

CODI@NC Pilot Queries

|  |
| --- |
| Work Plan  Query 3: Program Enrollment and Health Outcomes |
|  |

**MODIFICATION HISTORY**

|  |  |  |
| --- | --- | --- |
| **Date** | **Query Package Version** | **Description** |
| 8/29/2023 | B01 | Initial Beta Release. |
|  |  |  |
|  |  |  |

Table of Contents

[I. QUERY METADATA 4](#_Toc132377597)

[II. PURPOSE AND SCOPE 6](#_Toc132377598)

[III. query PACKAGE FOLDERS and FILES 8](#_Toc132377599)

[IV. RESPONDING TO THE QUERY PACKAGE 9](#_Toc132377600)

[V. Output Files 10](#_Toc132377601)

[VI. Query Output Table SHELLS 12](#_Toc132377602)

[VII. CODELIST 12](#_Toc132377603)

# 

# QUERY METADATA

|  |  |
| --- | --- |
| Query Name: | Query 3 builds on the output of Queries 1 and 2 by identifying a comparison group of individuals **not** participating in a CODI program, but who are similar in other ways to individuals who are. Query 3 then returns clinical and SDOH data for those individuals. The goal of Query 3 is to give a researcher enough information to answer the following question: Is enrollment in a program associated with a change in health outcomes, such as BMI, HbA1c, or blood pressure? |
| Due Date: | 08/31/2023 |
| 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 **NOT** enrolled in at least one CODI-participating organization or program in at least one year between the years of 2017–2022, but who has similar demographic characteristics to individuals whose LINKID is present in the Finder File created by Query 1   + 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.   + The demographic characteristics Query 3 considers are: age, sex, race, and ethnicity   This is a work plan for Query 3 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 (along with those from Queries 1 and 2) will be used by researchers to answer the following question (CODI@NC Pilot Use Case #2): Is enrollment in a program associated with a change in health outcomes, such as BMI, HbA1c, or blood pressure?  **Program Package Contents:** The query package includes the Work Plan, the SAS code which will produce a SAS data file, the Finder File produced by Query 1, and a condition codeset 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 Query3Output.sas7bdat  The query uses the Finder File produced by Query 1, as well as an ICD-10 codeset file as inputs.  **User instructions:** Replace the file paths/directories to match your local file structure and run the query.  **Output:** This will generate 1 output SAS data files, the Finder File named Query3Output.sas7bdat. Rename the file to include your organization name and send the SAS file to the Data Coordinating Center via Box. |

# PURPOSE AND SCOPE

This is the third and final query in the CODI@NC pilot query series. Queries 1 and 2 pulled demographic, geographic, clinical, SDOH, and program participation data for individuals who participated in CODI programs during the study period.

This query creates a comparison group of individuals who are demographically similar to those identified in Query 1, but do not participate in CODI programs. The query pulls clinical and SDOH data for these individuals so that it can be compared to corresponding data for individuals who do participate in CODI programs.

Once this query is complete, the results can be compared to the clinical and SDOH data of program participants pulled in Query 2 to answer the following question: Is enrollment in a program associated with a change in health outcomes, such as BMI, HbA1c, or blood pressure?

**Study Population & Timeframe:**

**Inclusion criteria** for this query includes the following:

* Any person **NOT** enrolled in at least one CODI-participating organization or program in at least one year between the years of 2017–2022, but who has similar demographic characteristics to individuals whose LINKID is present in the Finder File created by Query 1
  + 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.
  + The demographic characteristics Query 3 considers are: age, sex, race, and ethnicity

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 3 will query data from nine 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).

|  |  |  |
| --- | --- | --- |
| **Table** | **Variable** | **Description** |
| LINK | LINKID | The Link ID from the PPRL process |
| LINK | LINK\_ITERATION | An iteration of the record linkage 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. |
| SDOH\_EVIDENCE\_INDICATOR | PATID | A link back to the demographics table. |
| SDOH\_EVIDENCE\_INDICATOR | SDOH\_CATEGORY | A social topic area pertaining to circumstances which can determine health outcomes for an individual. |
| SDOH\_EVIDENCE\_INDICATOR | EVIDENCE\_DATE | A date on which a data owner, partner, or researcher has made an assertion indicating the presence of SDOH evidence. This date corresponds to the data partner's most recent determination of available evidence and does not necessarily match submission dates of any of the SDOH evidence. CODI is not expected to maintain a history of assertions, only one assertion based on the data partner's supplied evidence. |
| SDOH\_EVIDENCE\_INDICATOR | EVIDENCE\_TABLE\_NAME | A name of a table in the CODI schema in which there is some evidence pertaining to the CODI SDOH indicator category. The evidence may be a screening response (in PRO\_CM), or a reported problem (in CONDITION or DIAGNOSIS), or some other information stored in a CODI table. |
| SDOH\_EVIDENCE\_INDICATOR | EVIDENCE\_EXPLANATION | For indicator assertions without CODI data evidence; an explanation for the assertion. |
| SDOH\_EVIDENCE\_INDICATOR | EVIDENCE\_ROWID | An identifier for a specific row in the table referenced in the EVIDENCE\_TABLE\_NAME that contains evidence of a potential social determinant. |
| ASSET\_DELIVERY | ASSET\_DELIVERY\_ID | A unique identifier for the asset delivery |
| ASSET\_DELIVERY | PATID | A link back to the demographic table. |
| ASSET\_DELIVERY | PROGRAMID | A link back to the program table. |
| ASSET\_DELIVERY | ASSET\_PURPOSE | An intended purpose for the use of a monetary asset (e.g., health insurance or food). |
| ASSET\_DELIVERY | DELIVERY\_START\_DATE | A date the asset delivery began. |
| ASSET\_DELIVERY | DELIVERY\_END\_DATE | A date the asset delivery ended. |
| ASSET\_DELIVERY | DELIVERY\_FREQ | A number of times an asset is delivered each unit of time. |
| ASSET\_DELIVERY | DELIVERY\_FREQ\_UNIT | A unit of time used to describe how often an asset is delivered. For example, an asset delivered twice a week has a frequency of 2 and a unit of Weekly. An asset delivered every other week has a frequency of 0.5 and a unit of Weekly. |
| PREGNANCY | PATID | A link back to the demographics table. |
| PREGNANCY | LAST\_MENSES\_DATE | A date of the parent's last menstrual period. |
| PREGNANCY | ESTIMATED\_DELIVERY\_DATE | An estimated date of delivery. |
| PREGNANCY | DELIVERY\_DATE | An actual date of delivery. |
| PREGNANCY | FETUS\_COUNT | A number of fetuses involved in this pregnancy. |
| PREGNANCY | PRE\_PREGNANCY\_WT | A measure of the parent's weight (in pounds) before becoming pregnant. |
| PREGNANCY | PRE\_PREGNANCY\_BMI | A measure of the parent's body mass index before becoming pregnant. |
| PREGNANCY | DELIVERY\_WT | A measure of the parent's weight (in pounds) at delivery. |
| ENCOUNTER | PATID | Arbitrary person-level identifier used to link across tables. |
| ENCOUNTER | PAYER\_TYPE\_PRIMARY | Categorization of payer type for primary payer associated with the encounter. |
| ENCOUNTER | PAYER\_TYPE\_SECONDARY | Categorization of payer type for secondary payer associated with the encounter. |
| ENCOUNTER | ADMIT\_DATE | Encounter or admission date. |
| ENCOUNTER | ENC\_TYPE | Encounter type. |
| ENCOUNTER | ENCOUNTERID | A unique identifier for the encounter. |
| VITAL | VITALID | A unique identifier for the vitals measurement |
| VITAL | PATID | Arbitrary person-level identifier. Used to link across tables. |
| VITAL | ENCOUNTERID | Arbitrary encounter-level identifier. Not all vital sign measures will be associated with a healthcare encounter. |
| VITAL | MEASURE\_DATE | Date of vitals measure. |
| VITAL | HT | Height (in inches) measured by standing. Only populated if measure was taken on this date. If missing, this value should be null. Decimal precision is permissible. |
| VITAL | WT | Weight (in pounds). Only populated if measure was taken on this date. If missing, this value should be null. Decimal precision is permissible. |
| VITAL | DIASTOLIC | Diastolic blood pressure (in mmHg). Only populated if measure was taken on this date. If missing, this value should be null. |
| VITAL | SYSTOLIC | Systolic blood pressure (in mmHg). Only populated if measure was taken on this date. If missing, this value should be null. |
| VITAL | ORIGINAL\_BMI | BMI if calculated in the source system. Decimal precision is permissible. |
| CONDITION | PATID | Arbitrary person-level identifier. Used to link across tables. |
| CONDITION | ENCOUNTERID | Arbitrary encounter-level identifier used to link across tables. This is an optional field and should only be populated if the item was collected as part of a healthcare encounter. If more than one encounter association is present, this field should be populated with the ID of the encounter when the condition was first entered into the system. However, please note that many conditions may be recorded outside of an encounter context. |
| CONDITION | CONDITION | Some codes will contain leading zeroes, and different levels of decimal precision may also be present. This field is a character field, not numeric, to accommodate these coding conventions. Please populate the exact value of this diagnosis code, but remove any source-specific suffixes and prefixes. (Description updated in v3.1.) |
| CONDITION | CONDITION\_SOURCE | Please note: The “Patient-reported” category can include reporting by a proxy, such as patient’s family or guardian. |
| CONDITION | CONDITION\_TYPE | Condition code type. Please note: The “Other” category is meant to identify internal use ontologies and codes. |
| CONDITION | REPORT\_DATE | Date condition was noted, which may be the date when it was recorded by a provider or nurse, or the date on which the patient reported it. Please note that this date may not correspond to onset date. |
| CONDITION | ONSET\_DATE | The onset date concept here refers to "the date and time when problem (illness, disorder, or symptom) started" (ONC:MU Clinical Data Set, caDSR 4973971). This is a different concept than report date, which is the date on which the medical status was collected. An onset date should generally be considered independently of the observer or provider. However, the judgment of when a condition "started" depends on the disease, the frequency of visits, and many other factors. It is not clear that any facility or physician employs this field in a manner which can be trusted without validation during analysis. (New definition added in v3.1.) |
| CONDITION | RESOLVE\_DATE | Date condition was resolved, if resolution of a transient condition has been achieved. A resolution date is not generally expected for chronic conditions, even if the condition is managed. |
| DIAGNOSIS | PATID | Arbitrary person-level identifier. Used to link across tables. |
| DIAGNOSIS | ENCOUNTERID | Arbitrary encounter-level identifier. Used to link across tables. |
| DIAGNOSIS | DX | Diagnosis code. |
| DIAGNOSIS | DX\_TYPE | Diagnosis code type. We provide values for ICD and SNOMED code types. Other code types will be added as new terminologies are more widely used. |
| DIAGNOSIS | DX\_SOURCE | Classification of diagnosis source. We include these categories to allow some flexibility in implementation. The context is to capture available diagnoses recorded during a specific encounter. |
| DIAGNOSIS | DX\_ORIGIN | Source of the diagnosis information. |
| DIAGNOSIS | DX\_DATE | Date diagnosis was recorded, if known. |

1. **query PACKAGE FOLDERS and FILES**

Please see below for list of folders and included files.

| **Folder** | **Description** | **File Type** | **File Name** |
| --- | --- | --- | --- |
| …/Query 3 | Query program **instructions**. | PDF – Work Plan (this document) | CODI@NC\_Query3\_Workplan.pdf |
| …/Query 3 | **Query Master Program**. This is the only program that should be edited by users.  Note that Duke and UNC will run unique versions of the SAS code – make sure to run the version with your organization’s name in the filename. | SAS program | Query 3\_[Date]\_[name of org].sas |
| …/Query 3 | Condition **codelist.** The query uses this as an input to read in ICD-10 codes | Excel file | ConditionCodes.xlsx |
| …/Query 3 | **Finder File** of LINKIDs used as input (this is the consolidated output from Query 2) | SAS Dataset | UseCase1FinderFile.sas7bdat |
| …/Query 3 | **The output** generated by the program should go into this folder. The output will need to be returned to the data coordinating center. | SAS Dataset | See section V: Output Files  Output: Query3Output.sas7bdat  You will rename finder file to include the name of your organization:  Query3Output[name of org].sas7bdat |

# RESPONDING TO THE QUERY PACKAGE

1. Open the **Query 3** folder and open the SAS program ‘Query 3\_[Date]\_[name of org].sas’
   1. Note that Duke and UNC will run unique versions of the SAS code – make sure to run the version with your organization’s name in the filename.
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 **PCORSAS** libname statement: provide the directory path to where your PCOR tables are located.
5. After the **OUT** libname statement: provide the directory path to where the output file should be saved.
6. After the **ICDcodes** libname statement: provide the directory path to where the reference file is located.
7. After the **dir** statement: replace with the local file path for the excel file containing the ICD-10 condition codes
8. Save and run ‘Query 3\_[Date]\_[name of org].sas’. The programs are not to be altered in any way except as described above.
9. Once the program is run, the output will be a file named ‘Query3Output.sas7bdat’. Rename this file as ‘Query3Output[name of org].sas7bdat’
10. Return the finder file ‘Query3Output[name of org].sas7bdat’ via Duke Box.

**Running the program as a batch job**

Running the query as a batch job may reduce processing time. Use the following steps to run Query 2 as a batch job:

1. Within the SAS program text editor, update the locations for the "CODISAS" and "out" libraries as instructed.
2. Confirm the libraries are successfully assigned.
3. Save and close the SAS program.
4. Right click on the SAS program file then select the option "Batch Submit with SAS 9.#" (where # is your current SAS version).
   1. A log and lst file will be written to the directory that contains the SAS program. Outputs will be written to the specified "out" location.

# 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.

**File Returned to the Coordinating Center**

| **Folder Name** | **Included File Type** | **File Name** |
| --- | --- | --- |
| /Query 3 | SAS | Query3Output[name of org].sas7bdat |

# CODELIST

There is a codelist for Query 3. ConditionCodes.xls contains the ICD-10 codes for conditions of interest for Use Case #2, including:

* + Diabetes
  + Prediabetes
  + Pregnancy
  + Hypertension
  + HemoglobinA1C levels

# Finder File

The Finder File is the consolidated output of Query 1 (i.e., all of the LINKIDs of program participants from all Data Owners).

Query 3 uses this file as an input to inform the creation of a comparison group of individuals who do not participate in CODI programs. Clinical and SDOH data for those individuals will then be compared to that of the individuals in the Finder File.

The Finder File is a SAS datafile that should contain only LINKIDs (though if there is other data in the file, Query 3 will select only the LINKIDs from the file).