# **Gathering Data Status Report**

## Name: Andrea Plunkett Jackson

## Date: 02/09/2024

## Status Report for Gathering Data

* **Research Question :** What is the relationship between patients without health coverage that smoke regularly and those that have been told they have asthma?

| **Status** | Progress |
| --- | --- |
| **Variables of Interest** | * IDATE : Date of survey (interview) was completed by participant. The date is required in time series analysis as this is a continuous collection of data. This allows for both a display to track the participation (no other identification provided due to HIPPA regulations) and detect trends. * DISPCODE : Complete or Partial completed survey. Missing data impacts the overall quality of analysis so knowing the number of surveys not completed would provide insight into if the value of data completed is there. * HLTHPLN1: If the respondent has health insurance. Based on the research question described above our analysis in into if there is a relationship between those without insurance therefore this plays a vital role in addition to all bullet points below. * ASTHMA3: If they have ever been told had asthma. * ASTHNOW: If the contributor still has asthma. * SMOKE100: If the participant has ever smoked at least 100 cigarettes. * SMOKEDAY2: How often do they smoke. * STOPSMK2: Have they stopped smoking in the last 12 months. |
| **Data Set Preparation** | * Data type for variable IDATE = Date * Data type for variable DISPCODE = SMALLINT * Data type for all remaining variables listed below = TINYINT * HLTHPLN1 * ASTHMA3 * ASTHNOW * SMOKE100 * SMOKDAY2 * STOPSMK2 * Primary Key variable = ID. It has the INT data type and set to auto-increment * The revised data set and code book is attached below. |
| **Gathering Data\*** | * Screenshots in the Gathering Data section below provides the finished new table in SQL is shown below. |
| **Process for Gathering Data** | * CSV of data obtained * Review of research question and available data. * Combine IMONTH, IDAY, IYEAR into new IDATE column using formula =CONCAT(A2,"-",B2,"-",C2). * Copied new column IDATE into a separate Excel spreadsheet using Paste option of Values. Copied required data sets discussed in variable of interest to new spreadsheet (see below. * Replaced all blanks with value of 0.  I completed that through Excel.  Home > Find and Select (editing section) > Replace > under find what I left it blank and then replace with 0. * Created a new column labeled “ID” and assigned each record with number in sequential order beginning with 1. * Removed all headers from Excel spreadsheet to prepare for importation. * Created schema “ HeartMatters”. * Created table using aforementioned variable names and data types ( see updated codebook provided below). * Imported data from Excel into SQL table using Data Table Import Wizard. |

**Data Set Preparation**

****

**Click image to see full data set**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable Name** | **Description** | **Value Range** | **Label or Format** | **Type** |
| ID | Numerical Value of Participate listed | 1,………..,3999 | Numerical Ascending Order | INT |
| IDATE | Interview Date | 2017/01/03,…,2018/01/08 | YYYYMMDD | Date |
| DISPCODE | Final Deposition | 1100,1200 | Completed or Partial | Small INT (6) |
| HLTHPLN1 | Have health coverage | 1,2,7,9 | Yes, No, Don't know or not sure, Refused | Tiny INT(3) |
| ASTHMA3 | Told had asthma | 1,2,7,9 | Yes, No, Don't know or not sure, Refused | Tiny INT (3) |
| ASTHNOW | Still have asthma | 1,2,7,9 | Yes, No, Don't know or not sure, Refused | Tiny INT (3) |
| SMOKE100 | Smoked at least 100 cigarettes | 1,2,7,9 | Yes, No, Don't know or not sure, Refused | Tiny INT(3) |
| SMOKDAY2 | Frequency days now smoking | 1,2,3,7,9 | Every day, Some days, Not at all, Don't know or not sure, Refused | Tiny INT (3) |
| STOPSMK2 | Stopped smoking last 12 months | 1,2,7,9 | Yes, No, Don't know or not sure, Refused | Tiny Int (3) |

Revised LLCP 2017 Codebook

* **0 = no data available**

## \*Gathering Data

1. Create Table named Health research completed with 9 columns and type value shown with Primary Key being that of ID.

A screenshot of a computer

Description automatically generated

1. Import of data completed and confirmed using command SELECT \* FROM HeartMatters.Healthrearch.

A screenshot of a computer

Description automatically generated

1. Confirmed all data imported by using command SELECT count(\*) from Healthresearch table.

A screenshot of a computer

Description automatically generated