCORnet CDM 5 1 Mapping.xlsx

IV. OUTPUT FILES

SAS datasets generated in the cdm_v51 folder

File name	File description
death.sas7bdat	SAS dataset for PCORnet CDM Death table
demographic.sas7bdat	SAS dataset for PCORnet CDM Demographic table

File name	File description
lds_address_history	SAS dataset for PCORnet CDM Lds_address_history 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
provider.sas7bdat	SAS dataset for PCORnet CDM Provider 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.

Notes: yyyy in file names represents four-digit year

The following SAS datasets will be generated in the cdm_v51 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
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_v51/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

File name	File description
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

The following empty tables will be generated in the cdm_v51 or dsview folder:

- condition
- death_cause
- lab_result_cm
- med_admin
- obs_clin
- obs_gen
- vital
- prescribing
- pro_cm
- pcornet_trial
- immunization
- hash_token

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_v51_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 **SDSNF** option. This is the sub-directory where Skilled Nursing Facility (SNF) claim data are stored.
 - d) Set **SDOP** option. This is the sub-directory where Outpatient claim data are stored.
 - e) Set **SDCARR** option. This is the sub-directory where Carrier claim data are stored.
 - f) Set **SDMBSFD** option. This is the sub-directory where mbsf_d/mbsf_abcd data are stored.
 - g) Set **SDPDE** option. This is the sub-directory where Part D drug event data are stored.
 - h) Set **SDMP** option. This is the sub-directory where Medpar data are Stored.
- 7) Provide the file name prefixes (e.g. inptclms) for the Medicare SAS datasets, leaving off the version (e.g. j, k) and the four-digit year. You can replace them with 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 **IPBTBL** option. This is the inpatient claim base dataset name. Leave blank if you are transforming the MedPAR data.
 - c) Set **IPREVTBL** option. This is the inpatient revenue dataset name. Leave blank if you are transforming the MedPAR data.
 - d) Set **SNFTBL** option. This is the snf claim base dataset name. Leave blank if you are transforming the MedPAR data.
 - e) Set **SNFREVTBL** option. This is the snf revenue dataset name, leave blank if you are transforming the MedPAR data.
 - f) Set **MPTBL** option. This is the MedPAR dataset name, leave blank if you are transforming the inpatient claim data and snf claim data if available.
 - g) Set **OPTBL** option. This is the outpatient claim base dataset name.
 - h) Set **OPREVTBL** option. This is the outpatient revenue dataset name.
 - i) Set **CARTBL** option. This is the carrier claim base dataset name.
 - j) Set **CARLNTBL** option. This is the carrier line dataset name.
 - k) Set **PDETBL** option. This is the Part D drug event dataset name.
 - l) 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 if you used *_v6.sas read-in programs to create source tables from flat files. Set to L if source tables were created using *_v8.sas read-in programs. See Appendix A for an example of Short and 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, provider table and facility crosswalk table 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 **PDENYR** 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. Default is N if you want to generate tables only. Set to Y if you only want to combine the Annual tables to single views, set to B to create both permanent datasets and views.
- 9) Specify network id, network name, datamart id, datamart name and datamart platform that will be present in the HARVEST table.
- a. Set **NETID** option. This is the network ID.
 - b. Set **NETNM** option. This is the network name.
 - c. Set **DMID** option. This is the datamart ID.
 - d. Set **DMNM** option. This is the datamart name.
 - e. Set **DMPLFM** option. This is the datamart platform.
- 10) Save and run 'pcornet_cdm_etl_master_file'.
- 11) Check the log file for errors and warnings.
- 12) Review the output files.

13) Check the newly generated CDM tables in the cdm_v51 folder.

VI. APPENDIX A.

Short Column Name ¹	Long Column Name ²	Label
BENE_ID	BENE_ID	Encrypted CCW Beneficiary ID
RFRNC_YR	BENE_ENROLLMT_REF_YR	Reference Year
FIVEPCT	FIVE_PERCENT_FLAG	Medicare 5% Sample Indicator
EFIVEPCT	ENHANCED_FIVE_PERCENT_FLAG	Enhanced Medicare 5% Sample Indicator
COVSTART	COVSTART	Medicare Coverage Start Date
CRNT_BIC	CRNT_BIC_CD	Current Beneficiary Identification Code
STATE_CD	STATE_CODE	State code for beneficiary (SSA code)
CNTY_CD	BENE_COUNTY_CD	County code for beneficiary (SSA code)
BENE_ZIP	BENE_ZIP_CD	Zip code for beneficiary
AGE	BENE_AGE_AT_END_REF_YR	Age of beneficiary at end of year
BENE_DOB	BENE_BIRTH_DT	Beneficiary date of birth
V_DOD_SW	BENE_VALID_DEATH_DT_SW	Valid Date of Death Switch
DEATH_DT	BENE_DEATH_DT	Date of Death
NDI_DEATH_DT	NDI_DEATH_DT	National Death Index (NDI) Date of Death
SEX	BENE_SEX_IDENT_CD	Sex
RACE	BENE_RACE_CD	Beneficiary Race Code
RTI_RACE_CD	RTI_RACE_CD	Research Triangle Institute (RTI) Race Code
OREC	BENE_ENTLMT_RSN_ORIG	Original Reason for Entitlement Code
CREC	BENE_ENTLMT_RSN_CURR	Current Reason for Entitlement Code
ESRD_IND	BENE_ESRD_IND	End-stage Renal Disease (ESRD) Indicator
MS_CD	BENE_MDCR_STATUS_CD	Medicare Status Code
A_TRM_CD	BENE_PTA_TRMNTN_CD	Part A Termination Code
B_TRM_CD	BENE_PTB_TRMNTN_CD	Part B Termination Code
A_MO_CNT	BENE_HI_CVRAGE_TOT_MONS	Part A Months Count
B_MO_CNT	BENE_SMI_CVRAGE_TOT_MONS	Part B Months Count
BUYIN_MO	BENE_STATE_BUYIN_TOT_MONS	State Buy-In Coverage Count
HMO_MO	BENE_HMO_CVRAGE_TOT_MONS	HMO Coverage Count
BUYIN01	BENE_MDCR_ENTLMT_BUYIN_IND_01	Medicare Entitlement/Buy-In Indicator - January
BUYIN02	BENE_MDCR_ENTLMT_BUYIN_IND_02	Medicare Entitlement/Buy-In Indicator - February
BUYIN03	BENE_MDCR_ENTLMT_BUYIN_IND_03	Medicare Entitlement/Buy-In Indicator - March
BUYIN04	BENE_MDCR_ENTLMT_BUYIN_IND_04	Medicare Entitlement/Buy-In Indicator - April
BUYIN05	BENE_MDCR_ENTLMT_BUYIN_IND_05	Medicare Entitlement/Buy-In Indicator - May
BUYIN06	BENE_MDCR_ENTLMT_BUYIN_IND_06	Medicare Entitlement/Buy-In Indicator - June
BUYIN07	BENE_MDCR_ENTLMT_BUYIN_IND_07	Medicare Entitlement/Buy-In Indicator - July
BUYIN08	BENE_MDCR_ENTLMT_BUYIN_IND_08	Medicare Entitlement/Buy-In Indicator - August
BUYIN09	BENE_MDCR_ENTLMT_BUYIN_IND_09	Medicare Entitlement/Buy-In Indicator - September
BUYIN10	BENE_MDCR_ENTLMT_BUYIN_IND_10	Medicare Entitlement/Buy-In Indicator - October

Short Column Name ¹	Long Column Name ²	Label
BUYIN11	BENE_MDCR_ENTLMT_BUYIN_IND_11	Medicare Entitlement/Buy-In Indicator - November
BUYIN12	BENE_MDCR_ENTLMT_BUYIN_IND_12	Medicare Entitlement/Buy-In Indicator - December
HMOIND01	BENE_HMO_IND_01	HMO Indicator - January
HMOIND02	BENE_HMO_IND_02	HMO Indicator - February
HMOIND03	BENE_HMO_IND_03	HMO Indicator - March
HMOIND04	BENE_HMO_IND_04	HMO Indicator - April
HMOIND05	BENE_HMO_IND_05	HMO Indicator - May
HMOIND06	BENE_HMO_IND_06	HMO Indicator - June
HMOIND07	BENE_HMO_IND_07	HMO Indicator - July
HMOIND08	BENE_HMO_IND_08	HMO Indicator - August
HMOIND09	BENE_HMO_IND_09	HMO Indicator - September
HMOIND10	BENE_HMO_IND_10	HMO Indicator - October
HMOIND11	BENE_HMO_IND_11	HMO Indicator - November
HMOIND12	BENE_HMO_IND_12	HMO Indicator - December

¹Short Column Names will be displayed in the output file if using *_v6.sas to read in associated Medicare data.

²Long Column Names will be displayed in the output file if using *_v8.sas to read in associated Medicare data.