

CODI@NC Pilot Queries

|  |
| --- |
| Work Plan  Query 1: Program Enrollment and Demographic Data (Cohort Query) |
|  |

**MODIFICATION HISTORY**

|  |  |  |
| --- | --- | --- |
| **Date** | **Query Package Version** | **Description** |
| 6/16/2023 | B01 | Initial Beta Release. |
| 7/13/2023 | B02 | Release #2 includes the following changes:   * Removes PATID from the output and adds LINKID * Creates age bands for sampling of controls in Query 3 * Changed Session Count to be based on Session ID rather than Session Date * Removed Lat/Long from output |
|  |  |  |

Table of Contents

[I. QUERY METADATA 3](#_Toc132377597)

[II. PURPOSE AND SCOPE 4](#_Toc132377598)

[III. query PACKAGE FOLDERS and FILES 7](#_Toc132377599)

[IV. RESPONDING TO THE QUERY PACKAGE 7](#_Toc132377600)

[V. Output Files 8](#_Toc132377601)

[VI. Query Output Table SHELLS Error! Bookmark not defined.](#_Toc132377602)

[VII. CODELIST 8](#_Toc132377603)

# 

# QUERY METADATA

|  |  |
| --- | --- |
| Query Name: | Query 1, the Cohort Query, identifies the study population served by CODI-participating organizations/programs and returns program enrollment and demographic data. The output of the Query 1 is the “Finder File.” (in other words, Query 1 produces the “Finder File”).  Query 1 is the first of two queries which together answer the following question: What are the characteristics (demographic, geographic, SDOH, or health-related) of the population served by CODI-participating organization/programs? |
| Due Date: | 07/19/23 |
| Description | The goal of this query is to identify individuals who participate in programs offered by CODI NC network partners and create a data set defining the cohort for subsequent queries.  **Inclusion criteria** for this query includes the following:   * Any person enrolled in at least one CODI-participating organization or program in at least one year between the years of 2017–2022.   + Programs include any health or wellness program with Duke University, University of North Carolina, Durham County Department of Public Health, YMCA of the Triangle, Durham and Chapel Hill Parks & Recreation, or the North Carolina Coalition to End Homelessness.   + The timeframe of interest is as follows: January 1, 2017 – December 31, 2022.   **Exclusion criteria** includes the following:   * Persons with medical records on file with Duke University or the University of North Carolina who between 2017–2022 did NOT participate in any wellness programs at Duke University or the University of North Carolina or other programs with CODI-network organizations (e.g., Durham County Department of Public Health, YMCA of the Triangle, Durham and Chapel Hill Parks & Recreation, or the North Carolina Coalition to End Homelessness) will be excluded from analysis.   This is a work plan for Query 1 of the CODI@NC Project. The work plan describes the purpose of the query, program running instructions, and outputs. |
| Contact | Questions about this query package should be sent to:   * Stephanie Poley ([stephanie.poley@duke.edu](mailto:stephanie.poley@duke.edu)) * Alex Beede ([abeede@mitre.org](mailto:abeede@mitre.org)) |
| Additional instructions | Planned Use: Query results will be used by the CODI Data Coordinating Center to create a finder file that will extract additional data needed for Query 2.  Program Package Contents: The query package includes the Work Plan and the SAS code which will produce a SAS data file.  General instructions: The query should be executed against the most recent approved CODI Datamart and CDM tables ([CODI Data Dictionary](https://github.com/mitre/codi/blob/c1a25cb8f8d980f972e72de3cdb5ec2c635a4dfc/CODI%20Data%20Dictionary.xlsx)). This will generate an output SAS data file, named “UseCase1FinderFile.sas7bdat.“  User instructions: Replace the file path / directory to save the file on your machine.  Output: This will generate one output SAS data file, the Finder File named “UseCase1FinderFile.sas7bdat.” Send the SAS file to the Data Coordinating Center via Box. |

# PURPOSE AND SCOPE

The purpose of this query package is to begin to characterize the populations served by each CODI-participating organization/program. This query will help answer the following question: What are the characteristics (demographic, geographic, SDOH, or health-related) of the population served by CODI-participating organizations/programs?

Query 1 returns the LINKIDs for individuals who participated in CODI programs during the study period. These are the same as the LINKIDs that were generated by PPRL. Query 1 also returns demographic, geographic, and program participation data for those individuals. A subsequent query (Query 2) will use the LINKIDs returned by this query to pull health information for those individuals who both participated in a CODI program and have a health record at either Duke Health or UNC Health.

Finally, Query 1 creates age bands for the study population, so that a later query (Query 3) can use those age bands, plus the LINKIDs to identify non-program participants at Duke and UNC who can act as the control group for Use Case #2 (which compares key health metrics between CODI program participants and non-participants).

**Study Population & Timeframe:**

**Inclusion criteria** for this query (Query 1) includes the following:

* Any person enrolled in at least one CODI-participating organization or program in at least one year between the years of 2017–2022.
  + Programs include any health or wellness program with Duke University, University of North Carolina, Durham County Department of Public Health, YMCA of the Triangle, Durham and Chapel Hill Parks & Recreation, or the North Carolina Coalition to End Homelessness.
  + The timeframe of interest is as follows: January 1, 2017 – December 31, 2022.

**Exclusion criteria** includes the following:

* Persons with medical records on file with Duke University or the University of North Carolina who between 2017-2022 did NOT participate in any wellness programs at Duke University or the University of North Carolina or other programs with CODI-network organizations (e.g., Durham County Department of Public Health, YMCA of the Triangle, Durham and Chapel Hill Parks & Recreation, or the North Carolina Coalition to End Homelessness) will be excluded from analysis.

Output tables will be produced by running SAS programs against static local data marts. The output tables consist of patient-level data. The programs are written in SAS and designed to run against data marts conforming to CODI Data Model v4.2.1 using SAS 9.4 or higher.

**Queried Tables and Variables:**

Query 1 will query data from seven CODI Data Model tables. The variable descriptions below are pulled from the [CODI Data Dictionary](https://github.com/mitre/codi/blob/c1a25cb8f8d980f972e72de3cdb5ec2c635a4dfc/CODI%20Data%20Dictionary.xlsx). Not all of the data queried will be present in the output (e.g., PATIDs are used to link data across tables but will not be in the query output).

|  |  |  |
| --- | --- | --- |
| **Table** | **Variable** | **Description** |
| LINK | LINKID | The Link ID from the PPRL process |
| DEMOGRAPHIC | BIRTH\_DATE | Date of birth. |
| DEMOGRAPHIC | SEX | Sex assigned at birth. |
| DEMOGRAPHIC | RACE | Details of categorical definitions are available in the [CODI Data Dictionary](https://github.com/mitre/codi/blob/c1a25cb8f8d980f972e72de3cdb5ec2c635a4dfc/CODI%20Data%20Dictionary.xlsx). |
| DEMOGRAPHIC | HISPANIC | A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. |
| HOUSEHOLD\_LINK | HOUSEHOLDID | An identifier for a (virtual) household that is unique across CODI data marts in the same CODI network. More than one patient or participant (PATID) may be linked to the same household (HOUSEHOLDID) |
| CENSUS\_LOCATION | PERSON\_ID | (Same as PATID) |
| CENSUS\_LOCATION | LOC\_START | Beginning date address for individual is valid (if date is unknown, assign a date estimated to when electronic address data began collection. |
| CENSUS\_LOCATION | LOC\_END | Ending date address for individual is valid (dates prior to 01/01/2010 should default to 01/01/2010. If location record is a patient’s current location, leave null. Queries will use this criterion to identify a patient’s current address. |
| CENSUS\_LOCATION | GEOCODE | Numeric string of multiple census variables. The geocode can be up to 15 digits long with geographic information hierarchically documented through the string of digits: state(2) + county(3) + tract(6), blockgp(1) + block(3) = GeoCode(15)   When an address can be geocoded to a tract, the geocode is 11 digits long. When an address cannot be geocoded to a tract but can be geocoded to a county, the geocode is 5 digits long. Geocodes with block group and block information should be truncated to 11 digits to meet the requirements of a limited dataset (state(2) + county(3) + tract(6)). |
| CENSUS\_LOCATION | GEOCODE\_BOUNDARY\_YEAR | Indicating the census year for which geocode applies (year in which geocode is valid) |
| CENSUS\_LOCATION | GEOLEVEL | Indicates the specificity of the GEOCODE match  This can be assessed using logic that considers the length of the GEOCODE value (2 characters for state; 5 characters for county; 11 characters for census tract) Data partners should have the values T or C. Data partners should not map to B or G. |
| PROGRAM\_ENROLLMENT | PATID | A link back to the demographics table. |
| PROGRAM\_ENROLLMENT | PROGRAMID | A link back to the program this enrollment belongs to. |
| PROGRAM\_ENROLLMENT | ENROLLMENT\_DATE | A date on which the enrollment was performed. |
| PROGRAM\_ENROLLMENT | COMPLETION\_DATE | A date on which the individual who enrolled completed the program. |
| PROGRAM\_ENROLLMENT | DISPOSITION\_DESCRIPTION | A description of the circumstances under which an individual ended their participation in the program. For example, an individual might complete a program successfully, they might drop out, or they might move to a different state. |
| SESSION | INTERVENTION\_ACTIVITY | True if the session included performing at least moderate physical activity; moderate activity requires a moderate amount of effort (5-6 on a scale of 0 to 10) and noticeably accelerates the heart rate and breathing. |
| SESSION | INTERVENTION\_NUTRITION | True if the session included an activity designed to improve nutrition. |
| SESSION | INTERVENTION\_NAVIGATION | True if the session included a navigational service to access benefits or to overcome barriers to care. |
| SESSION | DOSE | A measure of the amount of time spent on this encounter. Researchers can compare the total dose to the prescribed total dose to assess the extent to which an individual completed a program. |
| SESSION | PROGRAMID | A link back to the program this session belongs to (if any). |
| SESSION | SESSION\_DATE | A date on which the session was conducted. |
| PROGRAM | PROGRAMID |  |
| PROGRAM | PROGRAM\_NAME | A name of the program (e.g., Girls on the Run). |
| PROGRAM | PROGRAM\_DESCRIPTION | A description of the program. |
| PROGRAM | AIM\_NUTRITION | True if the aim of the program includes improving nutrition. |
| PROGRAM | AIM\_ACTIVITY | True if the aim of the program includes improving physical activity. |
| PROGRAM | AIM\_WEIGHT | True if the aim of the program includes improving weight status. |
| PROGRAM | PRESCRIBED\_TOTAL\_DOSE | A total amount of time (in hours) an individual should spend in the program. This field should equal DURATION x FREQUENCY x LENGTH (weeks x sessions/week x hours/session). |

1. **query PACKAGE FOLDERS and FILES**

Please see below for list of folders and included files.

| **Folder** | **Description** | **File Type** | **File Name** |
| --- | --- | --- | --- |
| /Query 1 | Query program **instructions**. | PDF – Work Plan (this document) | work\_plan.pdf |
| /Query 1 | **Query Master Program**. This is the only program that should be edited by users. | SAS program | CODI\_Query1July13.sas |
| /Query 1 | **The output** generated by the program should go into this folder. The finder file from this folder will need to be returned to the data coordinating center. | SAS Dataset | See section V: Output Files  Finder file: UseCase1FinderFile.sas7bdat  You will rename this finder file to include the name of your organization:  UseCase1FinderFile [name of org].sas7bdat |

# RESPONDING TO THE QUERY PACKAGE

1. Open the **Query 1** folder and open the SAS program “CODI\_Query1July13.sas”
2. Modify the directory paths as follows. For reasons of compatibility and standardization, directory paths must meet the following criteria:
   * + DO use forward slashes (e.g. /) which are always compatible on both UNIX and WINDOWS.
     + DO use end of path separators (e.g. /xyz/ and not /xyz) which are assumed by many programs.
     + DO use beginning of path separators on UNIX (e.g. /xyz).
     + DO NOT use beginning of path separators on WINDOWS (e.g. P:/xyz not /P:/xyz).
     + DO NOT surround directory paths with quotes (e.g. /xyz/ not "/xyz/").
3. After the **CODISAS** libname statement: provide the directory path to where your CODI datasets are located.
4. After the **OUT** libname statement: provide the directory path to where the output file should be saved.
5. Save and run “CODI\_Query1July13.sas”. The programs are not to be altered in any way except as described above.
6. Once the program is run, the output will be a finder file named “UseCase1FinderFile.sas7bdat”. Rename this file as “UseCase1FinderFile\_[name of org].sas7bdat”
7. Return the finder file “UseCase1FinderFile\_[name of org].sas7bda”t via Duke Box.

# Output Files

Please review your output to confirm that the layout and contents conform to the layouts in this work plan and the results you would expect. **SITES SHOULD ONLY RETURN THE FINDER FILE.**

**File Returned to the Coordinating Center**

| **Folder Name** | **Included File Type** | **File Name** |
| --- | --- | --- |
| /Query 1 | SAS | Finder file: UseCase1FinderFile\_[name of org].sas7bdat |

# CODELIST

There is no codelist for Query 1.