

Medicaid/CHIP Transformation to the Sentinel Common Data Model Version 8.2.0

User Documentation

Principal Investigator: Bradley Hammill, DrPH¹

Prepared by Michael Stagner¹; Steven J. Lippmann, PhD¹; Pratap Adhikari, MS¹; Jessica E. Pritchard, PhD¹; Bradley Hammill, DrPH¹

Author Affiliations: ¹Department of Population Health Sciences, Duke University School of Medicine, Durham, NC

Version 3.0.0

August 22, 2024

The Sentinel System is sponsored by the [U.S. Food and Drug Administration \(FDA\)](#) to proactively monitor the safety of FDA-regulated medical products and complements other existing FDA safety surveillance capabilities. The Sentinel System is one piece of FDA's [Sentinel Initiative](#), a long-term, multi-faceted effort to develop a national electronic system. Sentinel Collaborators include Data and Academic Partners that provide access to healthcare data and ongoing scientific, technical, methodological, and organizational expertise. The Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number 75F40119D10037.

History of Modifications

Version	Date	Modification	Author
1.0.0	07/07/2022	Original Version	Duke University Department of Population Health Sciences
1.1.0	12/16/2022	Addition of mother-infant linkage information	Duke University Department of Population Health Sciences
2.0.0	10/11/2023	Version for ETL2 and SCDM 8.1.0	Duke University Department of Population Health Sciences
3.0.0	08/22/2024	Version for ETL3 and SCDM 8.2.0	Duke University Department of Population Health Sciences

Table of Contents

- 1. Purpose.....1
- 2. Disclaimers1
- 3. Program Package.....1
 - 3.1. Folder Structure1
 - 3.2. SAS Programs..... 3
 - 3.2.1. Phase A ETL Main Program 3
 - 3.2.1.1. 000_run_etl.sas..... 3
 - 3.2.2. Phase A/B ETL Initialization and Set-up 3
 - 3.2.2.1. 001_etl_info.sas..... 3
 - 3.2.2.2. 002_options_libs.sas..... 3
 - 3.2.2.3. 003_formats.sas 3
 - 3.2.2.4. 004_remote_macros.sas 4
 - 3.2.2.5. 005_other_macros.sas 4
 - 3.2.2.6. 006_check_etl_info.sas 4
 - 3.2.2.7. 007_copy_prior.sas 4
 - 3.2.2.8. 008_prep_nppes 4
 - 3.2.2.9. 009_dq_exclude 4
 - 3.2.3. Phase A ETL Transformation of Beneficiary Information 4
 - 3.2.3.1. 010_etl_de_enr.sas 4
 - 3.2.3.2. 011_etl_de_dem.sas 4
 - 3.2.3.3. 012_etl_de_dth.sas 4
 - 3.2.4. Phase A ETL Transformation of Prescription Drug and Medical Utilization Files .. 4
 - 3.2.4.1. 020_etl_rx.sas 4
 - 3.2.4.2. 021_etl_ip.sas..... 4
 - 3.2.4.3. 022_etl_lt.sas 5
 - 3.2.4.4. 023_etl_ot.sas 5
 - 3.2.5. Phase A ETL Pre-Combination Data Fixes..... 5
 - 3.2.5.1. 030_pre_combine_fix.sas..... 5
 - 3.2.6. Phase A ETL Combination and Reconciliation of Intermediate Files..... 5
 - 3.2.6.1. 040_combine_enr.sas 5
 - 3.2.6.2. 041_combine_dem.sas 5
 - 3.2.6.3. 042_combine_dth.sas 5
 - 3.2.6.4. 043_combine_cod.sas 5
 - 3.2.6.5. 050_combine1_enc.sas..... 5
 - 3.2.6.6. 051_combine1_dia.sas..... 5
 - 3.2.6.7. 052_combine1_pro.sas..... 5
 - 3.2.6.8. 053_combine_dis.sas 5
 - 3.2.6.9. 054_combine2_enc.sas 5
 - 3.2.6.10. 055_combine2_dia.sas..... 6
 - 3.2.6.11. 056_combine2_pro.sas 6

3.2.7.	Phase A ETL Creation and Application of ID Crosswalks and Creation of ID SCDM Tables	6
3.2.7.1.	060_make_x_patid.sas	6
3.2.7.2.	061_make_x_encid.sas	6
3.2.7.3.	062_make_x_facid.sas	6
3.2.7.4.	063_make_x_provid.sas	6
3.2.7.5.	070_apply_x_enr.sas	6
3.2.7.6.	071_apply_x_dth.sas	6
3.2.7.7.	072_apply_x_cod.sas	6
3.2.7.8.	073_apply_x_dis.sas	6
3.2.7.9.	074_apply_x_enc.sas	6
3.2.7.10.	075_apply_x_dia.sas	7
3.2.7.11.	076_apply_x_pro.sas	7
3.2.8.	Phase A ETL Dataset Finalization	7
3.2.8.1.	080_finalize_enr.sas	7
3.2.8.2.	081_finalize_dem.sas	7
3.2.8.3.	082_finalize_dth.sas	7
3.2.8.4.	083_finalize_cod.sas	7
3.2.8.5.	084_finalize_dis.sas	7
3.2.8.6.	085_finalize_enc.sas	7
3.2.8.7.	086_finalize_dia.sas	7
3.2.8.8.	087_finalize_pro.sas	7
3.2.8.9.	088_finalize_provid.sas	7
3.2.8.10.	089_finalize_fac.sas	7
3.2.9.	Phase A/B ETL Processing Review	8
3.2.9.1.	090_review.sas	8
3.2.10.	Phase B ETL Main Program	8
3.2.10.1.	200_run_mil.sas	8
3.2.11.	Phase B ETL Mother-Infant Linkage	8
3.2.11.1.	201_etl_mil.sas	8
3.2.12.	Phase B ETL Dataset Finalization	8
3.2.12.1.	202_finalize_mil.sas	8
4.	Program Execution	8
4.1.	Define ETL-specific Information	8
4.1.1.	User Input for 000_run_etl.sas	8
4.1.2.	User Input for 001_etl_info.sas	11
4.1.2.1.	Input Block 1: Phase A and Phase B ETL Information for Source DE file(s)	11
4.1.2.2.	Input Block 2: Phase A and Phase B ETL Information for Source NDI file(s)	13
4.1.3.	User Input for 002_options_libs.sas	14
4.1.4.	User Input for 009_dq_exclude.sas	16
4.1.5.	User Input for 200_run_mil.sas	18
4.2.	Run the ETL Package	18

5.	Output Files	18
5.1.	Intermediate SCDM SAS Datasets	18
5.2.	Final SCDM SAS Datasets	19
5.3.	SAS Index Files.....	19
5.4.	Identifier Crosswalk Datasets	20
5.5.	NPPES Extract.....	20
5.6.	Log and Listing Files, and Run Information Tables	20

Index of Tables

Table 1. ETL directory structure	2
Table 2. Macro variables set by the user within 000_run_etl.sas	8
Table 3. Input block: specific information entered on each line	12
Table 4. Input block: specific information entered on each line	13
Table 5. Macro variables set by the user within 002_options_libs.sas	14
Table 6. SAS variables required for the dq_atlas_exclusions table	16
Table 7. Sample dq_atlas_exclusions table.....	17
Table 8. Macro variables output by the 009_dq_exclude.sas program.....	17
Table 9. Macro variables set by the user within 200_run_mil.sas	18
Table 10. Intermediate SCDM SAS datasets.....	18
Table 11. Final SCDM SAS datasets	19
Table 12. SAS index files	20
Table 13. Identifier crosswalk datasets.....	20
Table 14. NPPES extract	20
Table 15. Log and listing files, and run information tables	20

Abbreviations

Abbreviation	Description
CHIP	Children's Health Insurance Program
CCW	Chronic Conditions Warehouse
CMC	Comprehensive Managed Care
CMS	Centers for Medicare & Medicaid Services
CY	Calendar year
DE	Demographic and eligibility
DPHS	Duke University Department of Population Health Sciences
DQ	Data Quality (in the context of the DQ Atlas)
DUA	Data Use Agreement
ETL	Extract, transform, and load
FFS	Fee-for-service
ID	Identifier
IP	Inpatient
LT	Long-term care
MC	Managed care
MESF	Medicaid Enrollee Supplemental File
MIL	Mother-Infant Linkage
NDI	National Death Index
NPI	National Provider Identifier
NPPES	National Plan and Provider Enumeration System
OT	Other Services
QA	Quality Assurance (in the context of the QA package)
QC	Quality Control
RIF	Research Identifiable File
RX	Pharmacy
SCDM	Sentinel Common Data Model
TAF	T-MSIS Analytic Files
T-MSIS	Transformed Medicaid Statistical Information System
VRDC	Virtual Research Data Center

1. Purpose

This document describes the SAS program packages for the Phase A—that enables the transformation of Medicaid and Children’s Health Insurance Program (CHIP) research-identifiable files (RIFs) into the core tables¹ of the [Sentinel Common Data Model \(SCDM\) v8.2.0 format](#)—and Phase B—that links live deliveries and children in the Mother-Infant Linkage [MIL] table—ETLs. It also provides the instructions on how to run each program package. Note that in order to create an MIL table, a user must execute the Mother-Infant Identification package after completing Phase A. Additional information on programming for Phase A and Phase B can be found in this document.

Please see the accompanying Technical Specifications document for detailed information about the logic and rules for transformation of these data into the SCDM. In that document, as below, the phrase “source files” refers to the original Medicaid/CHIP RIF data.

2. Disclaimers

The programs in this package are specific to the processing of the 100% Medicaid/CHIP data within the Center for Medicare & Medicaid Services’ (CMS) [Virtual Research Data Center \(VRDC\)](#). Codebooks for the current data structure can be found at the [Chronic Conditions Data Warehouse](#) data portal. These programs will not work without modification for the transformation of Medicaid/CHIP data that are structured differently, although the programs can serve as a basis for adaptation to run with different structures or in other environments.

The VRDC user accounts provisioned for the Duke Department of Population Health Sciences (DPHS) for use in the Sentinel program have additional features that are not available, by default, to other VRDC users. This includes extra memory that can accommodate very large SAS hash objects to enable faster data merging on large datasets. DPHS has purchased additional disk space to accommodate large ETL files. VRDC users without additional memory and disk space will not be able to run the programs on the 100% Medicaid/CHIP data, although they may be able to run the programs on smaller samples of the data. DPHS VRDC user accounts are configured to access a large, dedicated SAS WORK space, which reduces the likelihood of job interruptions. VRDC users can only transform data to which they have approved access.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

3. Program Package

3.1. Folder Structure

There is no required folder structure for this package, but the user may wish to set up distinct folders for each purpose or type of file listed below. The table below summarizes the folder structure used by DPHS, which is all located within an ETL-specific parent directory.

¹ Core tables include ENROLLMENT, DEMOGRAPHIC, DISPENSING, ENCOUNTER, DIAGNOSIS, PROCEDURE, DEATH, FACILITY, and PROVIDER

Table 1. ETL directory structure

Folder Name	Description	Defined in...
/programs	<p>Program files folder</p> <p>All SAS programs from this package must be located in a single directory.</p>	000_run_etl.sas
/dev	<p>Intermediate and final SCDM SAS datasets from development (testing) Phase A ETL runs</p> <p>This package allows users to test the transformation programs on a 5% sample of the Medicaid/CHIP data. If this option is utilized, all output SAS datasets for development (testing) ETL runs are stored in this directory.</p> <p>The 5% random sample is determined by selecting beneficiaries where the last two digits of BENE_ID are equal to 05, 20, 45, 70, or 95. The methodology for determining this 5% sample is based off of the process used by CMS to create the enhanced Medicare 5% sample indicator (https://resdac.org/cms-data/variables/enhanced-medicare-5-sample-indicator)</p>	
/prelim	<p>Intermediate SCDM SAS datasets from production (final) Phase A and Phase B ETL runs</p> <p>This directory stores the intermediate SCDM datasets that result from the execution of this package, when not in testing or development mode. It also stores ETL-specific ID crosswalk files and provider information extracted from the National Plan and Provider Enumeration System (NPPES).</p>	002_options_lib.sas
/final	<p>Final SCDM SAS datasets from production (final) Phase A and Phase B ETL runs</p> <p>This directory stores the final SCDM datasets that result from the execution of this package, when not in testing or development mode.</p>	
/info/dev /info/mil /info/prod	<p>Information about development (testing) and production (final) Phase A and Phase B ETL runs</p> <p>Execution of this package produces SAS log and listing files and reference datasets that capture information about runtimes, dataset counts, and cross-table checks for each run of the program package. Separate sub-directories distinguish between development (/dev) and production Phase A (/prod) and Phase B (/mil) runs.</p>	

Folder Name	Description	Defined in...
	Temporary SAS datasets	
/temp	This package utilizes remote SAS sessions to parallel process certain stages of the transformation simultaneously; this approach necessitates a directory that is accessible by all remote sessions and the parent session to store temporary datasets. Users cannot rely on SAS work directories for this purpose. This directory is not intended for long-term data storage, so it is regularly cleared by different programs in this package. As a result, this directory must be distinct from any other directory listed above.	
	DQ Atlas-based exclusions	
/dq_atlas	This directory contains a SAS table of jurisdiction/year/plan exclusions named dq_atlas_exclusions.sas7bdat.	

3.2. SAS Programs

This program package includes the following SAS programs. [Section 5](#) describes program outputs. Related programs are grouped by their functions below.

The Phase A ETL (core tables, programs 000, 010-089) and Phase B ETL (MIL, programs 200-202) are separate, sequential processes. The Phase B ETL requires both the running of the Phase A ETL programming and the completion of the Phase A Quality Assurance (QA) package.

3.2.1. Phase A ETL Main Program

3.2.1.1. *000_run_etl.sas*

This master program controls the program flow of the Phase A ETL process. Users must edit this program to provide the programming root directory and to set the flags that control the Phase A ETL process, which are defined in [Section 4.1.1](#). This program is the first of three programs that the user must edit before running the Phase A ETL process. This is the only program that is executed by the user for the Phase A ETL.

3.2.2. Phase A/B ETL Initialization and Set-up

3.2.2.1. *001_etl_info.sas*

The user must edit this program to provide detailed information about the years/months of source data to be included in the Phase A and Phase B ETL processes. [Section 4.1.2](#) defines the required input. This program is the second of three programs that require editing by the user before running the Phase A or Phase B ETL.

3.2.2.2. *002_options_libs.sas*

The user must edit this program to define the SAS library names and locations used during the Phase A and Phase B ETL processes and to set options relevant to the Phase A and Phase B ETL processing. [Section 4.1.3](#) details these options. One of those options is whether to execute a development/testing run or a production run. This program is the third of three programs that require editing by the user before running the Phase A or Phase B ETL.

3.2.2.3. *003_formats.sas*

This program defines the SAS formats needed for the Phase A and Phase B ETL processes.

3.2.2.4. *004_remote_macros.sas*

This program contains the macros necessary for remote submission of jobs, via SAS/GRID or locally.

3.2.2.5. *005_other_macros.sas*

This program contains macros common to multiple Phase A and Phase B ETL programs, except macros related to remote submission.

3.2.2.6. *006_check_etl_info.sas*

This is a quality-control program that checks user inputs. It outputs user-specified Phase A and Phase B ETL information and options from the [001_etl_info.sas](#) and [002_options_libs.sas](#) programs to the listing.

3.2.2.7. *007_copy_prior.sas*

This program copies intermediate files from the prior Phase A ETL's preliminary dataset library, if desired, as designated by the user in [001_etl_info.sas](#).

3.2.2.8. *008_prep_nppes*

This program extracts and saves provider and specialty information from the NPPES data and prepares it for later merging onto the SCDM PROVIDER and FACILITY tables during the Phase A ETL process (See [062_make_x_facid.sas](#) and [063_make_x_provid.sas](#) below).

3.2.2.9. *009_dq_exclude*

This program creates two macro variables (one for fee-for-service [FFS] and one for comprehensive managed care [CMC] claims) for every data year in the Phase A ETL. These variables are lists of jurisdictions to be excluded by year/plan. The program uses the SAS table available in the /dq_atlas folder to create these variables. The format for this SAS table (dq_atlas_exclusions.sas7bdat) is described in [Section 4.1.4](#).

3.2.3. Phase A ETL Transformation of Beneficiary Information

3.2.3.1. *010_etl_de_enr.sas*

This program transforms source Demographic and Eligibility (DE) enrollment data into intermediate SCDM ENROLLMENT tables.

3.2.3.2. *011_etl_de_dem.sas*

This program transforms source DE demographics data into intermediate SCDM DEMOGRAPHIC tables.

3.2.3.3. *012_etl_de_dth.sas*

This program transforms source DE and Medicaid Enrollee Supplemental File (MESF) National Death Index (NDI) death data into intermediate SCDM DEATH and cause-of-death (COD) tables.

3.2.4. Phase A ETL Transformation of Prescription Drug and Medical Utilization Files

3.2.4.1. *020_etl_rx.sas*

This program transforms source monthly RX files (outpatient medication dispensing) into intermediate SCDM DISPENSING tables.

3.2.4.2. *021_etl_ip.sas*

This program transforms source monthly inpatient (IP) claims files into intermediate SCDM ENCOUNTER, DIAGNOSIS, and PROCEDURE tables.

3.2.4.3. *022_etl_lt.sas*

This program transforms source monthly long-term care (LT) claims files into intermediate SCDM ENCOUNTER, DIAGNOSIS, and PROCEDURE tables.

3.2.4.4. *023_etl_ot.sas*

This program transforms source monthly other services (OT) claims files into intermediate SCDM ENCOUNTER, DIAGNOSIS, and PROCEDURE tables.

3.2.5. Phase A ETL Pre-Combination Data Fixes

3.2.5.1. *030_pre_combine_fix.sas*

This program updates intermediate files with specific fixes related to 2020 BENE_IDs, 2016 Rhode Island postal codes, and death dates prior to birth before combination across years/source types.

3.2.6. Phase A ETL Combination and Reconciliation of Intermediate Files

3.2.6.1. *040_combine_enr.sas*

This program combines intermediate SCDM ENROLLMENT tables into a single table.

3.2.6.2. *041_combine_dem.sas*

This program combines intermediate SCDM DEMOGRAPHIC tables into a single table.

3.2.6.3. *042_combine_dth.sas*

This program combines intermediate SCDM DEATH tables into a single table.

3.2.6.4. *043_combine_cod.sas*

This program combines intermediate SCDM COD tables into a single table.

3.2.6.5. *050_combine1_enc.sas*

This program combines the intermediate yearly SCDM ENCOUNTER tables for IP, LT, and OT into a set of source-specific tables. The program also identifies and creates single stays from consecutive IP or LT claims, when appropriate, and creates a crosswalk that allows encounter IDs and admissions dates on related diagnosis and procedure records to be updated (see [051_combine1_dia.sas](#) and [052_combine1_pro.sas](#) below).

3.2.6.6. *051_combine1_dia.sas*

This program combines the intermediate IP, LT, and OT SCDM DIAGNOSIS tables from each year into a set of yearly combined tables. The program also updates encounter IDs and admissions dates for records determined to be part of the same stay in [050_combine1_enc.sas](#).

3.2.6.7. *052_combine1_pro.sas*

This program combines the intermediate IP, LT, and OT SCDM PROCEDURE tables from each year into a set of yearly combined tables. The program also updates encounter IDs and admissions dates for records determined to be part of the same stay in [050_combine1_enc.sas](#).

The following four programs are configured to run in parallel and then wait until all programs in this set are finished before proceeding.

3.2.6.8. *053_combine_dis.sas*

This program combines the intermediate SCDM DISPENSING tables into a single table.

3.2.6.9. *054_combine2_enc.sas*

This program combines all of the source-specific combined SCDM ENCOUNTER tables into a single table.

3.2.6.10. 055_combine2_dia.sas

This program combines all of the yearly combined SCDM DIAGNOSIS tables into a single table.

3.2.6.11. 056_combine2_pro.sas

This program combines all of the yearly combined SCDM PROCEDURE tables into a single table.

3.2.7. Phase A ETL Creation and Application of ID Crosswalks and Creation of ID SCDM Tables

The following “make_x” programs are configured to run in parallel and then wait until all programs in this set are finished before proceeding.

3.2.7.1. 060_make_x_patid.sas

This program replaces the original PatID values in the combined SCDM DEMOGRAPHIC table with consecutive integers and creates a PatID crosswalk file containing each pair of values.

3.2.7.2. 061_make_x_encid.sas

This program replaces the original EncounterID values in the combined SCDM ENCOUNTER table with consecutive integers and creates an EncounterID crosswalk file containing each pair of values.

3.2.7.3. 062_make_x_facid.sas

This program identifies unique FacilityID values in the combined SCDM ENCOUNTER table and creates a crosswalk file that associates these values with consecutive integers. This program also merges facility location data from the NPPES extract (*see 008_prep_NPPES.sas above*) and creates the SCDM FACILITY table.

3.2.7.4. 063_make_x_provid.sas

This program identifies unique ProviderID values in (a) the combined SCDM DISPENSING table, (b) the combined SCDM DIAGNOSIS table, and (c) the combined SCDM PROCEDURE table and creates a crosswalk file that associates these values with consecutive integers. This program also merges provider specialty data from the NPPES extract (*see 008_prep_nppes.sas above*) and creates the SCDM PROVIDER table.

The following “apply_x” programs are configured to run in parallel and then wait until all programs in this set are finished before proceeding.

3.2.7.5. 070_apply_x_enr.sas

This program applies the PatID crosswalk to the combined SCDM ENROLLMENT table.

3.2.7.6. 071_apply_x_dth.sas

This program applies the PatID crosswalk to the combined SCDM DEATH table.

3.2.7.7. 072_apply_x_cod.sas

This program applies the PatID crosswalk to the combined SCDM CAUSE_OF_DEATH table.

3.2.7.8. 073_apply_x_dis.sas

This program applies the PatID and ProviderID crosswalks to the combined SCDM DISPENSING table.

3.2.7.9. 074_apply_x_enc.sas

This program applies the PatID and FacilityID crosswalks to the combined SCDM ENCOUNTER table.

3.2.7.10. 075_apply_x_dia.sas

This program applies the PatID, EncounterID, and ProviderID crosswalks to the combined SCDM DIAGNOSIS table.

3.2.7.11. 076_apply_x_pro.sas

This program applies the PatID, EncounterID, and ProviderID crosswalks to the combined SCDM PROCEDURE table.

3.2.8. Phase A ETL Dataset Finalization

Note for 080-089 programs: The data are already implicitly sorted correctly at this stage. These programs validate the sort order and update the SAS dataset metadata to list the validated sort order.

The following “finalize” programs are configured to run in parallel and then wait until all programs in this set are finished before proceeding.

3.2.8.1. 080_finalize_enr.sas

This program finalizes the combined SCDM ENROLLMENT table by ensuring the data are properly sorted per SCDM specifications. It also indexes the table by PatID.

3.2.8.2. 081_finalize_dem.sas

This program finalizes the combined SCDM DEMOGRAPHIC table by ensuring the data are properly sorted. It also indexes the table by PatID.

3.2.8.3. 082_finalize_dth.sas

This program finalizes the combined SCDM DEATH table by ensuring the data are properly sorted. It also indexes the table by PatID.

3.2.8.4. 083_finalize_cod.sas

This program finalizes the combined SCDM CAUSE_OF_DEATH table by ensuring the data are properly sorted. It also indexes the table by PatID.

3.2.8.5. 084_finalize_dis.sas

This program finalizes the combined SCDM DISPENSING table by ensuring the data are properly sorted. It also indexes the table by PatID.

3.2.8.6. 085_finalize_enc.sas

This program finalizes the combined SCDM ENCOUNTER table by ensuring the data are properly sorted.

3.2.8.7. 086_finalize_dia.sas

This program finalizes the combined SCDM DIAGNOSIS table by ensuring the data are properly sorted.

3.2.8.8. 087_finalize_pro.sas

This program finalizes the combined SCDM PROCEDURE table by ensuring the data are properly sorted.

3.2.8.9. 088_finalize_provid.sas

This program finalizes the SCDM PROVIDER table by ensuring the data are properly sorted. It also indexes the table by ProviderID.

3.2.8.10. 089_finalize_fac.sas

This program finalizes the SCDM FACILITY table by ensuring the data are properly sorted. It also indexes the table by FacilityID.

3.2.9. Phase A/B ETL Processing Review

3.2.9.1. 090_review.sas

This program checks the [log and listing files](#) for errors, warnings, and other important information, including dataset counts, timing results, etc.

3.2.10. Phase B ETL Main Program

3.2.10.1. 200_run_mil.sas

This master program controls the program flow of the MIL process. Users must edit this program to provide the programming root directory and to set the flags that control the MIL process, which are defined in [Section 4.1.1](#). This program is the first of three programs that the user must edit before running the Phase B ETL (the other programs being [001_etl_info.sas](#) and [002_options_libs.sas](#)). This is the only program that is executed by the user for the Phase B ETL.

3.2.11. Phase B ETL Mother-Infant Linkage

3.2.11.1. 201_etl_mil.sas

This program links live birth deliveries and children identified by Phase A ETL QA package and saves the links into an intermediate SCDM MOTHER_INFANT_LINKAGE table. Any linkages resulting in a single child being matched with more than one delivery/mother are excluded from the intermediate file.

3.2.12. Phase B ETL Dataset Finalization

3.2.12.1. 202_finalize_mil.sas

This program finalizes the combined SCDM MOTHER_INFANT_LINKAGE table and ensures the data are properly sorted per SCDM specifications.

4. Program Execution

The package is executed by submitting one program ([000_run_etl.sas](#) for the Phase A ETL, [200_run_mil.sas](#) for the Phase B ETL) that calls and executes the remaining programs. The Phase A ETL generates the SCDM core tables and must be run before the Phase B ETL, which creates the SCDM MIL table. Before submission, users must define the parameters of their computing environment and their ETL. Each of these parameters is described below.

4.1. Define ETL-specific Information

The user must define several macro variables and provide specific source data information prior to program package execution.

4.1.1. User Input for 000_run_etl.sas

Table 2. Macro variables set by the user within [000_run_etl.sas](#)

Macro Variable	Description
ETLPROG	Path to the directory containing this package's program files. This is the location of the "Programs files folder" described in Section 3.1 .
ETLINFO	Path to the directory for storing run information, including log/lst files and runtimes and runinfo datasets.
The flags below define the scope of the Phase A ETL processing.	
RUN_ALL	Set to Y to run the entire Phase A ETL process. If set to Y, the step-specific flags described below are ignored.

Macro Variable	Description
	Set to N to run specific parts of the Phase A ETL process. If set to N , then each of step-specific flags described below must be set to Y or N , as desired.
COPY_PRIOR	Set to Y to copy intermediate files from a prior Phase A ETL, if desired. Set to N to skip.
PREP_NPPES	Set to Y to extract provider and facility information from the NPPES to prepare for use in the SCDM PROVIDER and FACILITY tables. Set to N to skip.
DQ_EXCLUDE	Set to Y to generate lists of jurisdictions to exclude by year/plan. Set to N to skip.
ETL_DE_ENR	Set to Y to transform the source DE file(s) into intermediate SCDM ENROLLMENT table(s). Set to N to skip.
ETL_DE_DEM	Set to Y to transform the source DE file(s) into intermediate SCDM DEMOGRAPHIC table(s). Set to N to skip.
ETL_DE_DTH	Set to Y to transform the source DE file(s) to intermediate SCDM DEATH table(s). Set to N to skip.
ETL_RX	Set to Y to transform the source RX file(s) to intermediate SCDM DISPENSING table(s). Set to N to skip.
ETL_IP	Set to Y to transform the source IP file(s) to intermediate SCDM ENCOUNTER, DIAGNOSIS, and PROCEDURE tables. Set to N to skip.
ETL_LT	Set to Y to transform the source LT file(s) to intermediate SCDM ENCOUNTER, DIAGNOSIS, and PROCEDURE tables. Set to N to skip.
ETL_OT	Set to Y to transform the source OT file(s) to intermediate SCDM ENCOUNTER, DIAGNOSIS, and PROCEDURE tables. Set to N to skip.
COMBINE_ENR	Set to Y to combine the intermediate SCDM ENROLLMENT files into a single SCDM ENROLLMENT file. Set to N to skip.
COMBINE_DEM	Set to Y to combine the intermediate SCDM DEMOGRAPHIC files into a single SCDM DEMOGRAPHIC file. Set to N to skip.
COMBINE_DTH	Set to Y to combine the intermediate SCDM DEATH files into a single SCDM DEATH file. Set to N to skip.
COMBINE_COD	Set to Y to combine the intermediate SCDM COD files into a single SCDM CAUSE_OF_DEATH file. Set to N to skip.
COMBINE1_ENC	Set to Y to combine the intermediate yearly SCDM ENCOUNTER tables for IP, LT, and OT into a set of source-specific tables. Set to N to skip.

Macro Variable	Description
COMBINE1_DIA	Set to Y to combine the intermediate IP, LT, and OT SCDM DIAGNOSIS tables from each year into a set of yearly combined tables. Set to N to skip.
COMBINE1_PRO	Set to Y to combine the intermediate IP, LT, and OT SCDM PROCEDURE tables from each year into a set of yearly combined tables. Set to N to skip.
COMBINE_DIS	Set to Y to interleave the intermediate SCDM DISPENSING files into a single SCDM DISPENSING file. Set to N to skip.
COMBINE2_ENC	Set to Y to combine all of the source-specific combined SCDM ENCOUNTER tables into a single table. Set to N to skip.
COMBINE2_DIA	Set to Y to combine all of the yearly combined SCDM DIAGNOSIS tables into a single table. Set to N to skip.
COMBINE2_PRO	Set to Y to combine all of the yearly combined SCDM PROCEDURE tables into a single table. Set to N to skip.
MAKE_X_PATID	Set to Y to generate a PatID crosswalk and replace PatID on the combined SCDM DEMOGRAPHIC file. Set to N to skip.
MAKE_X_ENCID	Set to Y to generate an EncounterID crosswalk and replace EncounterID on the combined SCDM ENCOUNTER file. Set to N to skip.
MAKE_X_FACID	Set to Y to generate a FacilityID crosswalk and create the SCDM FACILITY file. Set to N to skip.
MAKE_X_PROVID	Set to Y to generate a ProviderID crosswalk and create the SCDM PROVIDER file. Set to N to skip.
APPLY_X_ENR	Set to Y to apply the PatID crosswalk to the combined SCDM ENROLLMENT file. Set to N to skip.
APPLY_X_DTH	Set to Y to apply the PatID crosswalk to the combined SCDM DEATH file. Set to N to skip.
APPLY_X_COD	Set to Y to apply the PatID crosswalk to the combined SCDM CAUSE_OF_DEATH file. Set to N to skip.
APPLY_X_DIS	Set to Y to apply the PatID and ProviderID crosswalks to the combined SCDM DISPENSING file. Set to N to skip.
APPLY_X_ENC	Set to Y to apply the PatID and FacilityID crosswalks to the combined SCDM ENCOUNTER file. Set to N to skip. <i>Note: EncounterID transformation is applied in the MAKE_X_ENCID step above.</i>
APPLY_X_DIA	Set to Y to apply the PatID, EncounterID, and ProviderID crosswalks to the combined SCDM DIAGNOSIS file.

Macro Variable	Description
	Set to N to skip.
APPLY_X_PRO	Set to Y to apply the PatID, EncounterID, and ProviderID crosswalks to the combined SCDM PROCEDURE file. Set to N to skip.
FINALIZE_ENR	Set to Y to finalize the combined SCDM ENROLLMENT file by ensuring sort order and creating a PatID index. Set to N to skip.
FINALIZE_DEM	Set to Y to finalize the combined SCDM DEMOGRAPHIC file by ensuring sort order and creating a PatID index. Set to N to skip.
FINALIZE_DTH	Set to Y to finalize the combined SCDM DEATH file by ensuring sort order and creating a PatID index. Set to N to skip.
FINALIZE_COD	Set to Y to finalize the combined SCDM CAUSE_OF_DEATH file by ensuring sort order and creating a PatID index. Set to N to skip.
FINALIZE_DIS	Set to Y to finalize the combined SCDM DISPENSING file by ensuring sort order and creating a PatID index. Set to N to skip.
FINALIZE_ENC	Set to Y to finalize the combined SCDM ENCOUNTER file by ensuring sort order. Set to N to skip.
FINALIZE_DIA	Set to Y to finalize the combined SCDM DIAGNOSIS file by ensuring sort order. Set to N to skip.
FINALIZE_PRO	Set to Y to finalize the combined SCDM PROCEDURE file by ensuring sort order. Set to N to skip.
FINALIZE_PROVID	Set to Y to finalize the combined SCDM PROVIDER file by ensuring sort order and creating a ProviderID index. Set to N to skip.
FINALIZE_FAC	Set to Y to finalize the combined SCDM FACILITY file by ensuring sort order and creating a FacilityID index. Set to N to skip.

4.1.2. User Input for 001_etl_info.sas

The program [001_etl_info.sas](#) allows users to provide detailed information about the source data transformed by this package. Three different input blocks require editing; all are structured text that is read by SAS as raw input. As such, this input must follow a `datalines` statement and precede a semicolon (;) and a `run` statement.

Users may flexibly define the ETL build type of the Phase A ETL by specifying whether to copy prior intermediate SCDM files or to (re)create them in the current process. The companion Technical Specifications document describes three general build types—complete rebuild, incremental build, hybrid build—that are supported by this functionality.

4.1.2.1. Input Block 1: Phase A and Phase B ETL Information for Source DE file(s)

This input block contains information about the DE files that populate the SCDM DEMOGRAPHICS, ENROLLMENT, DEATH, and CAUSE_OF_DEATH files, medical utilization files - IP, LT, and OT claims - that populate the SCDM ENCOUNTER, DIAGNOSIS, and PROCEDURE files, and the outpatient prescription medication dispensing (RX) files that

populate the SCDM DISPENSING files. The entire period of time from the minimum ETL date to the maximum ETL date must be covered by the lines of this input block. Each calendar year that is part of the ETL process must be represented by only one line of the input block.

Table 3. Input block: specific information entered on each line

Field Name	Description
YR	Year of DE file specified below. This value is the 4-digit file suffix of derived files.
MSTART	First month of information that is used for a specified year. The month is designated numerically. It is almost always 1 (January), unless the minimum ETL date occurs within this year and is set to something other than January. Full months are used because enrollment and benefits information is captured only at the monthly level (the DE Eligibility Dates file only contains dates related to enrollment, not benefits); mid-month dates cannot be specified.
MEND	Last month of information that is used for a specified year. The month is designated numerically. It is usually 12 (December), unless the maximum ETL date occurs within this year and is set to something other than December. Full months are used because enrollment and benefits information is captured only at the monthly level (the DE Eligibility Dates file only contains dates related to enrollment, not benefits); mid-month dates cannot be specified.
SRCLIB	The SAS libname for the source utilization files
METHOD_ENR	Set to ETL to create the intermediate SCDM ENROLLMENT file for this year from the source file. Set to COPY to copy the intermediate SCDM ENROLLMENT file for this year from the prior ETL.
METHOD_DEM	Set to ETL to create the intermediate SCDM DEMOGRAPHICS file for this year from the source file. Set to COPY to copy the intermediate SCDM DEMOGRAPHICS file for this year from the prior ETL.
METHOD_DTH	Set to ETL to create the intermediate SCDM DEATH file for this year from the source file. Set to COPY to copy the intermediate SCDM DEATH file for this year from the prior ETL.
METHOD_IP	Set to ETL to create the intermediate SCDM files (ENCOUNTER, DIAGNOSIS, PROCEDURE) from the IP source files for this time period. Set to COPY to copy the intermediate SCDM files (ENCOUNTER, DIAGNOSIS, PROCEDURE) based on the IP source files for this time period from the prior ETL.
METHOD_OT	Set to ETL to create the intermediate SCDM files (ENCOUNTER, DIAGNOSIS, PROCEDURE) from the OT source files for this time period. Set to COPY to copy the intermediate SCDM files (ENCOUNTER, DIAGNOSIS, PROCEDURE) based on the OT source files for this time period from the prior ETL.

Field Name	Description
METHOD_LT	Set to ETL to create the intermediate SCDM files (ENCOUNTER, DIAGNOSIS, PROCEDURE) from the LT source files for this time period. Set to COPY to copy the intermediate SCDM files (ENCOUNTER, DIAGNOSIS, PROCEDURE) based on the LT source files for this time period from the prior ETL.
METHOD_RX	Set to ETL to create the intermediate SCDM DISPENSING file from the source PDE files for this time period. Set to COPY to copy the intermediate SCDM DISPENSING file for this time period from the prior ETL.

The sample input block below instructs the program package to (a) copy intermediate SCDM ENROLLMENT, DEMOGRAPHIC, IP, OT, and LT-derived ENCOUNTER, DIAGNOSIS, and PROCEDURE, and DISPENSING files for 2014 to 2018 from the prior ETL, (b) create new intermediate SCDM DEATH files for 2014 to 2018, and (c) create new intermediate SCDM files from DE, IP, OT, LT, and RX files for 2019.

```
*YR MSTART MEND SRCLIB ENR DEM DTH IP OT LT RX;
datalines;
2014 1 12 TAFR14 COPY COPY ETL COPY COPY COPY COPY
2015 1 12 TAFR15 COPY COPY ETL COPY COPY COPY COPY
2016 1 12 TAFR16 COPY COPY ETL COPY COPY COPY COPY
2017 1 12 TAFR17 COPY COPY ETL COPY COPY COPY COPY
2018 1 12 TAFR18 COPY COPY ETL COPY COPY COPY COPY
2019 1 12 TAFR19 ETL ETL ETL ETL ETL ETL ETL
;
```

4.1.2.2. Input Block 2: Phase A and Phase B ETL Information for Source NDI file(s)

This input block contains information about the NDI files that populate the SCDM CAUSE_OF_DEATH file. Not all years of the ETL will necessarily have NDI data. Each calendar year that is part of the ETL process must be represented by only one line of the input block.

Table 4. Input block: specific information entered on each line

Field Name	Description
YR	Year of NDI file specified below. This value is the 4-digit file suffix of derived files.
MSTART	First month of information that is used for a specified year. The month is designated numerically. It is almost always 1 (January), unless the minimum ETL date occurs within this year and is set to something other than January. Full months are used because enrollment and benefits information is captured only at the monthly level (the DE Eligibility Dates file only contains dates related to enrollment, not benefits); mid-month dates cannot be specified.
MEND	Last month of information that is used for a specified year. The month is designated numerically. It is usually 12 (December), unless the maximum ETL date occurs within this year and is set to something other than December. Full months are used because enrollment and benefits information is captured only

Field Name	Description
	at the monthly level (the DE Eligibility Dates file only contains dates related to enrollment, not benefits); mid-month dates cannot be specified.
SRCDS	Fully-specified (library.dataset) source MES_NDI dataset name to use
METHOD_COD	Set to ETL to create the intermediate SCDM CAUSE_OF_DEATH file for this year from the source file. Set to COPY to copy the intermediate SCDM CAUSE_OF_DEATH file for this year from the prior ETL. This value need to be the same as METHOD_DTH for every year in Input Block 1 (above), since DeathDt in the SCDM DEATH table relies on both DE and NDI data.

The sample input block below instructs the program package to (a) copy intermediate SCDM CAUSE_OF_DEATH files for 2014 to 2018 from the prior ETL and (b) create new intermediate SCDM CAUSE_OF_DEATH files for 2019 to 2020.

```
*YR MSTART MEND SRCDS COD;
datalines;
2014 1 12 MES.MES_NDI_2014 COPY
2015 1 12 MES.MES_NDI_2015 COPY
2016 1 12 MES.MES_NDI_2016 COPY
2017 1 12 MES.MES_NDI_2017 COPY
2018 1 12 MES.MES_NDI_2018 COPY
2019 1 12 MES.MES_NDI_2019 ETL
2020 1 12 MES.MES_NDI_2020 ETL
;
run;
```

4.1.3. User Input for 002_options_libs.sas

Table 4 defines the macro variables set by the user within [002_options_libs.sas](#). Some pre-specified, general SAS options that can be modified are included at the end of this program.

Table 5. Macro variables set by the user within [002_options_libs.sas](#)

Macro Variable	Description
The variables below define Phase A and Phase B ETL-specific information.	
ETLNUM, ETLVRSN	Number and version of the current ETL (ETLNUM, ETLVRSN) and the prior approved ETL (LASTNUM, LASTVRSN). This package is designed to accommodate iterative versions within a given ETL to reflect changes to the ETL programs without a change to the underlying data.
LASTNUM, LASTVRSN	In the distributed code, these macro values are used as placeholder components within file paths that specify the locations of multiple SAS libraries. They are not needed if file paths are explicitly defined rather than constructed using these placeholder components.
MINDT	The minimum date for encounters and other information included in the current ETL process. Format as SAS date literal "DDMONYYYY"d (e.g., "01JAN2014"d).

Macro Variable	Description
MAXDT	The maximum date for encounters and other information included in the current ETL process. Format as SAS date literal "DDMONYYYY"d (e.g., "31DEC2018"d).
ETLDEV	Set to Y to run the ETL process on a 5% sample of Medicaid/CHIP beneficiaries for testing and development. Set to N for the production run of the Phase A ETL process, using all data available. Set to MIL for the production run of the Phase B ETL process, using all data available. The 5% random sample is determined by selecting beneficiaries where the last two digits of BENE_ID are equal to 05, 20, 45, 70, or 95. The methodology for determining this 5% sample is based off of the process used by CMS to create the enhanced Medicare 5% sample indicator (https://resdac.org/cms-data/variables/enhanced-medicare-5-sample-indicator)
DEMOCUT	Cut-off for excluding BENE_IDs due to number of records in a yearly DEMOG_ELIG_BASE file. An excessive number of demographic/eligibility base file records linked to a single BENE_ID indicate that more than one beneficiary is associated with that BENE_ID. Duke uses a cut-off value of 3.
The variables below define MIL-specific information.	
MCNCUT	Cut-off for excluding MSIS_CASE_NUMs due to number of associated BENE_IDs. An excessive number of BENE_IDs linked to a single MSIS_CASE_NUM indicates a source data issue with that MSIS_CASE_NUM. Duke uses a cut-off value of 18 (99.9% of all MSIS_CASE_NUMs from 2014-2018 are associated with 18 or fewer BENE_IDs).
DAYDIF	Number of days an infant's birth date can be pre- or post-delivery admission date when making mother-infant linkages. Duke uses a value of 3 days, creating a window that allows for out of hospital births or other data anomalies.
The variables below define remote SAS session processing information.	
NUMSESS	Number of remote sessions available to open for parallel processing
USEGRID	Set to Y if SAS/GRID is enabled in the computing environment and must be used for remote session management. Set to N to use the local machine for remote sessions.
GRIDSRV	Name of the SAS/GRID server system used for remote sessions; required if USEGRID = Y.
The SAS LIBNAMEs below specify the locations of SAS datasets for the Phase A and Phase B ETL processes.	
SAS LIBNAMEs for source data in the VRDC are pre-set for SAS Enterprise Guide. Enter additional SAS LIBNAMEs in this file, as desired.	
FINAL	Path to the directory that stores the final SCDM datasets from the current ETL's production run
DEV	Path to the directory that stores the intermediate and final SCDM datasets from the current ETL's development run
PRELIM	Path to the directory that stores the intermediate SCDM datasets from the current ETL's production run
TEMP	Path to the directory that stores temporary files for jobs submitted to remote sessions

Macro Variable	Description
INFODEV	Path to the directory that stores the SAS logs and listing files and tables containing run information from development runs
INFOPROD	Path to the directory that stores the SAS logs and listing files and tables containing run information from Phase A production runs
INFOMIL	Path to the directory that stores the SAS logs and listing files and tables containing run information from Phase B production runs
DQ	Path to the directory that stores the SAS table of jurisdiction/year/plan exclusions
LASTPROD	Path to the directory that stores the final SCDM datasets from the most recent prior ETL's production run. Files from this directory may be used for comparisons between current and prior ETL production runs.
LASTDEV	Path to the directory that stores the intermediate and final SCDM datasets from the most recent prior ETL's development run. Files from this directory may be copied for use in the current ETL development run. Files from this directory may also be used for comparisons between current and prior ETL development runs.
LASTPREL	Path to the directory that stores the intermediate SCDM datasets from the most recent prior ETL's production run. Files from this directory may be copied for use in the current ETL production run.

4.1.4. User Input for 009_dq_exclude.sas

Program [009_dq_exclude.sas](#) creates yearly macro variables lists of jurisdictions by plan to be excluded from the Phase A ETL. These macro variables are used by programs [010_etl_de_enr.sas](#), [020_etl_rx.sas](#), [021_etl_ip.sas](#), [022_etl_lt.sas](#), and [023_etl_ot.sas](#) to restrict source data files jurisdiction/year/plan.

In order to create these jurisdiction/year/plan exclusion lists, program [009_dq_exclude.sas](#) requires a SAS table named dq_atlas_exclusions.sas7bdat saved in the user-specified DQ SAS library. This SAS table includes one record per jurisdiction per year per plan to be excluded from the Phase A ETL. For the Medicaid/CHIP Sentinel DataMart, jurisdiction/year/plan exclusions are determined by Duke DPHS prior to the Phase A ETL programming process using data quality information from the Medicaid and CHIP DQ Atlas (<https://www.medicaid.gov/dq-atlas/welcome>). A copy of the dq_atlas_exclusions table used by DPHS is included with the Medicaid/CHIP ETL codepack, and a detailed explanation of the process used to determine exclusions for Medicaid/CHIP ETL can be found in the Sentinel TAF Characteristics Memo. It would also be possible to populate the dq_atlas_exclusions table based on a separate set of exclusion criteria, or even manually for testing purposes. Table 5 describes the variable names required for the dq_atlas_exclusions table.

Table 6. SAS variables required for the dq_atlas_exclusions table

Variable	Description
The variables below define Phase A ETL-specific information.	
STATE_CD	Two-letter USPS code of jurisdiction to be excluded (i.e. AL, AK, AZ, AR, CA, CO, CT, DE, DC, FL, GA, GU, HI, ID, IL, IN, IA, KS, KY, LA, ME, MD, MA, MI, MN, MS, MO, MT, NE, NV, NH, NJ, NM, NY, NC, ND, OH, OK, OR, PA, PR, RI, SC, SD, TN, TX, UT, VT, VA, WA, WV, WI, WY)
DATA_YEAR	Year (4-digit) to be excluded (e.g. 2014, 2018)

Variable	Description
PLAN	Code to signify whether to exclude all beneficiaries (fee-for-service and comprehensive managed care) for a jurisdiction/year (ALL) or only CMC beneficiaries (CMC). There is never a situation where only FFS beneficiaries are excluded.

*This document uses the two-letter jurisdiction abbreviations recognized by the United States government. See <https://secure.ssa.gov/poms.nsf/lnx/0901501010> for a complete list.

The sample SAS table below instructs the program package to exclude (a) all beneficiaries from Alabama for 2014-2018, (b) CMC beneficiaries from Colorado for 2014-2016, (c) CMC beneficiaries in Rhode Island for 2014 and 2017, and (d) all beneficiaries from Rhode Island for 2018.

Table 7. Sample dq_atlas_exclusions table

STATE	DATA_YEAR	PLAN
AL	2014	ALL
AL	2015	ALL
AL	2016	ALL
AL	2017	ALL
AL	2018	ALL
CO	2014	CMC
CO	2015	CMC
CO	2016	CMC
RI	2014	CMC
RI	2017	CMC
RI	2018	ALL

The outputs of program [009_dq_exclude.sas](#) are macro variable lists of jurisdictions to exclude from the Phase A ETL by year/plan. The names and structure of these variables are described below.

Table 8. Macro variables output by the [009_dq_exclude.sas](#) program

Macro Variable	Description
The variables below define Phase A ETL-specific information.	
ALL_[yyyy]	List of jurisdictions (as two-letter USPS codes) from which to exclude all beneficiaries in [yyyy] (e.g. AL, CO, DC, RI). Postal code values are determined by applying a format from program 003_formats.sas to the original jurisdiction names in the dq_atlas_exclusions table.
CMC_[yyyy]	List of jurisdictions (as two-letter USPS codes) from which to exclude comprehensive managed care beneficiaries in [yyyy]. Postal code values are determined by applying a format from program 003_formats.sas to the original jurisdiction names in the dq_atlas_exclusions table.

4.1.5. User Input for 200_run_mil.sas

Table 9. Macro variables set by the user within [200_run_mil.sas](#)

Macro Variable	Description
ETLPROG	Path to the directory containing this package's program files. This is the location of the "Programs files folder" described in Section 3.1.
ETLINFO	Path to the directory for storing run information, including log/lst files and runtimes and runinfo datasets.
SAS LIBNAMEs specific to the Phase B ETL	
QA	Path to the directory that stores the deliveries (R01_MOTHER_DELIVERIES) and infants (R02_INFANTS) tables output by the Phase A QA package
The flags below define the scope of the Phase B ETL processing.	
RUN_ALL	Set to Y to run the entire Phase B ETL process. If set to Y, the step-specific flags described below are ignored. Set to N to run specific parts of the Phase B ETL process. If set to N, then each of step-specific flags described below must be set to Y or N, as desired.
ETL_MIL	Set to Y to link live birth deliveries and children and save into intermediate MOTHER_INFANT_LINKAGE table. Set to N to skip.
FINALIZE_MIL	Set to Y to finalize the combined SCDM MOTHER_INFANT_LINKAGE file and ensure sort order. Set to N to skip.

4.2. Run the ETL Package

Run the [000_run_etl.sas](#) (Phase A ETL core tables) or [200_run_mil.sas](#) (Phase B ETL MIL) program in batch mode. The Phase A ETL and Phase B ETL are separate, sequential processes. The Phase B ETL requires both the running of the Phase A ETL programming and the completion of the Phase A Quality Assurance (QA) package. These are the only programs that requires execution, and they incorporate other programs as needed.

5. Output Files

The following files are generated by this ETL package.

5.1. Intermediate SCDM SAS Datasets

Multiple files of each type listed below are generated for any specific year [yyyy] or time period [suffix], according to the user input in [001_etl_info.sas](#). (See [Section 4.1.2](#) for inputs.) These have a file extension of .sas7bdat.

Table 10. Intermediate SCDM SAS datasets

Dataset Name	Description
COD_[yyyy]	Intermediate SCDM CAUSE_OF_DEATH file; derived from NDI table(s)
DEATH_[yyyy]	Intermediate SCDM DEATH file; derived from DE table(s)
DEMOGRAPHIC_[yyyy]	Intermediate SCDM DEMOGRAPHIC file; derived from DE table(s)

Dataset Name	Description
DISPENSING_[suffix]	Intermediate SCDM DISPENSING file; derived from RX claim(s)
DX_IP_[suffix]	Intermediate SCDM DIAGNOSIS file; derived from IP claim(s)
DX_OT_[suffix]	Intermediate SCDM DIAGNOSIS file; derived from OT claim(s)
DX_LT_[suffix]	Intermediate SCDM DIAGNOSIS file; derived from LT claim(s)
ENC_IP_[suffix]	Intermediate SCDM ENCOUNTER file; derived from IP claim(s)
ENC_OT_[suffix]	Intermediate SCDM ENCOUNTER file; derived from OT claim(s)
ENC_LT_[suffix]	Intermediate SCDM ENCOUNTER file; derived from LT claim(s)
ENROLLMENT_[yyyy]	Intermediate SCDM ENROLLMENT file; derived from DE table(s)
PX_IP_[suffix]	Intermediate SCDM PROCEDURE file; derived from IP claim(s)
PX_OT_[suffix]	Intermediate SCDM PROCEDURE file; derived from OT claim(s)
PX_LT_[suffix]	Intermediate SCDM PROCEDURE file; derived from LT claim(s)
MOTHER_INFANT_LINKAGE	Intermediate SCDM MOTHER_INFANT_LINKAGE file; contains only linked deliveries-infants, not all deliveries and infants identified by the Phase A QA package

5.2. Final SCDM SAS Datasets

These have a file extension of .sas7bdat. Each file contains a single final SCDM table.

Table 11. Final SCDM SAS datasets

Dataset Name	Description
CAUSE_OF_DEATH	Final SCDM CAUSE_OF_DEATH file
DEATH	Final SCDM DEATH file
DEMOGRAPHIC	Final SCDM DEMOGRAPHIC file
DIAGNOSIS	Final SCDM DIAGNOSIS file
DISPENSING	Final SCDM DISPENSING file
ENCOUNTER	Final SCDM ENCOUNTER file
ENROLLMENT	Final SCDM ENROLLMENT file
FACILITY	Final SCDM FACILITY file
PROCEDURE	Final SCDM PROCEDURE file
PROVIDER	Final SCDM PROVIDER file
MOTHER_INFANT_LINKAGE	Final SCDM MOTHER_INFANT_LINKAGE file

5.3. SAS Index Files

Although not required to be SCDM-compliant, indexes are created on PatID for many final SCDM SAS datasets. These have a file extension of .sas7bndx and are stored in the same directory as the final SCDM SAS datasets.

Table 12. SAS index files

Index Name	Description
CAUSE_OF_DEATH	Final SCDM CAUSE_OF_DEATH file index
DEATH	Final SCDM DEATH file index
DEMOGRAPHIC	Final SCDM DEMOGRAPHIC file index
DISPENSING	Final SCDM DISPENSING file index
ENROLLMENT	Final SCDM ENROLLMENT file index

5.4. Identifier Crosswalk Datasets

These have a file extension of .sas7bdat. Each file contains a crosswalk between source or intermediate identifiers and SCDM-compliant identifiers. These crosswalks are ETL-specific.

Table 13. Identifier crosswalk datasets

Dataset Name	Description
XWALK_ENCID	Crosswalk table pairing source or intermediate encounter IDs with final, SCDM-compliant EncounterIDs
XWALK_FACID	Crosswalk table pairing source or intermediate facility IDs with final, SCDM-compliant FacilityIDs
XWALK_PATID	Crosswalk table pairing source or intermediate patient IDs with final, SCDM-compliant PatIDs
XWALK_PROVID	Crosswalk table pairing source or intermediate provider IDs with final, SCDM-compliant ProviderIDs

5.5. NPPES Extract

This has a file extension of .sas7bdat. The file contains an extract of the NPPES containing provider and facility information that is used in the SCDM PROVIDER and FACILITY tables. Both providers and facilities extracted from the NPPES are identified by NPI. The source data is the NPPES file corresponding to last month-year in a Phase A ETL's data period. This file is ETL-specific.

Table 14. NPPES extract

Dataset Name	Description
NPPES	Extract of provider information, such as provider specialty, and facility information, such as facility location, that are used to populated the final SCDM PROVIDER and FACILITY tables

5.6. Log and Listing Files, and Run Information Tables

Log and listing files are run-specific. In the file names below, [yyyymmdd] represents the date that [000_run_etl.sas](#) or [200_run_mil.sas](#) was submitted.

Table 15. Log and listing files, and run information tables

File Name	Description
_etl_dev_[yyyymmdd].log	Log file for the development run
_etl_dev_[yyyymmdd].lst	Listing file for the development run
_etl_prod_[yyyymmdd].log	Log file for the production run
_etl_prod_[yyyymmdd].lst	Listing file for the production run
_etl_mil_[yyyymmdd].log	Log file for the MIL run
_etl_mil_[yyyymmdd].lst	Listing file for the MIL run
runtimes.sas7bdat	Table recording the start, stop, and runtime for each step of the Phase A or Phase B ETL processes

File Name	Description
runinfo.sas7bdat	Table recording information about the Phase A or Phase B ETL runs, including dataset counts, missing data counts, and cross-table checks