**DETECT Follow-up Interview Data**

1. Where is the data located:
   1. [Raw data](https://uthtmc.sharepoint.com/:f:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/DETECT%20Follow-up%20Inereview%20Data%20Shared/detect_public/data/raw?csf=1&web=1&e=EKNJPh)
   2. [Recoded data](https://uthtmc.sharepoint.com/:f:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/DETECT%20Follow-up%20Inereview%20Data%20Shared/detect_public/data?csf=1&web=1&e=VTay7m) with factor variables and coded to match levels in the [follow-up interview protocol](https://docs.google.com/document/d/1x_q04CktO3K5157EBylsXYGrd9MutSyyMlhjDcRUs-A/edit?usp=sharing). There is an older version of the follow-up interview protocol that has variables that are no longer used. These variables are still included in the analysis data frame.
      1. These 6 data files are the most up-to-date: aps\_investigations\_factor, clusca\_factor, geheal\_factor, obs\_factor, socdem\_factor, sr\_factor
         1. They are each saved as a CSV file and an RDS file. Use the RDS files as they will preserve the metadata and keep factor variables as factor variables
   3. Merged data frame: This is all of the above date files merged into one analysis data frame
      1. There is a merged\_follow\_up\_data CSV and RDS file in the [data folder](https://uthtmc.sharepoint.com/:f:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/DETECT%20Follow-up%20Inereview%20Data%20Shared/detect_public/data?csf=1&web=1&e=DOreiW)
2. RMD files for data cleaning/recoding: [qaqc folder](https://github.com/brad-cannell/detect_public/tree/main/qaqc) of the detect\_public GitHub repository
   1. Use the files that use the convention data\_01\_general\_recode\_factors.rmd. These are the most recent files.
   2. These files recoded all variables from categorical variables to numeric variables to match the follow-up interview protocol and then factor variables
      1. Naming convention:
         1. variable – variable from raw data
            1. Ex: sode\_school
         2. variable\_#cat – recoded variable from categorical to numeric where the numbers match the numbers in the follow-up interview protocol document
            1. Ex: sode\_school\_9cat

Levels from protocol:

1 = Some high school (no diploma)

2 = High school graduate

3 = Some college (no degree)

4 = Associate’s degree

5 = Bachelor’s degree (BA, AB, BS, etc.)

6 = Some graduate or professional school

(no degree)

7 = Graduate or professional school degree

(MS, MA, MD, PhD, etc.)

77 = Don’t know

99 = Refused

* + - 1. variable\_#cat\_f – recoded raw variable as a factor variable. The levels match what we see in the raw categorical variable
         1. Ex: sode\_school\_9cat\_f

1. [Composite Variables folder](https://uthtmc.sharepoint.com/:u:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/Composite%20Variables/Composite%20variables.Rmd?csf=1&web=1&e=wIlkXc) on teams
   1. RMD file adds composite variables to the general health (geheal) data frame
      1. This file added composite variables for PEG: A Three-Item Scale Assessing Pain Intensity and Interference, Geriatric Depression Score and UCLA 3-item Loneliness Scale
      2. There is also a CSV and RDS file that has these variables added to the geheal data frame
   2. A composite variable still needs to be made for the [SF-20 of the medical outcomes survey](https://www.rand.org/health-care/surveys_tools/mos/20-item-short-form.html)
   3. I have not saved a new merged data frame that contains these variables yet
   4. This file has not been added to GitHub and needs to be
2. Codebook
   1. I used Dr. Cannell’s [CodebookR](https://brad-cannell.github.io/codebookr/) package to create a codebook from each data frame that contains descriptive statistics for each variable
   2. [Codebook Folder](https://uthtmc.sharepoint.com/:f:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/Codebook?csf=1&web=1&e=6f8leA) in Teams
      1. RMD files
         1. [Sociodemographic Information](https://uthtmc.sharepoint.com/:u:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/Codebook/socdem_codebook.Rmd?csf=1&web=1&e=II5fJR)
         2. [General Health](https://uthtmc.sharepoint.com/:u:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/Codebook/geheal_codebook.rmd?csf=1&web=1&e=gzvLfK) (not including composite variables)
         3. [Clutter Scale](https://uthtmc.sharepoint.com/:u:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/Codebook/clsca_codebook.Rmd?csf=1&web=1&e=mb2ZHa)
         4. [Self-report](https://uthtmc.sharepoint.com/:u:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/Codebook/sr_codebook.Rmd?csf=1&web=1&e=SQ7vgb)
         5. [Observational Measures](https://uthtmc.sharepoint.com/:u:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/Codebook/obs_codebook.Rmd?csf=1&web=1&e=7Nddpp)
         6. [APS Investigations](https://uthtmc.sharepoint.com/:u:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/Codebook/aps_codebook.Rmd?csf=1&web=1&e=DyMry2)
         7. [Full merged data frame](https://uthtmc.sharepoint.com/:u:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/Codebook/data_01_merged_data_codebook.Rmd?csf=1&web=1&e=ftQMzU) (not including composite variables)
      2. Document files
         1. [Codebook for full merged data frame with comments](https://uthtmc.sharepoint.com/:w:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/Codebook/merged_data_codebook_w_comments.docx?d=wfbf2cd2916d344afada5c8329e0997f9&csf=1&web=1&e=wGeetB)
            1. This is the original codebook of the merged data frame with Dr. Cannell’s comments
         2. [Codebook for full merged data frame](https://uthtmc.sharepoint.com/:w:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/Codebook/merged_data_codebook.docx?d=wb01f7a823dd742ecbc2514b25483c9af&csf=1&web=1&e=nNoIDo)
            1. This is the revised codebook made from Full merged data frame file above. This made the changes discussed in the comments from the previous codebook.
         3. We decided that it may be better to have a codebook for each individual data frame since the merged data has so many variables. The individual Codebook documents are in the [Codebook folder](https://uthtmc.sharepoint.com/:f:/r/sites/SPHCannellResearch-DETECT/Shared%20Documents/DETECT/Codebook?csf=1&web=1&e=HYZgKw) on Teams
   3. Codebook files have not been added to GitHub
      1. I wasn’t sure if we want to create a new repository for these folders or just add a file to the detect-public repository