# CODI-PQ READ ME

The Health FFRDC developed open source queries that can be used to estimate prevalence among children and teens by weight category using convenience samples. This document describes how to organize and submit the CODI prevalence queries. The queries will be referenced as the CODI prevalence queries (CODI-PQs).

## Background

Individuals are likely to have data at multiple organizations. It is critical to be able to use linked information gathered from different places to construct a complete picture of an individual for research purposes.

CODI-PQ is a process whereby researchers create childhood obesity prevalence estimates from non-probability samples derived from EHR data.

The algorithms described in this document, were designed to

1. impute race for children and teens missing race information, and
2. create estimate childhood obesity prevalence by weight categories, including:

* **Underweight**: less than 5th percentile
* **Healthy Weight**: 5th percentile to less than the 85th percentile
* **Overweight**: 85th to less than the 95th percentile
* **Obesity**[[1]](#footnote-1): 95th percentile to less than 120 percent of the body mass index (BMI) value for the 95th percentile
* **Severe Obesity**: 120 percent or greater of the BMI value for the 95th percentile.

Throughout this document, we refer to two different data types:

(1) IQVIA - a large sample of data drawn from EHRs across the United States, based on the data set with which we tested the algorithm; and

(2) CODI - a smaller set of data derived from a distributed health data network operating CODI infrastructure.

Our algorithms can generate prevalence estimates derived from a large sample of data from across the United States (e.g., IQVIA data) at either the state or ZCTA-3 level. The CODI algorithms generate prevalence estimates at the county level.

The purpose of this document is to provide the guidance necessary for database programmers to utilize CODI-PQ. It provides a description of:

* CODI-PQ data inputs and pre-processing
* CODI-PQ processing
* CODI-PQ analytic output

The CODI-PQ and pre-processing algorithms described in this document were created and tested with IQVIA’s Ambulatory Electronic Medical Record (AEMR) data[[2]](#footnote-2) and synthetic data generated for CODI using Synthea.[[3]](#footnote-3) All statistical programs described in this document were created and tested using SAS 9.4 software (SAS Institute, Inc., Cary, North Carolina). The guidance provided in this document is implemented through open source programs.

## Audience

The primary audience for this document is the database programmer submitting CODI-PQ algorithms. This document is written with database programmers as a primary focus. Please refer to GitHub for further methodological details regarding the implementation of the algorithms.[[4]](#footnote-4)

## Assumptions

The CODI-PQ algorithms require careful review of the methodological details, EHR data from a large population, and American Community Survey data from 2018. The person using the query should have a working knowledge of SAS language and macros.

## See Also

This file works in conjunction with the Implementation Guide, also available at:

<https://github.com/NORC-UChicago/CODI-PQ>

The two documents work together to explain the methodology, files, inputs, and output.

## Document Organization

This document is organized as follows:

* Part 1 – SAS Setup
* Part 2 – Downloading and Organizing the Files
* Part 3 –Pre-processing CODI-PQ
* Part 4 –CODI-PQ
* Appendix A – ACS Pre-processing Input File Layout
* Appendix B – EHR Pre-processing Input File Layout
* Appendix C – ACS Pre-processing Output File Layout
* Appendix D – EHR Pre-processing Output File Layout
* Appendix E – County Specific Example Code
* Appendix F – ZCTA3 Specific Example Code
* Appendix G – Example CODI-PQ Output
* Appendix H – State FIPS Codes

## PART 1 - SAS Setup

All statistical programs described in this document were created and tested using SAS 9.4 software (SAS Institute, Inc., Cary, North Carolina) in a Windows environment. The guidance provided in this document is implemented through open source programs.

The programs require the following SAS features:

* BASE SAS
* SAS STAT
* The ability to import a file from csv into SAS
* The ability to export a file from csv into SAS

## PART 2 - Download and Organize the Files

Terminology:

* COUNTY-PQ refers to the query designed for child or teen with their state and county designation.
* ZCTA3-PQ refers to the query designed for child or teen with their state and the first three digits of their ZIP code tabulation area (ZCTA3).
* The query refers to requirements that spans across the two query types.

### 2.1 The Query

The steps in Part 2.1 are consistent, regardless of creating estimates with count or ZCTA3 data.

|  |  |  |
| --- | --- | --- |
| Step | Description | Details |
| 1 | Create a core directory (CORE) with the following subdirectories: | "..\0 SAS Programs"  "..\0 SAS Programs\Pre-Processing”  "..\0 SAS Programs\CODI PQ”  "..\00 Raw Data"  "..\01 Output"  "..\01 Output\SAS LOGS" |
| 2 | Determine the SAS programs and data files needed to download based on the data inputs. | for county, see part 2.2. For ZCTA3, see part 2.3. |
| 3 | Download and store programs. | a) place the Pre-Processing programs into the “0 SAS Programs\Pre-Processing” folder. The Pre-Processing suite of programs includes the Quickstart and Module1 through Module3.  b) place the CODI-PQ programs into the “0 SAS Programs\CODI-PQ” folder. The CODI-PQ package includes SAS programs Quickstart, Macro 1 through 4, and Module 1 and 2. |
| 4 | Download and store data. | a) place the ACS data file in the “00 Raw Data” folder.  b) place the Race Imputation file in the “00 Raw Data”  c) place the EHR data file in the “00 Raw Data” |

### 2.2 County Specifics

Table: County-Specific Details for use with Part 2.1

|  |  |  |
| --- | --- | --- |
| Title | Naming convention and details | Storage Directory |
| SAS Programs: Pre-processing for County | Module1-Pre\_Processing\_CODI\_PQ.sas  Module2-Pre\_Processing\_CODI\_PQ.sas  Module3-Pre\_Processing\_CODI\_PQ.sas  Quickstart-Pre\_Processing\_CODI\_PQ.sas | "..\0 SAS Programs\Pre-Processing” |
| SAS Programs: CODI-PQ for County | Macro1-CODI\_PQ.sas  Macro2-CODI\_PQ.sas  Macro3-CODI\_PQ.sas  Macro4-CODI\_PQ.sas  Module1-CODI\_PQ.sas  Module2-CODI\_PQ.sas  Quickstart-CODI\_PQ.sas | "..\0 SAS Programs\CODI PQ”  "..\00 Raw Data"  "..\01 Output"  "..\01 Output\SAS LOGS" |
| ACS data input file: for County | <https://sft.mitre.org/#/folder/6281923>  ACS\_State\_COUNTY.csv  See Appendix A for full details. | "..\00 Raw Data" |
| Pre-processing Input File: Race Imputation | <https://sft.mitre.org/#/folder/6281923>  Pre\_PQ\_Reg\_Est | "..\00 Raw Data" |
| EHR data input file: for County | Name of file may vary, stored as .csv.  See Appendix B for full details. | "..\00 Raw Data" |

### 2.3 ZCTA3 Specifics

Table: ZCTA3 Specific Details for use with Part 2.1

|  |  |  |
| --- | --- | --- |
| Title | Naming convention and details | Storage Directory |
| SAS Programs: Pre-processing for ZCTA3 | Module1-Pre\_Processing\_CODI\_PQ\_ZCTA3.sas  Module2-Pre\_Processing\_CODI\_PQ\_ZCTA3.sas  Module3-Pre\_Processing\_CODI\_PQ\_ZCTA3.sas  Quickstart-Pre\_Processing\_CODI\_PQ\_ZCTA3.sas | "..\0 SAS Programs\Pre-Processing” |
| SAS Programs: CODI-PQ for ZCTA3 | Macro1-CODI\_PQ\_ZCTA3.sas  Macro2-CODI\_PQ\_ZCTA3.sas  Macro3-CODI\_PQ\_ZCTA3.sas  Macro4-CODI\_PQ\_ZCTA3.sas  Module1-CODI\_PQ\_ZCTA3.sas  Module2-CODI\_PQ\_ZCTA3.sas  Quickstart-CODI\_PQ\_ZCTA3.sas | "..\0 SAS Programs\CODI PQ”  "..\00 Raw Data"  "..\01 Output"  "..\01 Output\SAS LOGS" |
| ACS data file: for ZCTA3 | <https://sft.mitre.org/#/folder/6281923>  ACS\_State\_ZCTA3.csv  See Appendix A for data format details. | "..\00 Raw Data" |
| Pre-processing Input File: Race Imputation | <https://sft.mitre.org/#/folder/6281923>  Pre\_PQ\_Reg\_Est | "..\00 Raw Data" |
| EHR data input file: for ZCTA3 | Name of file may vary, stored as .csv.  See Appendix B for data format details. | "..\00 Raw Data" |

## Part 3 – Change Selections and Submit the Pre-processing CODI-PQ

### 3.1 Change Pre-processing CODI-PQ Selections

The steps in part 3.1 are consistent, regardless of creating estimates with count or ZCTA3 data. Minimal knowledge of SAS syntax is needed to change the selections. In particular, the user is expected to know how to read SAS comments and edit SAS macro values.

SAS comments in the programs are written between slashes and asterisks. For example:

/\*read this as a comment\*/

Is an example of a comment. Comments may fall as a stand-alone line or before, in the middle of, or at the end of a row of SAS syntax.

The user is expected to update SAS macros in sections 1 through 3 based on the examples provided in this document as well as the comments within the program. Macro variables are SAS tools that enable users to dynamically modify the text in a SAS program through symbolic substitution.

For example, the statement:

%let try\_macro = abcd;

Creates a macro variable named try\_macro which is equal to the test string abcd. A semi-colon is used to end the text to include within the SAS macro variable. Below is an example from a Quickstart program.

/\*\*\*/ %LET BEGIN\_YEAR = 2015; /\*@Note: LONGITUDINAL Start year (ACCEPTABLE VALUES: 4-digit numeric year) \*\*\*/

Note the code starts with a comment to emphasize the row, followed by the SAS syntax “%LET BEGIN\_YEAR=” which must remain unchanged. After the equal sign is the text 2015;

To change the year, a user should edit the text after the equal sign and before the semicolon.

The text after the semicolon provides insights into the acceptable values for the macro variable.

Below are the steps and descriptions to customize the pre-processing algorithm. Note that the pre-prevalence algorithm should be submitted once and only once per file. As such, include the start and end years for the full file (three years maximum). Once pre-processing is complete, the user can submit an unlimited number of prevalence queries using the same pre-processed file each time. Using the same pre-processed data file in the prevalence query (Section 4) ensures consistency in the race imputation.

Table: Change Specifications, Processing Steps

|  |  |  |
| --- | --- | --- |
| Step | Description | Details |
| 1 | Open the Pre-processing Quickstart program | The Quickstart program is stored in the directory:  "..\0 SAS Programs\Pre-Processing” |
| 2 | Edit the SAS program within “SECTION 1: Input Folder and file names” | Follow the SAS programs and update the macro variable specifications, in particular: |

Table: Change Specifications, SAS Macro Variables

|  |  |  |
| --- | --- | --- |
| SAS Macro Variable | Details | Example |
| **ROOT\_PRE** | The core directory name (see part 2.1.1). | %let ROOT\_PRE = C:\Documents\CODI; |
| **PROGS\_PRE** | The directory name of the pre-processing SAS programs (e.g. core directory name plus “\0 SAS Programs\Pre-Processing”). | %let PROGS\_PRE = C:\Documents\CODI\0 SAS Programs  Pre-Processing; |
| **PRE\_DEST** | The folder name for output. | %let PRE\_DEST = CODI\_PQ;  /\* output would be stored in P:\Documents\CODI\01 Output\CODI\_PQ\*/ |
| **ACS\_FILENAME** | The American Community Survey file name from part 2.2 or 2.3. The file is in csv format. Do not include the extension in the file name. | %let ACS\_FILENAME = ACS\_State\_COUNTY;  Or  %let ACS\_FILENAME = ACS\_State\_ZCTA3; |
| **EHR\_FileNAME** | The child and teen level EHR file from part 2.2 or 2.3. Do not include the extension. | %let EHR\_FileNAME = EHR\_Stored\_here; |
| **LOG\_NAME\_PRE** | The name of the output SAS log. Quickstart\_Pre\_Processing\_CODI\_PQ. Users have the option to rename the log file name before it is created. | %let LOG\_NAME\_PRE = LogName;  /\*the SAS log will be stored in: P:\Documents\CODI\01 Output\SAS LOGS\LogName<Date and Time>.log. Note, the program automatically includes the date and time in all log file names\*/ |

Table: Change Specifications, Processing Steps, Continued

|  |  |  |
| --- | --- | --- |
| Step | Description | Details |
| 3 | Edit the SAS program within “SECTION 2: Beginning and End Year of longitudinal EHR data” | By default, the algorithm assumes pre-processing will occur on three years of data from 2016 to 2018. Edit the start and end year as appropriate, include all three years within your file as ***pre-processing start and end years should include all years on the file***. In contrast, the prevalence algorithm will request the specific year for each prevalence analysis (Part 4 below). |
| 4 | Edit the SAS program within “SECTION 3: Optional Output File Name Suffix” | The name of the output file. For example:  %LET EHR\_PRE\_Out = NEW;  /\*following the same example as above, the output from pre-processing will be stored as a SAS data file in: C:\Documents\CODI\01 Output\NEW\*/ |
| 5 | Save the Quickstart program. | SAS recommends saving all files before submitting the program. |

### 3.2 Submit Pre-processing CODI-PQ Programs

The pre-processing query should be submitted once and only once for each EHR data file. The SAS programs impute the race of children and teens with unknown race. As such, each time the program is submitted, new imputed race values are created and stored for each child. For consistency, we recommend submitting the pre-processing query once and only once for each EHR data file.

To submit a SAS query, use the commands within your SAS installation. For example, in SAS 9.4 for Windows, a user has several options to submit a program. First, make sure the cursor is not highlighting text. Next, either:

* Press F3 or F8 when the editor window is active.
* Click the **Submit** toolbar button.
* Enter submit in the command bar.
* Select **Run**then select**Submit**.

See sas.com for instructions with various SAS version numbers and installation.

Table: Pre-processing CODI-PQ Program Submission Processing Steps

|  |  |  |
| --- | --- | --- |
| Step | Description | Details |
| 1 | Submit the Quickstart program. | Submit the Quickstart program. The program completes all tasks within the data sets and proc statements in the Quickstart program and moves to the next SAS program automatically through an include statement. |
| 2 | Review the log. | Review the log for possible errors including words such as error, warning, and uninitialized. Assuming no errors, continue to Part 4. In the event of errors, reassess the location of the files and the file formats. |

## Part 4 – Change Selections and Submit the CODI-PQ

### 4.1 Change CODI-PQ Selections

The steps in part 4.1 are consistent, regardless of creating estimates with count or ZCTA3 data. Minimal knowledge of SAS syntax is needed to change the selections. See part 3.1 for more details.

Table: Change Specifications, Processing Steps

|  |  |  |
| --- | --- | --- |
| Step | Description | Details |
| 1 | Open the Quickstart program | The Quickstart program is stored in the directory:  "..\0 SAS Programs\CODI PQ” |
| 2 | Edit the SAS program within  “SECTION 1: Input Folder and file names”;  “SECTION 2: Subset data based on specifications INCLUDING YEAR, GEOGRAPHY, STATE, or STATE/COUNTY CODE” | Follow the SAS programs and update the macro variable specifications, in particular: |

Table: Change Specifications, Macro Variables

|  |  |  |
| --- | --- | --- |
| SAS Macro Variable | Details | Example |
| SECTION 1: Input Folder and file names | | |
| **ROOT\_PQ** | The core directory name (see section 2.1.1). | %let ROOT\_PQ = C:\Documents\CODI; |
| **PROGS\_PQ** | The directory name of the CODI-PQ SAS programs. | %let PROGS\_PQ= C:\Documents\CODI\0 SAS Programs\CODI\_PQ; |
| **PRE\_DEST** | The folder name for pre-processing output which is the input for the CODI-PQ programs. This should equal the same value as PRE\_DEST in section 3.1=. | %let PRE\_DEST = CODI\_PQ;  /\* output would be stored in P:\Documents\CODI\01 Output\CODI\_PQ\*/ |
| **EHR\_PRE\_OUT** | The child and teen level EHR file from section 3.1 and 3.2. This should equal the same value as what was specified in section 3.1. | %let EHR\_PRE\_OUT = NEW; |
| **LOG\_NAME** | The name of the output SAS log. Users have the option to rename the log file name before it is created. | %let LOG\_NAME = LogName;  /\*the SAS log will be stored in: P:\Documents\CODI\01 Output\SAS LOGS\LogName<Date and Time>.log. Note, the program automatically includes the date and time in all log file names\*/ |
| SAS Macro Variable | Details | Example |
| SECTION 2: Subset data based on specifications INCLUDING YEAR, GEOGRAPHY, STATE, or STATE/COUNTY CODE | | |
| **BEG\_YEAR** | Subsets the prevalence estimate based on the year of the obesity estimate. The prevalence estimates will include children and teen EHR data from this year and after. | /\*\*\*/ %LET BEG\_YEAR = 2016; /\*@Note: Beginning year of analysis (ACCEPTED VALUES: 4-Digit numeric, 2016-2018) \*\*\*/ |
| **END\_YEAR** | Subsets the prevalence estimate based on the year of the obesity estimate. The prevalence estimates will include children and teen EHR data from this year and before. | /\*\*\*/ %LET END\_YEAR = 2016; /\*@Note: End year of analysis (ACCEPTED VALUES: 4-Digit numeric, 2016-2018) \*\*\*/ |
| **STATE** | **County processing only**; Do you want to create State estimates? If the value is set to yes then the program will anticipate ***state codes*** for the GEO\_LIST algorithm and the algorithm will generate prevalence estimates from children reported across all states listed. If the value is set to no then the program will anticipate ***state and county codes*** in the GEO\_LIST macro and the algorithm will generate estimates from children reported across all counties listed. | /\*\*\*/ %LET STATE = N; /\*@Note: Include all geographical locations within requested state (see GEO\_LIST)? (ACCEPTED VALUES: Y/N) \*\*\*/ |
| **GEO\_LIST** | **County processing only**; If STATE = Y; then by default the program will include all geographical locations listed. If STATE = N; then by default the program will subset the prevalence estimate based on the state+county codes listed. | **If STATE=Y;** /\*\*\*/ %LET GEO\_LIST = %STR(‘08’);  **If STATE=N;** /\*\*\*/ %LET GEO\_LIST = %STR(‘08059’, ’08125’);  /\*@Note: IF STATE="Y" then populate with State FIPS code(s), otherwise populate with FIPS State+FIPS County code(s) (ACCEPTED VALUES: 2-digit state FIPS or 5-digit state FIPS+county FIPS (Must be surrounded by single quotation and comma delimited)) \*\*\*/ |
| **ALL\_STATES** | **ZCTA3 processing only**; Includes all states (including D.C.) in the prevalence estimate based on the location of the child or teen. If ALL\_STATES = N; then by default the program will subset the prevalence estimate based on the individual state or state+ZCTA3 values specified (in future step) | /\*@Note: Include all geographical locations in file? (ACCEPTED VALUES: Y/N) \*\*\*/ |
| **ALL\_AGES** | Subsets the prevalence estimate based on the age of the child or teen. The user may either choose to include all children and teens age 2 to 19 or alternatively may choose select age groups. Note: if ALL\_AGES = Y; then by default the program will include all children and teens age 2 to 19. If ALL\_AGES = N; then by default the program will subset the prevalence estimate based on the individual age ranges selected (in future step). | /\*\*\*/ %LET ALL\_AGES = Y; /\*@Note: Include all age ranges? (ACCEPTED VALUES: Y/N) \*\*\*/ |
| **ALL\_SEXES** | Subsets the prevalence estimate based on the sec of the child or teen. The user may either choose to include all male and female children and teens or alternatively may choose either males or females. Note: if ALL\_SEXES = Y; then by default the program will include both males and females. If ALL\_SEXES = N; then by default the program will subset the prevalence estimate based on the individual sex(es) selected (in future step). | /\*\*\*/ %LET ALL\_SEXES = Y; /\*@Note: Include all sex values? (ACCEPTED VALUES: Y/N) \*\*\*/ |
| **ALL\_RACES** | Subsets the prevalence estimate based on the race of the child or teen. The user may either choose to include all races or alternatively may choose select race(s). Inclusion or exclusion of imputed race is not impacted by the choice made in this step. Note: if ALL\_RACES = Y; then by default the program will include all races (White, Black, Asian, Other). If ALL\_RACES = N; then by default the program will subset the prevalence estimate based on the individual races selected (in future step). | /\*\*\*/ %LET ALL\_RACES = Y; /\*@Note: Include all race categories? (ACCEPTED VALUES: Y/N) \*\*\*/ |

Table: Change Specifications, Processing Steps, Continued

|  |  |  |
| --- | --- | --- |
| Step | Description | Details |
| 3 | Edit the SAS program within  “SECTION 3: Only complete section 3 for any "N" values listed in section 2“  “SECTION 4: Methodological option selections” | **EHR\_PRE\_Out** by default, the EHR data file after pre-processing will be named CODI\_PQ. Users have the option to rename this file before it is created. |

Table: Change Specifications, SAS Macro Variables

|  |  |  |
| --- | --- | --- |
| SAS Macro Variable Category | Details | Example |
| SECTION 3: Only complete section 3 for any "N" values listed in section 2 | | |
| **If ALL\_STATES = N (ZCTA3 Processing only)** | **ZCTA3 processing only;** GEO\_GROUP informs the program what level of geography is to be used in the GEO\_LIST macro variable. GEO\_LIST subsets the prevalence estimates based on the location of the child or teen. If GEO\_GROUP=STATE; then the program defaults to using state FIPS codes. If GEO\_GROUP=ZCTA3; then the program defaults to using state FIPS+ZCTA3 codes. Of note, values should be surrounded by single quotes and comma delimited. | /\*IF ALL\_STATES= N THEN SELECT STATE CODES OR STATE AND COUNTY CODES BELOW: \*\*\*/  /\*\*\* (ACCEPTED VALUES: SINGLE QUOTES SURROUNDING 2 OR 5-Digit CODES w/ "," BETWEEN MULTIPLE SELECTIONS, ) \*\*\*/  /\*IF ALL\_STATES = N THEN SELECT ONE OR MORE AGE CATEGORIES BELOW: \*\*\*/  /\*\*\*/ %LET GEO\_GROUP = STATE; /\*@Note: Level of geography (ACCEPTED VALUES: STATE/ZCTA3) \*\*\*/  /\*\*\*/ %LET GEO\_LIST = %STR('08'); /\*@Note: IF GEO\_GROUP="STATE" then populate with State FIPS code(s), If GEO\_GROUP="ZCTA3" then populate with FIPS State+FIPS County code(s) \*\*\*/ |
| **If ALL\_AGES = N;** | If ALL\_AGES is set to no, the age macros (2-4, 5-9, 10-14, 15-17, 18-19) subset the prevalence estimates based on the age of the child or teen and the responses to each individual age macro. Note that if ALL\_AGES is set to yes, then the SAS program does not review the age-specific macros. | %LET AGE\_2\_4 = N;  %LET AGE\_5\_9 = N;  %LET AGE\_10\_14 = Y;  %LET AGE\_15\_17 = Y;  %LET AGE\_18\_19 = Y; |
| **If ALL\_RACES = N;** | If ALL\_RACES is set to no, the race macros (White, Black, Asian, Other) subset the prevalence estimates based on the race or imputed race of the child or teen and the responses to each individual age macro. Note that if ALL\_RACES is set to yes, then the SAS program does not review the race-specific macros. | %LET RACE\_WHITE = N;  %LET RACE\_BLACK = Y;  %LET RACE\_ASIAN = Y;  %LET RACE\_OTHER = Y; |
| **If ALL\_SEXES = N;** | If ALL\_SEXES is set to no, the sex macros (male, female) subset the prevalence estimates based on the sex of the child or teen and the responses to each individual sex macro. Note that if ALL\_SEXES is set to yes, then the SAS program does not review the sex-specific macros. | %LET SEX\_MALE = N;  %LET SEX\_FEMALE = Y; |
| SAS Macro Variable Category | Details | Example |
| SECTION 4: Methodological option selections | | |
| **Race Imputation** | If IMP\_RACES is set to yes, then the query includes children and teens with imputed race values. Otherwise, if IMP\_RACES is set to no, then the children with imputed races are excluded. | %LET IMP\_RACES = Y; |
| **Age Adjustment** | If AGE\_ADJ is set to yes, then the query generates age adjusted prevalence estimates and standard errors. Otherwise, if AGE\_ADJ is set to no, age adjusted estimates are not generate. | %LET AGE\_ADJ = Y; |

Table: Change Specifications, Processing Steps, Continued

|  |  |  |
| --- | --- | --- |
| Step | Description | Details |
| 4 | Save the Quickstart program. | It is recommended to save the quickstart program before submitting the programs. |

### 4.2 Submit CODI-PQ and Review Prevalence Results

The prevalence query may be submitted multiple times with different user selections each time. The SAS program subsets the preprocessed file based on the user’s specifications, calculates statistical weights, prevalence estimates, and standard errors, then outputs the results into a csv file. As mentioned in part 3, for consistency, we recommend submitting the pre-processing query once and only once for each EHR data file whereas the prevalence query may be submitted an unlimited number of times.

To submit a SAS query, use the commands within your SAS installation. For example, in SAS 9.4 for Windows, a user has several options to submit a program. First, make sure the cursor is not highlighting text. Next, either:

* Press F3 or F8 when the editor window is active.
* Click the **Submit** toolbar button.
* Enter submit in the command bar.
* Select **Run**then select**Submit**.

See sas.com for instructions with various SAS version numbers and installation.

Table: CODI-PQ Submission Processing Steps

|  |  |  |
| --- | --- | --- |
| Step | Description | Details |
| 1 | Submit the CODI-PQ Quickstart program. | Submit the Quickstart program. The program completes all tasks within the data sets and proc statements in the Quickstart program and moves to the next SAS program automatically through an include statement. |
| 2 | Review the log. | Review the log for possible errors including words such as error, warning, and uninitialized. Assuming no errors, continue to step 3. In the event of errors, reassess the location of the files and the file formats. |
| 3 | Review the output. | Review the output for possible data suppression or errors. In the event of errors reassess the choices described in part 4.1 and re-submit. In the event of data suppression, consider expanding your selection criteria and re-submit. For example, if estimates cannot be created for a single year, consider using two or more years of data. See the Implementation Guide for more information about data suppression or errors. |

1. ACS Pre-processing Input File Layout

Table A.1 ACS File Layout, csv file

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| NAME | TYPE | LENGTH | VARNUM | FORMAT | FORMATL | INFORML |
| geoid | 2 | 12 | 1 | Character | 12 | 12 |
| b01001f\_001 | 1 | 8 | 2 | Number | 12 | 32 |
| b01001f\_003 | 1 | 8 | 3 | Number | 12 | 32 |
| b01001f\_004 | 1 | 8 | 4 | Number | 12 | 32 |
| b01001f\_005 | 1 | 8 | 5 | Number | 12 | 32 |
| b01001f\_006 | 1 | 8 | 6 | Number | 12 | 32 |
| b01001f\_007 | 1 | 8 | 7 | Number | 12 | 32 |
| b01001f\_018 | 1 | 8 | 8 | Number | 12 | 32 |
| b01001f\_019 | 1 | 8 | 9 | Number | 12 | 32 |
| b01001f\_020 | 1 | 8 | 10 | Number | 12 | 32 |
| b01001f\_021 | 1 | 8 | 11 | Number | 12 | 32 |
| b01001f\_022 | 1 | 8 | 12 | Number | 12 | 32 |
| b01001e\_001 | 1 | 8 | 13 | Number | 12 | 32 |
| b01001e\_003 | 1 | 8 | 14 | Number | 12 | 32 |
| b01001e\_004 | 1 | 8 | 15 | Number | 12 | 32 |
| b01001e\_005 | 1 | 8 | 16 | Number | 12 | 32 |
| b01001e\_006 | 1 | 8 | 17 | Number | 12 | 32 |
| b01001e\_007 | 1 | 8 | 18 | Number | 12 | 32 |
| b01001e\_018 | 1 | 8 | 19 | Number | 12 | 32 |
| b01001e\_019 | 1 | 8 | 20 | Number | 12 | 32 |
| b01001e\_020 | 1 | 8 | 21 | Number | 12 | 32 |
| b01001e\_021 | 1 | 8 | 22 | Number | 12 | 32 |
| b01001e\_022 | 1 | 8 | 23 | Number | 12 | 32 |
| b01001d\_001 | 1 | 8 | 24 | Number | 12 | 32 |
| b01001d\_003 | 1 | 8 | 25 | Number | 12 | 32 |
| b01001d\_004 | 1 | 8 | 26 | Number | 12 | 32 |
| b01001d\_005 | 1 | 8 | 27 | Number | 12 | 32 |
| b01001d\_006 | 1 | 8 | 28 | Number | 12 | 32 |
| b01001d\_007 | 1 | 8 | 29 | Number | 12 | 32 |
| b01001d\_018 | 1 | 8 | 30 | Number | 12 | 32 |
| b01001d\_019 | 1 | 8 | 31 | Number | 12 | 32 |
| b01001d\_020 | 1 | 8 | 32 | Number | 12 | 32 |
| b01001d\_021 | 1 | 8 | 33 | Number | 12 | 32 |
| b01001d\_022 | 1 | 8 | 34 | Number | 12 | 32 |
| b01001c\_001 | 1 | 8 | 35 | Number | 12 | 32 |
| b01001c\_003 | 1 | 8 | 36 | Number | 12 | 32 |
| b01001c\_004 | 1 | 8 | 37 | Number | 12 | 32 |
| b01001c\_005 | 1 | 8 | 38 | Number | 12 | 32 |
| b01001c\_006 | 1 | 8 | 39 | Number | 12 | 32 |
| b01001c\_007 | 1 | 8 | 40 | Number | 12 | 32 |
| b01001c\_018 | 1 | 8 | 41 | Number | 12 | 32 |
| b01001c\_019 | 1 | 8 | 42 | Number | 12 | 32 |
| b01001c\_020 | 1 | 8 | 43 | Number | 12 | 32 |
| b01001c\_021 | 1 | 8 | 44 | Number | 12 | 32 |
| b01001c\_022 | 1 | 8 | 45 | Number | 12 | 32 |
| b01001b\_001 | 1 | 8 | 46 | Number | 12 | 32 |
| b01001b\_003 | 1 | 8 | 47 | Number | 12 | 32 |
| b01001b\_004 | 1 | 8 | 48 | Number | 12 | 32 |
| b01001b\_005 | 1 | 8 | 49 | Number | 12 | 32 |
| b01001b\_006 | 1 | 8 | 50 | Number | 12 | 32 |
| b01001b\_007 | 1 | 8 | 51 | Number | 12 | 32 |
| b01001b\_018 | 1 | 8 | 52 | Number | 12 | 32 |
| b01001b\_019 | 1 | 8 | 53 | Number | 12 | 32 |
| b01001b\_020 | 1 | 8 | 54 | Number | 12 | 32 |
| b01001b\_021 | 1 | 8 | 55 | Number | 12 | 32 |
| b01001b\_022 | 1 | 8 | 56 | Number | 12 | 32 |
| b01001a\_001 | 1 | 8 | 57 | Number | 12 | 32 |
| b01001a\_003 | 1 | 8 | 58 | Number | 12 | 32 |
| b01001a\_004 | 1 | 8 | 59 | Number | 12 | 32 |
| b01001a\_005 | 1 | 8 | 60 | Number | 12 | 32 |
| b01001a\_006 | 1 | 8 | 61 | Number | 12 | 32 |
| b01001a\_007 | 1 | 8 | 62 | Number | 12 | 32 |
| b01001a\_018 | 1 | 8 | 63 | Number | 12 | 32 |
| b01001a\_019 | 1 | 8 | 64 | Number | 12 | 32 |
| b01001a\_020 | 1 | 8 | 65 | Number | 12 | 32 |
| b01001a\_021 | 1 | 8 | 66 | Number | 12 | 32 |
| b01001a\_022 | 1 | 8 | 67 | Number | 12 | 32 |
| b01001g\_001 | 1 | 8 | 68 | Number | 12 | 32 |
| b01001g\_003 | 1 | 8 | 69 | Number | 12 | 32 |
| b01001g\_004 | 1 | 8 | 70 | Number | 12 | 32 |
| b01001g\_005 | 1 | 8 | 71 | Number | 12 | 32 |
| b01001g\_006 | 1 | 8 | 72 | Number | 12 | 32 |
| b01001g\_007 | 1 | 8 | 73 | Number | 12 | 32 |
| b01001g\_018 | 1 | 8 | 74 | Number | 12 | 32 |
| b01001g\_019 | 1 | 8 | 75 | Number | 12 | 32 |
| b01001g\_020 | 1 | 8 | 76 | Number | 12 | 32 |
| b01001g\_021 | 1 | 8 | 77 | Number | 12 | 32 |
| b01001g\_022 | 1 | 8 | 78 | Number | 12 | 32 |
| b03002\_012 | 1 | 8 | 79 | Number | 12 | 32 |
| b03002\_013 | 1 | 8 | 80 | Number | 12 | 32 |
| b03002\_014 | 1 | 8 | 81 | Number | 12 | 32 |
| b03002\_015 | 1 | 8 | 82 | Number | 12 | 32 |
| b03002\_016 | 1 | 8 | 83 | Number | 12 | 32 |
| b03002\_017 | 1 | 8 | 84 | Number | 12 | 32 |
| b03002\_018 | 1 | 8 | 85 | Number | 12 | 32 |
| b03002\_019 | 1 | 8 | 86 | Number | 12 | 32 |
| b15001\_011 | 1 | 8 | 87 | Number | 12 | 32 |
| b15001\_017 | 1 | 8 | 88 | Number | 12 | 32 |
| b15001\_018 | 1 | 8 | 89 | Number | 12 | 32 |
| b15001\_019 | 1 | 8 | 90 | Number | 12 | 32 |
| b15001\_025 | 1 | 8 | 91 | Number | 12 | 32 |
| b15001\_026 | 1 | 8 | 92 | Number | 12 | 32 |
| b15001\_027 | 1 | 8 | 93 | Number | 12 | 32 |
| b15001\_033 | 1 | 8 | 94 | Number | 12 | 32 |
| b15001\_034 | 1 | 8 | 95 | Number | 12 | 32 |
| b15001\_052 | 1 | 8 | 96 | Number | 12 | 32 |
| b15001\_058 | 1 | 8 | 97 | Number | 12 | 32 |
| b15001\_059 | 1 | 8 | 98 | Number | 12 | 32 |
| b15001\_060 | 1 | 8 | 99 | Number | 12 | 32 |
| b15001\_066 | 1 | 8 | 100 | Number | 12 | 32 |
| b15001\_067 | 1 | 8 | 101 | Number | 12 | 32 |
| b15001\_068 | 1 | 8 | 102 | Number | 12 | 32 |
| b15001\_074 | 1 | 8 | 103 | Number | 12 | 32 |
| b15001\_075 | 1 | 8 | 104 | Number | 12 | 32 |

1. EHR Pre-processing Input File Layout
   1. Universal Input File

Table B.1 Race Imputation Input File for County and ZCTA3-Level Programs

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | TYPE | LENGTH | Variable Order | LABEL | FORMAT | FORMATL | FORMATD |
| Effect | 2 | 20 | 1 |  | Character | 0 | 0 |
| Race\_Resp | 1 | 8 | 2 |  | Number | 0 | 0 |
| State\_ZIP\_Model | 2 | 5 | 3 |  | Character | 0 | 0 |
| Sex\_Model | 2 | 4 | 4 |  | Character | 0 | 0 |
| Estimate | 1 | 8 | 5 |  | Number | 8 | 4 |

* 1. County

Table B.2 EHR File Layout for County-Level Programs, csv file

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable Order | Variable | Format | Description | Example Content |
| 1 | DOB: Date of birth | Date |  | 10/27/2007 |
| 2 | SEX: Sex | Character |  | female |
| 3 | RACE: Race | Character |  | 2106-2 |
| 4 | ETHNICITY: Ethnicity | Character |  | 2135-1 |
| 5 | LAT: Latitude | Number |  | 11.91259289 |
| 6 | LNG: Longitude | Number |  | -99.2082893 |
| 7 | STATE: Patient state | Number |  | 51 |
| 8 | ZIP: Zip Code | Character |  | 22182 |
| 9 | CENSUS\_TRACT: US CENSUS Tract | Character |  | 111 |
| 10 | COUNTY\_FIPS: County FIPS Code | Character |  | 059 |
| 11 | WEIGHT2016: Patient Weight (kg) in 2016 | Character |  | 27.0000000 |
| 12 | HEIGHT2016: Patient Height (cm) in 2016 | Character |  | 128.0000000 |
| 13 | BMI2016: Patient BMI in 2016 | Character |  | 16.118233 |
| 14 | WTCAT2016: Patient weight category 2016 | Character |  | Healthy Weight |
| 15 | AGEYR2016: Patient Age (whole year) in 2016 | Number |  | 9 |
| 16 | WEIGHT2017: Patient Weight (kg) in 2017 | Character |  | NI |
| 17 | HEIGHT2017: Patient Height (cm) in 2017 | Character |  | NI |
| 18 | BMI2017: Patient BMI in 2017 | Character |  | NI |
| 19 | WTCAT2017: Patient weight category 2017 | Character |  | NI |
| 20 | AGEYR2017: Patient Age (whole year) in 2017 | Number |  | NI |
| 21 | WEIGHT2018: Patient Weight (kg) in 2018 | Character |  | 33.26900000 |
| 22 | HEIGHT2018: Patient Height (cm) in 2018 | Character |  | 138.0800000 |
| 23 | BMI2018: Patient BMI in 2018 | Character |  | 17.449299 |
| 24 | WTCAT2018: Patient weight category 2018 | Character |  | Healthy Weight |
| 25 | AGEYR2018: Patient Age (whole year) in 2018 | Number |  | 11 |
| 26 | WEIGHT2019: Patient Weight (kg) in 2019 | Character |  | NI |
| 27 | HEIGHT2019: Patient Height (cm) in 2019 | Character |  | NI |
| 28 | BMI2019: Patient BMI in 2019 | Character |  | NI |
| 29 | WTCAT2019: Patient weight category 2019 | Character |  | NI |
| 30 | AGEYR2019: Patient Age (whole year) in 2019 | Number |  |  |
| 31 | HCLCNT | Number | Extra, not used in algorithm | 0 |
| 32 | HCLDATE | Date | Extra, not used in algorithm | 1/1/2020 |
| 33 | CFCNT | Number | Number of times Cystic Fibrosis was mentioned in EHR | 0 |
| 34 | CFDATE | Date | Date of first mention of Cystic Fibrosis in EHR | 1/1/2020 |
| 35 | SCDCNT | Number | Number of times Sickle Cell Disease was mentioned in EHR | 0 |
| 36 | SCDDATE | Date | Date of first mention of Sickle Cell Disease in EHR | 1/1/2020 |
| 37 | SBCNT | Number | Number of times Spina Bifida was mentioned in EHR | 0 |
| 38 | SBDATE | Date | Date of first mention of Spina Bifida in EHR | 1/1/2020 |
| 39 | ASTHMACNT | Number | Number of times Asthma was mentioned in EHR | 0 |
| 40 | ASTHMADATE | Date | Date of first mention of Asthma in EHR | 1/1/2020 |
| 41 | CELIACCNT | Number | Number of times Celiacs Disease was mentioned in EHR | 0 |
| 42 | CELIACDATE | Date | Date of first mention of Celiacs Disease in EHR | 1/1/2020 |
| 43 | SCZCNT | Number | Extra, not used in algorithm | 0 |
| 44 | SCZDATE | Date | Extra, not used in algorithm | 1/1/2020 |

* 1. ZCTA3

Table B.3 ZCTA3 EHR input file 1 of 2

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Order | Variable Name | Format | Example |
| 1 | PatId | Number | 123456 |
| 2 | Gender | Number | 0 |
| 3 | Age Group | Number | 1998 |
| 4 | Ethnicity | Character | CAUCASIAN |
| 5 | Hypercholesterolemia:Event Count | Number | 0 |
| 6 | Hypercholesterolemia:First Event Date | Date | 1/1/2020 |
| 7 | Cystic-fibrosis:Event Count | Number | 0 |
| 8 | Cystic-fibrosis:First Event Date | Date | 1/1/2020 |
| 9 | Sickle-cell:Event Count | Number | 0 |
| 10 | Sickle-cell:First Event Date | Date | 1/1/2020 |
| 11 | Spina-bifida:Event Count | Number | 0 |
| 12 | Spina-bifida:First Event Date | Date | 1/1/2020 |
| 13 | Asthma:Event Count | Number | 0 |
| 14 | Asthma:First Event Date | Date | 1/1/2020 |
| 15 | Celiac:Event Count | Number | 0 |
| 16 | Celiac:First Event Date | Date | 1/1/2020 |
| 17 | Schizophrenia:Event Count | Number | 0 |
| 18 | Schizophrenia:First Event Date | Date | 1/1/2020 |

Table B.4 ZCTA3 EHR input file 2 of 2

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Order | Variable Name | Format | Example |
| 1 | patid | Number | 123456789 |
| 2 | agedays2015 | Number | 2887 |
| 3 | sex2015 | Number | 1 |
| 4 | ht2015 | Number | 105.41 |
| 5 | wt2015 | Number | 17.8035 |
| 6 | ageyrs2015 | Number | 3 |
| 7 | bmi2015 | Number | 16.0229 |
| 8 | z2015 | Number | 0.4998 |
| 9 | p952015 | Number | 18.038 |
| 10 | wtcat2015 | Number | 2 |
| 11 | state2015 | Character | VA |
| 12 | zip32015 | Number | 228 |
| 13 | agedays2016 | Number | 3237 |
| 14 | sex2016 | Number | 1 |
| 15 | ht2016 | Number | 132.715 |
| 16 | wt2016 | Number | 27.3289 |
| 17 | ageyrs2016 | Number | 8 |
| 18 | bmi2016 | Number | 15.5161 |
| 19 | z2016 | Number | -0.3707 |
| 20 | p952016 | Number | 21.6257 |
| 21 | wtcat2016 | Number | 2 |
| 22 | state2016 | Character | KY |
| 23 | zip32016 | Number | 405 |
| 24 | agedays2017 | Number | 3710 |
| 25 | sex2017 | Number | 1 |
| 26 | ht2017 | Number | 138.43 |
| 27 | wt2017 | Number | 30.3 |
| 28 | ageyrs2017 | Number | 10 |
| 29 | bmi2017 | Number | 15.8118 |
| 30 | z2017 | Number | -0.5254 |
| 31 | p952017 | Number | 23.08 |
| 32 | wtcat2017 | Number | 2 |
| 33 | state2017 | Character | KY |
| 34 | zip32017 | Number | 405 |
| 35 | agedays2018 | Number | 3998 |
| 36 | sex2018 | Number | 1 |
| 37 | ht2018 | Number | 141 |
| 38 | wt2018 | Number | 32.3 |
| 39 | ageyrs2018 | Number | 10 |
| 40 | bmi2018 | Number | 16.2467 |
| 41 | z2018 | Number | -0.5215 |
| 42 | p952018 | Number | 24.0461 |
| 43 | wtcat2018 | Number | 2 |
| 44 | state2018 | Character | KY |
| 45 | zip32018 | Number | 405 |
| 46 | male | Number | 0 |
| 47 | female | Number | 10 |
| 48 | africanamerican | Number | 10 |
| 49 | asian | Number | 0 |
| 50 | caucasian | Number | 0 |
| 51 | hispanic | Number | 0 |
| 52 | other | Number | 0 |
| 53 | unknown | Number | 0 |

1. ACS Pre-processing Output File Layout
   1. County

Table C.1 ACS Pre-processing Output File Layout – County

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Order | Variable Name | Type | Length |
| 1 | Geography | 2 | 5 |
| 2 | County\_FIPS | 2 | 3 |
| 3 | State\_FIPS | 2 | 2 |
| 4 | TOTAL\_ACS\_POPULATION | 1 | 8 |
| 5 | AGE\_L5\_MALE\_WHITE | 1 | 8 |
| 6 | AGE\_5\_9\_MALE\_WHITE | 1 | 8 |
| 7 | AGE\_10\_14\_MALE\_WHITE | 1 | 8 |
| 8 | AGE\_15\_17\_MALE\_WHITE | 1 | 8 |
| 9 | AGE\_18\_19\_MALE\_WHITE | 1 | 8 |
| 10 | AGE\_L5\_FEMALE\_WHITE | 1 | 8 |
| 11 | AGE\_5\_9\_FEMALE\_WHITE | 1 | 8 |
| 12 | AGE\_10\_14\_FEMALE\_WHITE | 1 | 8 |
| 13 | AGE\_15\_17\_FEMALE\_WHITE | 1 | 8 |
| 14 | AGE\_18\_19\_FEMALE\_WHITE | 1 | 8 |
| 15 | AGE\_L5\_MALE\_BLACK | 1 | 8 |
| 16 | AGE\_5\_9\_MALE\_BLACK | 1 | 8 |
| 17 | AGE\_10\_14\_MALE\_BLACK | 1 | 8 |
| 18 | AGE\_15\_17\_MALE\_BLACK | 1 | 8 |
| 19 | AGE\_18\_19\_MALE\_BLACK | 1 | 8 |
| 20 | AGE\_L5\_FEMALE\_BLACK | 1 | 8 |
| 21 | AGE\_5\_9\_FEMALE\_BLACK | 1 | 8 |
| 22 | AGE\_10\_14\_FEMALE\_BLACK | 1 | 8 |
| 23 | AGE\_15\_17\_FEMALE\_BLACK | 1 | 8 |
| 24 | AGE\_18\_19\_FEMALE\_BLACK | 1 | 8 |
| 25 | AGE\_L5\_MALE\_ASIAN | 1 | 8 |
| 26 | AGE\_5\_9\_MALE\_ASIAN | 1 | 8 |
| 27 | AGE\_10\_14\_MALE\_ASIAN | 1 | 8 |
| 28 | AGE\_15\_17\_MALE\_ASIAN | 1 | 8 |
| 29 | AGE\_18\_19\_MALE\_ASIAN | 1 | 8 |
| 30 | AGE\_L5\_FEMALE\_ASIAN | 1 | 8 |
| 31 | AGE\_5\_9\_FEMALE\_ASIAN | 1 | 8 |
| 32 | AGE\_10\_14\_FEMALE\_ASIAN | 1 | 8 |
| 33 | AGE\_15\_17\_FEMALE\_ASIAN | 1 | 8 |
| 34 | AGE\_18\_19\_FEMALE\_ASIAN | 1 | 8 |
| 35 | AGE\_L5\_MALE\_OTHER | 1 | 8 |
| 36 | AGE\_5\_9\_MALE\_OTHER | 1 | 8 |
| 37 | AGE\_10\_14\_MALE\_OTHER | 1 | 8 |
| 38 | AGE\_15\_17\_MALE\_OTHER | 1 | 8 |
| 39 | AGE\_18\_19\_MALE\_OTHER | 1 | 8 |
| 40 | AGE\_L5\_FEMALE\_OTHER | 1 | 8 |
| 41 | AGE\_5\_9\_FEMALE\_OTHER | 1 | 8 |
| 42 | AGE\_10\_14\_FEMALE\_OTHER | 1 | 8 |
| 43 | AGE\_15\_17\_FEMALE\_OTHER | 1 | 8 |
| 44 | AGE\_18\_19\_FEMALE\_OTHER | 1 | 8 |
| 45 | AGE\_25\_64\_BACH\_GRAD | 1 | 8 |
| 46 | AGE\_25\_64\_BACH\_GRAD\_GTR20PERC | 1 | 8 |
| 47 | TOTAL\_LATIN | 1 | 8 |
| 48 | LATIN\_WHITE | 1 | 8 |
| 49 | LATIN\_BLACK | 1 | 8 |
| 50 | LATIN\_ASIAN | 1 | 8 |
| 51 | LATIN\_OTHER | 1 | 8 |

* 1. ZCTA3

Table C.2 ACS Pre-processing Output File Layout – ZCTA3

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Name | Type | Length | Variable Order |
| Geography | 2 | 5 | 1 |
| State\_Alpha | 2 | 2 | 2 |
| State\_FIPS | 2 | 5 | 3 |
| ZCTA3 | 2 | 3 | 4 |
| TOTAL\_ACS\_POPULATION | 1 | 8 | 5 |
| AGE\_L5\_MALE\_WHITE | 1 | 8 | 6 |
| AGE\_5\_9\_MALE\_WHITE | 1 | 8 | 7 |
| AGE\_10\_14\_MALE\_WHITE | 1 | 8 | 8 |
| AGE\_15\_17\_MALE\_WHITE | 1 | 8 | 9 |
| AGE\_18\_19\_MALE\_WHITE | 1 | 8 | 10 |
| AGE\_L5\_FEMALE\_WHITE | 1 | 8 | 11 |
| AGE\_5\_9\_FEMALE\_WHITE | 1 | 8 | 12 |
| AGE\_10\_14\_FEMALE\_WHITE | 1 | 8 | 13 |
| AGE\_15\_17\_FEMALE\_WHITE | 1 | 8 | 14 |
| AGE\_18\_19\_FEMALE\_WHITE | 1 | 8 | 15 |
| AGE\_L5\_MALE\_BLACK | 1 | 8 | 16 |
| AGE\_5\_9\_MALE\_BLACK | 1 | 8 | 17 |
| AGE\_10\_14\_MALE\_BLACK | 1 | 8 | 18 |
| AGE\_15\_17\_MALE\_BLACK | 1 | 8 | 19 |
| AGE\_18\_19\_MALE\_BLACK | 1 | 8 | 20 |
| AGE\_L5\_FEMALE\_BLACK | 1 | 8 | 21 |
| AGE\_5\_9\_FEMALE\_BLACK | 1 | 8 | 22 |
| AGE\_10\_14\_FEMALE\_BLACK | 1 | 8 | 23 |
| AGE\_15\_17\_FEMALE\_BLACK | 1 | 8 | 24 |
| AGE\_18\_19\_FEMALE\_BLACK | 1 | 8 | 25 |
| AGE\_L5\_MALE\_ASIAN | 1 | 8 | 26 |
| AGE\_5\_9\_MALE\_ASIAN | 1 | 8 | 27 |
| AGE\_10\_14\_MALE\_ASIAN | 1 | 8 | 28 |
| AGE\_15\_17\_MALE\_ASIAN | 1 | 8 | 29 |
| AGE\_18\_19\_MALE\_ASIAN | 1 | 8 | 30 |
| AGE\_L5\_FEMALE\_ASIAN | 1 | 8 | 31 |
| AGE\_5\_9\_FEMALE\_ASIAN | 1 | 8 | 32 |
| AGE\_10\_14\_FEMALE\_ASIAN | 1 | 8 | 33 |
| AGE\_15\_17\_FEMALE\_ASIAN | 1 | 8 | 34 |
| AGE\_18\_19\_FEMALE\_ASIAN | 1 | 8 | 35 |
| AGE\_L5\_MALE\_OTHER | 1 | 8 | 36 |
| AGE\_5\_9\_MALE\_OTHER | 1 | 8 | 37 |
| AGE\_10\_14\_MALE\_OTHER | 1 | 8 | 38 |
| AGE\_15\_17\_MALE\_OTHER | 1 | 8 | 39 |
| AGE\_18\_19\_MALE\_OTHER | 1 | 8 | 40 |
| AGE\_L5\_FEMALE\_OTHER | 1 | 8 | 41 |
| AGE\_5\_9\_FEMALE\_OTHER | 1 | 8 | 42 |
| AGE\_10\_14\_FEMALE\_OTHER | 1 | 8 | 43 |
| AGE\_15\_17\_FEMALE\_OTHER | 1 | 8 | 44 |
| AGE\_18\_19\_FEMALE\_OTHER | 1 | 8 | 45 |
| AGE\_25\_64\_BACH\_GRAD | 1 | 8 | 46 |
| AGE\_25\_64\_BACH\_GRAD\_GTR20PERC | 1 | 8 | 47 |
| TOTAL\_LATIN | 1 | 8 | 48 |
| LATIN\_WHITE | 1 | 8 | 49 |
| LATIN\_BLACK | 1 | 8 | 50 |
| LATIN\_ASIAN | 1 | 8 | 51 |
| LATIN\_OTHER | 1 | 8 | 52 |

1. EHR Pre-processing Output File Layout
   1. County
   2. ZCTA3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable Name | LENGTH | Variable Order | LABEL | FORMAT |
| PATID | 20 | 1 | PATID: Patient ID | Character |
| AGEYR2016 | 4 | 2 | AGE2016: Patient age (years) in 2016 | Number |
| AGE\_CATEGORIES2016 | 7 | 3 |  | Character |
| HEIGHT2016 | 6 | 4 | HEIGHT2016: Patient Height (cm) in 2016 | Number |
| WEIGHT2016 | 6 | 5 | WEIGHT2016: Patient Weight (kg) in 2016 | Number |
| BMI2016 | 6 | 6 | BMI2016: Patient BMI in 2016 | Number |
| WTCAT2016 | 50 | 7 |  | Character |
| STATE\_ALPHA2016 | 2 | 8 | STATE: Patient state in 2016 | Character |
| STATE\_FIPS2016 | 2 | 9 |  | Character |
| ZCTA32016 | 3 | 10 |  | Character |
| Geography2016 | 5 | 11 |  | Character |
| AGEYR2017 | 4 | 12 | AGE2017: Patient age (years) in 2017 | Number |
| AGE\_CATEGORIES2017 | 7 | 13 |  | Character |
| HEIGHT2017 | 6 | 14 | HEIGHT2017: Patient Height (cm) in 2017 | Number |
| WEIGHT2017 | 6 | 15 | WEIGHT2017: Patient Weight (kg) in 2017 | Number |
| BMI2017 | 6 | 16 | BMI2017: Patient BMI in 2017 | Number |
| WTCAT2017 | 50 | 17 |  | Character |
| STATE\_ALPHA2017 | 2 | 18 | STATE: Patient state in 2017 | Character |
| STATE\_FIPS2017 | 2 | 19 |  | Character |
| ZCTA32017 | 3 | 20 |  | Character |
| Geography2017 | 5 | 21 |  | Character |
| AGEYR2018 | 4 | 22 | AGE2018: Patient age (years) in 2018 | Number |
| AGE\_CATEGORIES2018 | 7 | 23 |  | Character |
| HEIGHT2018 | 6 | 24 | HEIGHT2018: Patient Height (cm) in 2018 | Number |
| WEIGHT2018 | 6 | 25 | WEIGHT2018: Patient Weight (kg) in 2018 | Number |
| BMI2018 | 6 | 26 | BMI2018: Patient BMI in 2018 | Number |
| WTCAT2018 | 50 | 27 |  | Character |
| STATE\_ALPHA2018 | 2 | 28 | STATE: Patient state in 2018 | Character |
| STATE\_FIPS2018 | 2 | 29 |  | Character |
| ZCTA32018 | 3 | 30 |  | Character |
| Geography2018 | 5 | 31 |  | Character |
| Race | 8 | 32 |  | Number |
| Sex | 8 | 33 |  | Character |
| HCLCNT | 3 | 34 | HCLCNT: Hypercholesterolemia: Event Count | Number |
| CFCNT | 3 | 35 | CFCNT: Cystic-Fibrosis: Event Count | Number |
| SCDCNT | 3 | 36 | SCDCNT: Sickle-Cell: Event Count | Number |
| SBCNT | 3 | 37 | SBCNT: Spina-Bifida: Event Count | Number |
| ASTHMACNT | 3 | 38 | ASTHMACNT: Asthma: Event Count | Number |
| CELIACCNT | 3 | 39 | CELIACCNT: Celiac: Event Count | Number |
| SCZCNT | 3 | 40 | SCZCNT: Schizophrenia: Event Count | Number |
| HCL | 3 | 41 | Hypercholesterolemia: Flag | Number |
| CF | 3 | 42 | Cistic-Fibrosis: Flag | Number |
| SCD | 3 | 43 | Sickle-Cell: Flag | Number |
| SB | 3 | 44 | Spina-Bifida: Flag | Number |
| ASTHMA | 3 | 45 | Asthma: Flag | Number |
| CELIAC | 3 | 46 | Celiac: Flag | Number |
| SCZ | 3 | 47 | Schizophreniz: Flag | Number |
| Imputed\_Race | 7 | 48 | Imputed Race | Character |
| Race\_Imputed | 8 | 49 | Imputed Race flag | Number |

1. County Specific Example Code
   1. Pre-Processing Quickstart

Appendix E.1 includes code to generate a pre-processed file using the Quickstart pre-processing program. Text highlighted in yellow has been reviewed and approved or reviewed and edited from its original values. The program uses the data inputs: ACS\_COUNTY, EHR\_COUNTY. The file processes EHR data between 2015 and 2018 and creates a SAS dataset named CODI\_Ready stored in the folder C:\Example\01 Output\Example\_Only. The SAS log is stored in C:\Example\01 Output\SAS LOGS\ This\_is\_the\_Name\_for\_the\_Log <plus date and time information>.log.

/\*Note: subsection of the full program. Be sure to only edit this section, but submit the full program. \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* -- PREPROCESSING ALGORITHM USER INPUT SECTION (PLEASE COMPLETE SECTIONS 1-3 BELOW) -- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* -- PLEASE UPDATE THE BLACK TEXT AFTER THE EQUAL SIGN ACCEPTED VALUES LISTED IN SAS NOTE) -- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*SECTION 1: Input Folder and file names \*\*\*/

/\*\*\*/ %LET ROOT\_PRE = P:\Example; /\*@Note: base directory (ACCEPTABLE VALUES: computer directory name) \*\*\*/

/\*\*\*/ %LET PROGS\_PRE = P:Example\0 SAS Programs; /\*@Note: where SAS programs are stored (ACCEPTABLE VALUES: computer directory name) \*\*\*/

/\*\*\*/ %LET PRE\_DEST = Example\_Only; /\*@Note: Suffix name for EHR Output folder (ACCEPTABLE VALUES: folder name (no puctuation) \*\*\*/

/\*\*\*/ %LET ACS\_FILENAME = ACS\_COUNTY; /\*@Note: ACS file name (ACCEPTABLE VALUES: file name, do not include ".csv") \*\*\*/

/\*\*\*/ %LET EHR\_FILENAME = EHR\_County; /\*@Note: EHR file name (ACCEPTABLE VALUES: file name, do not include ".csv") \*\*\*/

/\*\*\*/ %LET LOG\_NAME\_PRE = This\_is\_the\_Name\_for\_the\_Log; /\*@Note: SAS log file name prefix ACCEPTABLE VALUES: SAS file name (no punctuation) \*\*\*/

/\*SECTION 2: Beginning and End Year of longitudinal EHR data \*\*\*/

/\*\*\*/ %LET BEGIN\_YEAR = 2015; /\*@Note: LONGITUDINAL Start year (ACCEPTABLE VALUES: 4-digit numeric year) \*\*\*/

/\*\*\*/ %LET END\_YEAR = 2018; /\*@Note: LONGITUDINAL End year (ACCEPTABLE VALUES: 4-digit numeric year) \*\*\*/

/\*SECTION 3: OPTIONAL Output File Name Suffix \*\*\*/

/\*\*\*/ %LET EHR\_PRE\_Out = CODI\_Ready; /\*@Note: EHR output file name (ACCEPTABLE VALUES: SAS file name (no punctuation) \*\*\*/

/\*\*\*Note: ROOT\_PRE directory includes subfolders:

"..\0 SAS Programs"

"..\00 Raw Data"

"..\01 Output" and

"..\01 Output\SAS LOGS" \*\*\*/

/\*\*\*NOTE: SAS programs must be stored in the PROGS\_PRE directory including:

Module1-Pre\_Processing\_CODI\_PQ.sas

Module2-Pre\_Processing\_CODI\_PQ.sas

Module3-Pre\_Processing\_CODI\_PQ.sas \*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP\*\*\*\*\*\*\*\*\*/

/\*\*\* DO NOT EDIT BEYOND THIS POINT DO NOT EDIT BEYOND THIS POINT DO NOT EDIT BEYOND THIS POINT DO NOT EDIT BEYOND THIS POINT \*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*Note: subsection of the full program. Be sure to only edit this section, but submit the full program. \*/

* 1. CODI-PQ Prevalence Query Example

Appendix E.2 includes code to generate prevalence estimates using the quickstart program and data inputs: ACS\_COUNTY and CODI\_Ready. The file processes EHR data for a subpopulation and a given methodology and creates a csv file with output named XXX stored in the folder C:\Example\01 Output\Example\_Only. The SAS log is stored in C:\Example\01 Output\SAS LOGS\Prevalence\_2017.log.

Subpopulation: EHR records from 2017 including children 2 to 14 years of age who are either white or Asian, living in Jefferson County (059) Colorado (FIPS code = 08) or Yuma County (125) Colorado (FIPS code = 08) see: <https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/home/?cid=nrcs143_013697> to determine the correct value (08059 is combined state and County). Note that Sex\_Male and Sex\_Female have not been edited since the ALL\_SEXES is turned on to yes.

Methods: Include imputed race information and calculate the age adjusted prevalence.

/\*Note: subsection of the full program. Be sure to only edit this section, but submit the full program. \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* -- USER SELECTION CRITERA SECTIONS 1 through 4 -- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* -- PLEASE UPDATE THE BLACK TEXT AFTER THE EQUAL SIGN (ACCEPTED VALUES LISTED IN SAS NOTE) -- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*SECTION 1: Folder and file names \*\*\*/

/\*\*\*/ %LET Root\_PQ = P:\Example\3 WORK; /\*@Note: base directory (ACCEPTABLE VALUES: computer directory name) \*\*\*/

/\*\*\*/ %LET PROGS\_PQ= P:\Example\0 SAS Programs; /\*@Note: Location of SAS programs (ACCEPTABLE VALUES: computer directory name) \*\*\*/

/\*\*\*/ %LET PRE\_DEST= Example\_Only; /\*@Note: Suffix name of pre-processing output folder (ACCEPTABLE VALUES: folder name (no punctuations)) \*\*\*/

/\*\*\*/ %LET EHR\_PRE\_OUT= CODI\_Ready; /\*@Note: Suffix name of pre-processing output file (ACCEPTABLE VALUES: file name (no punctuations)) \*\*\*/

/\*\*\*/ %LET LOG\_NAME= This\_is\_Another\_Log; /\*@Note: Name for SAS log storage location \*\*\*/

/\*SECTION 2: Subset data based on specifications INCLUDING YEAR, GEOGRAPHY, STATE OR STATE/COUNTY CODE \*\*\*/

/\*\*\*/ %LET BEG\_YEAR = 2017; /\*@Note: Beginning year of analysis (ACCEPTED VALUES: 4-Digit numeric, 2016-2019) \*\*\*/

/\*\*\*/ %LET END\_YEAR = 2017; /\*@Note: End year of analysis (ACCEPTED VALUES: 4-Digit numeric, 2016-2019) \*\*\*/

/\*\*\*/ %LET STATE = N; /\*@Note: Include all geographical locations in file?(ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET GEO\_LIST = %STR('08059', '08125'); /\*@Note: IF STATE="Y" then populate with State FIPS code(s), otherwise populate with FIPS State+FIPS County code(s) \*\*\*/

/\*\*\*/ %LET ALL\_AGES = N; /\*@Note: Include all age ranges? (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET ALL\_SEXES = Y; /\*@Note: Include all sex values? (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET ALL\_RACES = N; /\*@Note: Include all race categories? (ACCEPTED VALUES: Y/N) \*\*\*/

/\*SECTION 3: Only complete section 3 for any "N" values listed in section 2 \*\*\*/

/\*IF ALL\_AGES = N THEN SELECT ONE OR MORE AGE CATEGORIES BELOW: \*\*\*/

/\*\*\*/ %LET AGE\_2\_4 = Y; /\*@Note: Age Range: 2 to 4 (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET AGE\_5\_9 = Y; /\*@Note: Age Range: 5 to 9 (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET AGE\_10\_14 = N; /\*@Note: Age Range: 10 to 14 (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET AGE\_15\_17 = N; /\*@Note: Age Range: 15 to 17 (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET AGE\_18\_19 = N; /\*@Note: Age Range: 18 to 19 (ACCEPTED VALUES: Y/N) \*\*\*/

/\*IF ALL\_RACES = N THEN SELECT ONE OR MORE RACE BELOW: \*\*\*/

/\*\*\*/ %LET RACE\_WHITE = Y; /\*@Note: White (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET RACE\_BLACK = N; /\*@Note: Black/African American (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET RACE\_ASIAN = Y; /\*@Note: Asian (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET RACE\_OTHER = N; /\*@Note: Other (ACCEPTED VALUES: Y/N) \*\*\*/

/\*IF ALL\_SEXES = N THEN SELECT MALE OR FEMALE BELOW: \*\*\*/

/\*\*\*/ %LET SEX\_MALE = N; /\*@Note: Sex: Male (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET SEX\_FEMALE = N; /\*@Note: Sex: Female (ACCEPTED VALUES: Y/N) \*\*\*/

/\*SECTION 4: Methodological option selections \*\*\*/

/\*\*\*/ %LET IMP\_RACES = Y; /\*@Note: Include imputed race values? (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET AGE\_ADJ = Y; /\*@Note: Produce age-adjusted estimates? (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*Note: Root directory includes subfolders:

"..\0 SAS Programs"

"..\00 Raw Data"

"..\1 Output" and

"..\1 Output\SAS LOGS" \*\*\*/

/\*\*\*NOTE: SAS programs must be stored in the PROGS directory including:

Macro1-CODI\_PQ.sas,

Macro2-CODI\_PQ.sas,

Macro3-CODI\_PQ.sas,

Macro4-CODI\_PQ.sas,

Module1-CODI\_PQ.sas, and

Module2-CODI\_PQ.sas \*\*\*/

/\*\*\*NOTE: query output is stored as a csv file in "..\1 Output" named after a time/date stamp and CODI\_Prevalence\_Query\_Report \*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP\*/

/\*\*\* DO NOT EDIT BEYOND THIS POINT DO NOT EDIT BEYOND THIS POINT DO NOT EDIT BEYOND THIS POINT DO NOT EDIT BEYOND THIS POINT \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*Note: subsection of the full program. Be sure to only edit this section, but submit the full program. \*/

1. ZCTA3 Example Code
   1. Pre-Processing ZCTA3 Example

Appendix F.1 includes code to generate a pre-processed file. Text highlighted in yellow has been reviewed and approved or reviewed and edited from its original values. The program uses the data inputs: ACS\_ZCTA3, EHR\_ZCTA3, and Patient\_Condition, processes EHR data between 2016 and 2019, and creates a SAS dataset named CODI\_ZCTA\_Preprocessed stored in the folder C:\Common\_Dir\01 Output\CODI\_PQ\_ZCTA3. The SAS log is stored in C:\Common\_Dir\01 Output\SAS LOGS\Quickstart\_2016\_19 <plus date and time information>.log.

/\*Note: subsection of the full program. Be sure to only edit this section, but submit the full program. \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* -- PREPROCESSING ALGORITHM USER INPUT SECTION (PLEASE COMPLETE SECTIONS 1-3 BELOW) -- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* -- PLEASE UPDATE THE BLACK TEXT AFTER THE EQUAL SIGN (ACCEPTED VALUES LISTED IN SAS NOTE) -- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*SECTION 1: Input Folder and file names \*\*\*/

/\*\*\*/ %LET ROOT\_PRE = C:\Common\_Dir; /\*@Note: base directory (ACCEPTABLE VALUES: computer directory name) \*\*\*/

/\*\*\*/ %LET PROGS\_PRE = C:\Common\_Dir\0 SAS Programs\;/\*@Note: where SAS programs are stored (ACCEPTABLE VALUES: computer directory name) \*\*\*/

/\*\*\*/ %LET PRE\_DEST = CODI\_PQ\_ZCTA3; /\*@Note: Suffix name for EHR Output folder (ACCEPTABLE VALUES: folder name (no puctuation) \*\*\*/

/\*\*\*/ %LET ACS\_FILENAME= ACS\_ZCTA3; /\*@Note: ACS file name (ACCEPTABLE VALUES: file name, do not include ".csv") \*\*\*/

/\*\*\*/ %LET EHR\_FILENAME = EHR\_ZCTA3; /\*@Note: EHR file name (ACCEPTABLE VALUES: file name, do not include ".csv") \*\*\*/

/\*\*\*/ %LET EHR\_COND\_FILENAME = Patient\_Condition; /\*@Note: EHR with condition data file name (ACCEPTABLE VALUES: file name, do not include ".csv") \*\*\*/

/\*\*\*/ %LET LOG\_NAME\_PRE = Quickstart\_2016\_19; /\*@Note: SAS log file name prefix ACCEPTABLE VALUES: SAS file name (no punctuation) \*\*\*/

/\*SECTION 2: Beginning and End Year of longitudinal EHR data \*\*\*/

/\*\*\*/ %LET BEGIN\_YEAR = 2016; /\*@Note: LONGITUDINAL Start year (ACCEPTABLE VALUES: 4-digit numeric year) \*\*\*/

/\*\*\*/ %LET END\_YEAR = 2019; /\*@Note: LONGITUDINAL End year (ACCEPTABLE VALUES: 4-digit numeric year) \*\*\*/

/\*SECTION 3: OPTIONAL Output File Name Suffix \*\*\*/

/\*\*\*/ %LET EHR\_PRE\_Out = CODI\_ZCTA\_preprocessed; /\*@Note: EHR output file name (ACCEPTABLE VALUES: SAS file name (no punctuation) \*\*\*/

/\*\*\*Note: ROOT\_PRE directory includes subfolders:

"..\0 SAS Programs"

"..\00 Raw Data"

"..\01 Output" and

"..\01 Output\SAS LOGS" \*\*\*/

/\*\*\*NOTE: SAS programs must be stored in the PROGS\_PRE directory including:

Module1-Pre\_Processing\_CODI\_PQ.sas

Module2-Pre\_Processing\_CODI\_PQ.sas

Module3-Pre\_Processing\_CODI\_PQ.sas \*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP\*\*\*\*\*\*\*\*\*/

/\*\*\* DO NOT EDIT BEYOND THIS POINT DO NOT EDIT BEYOND THIS POINT DO NOT EDIT BEYOND THIS POINT DO NOT EDIT BEYOND THIS POINT \*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*Note: subsection of the full program. Be sure to only edit this section, but submit the full program. \*/

* 1. CODI-PQ ZCTA3 Example

Appendix F.2 includes code to generate prevalence estimates using the data inputs: ACS\_ZCTA3 and CODI\_ZCTA\_Preprocessed. The file processes EHR data for a subpopulation and a given methodology and creates a csv file with output named XXX stored in the folder C:\Common\_Dir\01 Output\CODI\_PQ\_ZCTA3. The SAS log is stored in C:\Common\_Dir\01 Output\SAS LOGS\Prevalence\_2019.log.

Subpopulation: EHR records from 2019 including pre-teen and teen (10 to 19 years of age) girls and young women, living in Colorado (FIPS code = 08) and within ZCTA3 of 801 (08801 is combined state and ZCTA3). Note that Race\_White, Race\_Black, Race\_Asian, and Race\_Other have not been edited since the ALL\_RACE is turned on to yes.

Methods: Only use records with race information (do not include girls and young women with imputed race), and calculate the age adjusted prevalence.

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* -- USER SELECTION CRITERA SECTIONS 1 through 4 -- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* -- PLEASE UPDATE THE BLACK TEXT AFTER THE EQUAL SIGN (ACCEPTED VALUES LISTED IN SAS NOTE) -- \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*SECTION 1: Folder and file names \*\*\*/

/\*\*\*/ %LET Root\_PQ = C:\Common\_Dir; /\*@Note: base directory (ACCEPTABLE VALUES: computer directory name) \*\*\*/

/\*\*\*/ %LET PROGS\_PQ = C:\Common\_Dir\0 SAS Programs; /\*@Note: Location of SAS programs (ACCEPTABLE VALUES: computer directory name) \*\*\*/

/\*\*\*/ %LET PRE\_DEST = CODI\_PQ\_ZCTA3; /\*@Note: Suffix name of pre-processing output folder (ACCEPTABLE VALUES: folder name (no punctuations)) \*\*\*/

/\*\*\*/ %LET EHR\_PRE\_OUT = CODI\_ZCTA\_preprocessed; /\*@Note: Suffix name of pre-processing output file (ACCEPTABLE VALUES: file name (no punctuations)) \*\*\*/

/\*\*\*/ %LET LOG\_NAME\_PQ = Prevalence\_2019; /\*@Note: Name for SAS log storage location \*\*\*/

/\*SECTION 2: Subset data based on specifications INCLUDING YEAR, GEOGRAPHY, STATE OR STATE/COUNTY CODE \*\*\*/

/\*\*\*/ %LET BEG\_YEAR = 2019; /\*@Note: Beginning year of analysis (ACCEPTED VALUES: 4-Digit numeric, 2016-2018) \*\*\*/

/\*\*\*/ %LET END\_YEAR = 2019; /\*@Note: End year of analysis (ACCEPTED VALUES: 4-Digit numeric, 2016-2018) \*\*\*/

/\*\*\*/ %LET ALL\_STATES = N; /\*@Note: Include all geographical locations in file?(ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET ALL\_AGES = N; /\*@Note: Include all age ranges? (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET ALL\_SEXES = N; /\*@Note: Include all sex values? (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET ALL\_RACES = Y; /\*@Note: Include all race categories? (ACCEPTED VALUES: Y/N) \*\*\*/

/\*SECTION 3: Only complete section 3 for any "N" values listed in section 2 \*\*\*/

/\*IF ALLSTATES = N THEN SELECT ONE OR MORE AGE CATEGORIES BELOW: \*\*\*/

/\*\*\*/ %LET GEO\_GROUP = ZCTA3; /\*@Note: Level of geography (ACCEPTED VALUES: STATE/ZCTA3) \*\*\*/

/\*\*\*/ %LET GEO\_LIST = %STR('08801'); /\*@Note: IF GEO\_GROUP="STATE" then populate with State FIPS code(s), If GEO\_GROUP="ZCTA3" then populate with FIPS State+FIPS County code(s) \*\*\*/

/\*IF ALL\_AGES = N THEN SELECT ONE OR MORE AGE CATEGORIES BELOW: \*\*\*/

/\*\*\*/ %LET AGE\_2\_4 = N; /\*@Note: Age Range: 2 to 4 (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET AGE\_5\_9 = N; /\*@Note: Age Range: 5 to 9 (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET AGE\_10\_14 = Y; /\*@Note: Age Range: 10 to 14 (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET AGE\_15\_17 = Y; /\*@Note: Age Range: 15 to 17 (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET AGE\_18\_19 = Y; /\*@Note: Age Range: 18 to 19 (ACCEPTED VALUES: Y/N) \*\*\*/

/\*IF ALL\_RACES = N THEN SELECT ONE OR MORE RACE BELOW: \*\*\*/

/\*\*\*/ %LET RACE\_WHITE = N; /\*@Note: White (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET RACE\_BLACK = N; /\*@Note: Black/African American (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET RACE\_ASIAN = Y; /\*@Note: Asian (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET RACE\_OTHER = N; /\*@Note: Other (ACCEPTED VALUES: Y/N) \*\*\*/

/\*IF ALL\_SEXES = N THEN SELECT MALE OR FEMALE BELOW: \*\*\*/

/\*\*\*/ %LET SEX\_MALE = N; /\*@Note: Sex: Male (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET SEX\_FEMALE = Y; /\*@Note: Sex: Female (ACCEPTED VALUES: Y/N) \*\*\*/

/\*SECTION 4: Methodological option selections \*\*\*/

/\*\*\*/ %LET IMP\_RACES = N; /\*@Note: Include imputed race values? (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*/ %LET AGE\_ADJ = Y; /\*@Note: Produce age-adjusted estimates? (ACCEPTED VALUES: Y/N) \*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*Note: Root directory includes subfolders:

"..\0 SAS Programs"

"..\00 Raw Data"

"..\1 Output" and

"..\1 Output\SAS LOGS" \*\*\*/

/\*\*\*NOTE: SAS programs must be stored in the PROGS directory including:

Macro1-CODI\_PQ.sas,

Macro2-CODI\_PQ.sas,

Macro3-CODI\_PQ.sas,

Macro4-CODI\_PQ.sas,

Module1-CODI\_PQ.sas, and

Module2-CODI\_PQ.sas \*\*\*/

/\*\*\*NOTE: query output is stored as a csv file in "..\1 Output" named after a time/date stamp and CODI\_Prevalence\_Query\_Report \*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP STOP\*/

/\*\*\* DO NOT EDIT BEYOND THIS POINT DO NOT EDIT BEYOND THIS POINT DO NOT EDIT BEYOND THIS POINT DO NOT EDIT BEYOND THIS POINT \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

1. Example CODI-PQ Output

Once complete, CODI-PQ generates a prevalence table as a csv file. Table G.1 provides an overview of the variables included, and Table G.2 provides example output based on synthetic data. Note in the sample provided in Table 2, descriptive information about the CODI-PQ user inputs, error codes, sources of technical documentation, caveats, and a recommendation citation begins with the row labeled order 3 and continues through order 19. The number of rows output will vary based on the criteria selected.

Table G.1. CODI-PQ Output Data Dictionary

|  |  |
| --- | --- |
| **Column** | **Description** |
| Order | Row order |
| Weight Category | The weight category based on BMI percentile. |
| Sample | The observed (or unadjusted, or crude) count of children and teens in the study population. |
| Population | The weighted (or adjusted) count of the study population. |
| Crude Prevalence | The observed (or unadjusted, or crude) prevalence in the study population. |
| Crude Prevalence Standard Error | The observed (or unadjusted, or crude) standard error in the study population. |
| Weighted Prevalence | Prevalence estimate based on weighted counts.  The use of sample weights is highly recommended. A sample weight is assigned to each sample person. It is a measure of the number of children and teens in the population represented by that sample person. See implementation guide, Appendix A. Sample Weights for more information. |
| Weighted Prevalence Standard Error | Standard error estimates based on weighted counts. See implementation guide, Appendix A. Variance Estimation for more information. |
| Age-adjusted Prevalence | Prevalence estimate based on weighted, age-adjusted counts. See implementation guide, Appendix A. Age Adjustment for more information. |

Table G.2. Example Synthetic Data

| Order | Weight Category | Sample | Population | Crude Prevalence | Crude Prevalence Standard Error | Weighted Prevalence | Weighted Prevalence Standard Error | Age-adjusted Prevalence | Age-adjusted Prevalence Standard Error |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | (1) Underweight (<5th percentile) | 61 | 5,801 | 4.16 | 0.25 | 4.80 | 0.82 | 4.80 | 0.82 |
| 1 | (2) Healthy Weight (5th to <85th percentile) | 824 | 82,020 | 56.17 | 1.3 | 67.88 | 3.58 | 67.88 | 3.58 |
| 1 | (3) Overweight (85th to <95th percentile) | 251 | 14,752 | 17.11 | 0.98 | 12.21 | 2.42 | 12.21 | 2.42 |
| 1 | (4) Obesity (>95th percentile) | 331 | 18,255 | 22.56 | 1.09 | 15.11 | 2.89 | 15.11 | 2.89 |
| 1 | (4b) Severe Obesity (>120% of the 95th percentile) | 100 | 6,085 | 7.52 | 1.01 | 5.25 | 2.22 | 5.25 | 2.22 |
| 2 | Totals: | 1,467 | 120,828 |  |  |  |  |  |  |
| 3 | Dataset: IQVIA, 2016-2018 |  |  |  |  |  |  |  |  |
| 4 | Query Parameters: AGE RACE SEX GEOGRAPHY YEAR |  |  |  |  |  |  |  |  |
| 5 | AGE: (10 - 14, 15 - 17, 18 - 19) |  |  |  |  |  |  |  |  |
| 6 | SEX: (Female) |  |  |  |  |  |  |  |  |
| 7 | RACE: (White, Black, Asian, Other) |  |  |  |  |  |  |  |  |
| 8 | RACE Suppressed: No. |  |  |  |  |  |  |  |  |
| 9 | RACE Imputed: People with unknown race were excluded. |  |  |  |  |  |  |  |  |
| 10 | Geography: (08810) |  |  |  |  |  |  |  |  |
| 11 | Year: 2019 |  |  |  |  |  |  |  |  |
| 12 | Weighting cells were collapsed for: (Geography) |  |  |  |  |  |  |  |  |
| 13 | AGE adjusted?: (Yes) |  |  |  |  |  |  |  |  |
| 13 | Error Codes: (None) |  |  |  |  |  |  |  |  |
| 14 | Technical Documentation: See https://github.com/NORC-UChicago/CODI-PQ for more information and full details on data sources and methodologies. |  |  |  |  |  |  |  |  |
| 15 | Query Date: Friday, 1 July 2020 4:09:46 PM |  |  |  |  |  |  |  |  |
| 15 | Suggested Citation: AEMR-US version 5 OMOP 5 [Aug 2019 Release] accessed through the E360TM Software-as-a-Service (SaaS) Platform. Accessed through prevalence query on Friday, 1 January 1960 4:09:46 PM |  |  |  |  |  |  |  |  |
| 16 | Caveats |  |  |  |  |  |  |  |  |
| 17 | Children with either missing or invalid age, sex, height, weight, or geography are not included in counts and prevalence estimates. |  |  |  |  |  |  |  |  |
| 18 | The method used to calculate the standard errors are documented in the technical documentation. |  |  |  |  |  |  |  |  |
| 19 | The population estimates are based on age-race-sex-location specific counts from the 2014-2018 American Community Survey Five-year Estimates released by the Census Bureau on December 19, 2019. |  |  |  |  |  |  |  |  |

1. State FIPS Codes

Note: for a list of all state and county codes visit: USDA’s website <https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/home/?cid=nrcs143_013697>

Table H.1 State FIPS Codes

| **Name** | **Postal Code** | **FIPS** |
| --- | --- | --- |
| Alabama | AL | 01 |
| Alaska | AK | 02 |
| Arizona | AZ | 04 |
| Arkansas | AR | 05 |
| California | CA | 06 |
| Colorado | CO | 08 |
| Connecticut | CT | 09 |
| Delaware | DE | 10 |
| District of Columbia | DC | 11 |
| Florida | FL | 12 |
| Georgia | GA | 13 |
| Hawaii | HI | 15 |
| Idaho | ID | 16 |
| Illinois | IL | 17 |
| Indiana | IN | 18 |
| Iowa | IA | 19 |
| Kansas | KS | 20 |
| Kentucky | KY | 21 |
| Louisiana | LA | 22 |
| Maine | ME | 23 |
| Maryland | MD | 24 |
| Massachusetts | MA | 25 |
| Michigan | MI | 26 |
| Minnesota | MN | 27 |
| Mississippi | MS | 28 |
| Missouri | MO | 29 |
| Montana | MT | 30 |
| Nebraska | NE | 31 |
| Nevada | NV | 32 |
| New Hampshire | NH | 33 |
| New Jersey | NJ | 34 |
| New Mexico | NM | 35 |
| New York | NY | 36 |
| North Carolina | NC | 37 |
| North Dakota | ND | 38 |
| Ohio | OH | 39 |
| Oklahoma | OK | 40 |
| Oregon | OR | 41 |
| Pennsylvania | PA | 42 |
| Rhode Island | RI | 44 |
| South Carolina | SC | 45 |
| South Dakota | SD | 46 |
| Tennessee | TN | 47 |
| Texas | TX | 48 |
| Utah | UT | 49 |
| Vermont | VT | 50 |
| Virginia | VA | 51 |
| Washington | WA | 53 |
| West Virginia | WV | 54 |
| Wisconsin | WI | 55 |
| Wyoming | WY | 56 |

1. Note: prevalence estimates of obesity will include two categories: those that are category 4 and 4a. [↑](#footnote-ref-1)
2. IQVIA’s Ambulatory EMR (AEMR) data contains the clinical data of approximately 74 million patients since 2006 and 22 million per year from all 50 states recorded by over 100,000 providers who are affiliated with over 800 ambulatory large practices and physician networks. About 40% of the contributing physicians are primary care practitioners, and the remaining 60% are specialists. These records capture key clinical variables such as laboratory values, weight, and blood pressure to prescriptions, diagnoses, hospital metrics, and therapeutic outcomes. AEMR is also used to connect patient vitals, health behaviors, and risk factors to diagnosis and treatment and develop insights based on provider treatment decisions and prescribed medications (in contrast to dispensed medications).

   Source: AEMR-US version 5 OMOP 5 (Aug 2019 release) accessed through the E360TM Software-as-a-Service (SaaS) Platform. [↑](#footnote-ref-2)
3. <https://synthetichealth.github.io/synthea/> [↑](#footnote-ref-3)
4. <https://github.com/NORC-UChicago/CODI-PQ> [↑](#footnote-ref-4)