

Instructor Biography

Jay Iyengar is director of Data Systems Consultants LLC. He is a SAS consultant, trainer, and SAS Certified Advanced Programmer. He was co-leader and organizer of the Chicago SAS Users Group (WCSUG) from 2015-19.

He's presented papers and training seminars at SAS Global Forum (SGF), and other regional and local SAS users groups. His main industry experience is in Health care, Public Health and Pharmaceutical. He obtained his Bachelor's in Public Policy and Economics from Syracuse University and his master's degree from the American University

Using National Healthcare Survey Data for Research and Analysis using SAS

Jay Iyengar, Data Systems Consultants LLC

Western Users of SAS Software Virtual Training 2026

April 22, 2026

Introduction

NHANES – National Health and Nutrition Examination Survey

MCBS – Medicare Current Beneficiaries Survey

MEPS – Medical Expenditure Panel Survey

NHIS – National Health Interview Survey

NHANES Survey

NHANES (National Health and Nutrition Examination Survey)

- The purpose of the survey is to assess the health and nutritional status of adults and children in the U.S.
- Administered and conducted by the CDC (Centers for Disease Control)

The CDC – Centers for Disease Control

- National Center on Health Statistics (NCHS)
 - Federal Statistical Agency located in Hyattsville MD

NHANES is conducted by NCHS, Division of NHANES

NCHS is located in Hyattsville MD

History of the NHANES Survey

First National Examination Survey (NHES I) – 1959-62

Second National Examination Survey (NHES II) - 1963-65

Third National Examination Survey (NHES III) - 1966-70

NHANES I (1971-74)

NHANES II (1988-94)

Continuous NHANES survey (1999-present)

The continuous survey is conducted in 2 year survey cycles.

NHANES Survey Overview

Uses of NHANES

- Determine the extent of major diseases (chronic diseases).
- Examine risk factors associated with disease.
- To formulate sound public health policy.
- Direct and design health programs and services.
- Compute public health measures – Prevalence and Incidence.

Survey Sample Design

NHANES Uses a Stratified, Multi-Year, Clustered, Four-Stage Sample Design

Four Stages of Sample Design

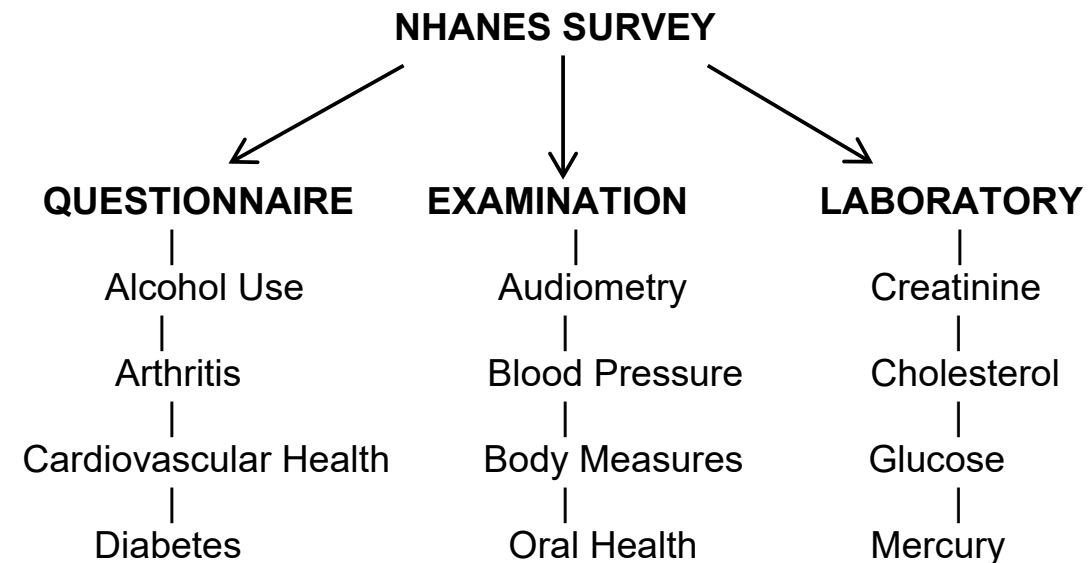
- Counties/Census Tracts – Primary Sampling Unit (PSU)
- Census Blocks
- Dwelling Units (Households)
- Persons within Households

Oversamples Minority and Underrepresented Groups in Survey Research

- Hispanics, African-Americans, at or Below 185% of Poverty Level, Over age 80

Structure of the NHANES Survey

Three Sections: Questionnaire, Exam, and Lab.
A dozen or more survey components within each of the sections



NHANES data production and use

The Division of NHANES is divided up into 2 branches; Operations (OB) and Informatics (IB)

NHANES Informatics Branch

- data cleaning of survey data
- performing QA and QC on the data.
- production of survey data files

The NHANES IB consists of IT specialists, SAS programmers, Project managers, Statisticians, and Epidemiologists all working on the survey.

NHANES data is used by researchers and statisticians, and also by pharma companies to do observational research.

The NHANES Questionnaire

Survey respondents are interviewed in the field

Selected survey respondents are required to complete the NHANES Questionnaire.

The NHANES Questionnaire consists of two parts.

1 – The Household Interview

2 – The Examination Component

The Examination Component is conducted in a Mobile Examination Center (MEC)

NHANES Data Files

Survey data files and codebooks are produced and released to the public.

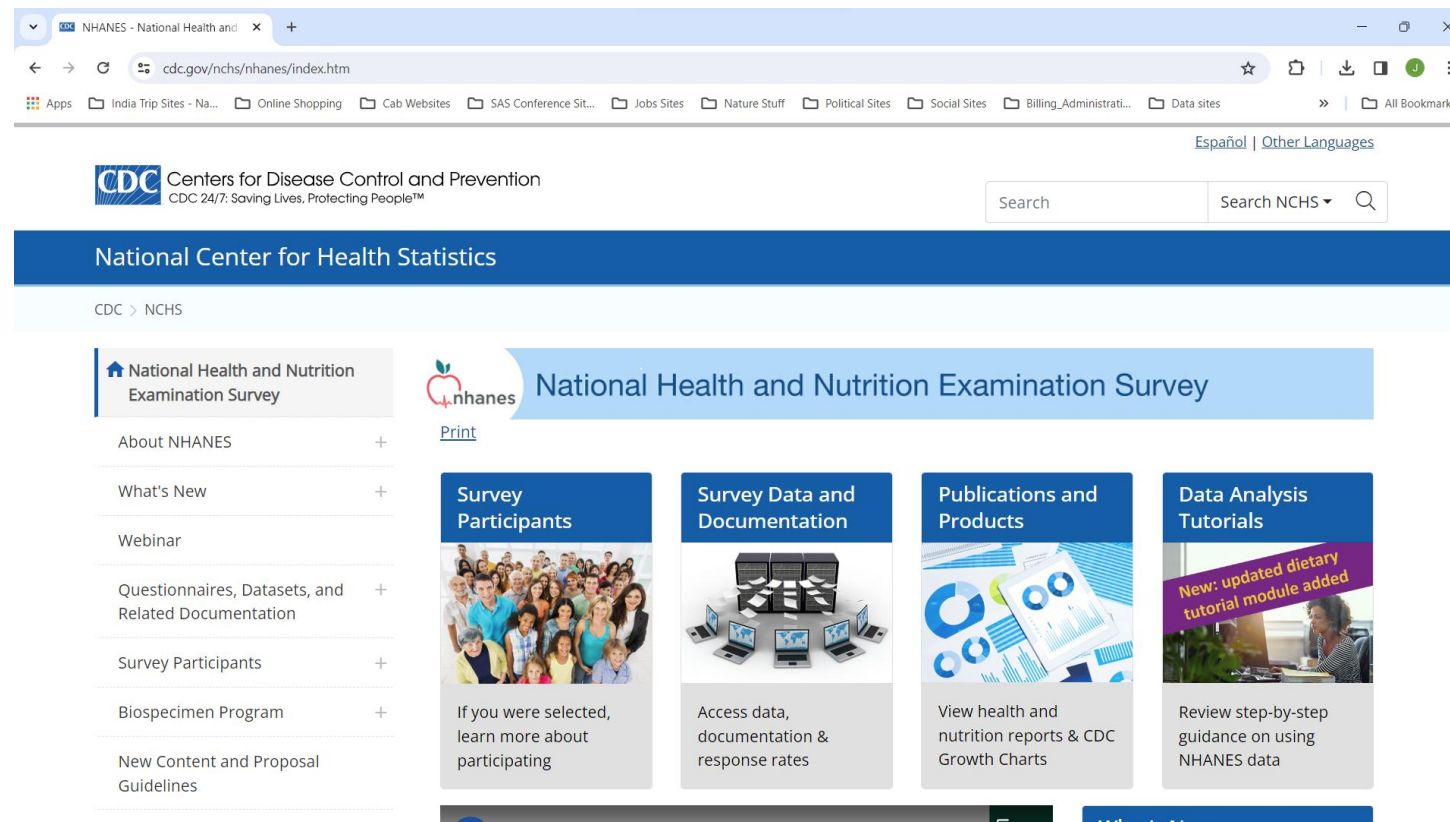
NHANES data files and codebooks are uploaded to and posted on the internet.

Data files and codebooks are produced for each survey component within each survey section (Questionnaire, Exam and Lab)

Data files, codebooks and documentation are released in 2-year survey cycles.

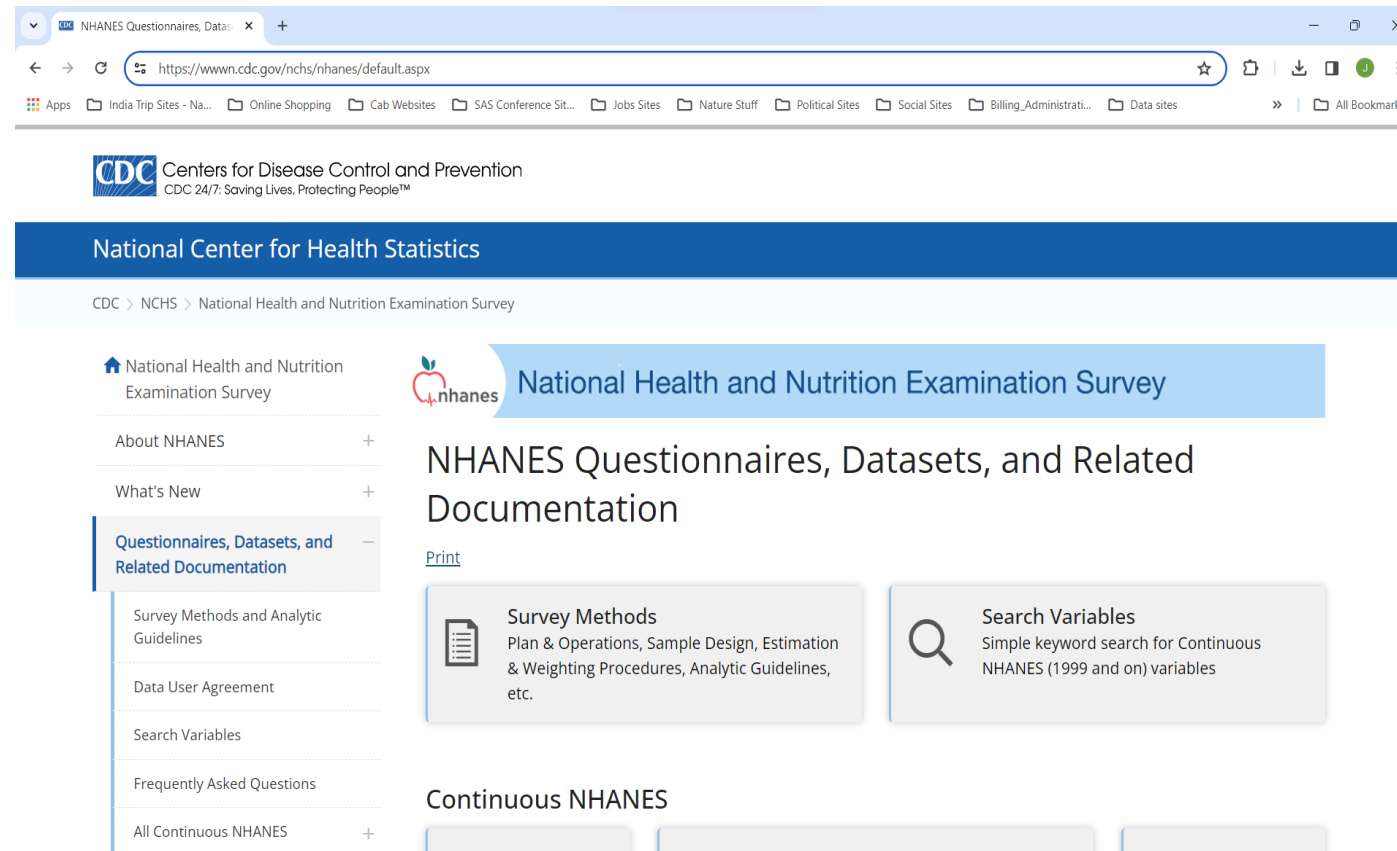
Data files and documentation can be downloaded from the Internet.

The NHANES Homepage



<https://www.cdc.gov/nchs/nhanes/index.htm>

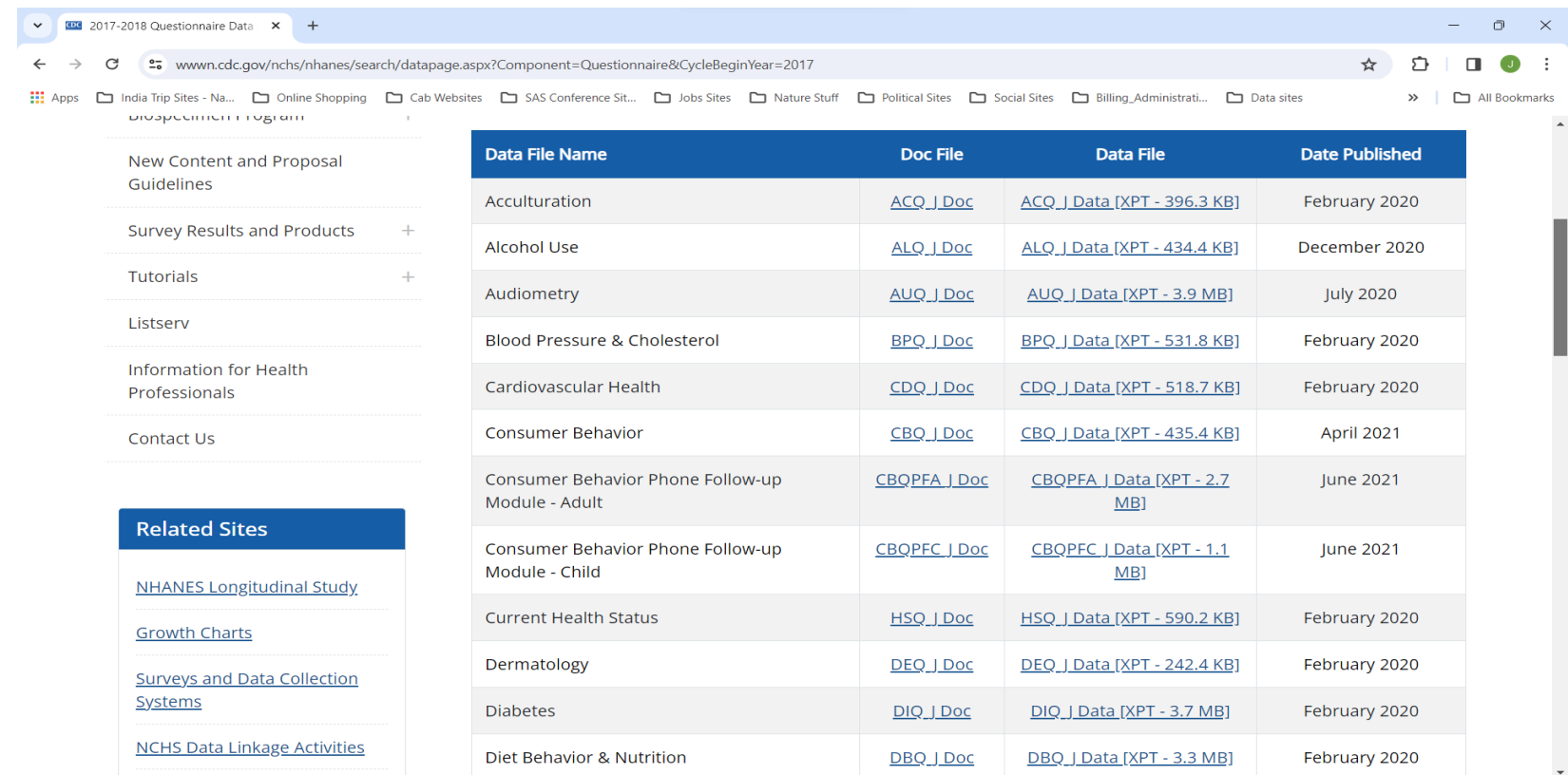
NHANES data sets and documentation



<https://wwwn.cdc.gov/nchs/nhanes/default.aspx>

NHANES datafiles

NHANES data files are released
in SAS Transport (.xpt) format



The screenshot shows a web browser window with the URL www.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Questionnaire&CycleBeginYear=2017. The page displays a table of NHANES data files. The table has four columns: Data File Name, Doc File, Data File, and Date Published. The data files are listed in descending order of publication date. The table includes links to download the data files in XPT format. The left sidebar contains navigation links for New Content and Proposal Guidelines, Survey Results and Products, Tutorials, Listserv, Information for Health Professionals, and Contact Us. The bottom section, titled 'Related Sites', includes links to the NHANES Longitudinal Study, Growth Charts, Surveys and Data Collection Systems, and NCHS Data Linkage Activities.

Data File Name	Doc File	Data File	Date Published
Acculturation	ACQ_J_Doc	ACQ_J_Data [XPT - 396.3 KB]	February 2020
Alcohol Use	ALQ_J_Doc	ALQ_J_Data [XPT - 434.4 KB]	December 2020
Audiometry	AUQ_J_Doc	AUQ_J_Data [XPT - 3.9 MB]	July 2020
Blood Pressure & Cholesterol	BPQ_J_Doc	BPQ_J_Data [XPT - 531.8 KB]	February 2020
Cardiovascular Health	CDQ_J_Doc	CDQ_J_Data [XPT - 518.7 KB]	February 2020
Consumer Behavior	CBQ_J_Doc	CBQ_J_Data [XPT - 435.4 KB]	April 2021
Consumer Behavior Phone Follow-up Module - Adult	CBQPFA_J_Doc	CBQPFA_J_Data [XPT - 2.7 MB]	June 2021
Consumer Behavior Phone Follow-up Module - Child	CBQPFC_J_Doc	CBQPFC_J_Data [XPT - 1.1 MB]	June 2021
Current Health Status	HSQ_J_Doc	HSQ_J_Data [XPT - 590.2 KB]	February 2020
Dermatology	DEQ_J_Doc	DEQ_J_Data [XPT - 242.4 KB]	February 2020
Diabetes	DIQ_J_Doc	DIQ_J_Data [XPT - 3.7 MB]	February 2020
Diet Behavior & Nutrition	DBQ_J_Doc	DBQ_J_Data [XPT - 3.3 MB]	February 2020

Test your knowledge

Which of the following is not a sampling unit in the NHANES Sample Design?

- A. Census Tracts
- B. Households
- C. Zip codes
- D. Blocks in Census Tracts

Demonstration I

Demonstration I - SAS On-demand for academics (SODA).

I – Download the MCQ data file and codebook from the NHANES Website

II - Read and convert the MCQ SAS Transport File into a SAS data set.

III – Run Proc Contents to get a listing of the variables on the MCQ data set.

PROC CONTENTS of the MCQ file

Data Set Name	SDS.P_MCQ	Observations	14986
Member Type	DATA	Variables	63
Engine	V9	Indexes	0
Created	08/07/2024 14:10:02	Observation Length	504
Last Modified	08/07/2024 14:10:02	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label	Medical Conditions		
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	58
First Data Page	1
Max Obs per Page	259
Obs in First Data Page	236
Number of Data Set Repairs	0
Filename	/home/iyenj/Training_Course_Files/Natl_Health_Survey_data/Nhanes/Data/p_mcq.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	19333815408
Access Permission	rw-r--r--
Owner Name	iyenj
File Size	7MB
File Size (bytes)	7733248

Alphabetic List of Variables and Attributes			
#	Variable	Type	Label
7	AGQ030	Num	8 Did SP have episode of hay fever/past yr
11	MCD093	Num	8 Year receive blood transfusion
18	MCD180B	Num	8 Age when told you had heart failure
20	MCD180C	Num	8 Age when told had coronary heart disease
22	MCD180D	Num	8 Age when told you had angina pectoris
24	MCD180E	Num	8 Age when told you had heart attack
26	MCD180F	Num	8 Age when told you had a stroke
33	MCD180L	Num	8 Age when told you had a liver condition
29	MCD180M	Num	8 Age when told you had thyroid problem
2	MCQ010	Num	8 Ever been told you have asthma
3	MCQ025	Num	8 Age when first had asthma
4	MCQ035	Num	8 Still have asthma
5	MCQ040	Num	8 Had asthma attack in past year
6	MCQ050	Num	8 Emergency care visit for asthma/past yr
8	MCQ053	Num	8 Taking treatment for anemia/past 3 mos
9	MCQ080	Num	8 Doctor ever said you were overweight
10	MCQ092	Num	8 Ever receive blood transfusion
12	MCQ149	Num	8 Menstrual periods started yet?
13	MCQ151	Num	8 Age in years at first menstrual period
16	MCQ195	Num	8 Which type of arthritis was it?
47	MCQ220	Num	8 Ever told you had cancer or malignancy
34	MCQ500	Num	8 Ever told you had any liver condition
41	MCQ520	Num	8 Abdominal pain during past 12 months?
42	MCQ530	Num	8 Where was the most uncomfortable pain
43	MCQ540	Num	8 Ever seen a DR about this pain
44	MCQ550	Num	8 Has DR ever said you have gallstones
45	MCQ560	Num	8 Ever had gallbladder surgery?
46	MCQ570	Num	8 Age when 1st had gallbladder surgery?
15	MCQ160A	Num	8 Doctor ever said you had arthritis
17	MCQ160B	Num	8 Ever told had congestive heart failure
19	MCQ160C	Num	8 Ever told you had coronary heart disease
21	MCQ160D	Num	8 Ever told you had angina/angina pectoris
23	MCQ160E	Num	8 Ever told you had heart attack
63	OSQ230	Num	8 Any metal objects inside your body?
14	RHD018	Num	8 Estimated age in months at menarche
1	SEQN	Num	8 Respondent sequence number

MCQ File - Key variables

SEQN – RESPONDENT SEQUENCE NUMBER

MCQ160B – EVER TOLD YOU HAD CONGESTIVE HEART FAILURE

MCQ160C – EVER TOLD YOU HAD CORONARY HEART DISEASE

MCQ160E – EVER TOLD YOU HAD HEART ATTACK

MCD180B – AGE WHEN TOLD YOU HAVE HEART FAILURE

MCD180C – AGE WHEN TOLD YOU HAVE CORONARY HEART DISEASE

MCD180E – AGE WHEN TOLD YOU HAD A HEART ATTACK

PROC CONTENTS of the Demographics File

Data Set Name	SDS.P_DEMO	Observations	15560
Member Type	DATA	Variables	29
Engine	V9	Indexes	0
Created	08/07/2024 14:10:02	Observation Length	232
Last Modified	08/07/2024 14:10:02	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	28
First Data Page	1
Max Obs per Page	564
Obs in First Data Page	537
Number of Data Set Repairs	0
Filename	/home/iyenj/Training_Course_Files/Natl_Health_Survey_data/Nhanes/Data/p_demo.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	19346014710
Access Permission	rw-r--r--
Owner Name	iyenj
File Size	4MB
File Size (bytes)	3801088

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
24	AIALANGA	Num	8			Language of ACASI Interview
10	DMDBORN4	Num	8			Country of birth
12	DMDEDUC2	Num	8			Education level - Adults 20+
13	DMDMARTZ	Num	8			Marital status
11	DMDYRUSZ	Num	8			Length of time in US
20	FIAINTRP	Num	8			Interpreter used in Family Interview?
18	FIALANG	Num	8			Language of Family Interview
19	FIAPROXY	Num	8			Proxy used in Family Interview?
29	INDFMPIR	Num	8			Ratio of family income to poverty
23	MAINTRP	Num	8			Interpreter used in MEC Interview?
21	MIALANG	Num	8			Language of MEC Interview
22	MIAPROXY	Num	8			Proxy used in MEC Interview?
4	RIAGENDR	Num	8			Gender
6	RIDAGEMN	Num	8			Age in months at screening - 0 to 24 mos
5	RIDAGEYR	Num	8			Age in years at screening
9	RIDEXMON	Num	8			Six month time period
14	RIDEXPRG	Num	8			Pregnancy status at exam
7	RIDRETH1	Num	8			Race/Hispanic origin
8	RIDRETH3	Num	8			Race/Hispanic origin w/ NH Asian
3	RIDSTATR	Num	8			Interview/Examination status
2	SDDSRVYR	Num	8			Data release cycle
27	SDMVPSU	Num	8			Masked variance pseudo-PSU
28	SDMVSTRA	Num	8			Masked variance pseudo-stratum
1	SEQN	Num	8			Respondent sequence number
17	SIAINTRP	Num	8			Interpreter used in SP Interview?
15	SIALANG	Num	8			Language of SP Interview
16	SIAPROXY	Num	8			Proxy used in SP Interview?
25	WTINTPRP	Num	8	15.6	15.6	Full sample interview weight
26	WTMECPRP	Num	8	15.6	15.6	Full sample MEC exam weight

Demographics File - Key variables

SEQN – RESPONDENT SEQUENCE NUMBER

RIAGENDR – GENDER

RIDAGEYR – AGE IN YEARS AT SCREENING

RIDRETH1 – RACE/HISPANIC ORIGIN

RIDRETH3 – RACE/HISPANIC ORIGIN WITH NH ASIAN

SDMVPSU – Masked Variance PSU

SDMVSTRA – Masked Variance STRA

WTINTPRP – FULL SAMPLE INTERVIEW WEIGHT

WTMECPRP – FULL SAMPLE MEC EXAM WEIGHT

Demonstration II

Demonstration using SAS On-Demand for Academics (SODA)

- I – Merge P_MCQ and P_DEMO files by SEQN to get all variables on one SAS data set.
- II – Generate listing of observations on new data set, the first 25 observations.
- III – Generate one-way and two-way frequency tables on key variables using PROC FREQ

SAS Code - Data manipulation and validation

Merge MCQ and DEMO files by SEQN to get variables on one data set and produce sample listing

```
Proc SQL;  
  Create Table MCQ_ALL as  
  Select A.Seqn, MCD180B, MCD180C, MCD180E, MCQ160B, MCQ160C, MCQ160E,  
         RIAGENDR, RIDAGEYR, RIDRETH1, RIDRETH3, SDMVPSU, SDMVSTRA, WTINTPRP  
  From SDS.P_MCQ as A, SDS.P_Demo as B  
  Where A.Seqn = B.Seqn;  
Quit;  
  
Proc Print Data=MCQ_ALL(Obs=25);  
  ID SEQN;  
  Var MCQ160B MCQ160C MCQ160E MCD180B MCD180C MCD180E RIAGENDR RIDAGEYR RIDRETH1;  
  Title 'Medical Conditions Questionnaire Data File';  
  Title 'NHANES 2017-2020';  
Run;
```

SAS Output – Proc Print

Sample Listing - First 25 Observations

SEQN	MCQ160B	MCQ160C	MCQ160E	MCD180B	MCD180C	MCD180E	RIAGENDR	RIDAGEYR	RIDRETH1
109263	.	.	.	.	.	.	1	2	5
109264	.	.	.	.	.	.	2	13	1
109265	.	.	.	.	.	.	1	2	3
109266	2	2	2	.	.	.	2	29	5
109267	2	2	2	.	.	.	2	21	2
109268	.	.	.	.	.	.	2	18	3
109269	.	.	.	.	.	.	1	2	2
109270	.	.	.	.	.	.	2	11	4
109271	2	2	2	.	.	.	1	49	3
109273	2	2	2	.	.	.	1	36	3
109274	2	2	2	.	.	.	1	68	5
109275	.	.	.	.	.	.	1	6	3
109277	.	.	.	.	.	.	2	12	1
109278	.	.	.	.	.	.	2	6	2
109279	.	.	.	.	.	.	2	17	3
109280	.	.	.	.	.	.	2	2	1
109282	2	2	2	.	.	.	1	76	3
109283	1	1	1	48	48	48	1	52	5
109284	2	2	2	.	.	.	2	44	1
109285	.	.	.	.	.	.	1	16	4
109286	2	2	2	.	.	.	2	33	5
109287	.	.	.	.	.	.	2	11	3
109288	.	.	.	.	.	.	1	5	2
109289	.	.	.	.	.	.	1	2	5
109290	2	2	2	.	.	.	2	68	4

SAS code – One-way Frequencies and Crosstabs

`/* One-Way Frequencies */`

```
Proc Freq Data = MCQ_ALL;  
  Tables MCQ160B MCQ160C MCQ160E MCD180B MCD180C  
         MCD180E RIAGENDR RIDAGEYR RIDRETH1/ List Missing;  
  
  Format RIAGENDR Gndr. RIDAGEYR Age. RIDRETH1 RaceEth.  
         MCQ160B MCQ160C MCQ160E AnsKey.;  
  
  Title1 'Medical Conditions Questionnaire Variables';  
  Title2 'UnWeighted One-way Frequencies';  
Run;
```

`/* CrossTab (Two-way) Frequencies */`

```
Proc Freq Data = MCQ_ALL;  
  Tables MCQ160B*RIAGENDR MCQ160B*RIDAGEYR  
         MCQ160C*RIAGENDR MCQ160C*RIDAGEYR  
         MCQ160E*RIAGENDR MCQ160E*RIDAGEYR /  
         List Missing;  
  
  Format RIAGENDR Gndr. RIDAGEYR Age.  
         MCQ160B MCQ160C MCQ160E AnsKey.;  
  
  Title2 'UnWeighted Crosstab Frequencies';  
Run;
```

SAS Output – One-way Frequency Tables unweighted

RIAGENDR	Frequency	Percent	Cumulative Frequency
Male	7429	49.57	7429
Female	7557	50.43	14986

Age in years at screening				
RIDAGEYR	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0-24	6439	42.97	6439	42.97
25-34	1433	9.56	7872	52.53
35-44	1414	9.44	9286	61.96
45-54	1452	9.69	10738	71.65
55-64	1837	12.26	12575	83.91
65-74	1340	8.94	13915	92.85
75-79	389	2.60	14304	95.45
80+	682	4.55	14986	100.00

Race/Hispanic origin				
RIDRETH1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Mexican American	1899	12.67	1899	12.67
Other Hispanic	1468	9.80	3367	22.47
Non-Hispanic White	5042	33.64	8409	56.11
Non-Hispanic Black	3974	26.52	12383	82.63
Other Race - Incl MultiRacial	2603	17.37	14986	100.00

Ever told you had heart attack			
MCQ160E	Frequency	Percent	Cumulative Freq
Missing	5754	38.40	5754
Yes	432	2.88	6186
No	8786	58.63	14972
Don't Know	14	0.09	14986

Ever told had congestive heart failure				
MCQ160B	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Missing	5754	38.40	5754	38.40
Yes	361	2.41	6115	40.80
No	8848	59.04	14963	99.85
Don't Know	23	0.15	14986	100.00

Ever told you had coronary heart disease				
MCQ160C	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Missing	5754	38.40	5754	38.40
Yes	423	2.82	6177	41.22
No	8781	58.59	14958	99.81
Don't Know	28	0.19	14986	100.00

SAS Output – Crosstab Frequency Tables unweighted

MCQ160B	RIAGENDR	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Missing	Male	2950	19.69	2950	19.69
Missing	Female	2804	18.71	5754	38.40
Yes	Male	205	1.37	5959	39.76
Yes	Female	156	1.04	6115	40.80
No	Male	4260	28.43	10375	69.23
No	Female	4588	30.62	14963	99.85
Don't Know	Male	14	0.09	14977	99.94
Don't Know	Female	9	0.06	14986	100.00

MCQ160B	RIDAGEYR	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Missing	0-24	5754	38.40	5754	38.40
Yes	0-24	1	0.01	5755	38.40
Yes	25-34	3	0.02	5758	38.42
Yes	35-44	13	0.09	5771	38.51
Yes	45-54	37	0.25	5808	38.76
Yes	55-64	77	0.51	5885	39.27
Yes	65-74	111	0.74	5996	40.01
Yes	75-79	46	0.31	6042	40.32
Yes	80+	73	0.49	6115	40.80
No	0-24	684	4.56	6799	45.37
No	25-34	1430	9.54	8229	54.91
No	35-44	1401	9.35	9630	64.26
No	45-54	1413	9.43	11043	73.69
No	55-64	1752	11.69	12795	85.38
No	65-74	1224	8.17	14019	93.55
No	75-79	343	2.29	14362	95.84
No	80+	601	4.01	14963	99.85
Don't Know	45-54	2	0.01	14965	99.86
Don't Know	55-64	8	0.05	14973	99.91
Don't Know	65-74	5	0.03	14978	99.95
Don't Know	80+	8	0.05	14986	100.00

Analyzing the Survey data files

To properly analyze NHANES public-use data files, you need to incorporate the survey's sample design properties and use survey weights.

This is required for a valid statistical analysis involving estimating population parameters, performing hypothesis or significance testing, and to compute population estimates

You fill this requirement by using the relevant variables from the demographics file.

SAS has procedures specifically designed for survey data analysis. Examples of these procedures are PROC SURVEYFREQ, PROC SURVEYMEANS, PROC SURVEYREG.

Demonstration III

Demonstration III -SAS On-Demand for Academics (SODA)

I - Applying survey weights to the MCQ data file

II - Account for the Sample Design properties, Strata and PSU's

III – Generate weighted frequencies

SAS Code – PROC SURVEYFREQ

Specify Strata and Primary Sampling Unit variables

Specify subpopulations of interest in TABLE statement;

```
Proc Surveyfreq data=MCQ_ALL;  
  Strata SDMVSTRA;  
  Cluster SDMVPSU;  
  Weight WTINTPRP;  
  Table MCQ160B*RIAGENDR MCQ160B*RIDAGEYR  
        MCQ160C*RIAGENDR MCQ160C*RIDAGEYR / row col wchisq wllchisq nostd chisq chisq1;  
  
  Format RIAGENDR Gndr. RIDAGEYR Age. MCQ160B MCQ160C AnsKey.;  
run;
```

SAS Output – Weighted Frequencies

Data Summary	
Number of Strata	24
Number of Clusters	49
Number of Observations	14986
Sum of Weights	318804183

Table of MCQ160B by RIAGENDR						
MCQ160B	RIAGENDR	Frequency	Weighted Frequency	Percent	Row Percent	Column Percent
Yes	Male	205	3528572	1.4675	56.1246	3.0495
	Female	156	2758456	1.1472	43.8754	2.2114
	Total	361	6287028	2.6147	100.0000	
No	Male	4260	111953219	46.5600	47.8891	96.7543
	Female	4588	121822706	50.6646	52.1109	97.6608
	Total	8848	233775925	97.2245	100.0000	
Don't Know	Male	14	227026	0.0944	58.7330	0.1962
	Female	9	159513	0.0663	41.2670	0.1279
	Total	23	386540	0.1608	100.0000	
Total	Male	4479	115708818	48.1219		100.0000
	Female	4753	124740675	51.8781		100.0000
	Total	9232	240449493	100.0000		
Frequency Missing = 5754						

Table of MCQ160C by RIAGENDR						
MCQ160C	RIAGENDR	Frequency	Weighted Frequency	Percent	Row Percent	Column Percent
Yes	Male	302	6985705	2.9053	68.7102	6.0373
	Female	121	3181209	1.3230	31.2898	2.5503
	Total	423	10166914	4.2283	100.0000	
No	Male	4165	108593429	45.1627	47.2426	93.8506
	Female	4616	121269911	50.4347	52.7574	97.2176
	Total	8781	229863340	95.5973	100.0000	
Don't Know	Male	12	129684	0.0539	30.9332	0.1121
	Female	16	289556	0.1204	69.0668	0.2321
	Total	28	419240	0.1744	100.0000	
Total	Male	4479	115708818	48.1219		100.0000
	Female	4753	124740675	51.8781		100.0000
	Total	9232	240449493	100.0000		
Frequency Missing = 5754						

SAS Output – Weighted Freqs (con't)

MCQ160B by RIDAGEYR

Table of MCQ160B by RIDAGEYR						
MCQ160B	RIDAGEYR	Frequency	Weighted Frequency	Percent	Row Percent	Column Percent
Yes	0-24	1	7088	0.0029	0.1127	0.0320
	25-34	3	129918	0.0540	2.0665	0.2943
	35-44	13	250789	0.1043	3.9890	0.6276
	45-54	37	575306	0.2393	9.1507	1.4799
	55-64	77	1754139	0.7295	27.9009	3.9257
	65-74	111	1632193	0.6788	25.9613	5.3472
	75-79	46	803438	0.3341	12.7793	9.1301
	80+	73	1134157	0.4717	18.0396	10.0297
	Total	361	6287028	2.6147	100.0000	
No	0-24	684	22148843	9.2114	9.4744	99.9680
	25-34	1430	44009493	18.3030	18.8255	99.7057
	35-44	1401	39711942	16.5157	16.9872	99.3724
	45-54	1413	38280809	15.9205	16.3750	98.4704
	55-64	1752	42721658	17.7674	18.2746	95.6087
	65-74	1224	28856508	12.0011	12.3437	94.5363
	75-79	343	7996479	3.3256	3.4206	90.8699
	80+	601	10050193	4.1798	4.2991	88.8773
	Total	8848	233775925	97.2245	100.0000	
Don't Know	0-24	0	.	.	.	.
	25-34	0	.	.	.	.
	35-44	0	.	.	.	.
	45-54	2	19321	0.0080	4.9984	0.0497
	55-64	8	208063	0.0865	53.8272	0.4656
	65-74	5	35561	0.0148	9.1997	0.1165
	75-79	0	.	.	.	.
	80+	8	123595	0.0514	31.9747	1.0930
	Total	23	386540	0.1608	100.0000	
Total	0-24	685	22155931	9.2144		100.0000
	25-34	1433	44139412	18.3570		100.0000
	35-44	1414	39962731	16.6200		100.0000
	45-54	1452	38875436	16.1678		100.0000
	55-64	1837	44683861	18.5835		100.0000
	65-74	1340	30524261	12.6947		100.0000
	75-79	389	8799917	3.6598		100.0000
	80+	682	11307945	4.7028		100.0000
	Total	9232	240449493	100.0000		
Frequency Missing = 5754						

SAS Output – Weighted Freqs (con't)

MCQ160C by RIDAGEYR

Table of MCQ160C by RIDAGEYR						
MCQ160C	RIDAGEYR	Frequency	Weighted Frequency	Percent	Row Percent	Column Percent
Yes	0-24	1	34478	0.0143	0.3391	0.1556
	25-34	1	8968	0.0037	0.0882	0.0203
	35-44	9	454541	0.1890	4.4708	1.1374
	45-54	33	606463	0.2522	5.9651	1.5600
	55-64	94	3123182	1.2989	30.7191	6.9895
	65-74	131	3237497	1.3464	31.8435	10.6063
	75-79	54	1041081	0.4330	10.2399	11.8306
	80+	100	1660702	0.6907	16.3344	14.6862
No	Total	423	10166914	4.2283	100.0000	
	0-24	683	22068950	9.1782	9.6009	99.6074
	25-34	1431	44121159	18.3494	19.1945	99.9586
	35-44	1404	39493676	16.4249	17.1814	98.8263
	45-54	1418	38240222	15.9036	16.6361	98.3660
	55-64	1737	41484402	17.2529	18.0474	92.8398
	65-74	1205	27255305	11.3351	11.8572	89.2906
	75-79	330	7682187	3.1949	3.3421	87.2984
Don't Know	80+	573	9517439	3.9582	4.1405	84.1660
	Total	8781	229863340	95.5973	100.0000	
	0-24	1	52503	0.0218	12.5234	0.2370
	25-34	1	9285	0.0039	2.2147	0.0210
	35-44	1	14514	0.0060	3.4619	0.0363
	45-54	1	28751	0.0120	6.8579	0.0740
	55-64	6	76276	0.0317	18.1939	0.1707
	65-74	4	31459	0.0131	7.5038	0.1031
Total	75-79	5	76649	0.0319	18.2828	0.8710
	80+	9	129803	0.0540	30.9615	1.1479
	Total	28	419240	0.1744	100.0000	
	0-24	685	22155931	9.2144		100.0000
	25-34	1433	44139412	18.3570		100.0000
	35-44	1414	39962731	16.6200		100.0000
	45-54	1452	38875436	16.1678		100.0000
	55-64	1837	44683861	18.5835		100.0000
	65-74	1340	30524261	12.6947		100.0000
	75-79	389	8799917	3.6598		100.0000
	80+	682	11307945	4.7028		100.0000
	Total	9232	240449493	100.0000		
Frequency Missing = 5754						

Part II - MCBS

MCBS

Medicare Current Beneficiaries Survey

Medicare Current Beneficiaries Survey (MCBS)

The Survey is conducted by CMS – The Centers for Medicare and Medicaid Services.

CMS is part of the Department of Health and Human Services (HHS)

MCBS – continuous, multi-purpose longitudinal survey

Nationwide Survey of Medicare beneficiaries

- Medicare beneficiaries aged 65 and over.
- Medicare beneficiaries under 65 with disabilities

History of the MCBS

MCBS was implemented in 1991 to serve as a source of information for administering Medicare, estimating healthcare expenditures and on beneficiaries health and well-being.

MCBS was implemented in response to the rapid increase in healthcare costs and an increase in the elderly population with disabilities.

The MCBS has been continuously conducted since 1991

The first interview was conducted in September 1991.

Purpose of the MCBS

MCBS is designed to aid CMS in administering, monitoring and evaluating the Medicare program.

The primary focus of the MCBS is the impact of Medicare on beneficiaries.

MCBS provides important information on beneficiaries which is not otherwise collected and recorded in Medicare administrative data (claims files).

It plays an essential role in the monitoring of Medicare beneficiaries health status.

Data Collection Procedures

Interviews are conducted in households and facilities using CAPI – Computer Assisted Personal Interviews

During the Covid-19 Pandemic, interviewing and data collection were exclusively performed by phone interviews in 2020 and most of 2021.

Post-pandemic - the survey is administered using a combination of in-person and phone interviewing.

CMS contracts with survey research organizations to conduct and administer the survey

- WESTAT
- NORC (National Opinion Research Center).

Data Collection Procedures (con't)

Data is collected and released in three segments per year; Fall, Winter and Summer

Fall segment

- Contains the full sample of survey respondents.

Winter Segment

- Contains a smaller subset of survey respondents.

Summer Segment

- Contains the smallest subset of survey respondents

Accessing MCBS Data

MCBS Limited Data Set (LDS)

- Requires a data use agreement (DUA)
- Contains Protected Health Information (PHI)
- Must submit an application to RESDAC – Research Data Assistance Center.

MCBS Survey File Public Use File (PUF's)

- No data use agreement required
- Applies de-identification protections
- Contains no protected health information (PHI)
- Can be downloaded from the CMS MCBS website.

MCBS Interviews

Two Types of Interviews

Community – Conducted within beneficiaries or respondent's residences or homes

Facility – Conducted within healthcare facilities where beneficiaries are receiving services.

MCBS Limited Data Set – Includes data Collected from Community and Facility Interviews

MCBS Survey File PUF – Includes data Collected from Community Interviews only.

Contents of the survey data files

Access to Care information

- Includes many questions about Insurance coverage and beneficiaries sources of payment.
- Beneficiaries knowledge of Medicare
- Attitudes towards Medicare
- Satisfaction towards Medicare
- Demographic data

Cost and Use information

- Mainly includes questions about healthcare utilization.

Sample Design

Rotating Panel Sample Design

MCBS Panel

- Annual statistical sample of all Medicare enrollees.
- Interviewed up to 3 times a year (Fall, Winter, Summer)

Panels are retired at the end of Winter round

- Outgoing panel

New panel is added at the start of Fall round

- Incoming panel

Sample Design (con't)

Three-stage Cluster Sample Design

Primary Sampling Units (PSU's)

- Major geographic areas
- Metropolitan areas or groups of rural counties

Secondary Sampling Units (SSU's)

- Census tracts
- Groups of Census tracts

Ultimate Sampling Units (USU's)

- Medicare beneficiaries

Sample Design (con't)

Within the Secondary Sampling Unit, a sample of Medicare beneficiaries is drawn and selected.

The sample is drawn from Medicare administrative enrollment data files.

The sample is a randomly selected cross-section of all beneficiaries.

- Beneficiaries had to be enrolled in Part A or Part B at some point during the calendar year.
- Beneficiaries had to be residing within the continental united states (48 states)

Test your knowledge

Primary Sampling Units of the MCBS are which of the following?

- A. Major Geographic Areas
- B. Metropolitan Areas
- C. Census Tracts
- D. Both A and B

MCBS Survey File PUF

For the Survey PUF, Medicare beneficiaries are selected as part of one of two groups of beneficiaries

Ever enrolled

- beneficiaries who were ever enrolled in Medicare at any point during the calendar year.

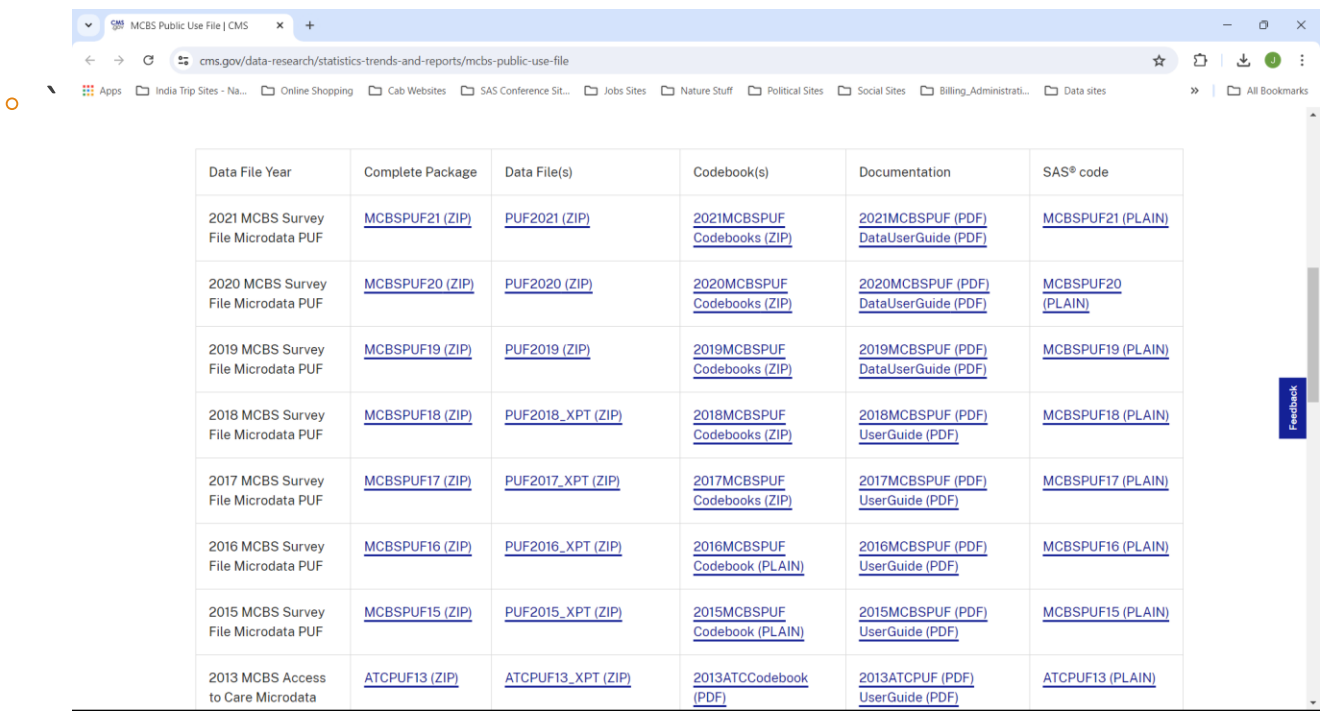
Continuously enrolled

- beneficiaries who were enrolled in the Medicare program throughout the calendar year

There is overlap between the two groups.

Accessing the MCBS Survey File PUF

<https://www.cms.gov/data-research/statistics-trends-and-reports/mcbs-public-use-file>



Data File Year	Complete Package	Data File(s)	Codebook(s)	Documentation	SAS® code
2021 MCBS Survey File Microdata PUF	MCBSPUF21 (ZIP)	PUF2021 (ZIP)	2021MCBSPUF Codebooks (ZIP)	2021MCBSPUF (PDF) DataUserGuide (PDF)	MCBSPUF21 (PLAIN)
2020 MCBS Survey File Microdata PUF	MCBSPUF20 (ZIP)	PUF2020 (ZIP)	2020MCBSPUF Codebooks (ZIP)	2020MCBSPUF (PDF) DataUserGuide (PDF)	MCBSPUF20 (PLAIN)
2019 MCBS Survey File Microdata PUF	MCBSPUF19 (ZIP)	PUF2019 (ZIP)	2019MCBSPUF Codebooks (ZIP)	2019MCBSPUF (PDF) DataUserGuide (PDF)	MCBSPUF19 (PLAIN)
2018 MCBS Survey File Microdata PUF	MCBSPUF18 (ZIP)	PUF2018_XPT (ZIP)	2018MCBSPUF Codebooks (ZIP)	2018MCBSPUF (PDF) UserGuide (PDF)	MCBSPUF18 (PLAIN)
2017 MCBS Survey File Microdata PUF	MCBSPUF17 (ZIP)	PUF2017_XPT (ZIP)	2017MCBSPUF Codebooks (ZIP)	2017MCBSPUF (PDF) UserGuide (PDF)	MCBSPUF17 (PLAIN)
2016 MCBS Survey File Microdata PUF	MCBSPUF16 (ZIP)	PUF2016_XPT (ZIP)	2016MCBSPUF Codebook (PLAIN)	2016MCBSPUF (PDF) UserGuide (PDF)	MCBSPUF16 (PLAIN)
2015 MCBS Survey File Microdata PUF	MCBSPUF15 (ZIP)	PUF2015_XPT (ZIP)	2015MCBSPUF Codebook (PLAIN)	2015MCBSPUF (PDF) UserGuide (PDF)	MCBSPUF15 (PLAIN)
2013 MCBS Access to Care Microdata	ATCPUF13 (ZIP)	ATCPUF13_XPT (ZIP)	2013ATCCodebook (PDF)	2013ATCPUF (PDF) UserGuide (PDF)	ATCPUF13 (PLAIN)

The Survey PUF can be downloaded from the CMS Website.

MCBS Survey File PUF Filetypes

The files can be downloaded as zip files as one complete package or as individual files.

The package contains the following individual files.

Description	Filename	File Format
Data files	PUF2020	SAS Transport and CSV Formats
Codebook	MCBSPUF_2020_1_fall	Text File Format
Documentation	2020MCBSDataUserGuide	PDF Format
SAS Code	MCBSPUF2020	Text File Format

MCBS Survey File PUF – File Segments

The MCBS Survey File PUF for a specific calendar year contains 3 data files

- One file for each segment (Fall, Summer, Winter).

Each annual PUF contains 3 separate codebooks

- One codebook for each segment (Fall, Summer, and Winter).

MCBS – Code and data specs

The SAS Code contained in the zip file reads and converts the SAS transport files.

It also creates formats for all the variables on the 3 SAS data sets and stores them in a format library. Then the code applies the formats to each of the variables.

The data file for the fall segment contains demographic variables

There's no need to create a SAS program to read and convert the files, create and apply the formats, and merge the files with a demo file to access the demographic variables.

Data Files – Types of variables - by Segment

Fall Segment		
Variable Prefix	Description	Number of Variables
ADM_	Administrative Data	22
INS_	Insurance status, coverage and type	11
INT_	Interview Characteristics	2
DEM_	Age, Sex and Race Groups	11
ACC_	Access to, use, and satisfaction with health care	19
HLT_	Health conditions and limitations in activities of daily living	88
PRV_	Preventive care and physical activity	28
RSK_	Health behavior risk factors	6
FAL_	Falls	12
HOU_	Housing characteristics	30
MA_	Medicare Advantage supplement	6
PUFF	PUF ever enrolled weights (fall)	101
TOTAL		336

Winter Segment		
Variable Prefix	Description	Number of Variables
ACW_	Access to care (winter round)	86
KNW	Medicare Knowledge	12
PUFW	PUF Ever enrolled weights (winter)	101
TOTAL		199

Summer Segment		
Variable Prefix	Description	Number of Variables
CVS_	Covid_19 Vaccine Update	2
FIS	Food Insecurity	4
RXS	Prescription medication use and access	29
PUFS	PUF ever enrolled weights (summer)	101
TOTAL		136

Demonstration I

Demonstration I - SAS On Demand for Academics (SODA)

Step 1 – Convert the SAS Transport files into SAS data sets.

Step 2 – Define formats and create a format Library.

Step 3 – Apply variables formats and labels. Run PROC CONTENTS

SAS Code – Importing and converting

```
%Let Datadir = /home/iyenj/Training_Course_Files/Natl_Health_Survey_data/MCBS/Data;

LIBNAME SDS "&Datadir";

LIBNAME FMT_L "/home/iyenj/Training_Course_Files/Natl_Health_Survey_data/MCBS/Formats";

FILENAME FALL_F "&Datadir./PUF2020_1_FALL.xpt";
FILENAME WINTER_F "&Datadir./PUF2020_2_WINTER.xpt";
FILENAME SUMMER_F "&Datadir./PUF2020_3_SUMMER.xpt";

PROC CIMPORT DATA=FALL_F INFILE=FALL_F;
PROC CIMPORT DATA=WINTER_F INFILE=WINTER_F;
PROC CIMPORT DATA=SUMMER_F INFILE=SUMMER_F;
RUN;

OPTIONS FMTSEARCH=(FMT_L FMT_L.FORMATS);
```

Proc Contents – Fall Segment

Data Set Name	SDS.PUF2020_1_FALL	Observations	13916
Member Type	DATA	Variables	339
Engine	V9	Indexes	0
Created	08/10/2024 13:43:46	Observation Length	2712
Last Modified	08/10/2024 13:43:46	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	291
First Data Page	1
Max Obs per Page	48
Obs in First Data Page	26
Number of Data Set Repairs	0
Filename	/home/iyenj/Training_Course_Files/Natl_Health_Survey_data/MCBS/Data/puf2020_1_fall.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	12889241113
Access Permission	rw-r--r--
Owner Name	iyenj
File Size	37MB
File Size (bytes)	38273024

Proc Contents – Fall Segment (con’t)

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
54	ACC COLAGNCY	Num	8	YESFMT.	Collection agency due to medical bills
52	ACC HCDELAY	Num	8	YESFMT.	Last year ever delay in care due to cost
51	ACC HCTRBMY	Num	8	INDPUF.	Troub getting care: cost, insuff. money
50	ACC HCTROUBL	Num	8	YESFMT.	Has SP had troub get needed health care?
57	ACC MCAVAIL	Num	8	SATFFMT.	Satisfaction: medical care night/wkend
8	ADM DUAL FLAG YR	Num	8	DUALFMT.	Dual eligibility flag for the year
9	ADM FFS FLAG YR	Num	8	FFSFMT.	Fee-for-Service flag for the year
21	ADM H ACTDAY	Num	8	IPDAYFMT.	Number IP covered days current yr (FFS)
20	ADM H ACTSTY	Num	8	IPSTYFMT.	Number inpatient stays current yr (FFS)
39	DEM AGE	Num	8	AGE2GRP.	Age group
48	DEM CBSA	Num	8	METRO.	Metro residence designated by CBSA
43	DEM EDU	Num	8	EDU.	Highest grade completed
45	DEM INCOME	Num	8	INCOM25F.	Income group
46	DEM INCSRCE	Num	8	INCSRCEF.	Source of income data
196	FAL FALBRUIS	Num	8	INDPUF.	Bruise from fall
198	FAL FALCDOH	Num	8	INDPUF.	Concuss/dislocate/other injury from fall
197	FAL FALCUT	Num	8	INDPUF.	Cut from fall
194	FAL FALFRACT	Num	8	INDPUF.	Fracture from fall
191	FAL FALLANY	Num	8	YESFMT.	Fallen down (past year)
81	HLT ALZDEM	Num	8	YESFMT.	Ever told Alzheimer's or dx of dementia
111	HLT BMI CAT	Num	8	BMIFMT.	BMI categories (kg/m^2)
109	HLT COMPHLTH	Num	8	COMPFMT.	General health compared to one year ago
137	HLT DASKRIDE	Num	8	YESFMT.	Ask others for rides
138	HLT DGIVUPDR	Num	8	YESNOMOB.	Given up driving altogether
116	HLT DIACTRLP	Num	8	DIACTRPF.	Blood sugar well controlled
117	HLT DIAHYPO	Num	8	YESFMT.	Ever experience hypoglycemia
112	HLT DIAINSUL	Num	8	YESFMT.	Takes insulin
113	HLT DIAMEDS	Num	8	YESFMT.	Takes prescribed diabetes pill/medicine
118	HLT DIAMNGE	Num	8	YESFMT.	Had diabetes self management course
228	HOU CAREPART	Num	8	YESFMT.	Provide Care in Diff Part of Housing

210	HOU DWELLING	Num	8	DWELFMT.	Description of Home
206	HOU D COMPHH	Num	8	COMPHH.	Household Composition Code
208	HOU D HHGE50	Num	8	HHLT.	Num in HH 50 and Over (May Include SP)
209	HOU D HHLT18	Num	8	HHLT.	Num <18 related to SP (May Include SP)
30	INS D MADV	Num	8	D MADV.	Enrolled in Medicare Advantage
31	INS D PRIVNUM	Num	8	YESFMT.	One or more private plans (excl. RDS)
32	INS D PRIVNUMC	Num	8	PRIVNUM.	Number of private plans (excl. RDS)
33	INS D PVESI	Num	8	YESFMT.	Has employer-sponsored insurance
34	INS D PVSELF	Num	8	YESFMT.	Has self-purchased private insurance
37	INT LANG	Num	8	LANG.	Language of interview
38	INT SPPROXY	Num	8	PROXY.	Self - respondent or proxy
233	MA MADVDENT	Num	8	YESFMT.	Mcare Adv plan covers dental
234	MA MADVEYE	Num	8	YESFMT.	Mcare Adv plan covers eye exams
235	MA MADVNH	Num	8	YESFMT.	Mcare Adv plan covers nursing home
183	PRV BASKORAL	Num	8	YESFMT.	Ever had oral cancer exam
158	PRV BCTAKEN	Num	8	TIME4FMT.	How long since blood cholesterol taken
162	PRV BLOODTST	Num	8	YESFMT.	PSA prostate blood test (past year)
157	PRV BPTAKEN	Num	8	TIME2FMT.	How long since BP taken
172	PRV COLCARD	Num	8	YESFMT.	Sent card in for most recent FOBT
240	PUFF001	Num	8	CONTIN.	PUF Fall Ever Enrolled rep wt 1
241	PUFF002	Num	8	CONTIN.	PUF Fall Ever Enrolled rep wt 2
242	PUFF003	Num	8	CONTIN.	PUF Fall Ever Enrolled rep wt 3
243	PUFF004	Num	8	CONTIN.	PUF Fall Ever Enrolled rep wt 4
244	PUFF005	Num	8	CONTIN.	PUF Fall Ever Enrolled rep wt 5
245	PUFF006	Num	8	CONTIN.	PUF Fall Ever Enrolled rep wt 6
246	PUFF007	Num	8	CONTIN.	PUF Fall Ever Enrolled rep wt 7
247	PUFF008	Num	8	CONTIN.	PUF Fall Ever Enrolled rep wt 8
248	PUFF009	Num	8	CONTIN.	PUF Fall Ever Enrolled rep wt 9
249	PUFF010	Num	8	CONTIN.	PUF Fall Ever Enrolled rep wt 10
239	PUFFWGT	Num	8	CONTIN.	PUF Fall Ever Enrolled Full Sample Wt
1	PUF ID	Num	8	PUFFMT.	Unique PUF Identification Number
189	RSK ALCDAY	Num	8	DRINKSPD.	Number of alcohol drinks/day (past yr)
187	RSK ALCLIFE	Num	8	YESFMT.	Has had at least 12 drinks in life
188	RSK D ALCFRQ	Num	8	DRINKDAY.	Number of days/month drink alcohol
185	RSK EVRSMKAL	Num	8	YESFMT.	Ever smoke cigarette/cigar/tobacco
190	RSK FFDRNK	Num	8	DRINKSPM.	Number days/mo has 5/4+ drinks (past yr)
186	RSK SMKKNOWAL	Num	8	YESFMT.	Currently smoke cigarette/cigar/tobacco
2	SURVEYR	Num	8	SVYRFMT.	Survey year
3	VERSION	Num	8	VERSFM.	Version number

MCBS Data files – Key variables

PUF_ID – Unique Beneficiary Number. Can be used to link the 3 segment files within a specific year.

DEM_SEX – Gender of the Respondent / Beneficiary

DEM_INCOME – Income of the Respondent / Beneficiary

DEM_RACE – Race of the Respondent / Beneficiary

DEM_AGE – Age of the Respondent / Beneficiary

ADM_H_INPSW – 1+ Inpatient Discharge current year

ADM_H_OUTSW – 1+ Outpatient Stays current year

ADM_H_SNFSW – 1+ SNF Admission current year

ADM_H_ACTSTAY – Number of Inpatient Stays current year

Demonstration II

Demonstration II - SAS On Demand for Academics (SODA)

Step 1 – Manipulate the Fall segment PUF File. Select the variables we need.

Step 2 – Produce one-way frequencies for Key Variables on demographics and health care utilization

SAS Code – One-Way Frequencies

```
Libname SDS '/home/iyenj/Training_Course_Files/Natl_Health_Survey_data/MCBS/Data';
```

```
        *Pull over necessary variables from Fall Segment;
Proc SQL;
  Select Name into : Varlist Separated By ' '
  From Dictionary.Columns
  Where Libname='SDS' and Memname='PUF2020_1_FALL' and
        Substr(Name, 1, 3) In ('DEM', 'ACC', 'PUFF', 'ADM');
Quit;
```

```
%Put &Varlist;
```

```
        *Select variables from SAS data set;
Data PUF2020_1_FALL;
  Set SDS.PUF2020_1_FALL (Keep=PUF_ID SURVEYYR &Varlist);
Run;
```

```
        *Print out sample listing of records;
Proc Print Data = PUF2020_1_FALL (Obs=25);
  ID PUF_ID;
  Var SURVEYYR ADM_H_ACTSTY ADM_H_INPSW ADM_H_OUTSW
      ADM_PARTD ADM_H_SNFSW DEM_AGE DEM_RACE DEM_INCOME
      DEM_SEX;
  Title'MCBS 2020 PUF FALL SEGMENT';
Run;
```

```
        *One-way Unweighted Frequencies;
Proc Freq Data = PUF2020_1_FALL;
  Tables ADM_H_ACTSTY ADM_H_INPSW ADM_H_OUTSW
      ADM_PARTD ADM_H_SNFSW DEM_AGE DEM_RACE
      DEM_INCOME DEM_SEX / List Missing;
  Title 'One-way Freqs - Fall Segment';
Run;
```

SAS Output – One-Way Frequencies

PUF Fall Segment –Demographic variables – One Way Freq

Gender				
DEM_SEX	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1:Male	6345	45.59	6345	45.59
2:Female	7571	54.41	13916	100.00

Income group				
DEM_INCOME	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1:<\$25,000	5125	36.83	5125	36.83
2:>=\$25,000	8791	63.17	13916	100.00

Race/ethnicity group				
DEM_RACE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1:Non-Hispanic white	10266	73.77	10266	73.77
2:Non-Hispanic black	1393	10.01	11659	83.78
3:Hispanic	1418	10.19	13077	93.97
4:Other	839	6.03	13916	100.00

SAS Output – One-way Freqs

Fall PUF – Admin variables – Unweighted Frequencies

1+ SNF admission current yr (FFS)				
ADM_H_SNFSW	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1:Yes	175	1.26	175	1.26
2:No	13741	98.74	13916	100.00

1+ outpatient visit current yr (FFS)				
ADM_H_OUTSW	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1:Yes	5509	39.59	5509	39.59
2:No	8407	60.41	13916	100.00

Part D plan flag for the year				
ADM_PARTD	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1:Yes	11195	80.45	11195	80.45
2:No	2721	19.55	13916	100.00

1+ inpatient discharge current yr (FFS)				
ADM_H_INPSW	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1:Yes	1023	7.35	1023	7.35
2:No	12893	92.65	13916	100.00

Number inpatient stays current yr (FFS)				
ADM_H_ACTSTY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0:No stay	12933	92.94	12933	92.94
1:One stay	726	5.22	13659	98.15
2:Two stays	169	1.21	13828	99.37
3:Three stays	53	0.38	13881	99.75
4:Four or more stays	35	0.25	13916	100.00

Statistical analysis – Sample design properties

To produce statistically valid population estimates and other parameters, confidence intervals, and statistical significance tests, you need to take the sample design properties into account. The MCBS uses a complex stratified cluster sample.

Researchers need to use sample weights and replicate weights which are contained in the MCBS Public Use File. PROC SURVEYFREQ is the relevant SAS construct which allows referencing the sample design properties

Fall Segment File	
Type of Weight	Variable
Sample Weights	PUFFWGT
Replicate Weights	PUFF01 – PUFF100

Demonstration III

Demonstration III - SAS On Demand for Academics (SODA)

Step 1 – Use PROC SURVEYFREQ to generate one-way weighted frequency tables.

Apply survey weights and replicate weights.

Step 2 – Use PROC SURVEYFREQ to generate two-way weighted frequency tables.

Apply survey weights and replicate weights.

SAS Code – PROC SURVEYFREQ

```
*One-way Weighted Frequencies;
Proc SurveyFreq Data = PUF2020_1_FALL VARMETHOD = brr (fay=.30);
  Weight PUFFWGT;
  Repweight PUFF001-PUFF100;
  Table ADM_H_INPSW / row chisq lrchisq;
  Title 'One-way Weighted Frequencies';
  Title2 '1 or more inpatient discharge current year (FFS)';
Run;

*Two-way Crosstabulations - Weighted Frequencies;
Proc SurveyFreq Data = PUF2020_1_FALL VARMETHOD = brr (fay=.30);
  Weight PUFFWGT;
  Repweight PUFF001-PUFF100;
  Table ADM_H_INPSW*DEM_AGE / row chisq lrchisq;
  Title 'Two-way Crosstabulations - Weighted Frequencies';
  Title2 '1 or more inpatient discharge current year (FFS) By Age Group';
Run;
```

SAS Output – One-way Weighted frequencies

Data Summary	
Number of Observations	13916
Sum of Weights	59754631.2

Sample of Medicare Beneficiaries (Respondents) = 13916

Population of Medicare Beneficiaries = 59,754,631 (Population estimate)

Variance Estimation	
Method	BRR
Replicate Weights	PUF2020_1_FALL
Number of Replicates	100
Fay Coefficient	0.300

1+ inpatient discharge current yr (FFS)					
ADM_H_INPSW	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent
1:Yes	1023	4023554	148331	6.7335	0.2483
2:No	12893	55731077	189561	93.2665	0.2483
Total	13916	59754631	121409	100.0000	

Population of Medicare Beneficiaries with 1 or more inpatient discharges = 4,023,554 (6.73)

Weighted Percents are close to UnWeighted Percents

SAS Output – Two-way Weighted frequencies

Data Summary	
Number of Observations	13916
Sum of Weights	59754631.2

Variance Estimation	
Method	BRR
Replicate Weights	PUF2020_1_FALL
Number of Replicates	100
Fay Coefficient	0.300

Table of ADM_H_INPSW by DEM_AGE								
ADM_H_INPSW	DEM_AGE	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent
1:Yes	1:Age Group <65	178	602039	48502	1.0075	0.0813	14.9629	1.1202
	2:Age Group [65,75)	239	1531046	104566	2.5622	0.1745	38.0521	1.7023
	3:Age Group >=75	606	1890469	84187	3.1637	0.1414	46.9850	1.7943
	Total	1023	4023554	148331	6.7335	0.2483	100.0000	
2:No	1:Age Group <65	2151	7305886	59311	12.2265	0.0980	13.1092	0.0967
	2:Age Group [65,75)	4599	29471675	128763	49.3212	0.2034	52.8819	0.1750
	3:Age Group >=75	6143	18953516	123929	31.7189	0.1833	34.0089	0.1730
	Total	12893	55731077	189561	93.2665	0.2483	100.0000	
Total	1:Age Group <65	2329	7907925	41247	13.2340	0.0694		
	2:Age Group [65,75)	4838	31002722	86525	51.8834	0.1114		
	3:Age Group >=75	6749	20843984	89459	34.8826	0.1166		
	Total	13916	59754631	121409	100.0000			

Interpreting the output and results

Are there any significant findings in the output?

2.5% of beneficiaries (65-74) had 1 or more inpatient discharges.

3.18% of beneficiaries (75+) had 1 or more inpatient discharges.

Of those who had 1 or more inpatient discharges,

38% were from 65-75 and 46% were 75+

Part III – MEPS

MEPS

Medical Expenditure Panel Survey

Medical Expenditure Panel Survey

The Medical Expenditure Panel Survey (MEPS) is a set of large-scale national health care surveys.

MEPS is conducted and administered by the Agency for Healthcare Research and Quality (AHRQ).

AHRQ is a part of the US Dept. of Health and Human Services (HHS).

History of Medical Expenditure Surveys

Medical Expenditure surveys began in the 1970's

National Medical Care Expenditure Surveys (NMCES) - 1977

- The first medical expenditure survey
- 14,000 Households

National Medical Expenditure Survey (NMES) - 1987

- 16,000 Households

Origin of MEPS

1996

Medical Expenditure Panel Survey – Household Component – MEPS-HC

- Provides more timely information.
- Survey conducted continuously instead of once every 10 years.
- New panel or sample introduced every year.

Medical Expenditure Panel Survey – Insurance Component – MEPS-IC

- Based on the National Employer Health Insurance Survey (NEHIS)

MEPS-Household Component (HC)

In the MEPS Household Component, data is collected for families and individuals in selected communities.

The sample of households is drawn from a nationally representative subsample of households

The nationally representative subsample of households consists of households who participated in the prior year's National Health Interview Survey (NHIS).

The NHIS is a survey administered by the National Center on Health Statistics (NCHS).

MEPS-HC

MEPS Household Component collects data from each member of the household on the following topics.

Demographic Characteristics

Health Status

Charges

Access to Care

Health Insurance Coverage

Employment

Health Conditions

Use of Medical Services

Sources of Payments

Satisfaction with Care

Income

MEPS - Insurance Component

Medical Expenditure Panel Survey – Insurance Component.

Collects data on the health insurance plans offered by employers.

Draws a sample of public and private sector employers.

IC data files are not released to the public.

MEPS-Insurance Component

MEPS – Insurance Component collects information on

Number of private insurance plans

Plan Premiums

Eligibility requirements

Employer characteristics

Types of private insurance plans

Employer/employee contributions,

Plan Benefits

MEPS-Medical Provider Component

MEPS – Medical Provider Component (MPC)- A Survey of Health care Providers.

Collects data on hospitals, physicians, Home Health providers and pharmacies.

Uses providers identified by respondents of MEPS-HC.

Its purpose is to supplement and/or replace information received from MEPS-HC respondents

MEPS-MPC data files are not publicly available.

Collected data is incorporated in the MEPS-HC data files.

MEPS Sample Design

The MEPS Sample Design is dependent on the Sample Design of the National Health Interview Survey (NHIS).

This is because the MEPS-HC sample is a subsample of the households selected from the prior year NHIS sample.

NHIS Sample Design

- Stratified Multistage Sampling Design

MEPS – NHIS Sample Design

NHIS Sample Design

- Sampling Units – geographic Entities

Primary Sampling Units (PSU's)

- Counties or groups of counties

Secondary Sampling Units (SSU's)

- Areas with counties (blocks or groups of blocks)

Medical Expenditure Panel Survey

NHIS Panels

- SSU's are split up into 4 panels (panels 1 – 4).

Final Sampling Units

- Households are selected from SSU's and/or panels 1 – 4

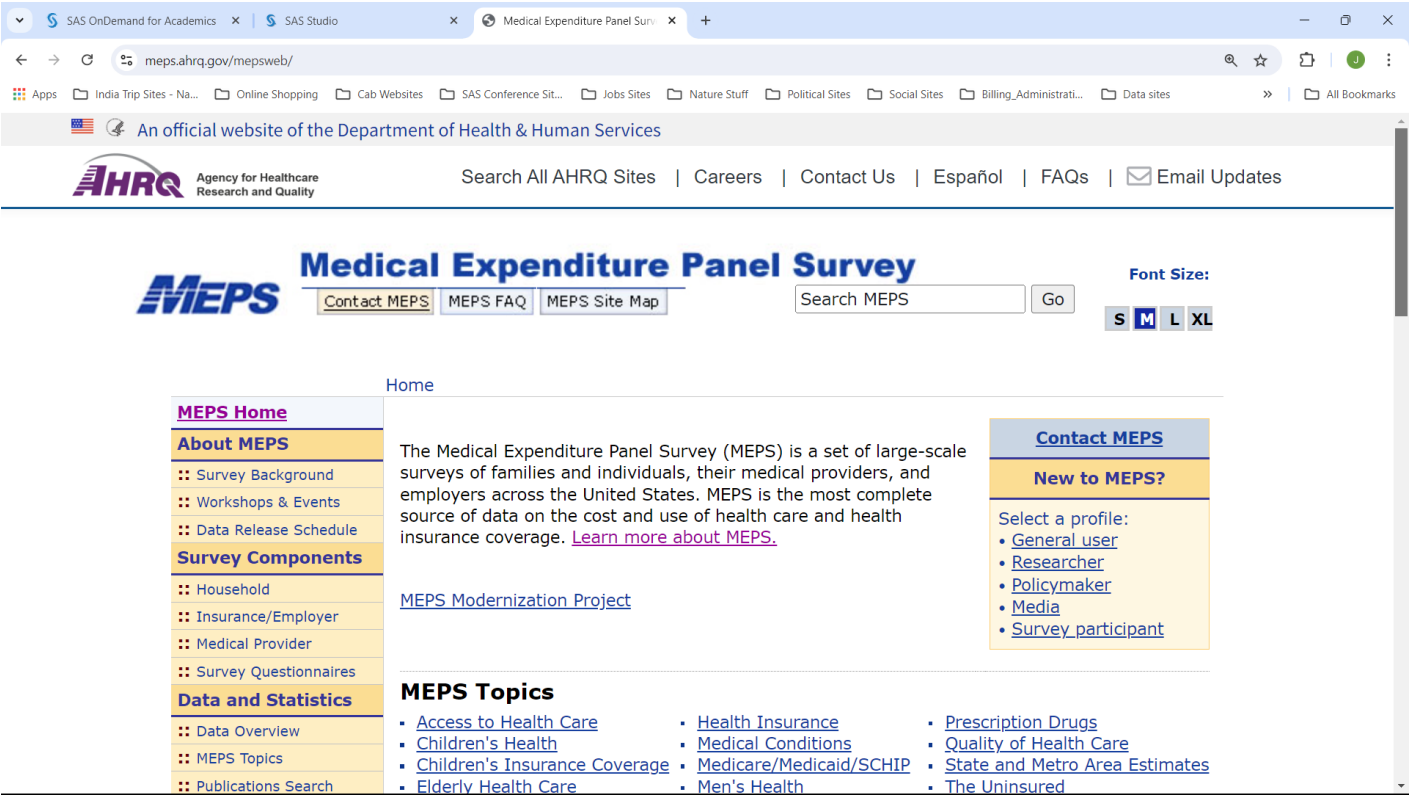
Households – Dwelling Units (DU's)

Families within Households – Reporting Units (RU's)

Individuals within RU's are interviewed.

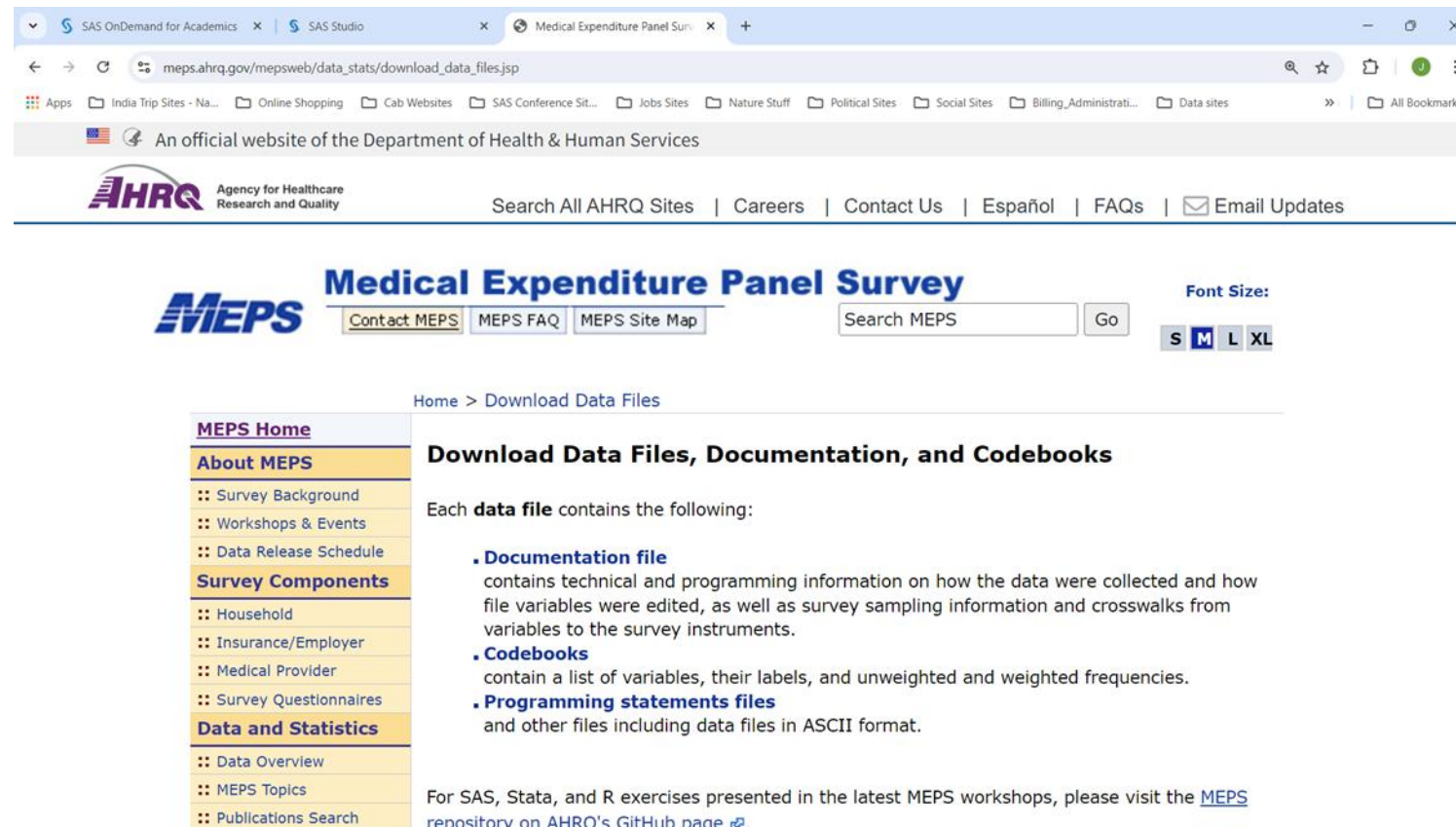
Minorities (Blacks and Latino persons) are disproportionately sampled (oversampled).

MEPS Home Page



www.meps.ahrq.gov

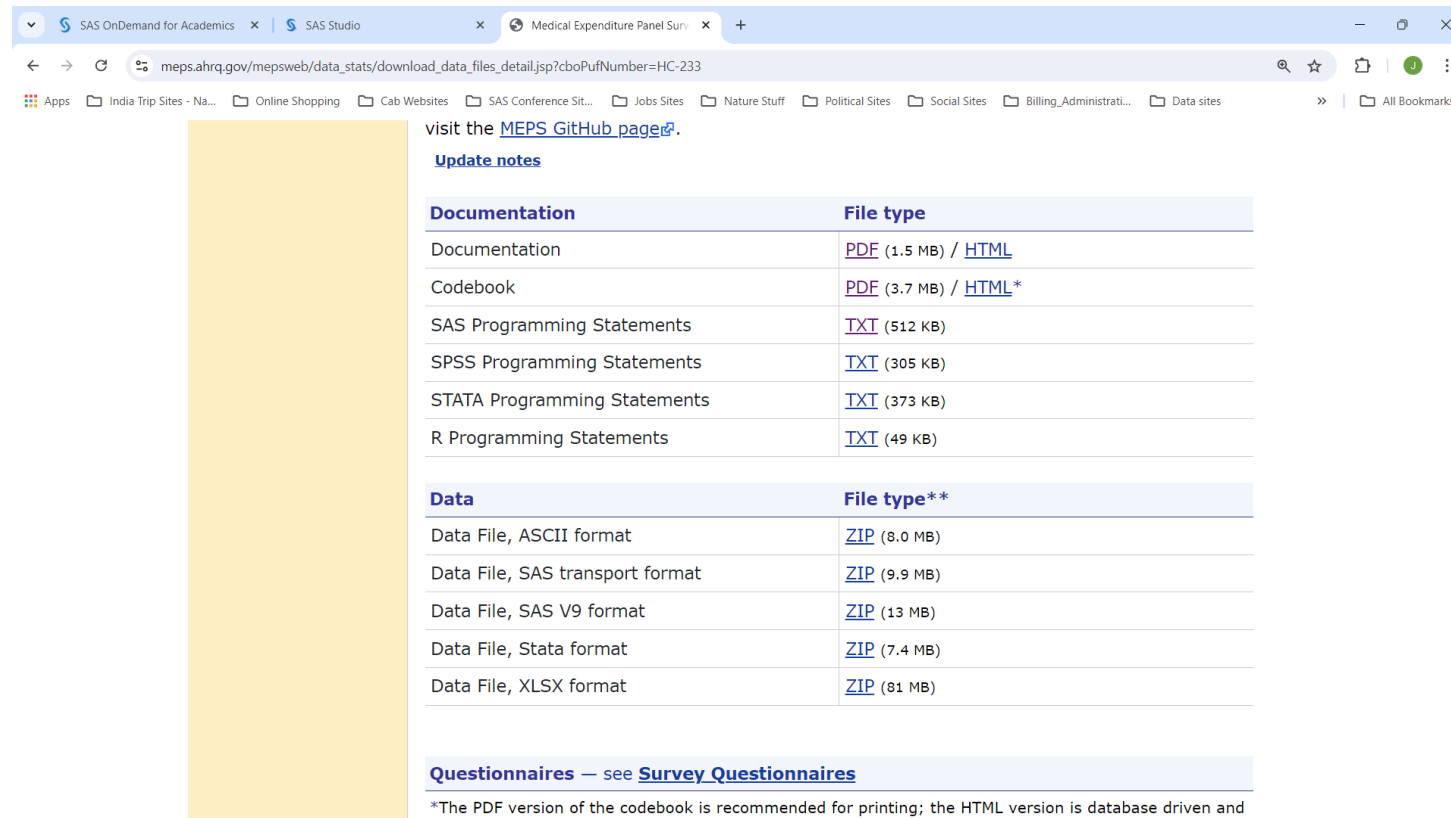
Accessing MEPS data files



You can download MEPS-HC data files directly from the website.

For the MEPS Household Component, along with the data file, you can download documentation, codebooks, and code used to extract the data.

MEPS data files and documentation



The screenshot shows a web browser window with the URL meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-233. The page content includes a yellow sidebar, a text prompt to visit the [MEPS GitHub page](#), and two tables. The first table, titled 'Documentation', lists various file types and their sizes. The second table, titled 'Data', lists data file formats and their sizes. At the bottom, there is a link to 'Questionnaires' and a note about the codebook.

visit the [MEPS GitHub page](#).

[Update notes](#)

Documentation	File type
Documentation	PDF (1.5 MB) / HTML
Codebook	PDF (3.7 MB) / HTML*
SAS Programming Statements	TXT (512 KB)
SPSS Programming Statements	TXT (305 KB)
STATA Programming Statements	TXT (373 KB)
R Programming Statements	TXT (49 KB)

Data	File type**
Data File, ASCII format	ZIP (8.0 MB)
Data File, SAS transport format	ZIP (9.9 MB)
Data File, SAS V9 format	ZIP (13 MB)
Data File, Stata format	ZIP (7.4 MB)
Data File, XLSX format	ZIP (81 MB)

[Questionnaires](#) — see [Survey Questionnaires](#)

*The PDF version of the codebook is recommended for printing; the HTML version is database driven and

Code to convert and format the data files is available in SAS, SPSS, STATA and R.

The data files are available in multiple formats also. End users can download data files in ASCII, SAS Transport, SAS V9, Stata, and Excel Formats.

MEPS Data Files

For the Household Component, there are many data files available for download. These are a few of the available files.

Data File	Filename	
The Full-Year Consolidated data file	MEPS HC-233	The five other files are subsets of the Full-Year Consolidated file.
Person Round Plan File	MEPS HC-232	
Medical Conditions File	MEPS HC-231	Their variables are also found on the consolidated file
Prescribed Medicines File	MEPS HC-229A	
Hospital Inpatient Visits File	MEPS HC-229D	
Outpatient Visits File	MEPS HC-229F	

MEPS – HC - Filesizes

File	Number of Variables
The Full-Year Consolidated data file	1488
Person Round Plan File	55
Medical Conditions File	32
Prescribed Medicines File	66
Hospital Inpatient Visits File	50
Outpatient Visits File	57

The six files are just a sample of the files available in the MEPS Household Component survey

MEPS Data Files – Key variables

PID – Person Number

DUID- Panel # + Encrypted Du Identifier

DUPERSID - Person Id (DUID + PID)

PERWT21F- Expenditure File Person Weight, 2021

VARPSU- Variance Estimation PSU, 2021

VARSTR- Variance Estimation Stratum, 2021

IPTC21X- Inpatient Total Chg For Event (IPFTC21X + IPDTC21X)

IPXP21X- Inpatient Total Exp For Event (IPFXP21X + IPDXP21X)

Demonstration I

Demonstration I – SAS On-Demand for Academics

- I. Reading and converting Full-Year Consolidated file – demographic variables only.
- II. Reading and converting the Hospital Inpatient Visits file

SAS Code – Reading and Converting data

```
Libname Puflib '/Home/Iyenj/Training_course_files/Natl_health_survey_data/Meps/Data';
Filename In1 '/Home/Iyenj/Training_course_files/Natl_health_survey_data/Meps/Sas Transport Files/H233.Ssp';

Options Fmtsearch=(puflib);

Proc CImport Data=work.H233 Infile=in1;
Run;

Proc Contents Data=puflib.H233;
    Title 'List Of Variables In Meps H233 Sas Dataset';
Run;

Proc Print Data=puflib.H233 (Obs=20);
    Var Age31x Age42x Age53x Age21x Agelast Dobmm Dobby Sex Racev1x Racev2x Raceax Racebx Racewx
        Racethx Hispanx Hispncat Marry31x Marry42x Marry53x Marry21x;
    Title1 'Meps Full Year Consolidated File - H233 - First 20 Observations';
Run;

Proc Freq Data=puflib.H233;
    Tables Sex Racev1x Racev2x Raceax Racebx Racewx Racethx
        Hispanx Hispncat Marry31x Marry42x Marry53x Marry21x / List Missing;
    Title1 'Meps Full Year Consolidated File - H233 - Frequency Tables';
Run;
```

Proc Contents – Hospital Inpatient File

Data Set Name	PUFLIB.H229D	Observations	2557
Member Type	DATA	Variables	50
Engine	V9	Indexes	0
Created	08/21/2024 22:02:44	Observation Length	416
Last Modified	08/21/2024 22:02:44	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	9
First Data Page	1
Max Obs per Page	314
Obs in First Data Page	290
Number of Data Set Repairs	0
Filename	/home/iyenj/Training_Course_Files/Natl_Health_Survey_data/MEPS/Data/h229d.sas7bdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	25830033137
Access Permission	rw-r--r--
Owner Name	iyenj
File Size	1MB
File Size (bytes)	1310720

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
18	ANYOPER	Num	8	ANYOPERF.	ANY OPERATIONS OR SURGERIES PERFORMED
19	DSCHPMED	Num	8	DSCHPMEDF.	MEDICINES PRESCRIBED AT DISCHARGE
1	DUID	Num	8	DUIDF.	PANEL # + ENCRYPTED DU IDENTIFIER
3	DUPERSID	Char	10	\$DUPERSIDF.	PERSON ID (DUID + PID)
15	EMERROOM	Num	4	EMERROOMF.	DID STAY BEGIN WITH EMERGENCY ROOM VISIT
6	ERHEVIDX	Char	16	\$ERHEVIDXF.	EVENT ID FOR CORRESPONDING EMER RM VISIT
5	EVENTRN	Num	8	EVENTRNF.	EVENT ROUND NUMBER
4	EVENTIDX	Char	16	\$EVENTIDXF.	EVENT ID
7	FFEEIDX	Char	12	\$FFEEIDXF.	FLAT FEE ID
20	FFIPTYPE	Num	8	FFIPTYPEF.	FLAT FEE BUNDLE
47	IMPFLAG	Num	8	IMPFLAGF.	IMPUTATION STATUS
11	IPBEGMM	Num	8	IPBEGMMF.	EVENT START DATE - MONTH
10	IPBEGYR	Num	8	IPBEGYRF.	EVENT START DATE - YEAR
37	IPDMD21X	Num	8	IPDMD21XF.	DOCTOR AMOUNT PAID, MEDICAID (IMPUTED)
36	IPDMR21X	Num	8	IPDMR21XF.	DOCTOR AMOUNT PAID, MEDICARE (IMPUTED)
41	IPDOF21X	Num	8	IPDOF21XF.	DOCTOR AMOUNT PAID, OTH FEDERAL (IMPUTED)
44	IPDOT21X	Num	8	IPDOT21XF.	DOCTOR AMOUNT PAID, OTH INSUR (IMPUTED)
38	IPDPV21X	Num	8	IPDPV21XF.	DOCTOR AMOUNT PAID, PRIV INSUR (IMPUTED)
35	IPDSF21X	Num	8	IPDSF21XF.	DOCTOR AMOUNT PAID, FAMILY (IMPUTED)
42	IPDSL21X	Num	8	IPDSL21XF.	DOCTOR AMOUNT PAID, STATE/LOC GOV (IMPUTED)
46	IPDTC21X	Num	8	IPDTC21XF.	TOTAL DOCTOR CHARGE (IMPUTED)
40	IPDTR21X	Num	8	IPDTR21XF.	DOCTOR AMOUNT PAID, TRICARE (IMPUTED)
39	IPDVA21X	Num	8	IPDVA21XF.	DOCTOR AMOUNT PAID, VETERANS/CHAMPVA (IMPUTED)
43	IPDWC21X	Num	8	IPDWC21XF.	DOCTOR AMOUNT PAID, WORKERS COMP (IMPUTED)
45	IPDXP21X	Num	8	IPDXP21XF.	DOCTOR SUM PAYMENTS IPDSF21X - IPDOT21X
13	IPENDMM	Num	8	IPENDMMF.	EVENT END DATE - MONTH
12	IPENDYR	Num	8	IPENDYRF.	EVENT END DATE - YEAR
25	IPFMD21X	Num	8	IPFMD21XF.	FACILITY AMOUNT PAID, MEDICAID (IMPUTED)
24	IPFMR21X	Num	8	IPFMR21XF.	FACILITY AMOUNT PAID, MEDICARE (IMPUTED)
29	IPFOF21X	Num	8	IPFOF21XF.	FACILITY AMOUNT PAID, OTH FEDERAL (IMPUTED)
32	IPFOT21X	Num	8	IPFOT21XF.	FACILITY AMOUNT PAID, OTH INSUR (IMPUTED)
26	IPFPV21X	Num	8	IPFPV21XF.	FACILITY AMOUNT PAID, PRIV INSUR (IMPUTED)
23	IPFSF21X	Num	8	IPFSF21XF.	FACILITY AMOUNT PAID, FAMILY (IMPUTED)
30	IPFSL21X	Num	8	IPFSL21XF.	FACILITY AMOUNT PAID, STATE/LOC GOV (IMPUTED)
34	IPFTC21X	Num	8	IPFTC21XF.	TOTAL FACILITY CHARGE (IMPUTED)
28	IPFTR21X	Num	8	IPFTR21XF.	FACILITY AMOUNT PAID, TRICARE (IMPUTED)
27	IPFVA21X	Num	8	IPFVA21XF.	FACILITY AMOUNT PAID, VETERANS/CHAMPVA (IMPUTED)
31	IPFWC21X	Num	8	IPFWC21XF.	FACILITY AMOUNT PAID, WORKERS COMP (IMPUTED)
33	IPFXP21X	Num	8	IPFXP21XF.	FACILITY SUM PAYMENTS IPFSF21X - IPFOT21X
22	IPTC21X	Num	8	IPTC21XF.	TOTAL CHG FOR EVENT (IPFTC21X + IPDTC21X)
21	IPXP21X	Num	8	IPXP21XF.	TOTAL EXP FOR EVENT (IPFXP21X + IPDXP21X)
9	MPCDATA	Num	8	MPCDATAF.	MPC DATA FLAG
14	NUMNIGHX	Num	8	NUMNIGHXF.	# OF NIGHTS IN HOSPITAL - EDITED/IMPUTED
8	PANEL	Num	8	PANELF.	PANEL NUMBER
48	PERWT21F	Num	8	PERWT21FF.	EXPENDITURE FILE PERSON WEIGHT, 2021
2	PID	Num	4	PIDF.	PERSON NUMBER
17	RSNINHOS	Num	8	RSNINHOSF.	REASON ENTERED HOSPITAL
16	SPECCOND	Num	8	SPECCONDF.	HOSPITAL STAY RELATED TO CONDITION
50	VARPSU	Num	8	VARPSUF.	VARIANCE ESTIMATION PSU, 2021
49	VARSTR	Num	8	VARSTRF.	VARIANCE ESTIMATION STRATUM, 2021

Test your knowledge

MEPS survey data files for the Household Component are not available in which of the following formats?

- A. SAS Transport File
- B. Excel
- C. ASCII
- D. SQL Server

MEPS Primary Keys

There are multiple identifiers which can be used as key variables to link the full-year consolidated file with the conditions and events files.

To link files, the appropriate variable to use is DUPERSID, which is the combination of the Dwelling Unit Identifier and the Person Identifier.

In the full-year file, each observation represents a unique respondent or sample person.

In the event and condition files, each observation represents a unique event or unique condition, with multiple events or conditions per person.

MEPS – data file structure

Full-Year Consolidated File – DUPERSID is a unique identifier

Hospital Inpatient Stays File – DUPERSID is not a unique identifier.

Different files are structured at different levels; respondent level vs. event level

We need to aggregate the hospital inpatient stays file at the person-level in order to link it to the full-year consolidated file.

Demonstration II

Demonstration II using SAS On-demand for Academics

1. Subset the Full-Year Consolidated File - H233 to select only demographic variables.
2. Aggregate the Hospital Inpatient file – H229D at the person level.
3. Merge Full-Year Consolidated file with Hospital Inpatient file.

SAS Code – Merging data files

*Demographic variables from Full Year Consol File;

```
Proc Sort Data = PUFLIB.h233;
```

```
  Out = h233_DEMO
```

```
  (KEEP=DUPERSID AGE21X SEX RACEV1X RACETHX HISPNCAT MARRY21X)
```

```
  By DUPERSID;
```

```
Run;
```

```
Proc Sort Data = PUFLIB.H229D;
```

```
  Out = H229D
```

```
  (KEEP=DUPERSID NUMNIGHX IPFXP21X IPFTC21X PERWT21F VARSTR VARPSU)
```

```
  By DUPERSID;
```

```
Run;
```

*Aggregate Hosp Inpatient File at the Person Level;

```
Data H229D_Agg;
```

```
  Set H229D;
```

```
  By DUPERSID;
```

```
  If First.DUPERSID Then Do;
```

```
    Num_Nights=0; IP_Fac_AmtPaid=0; IP_Fac_Chgs=0;
```

```
  End;
```

```
  Num_Nights+NUMNIGHX; *Number of Nights in Hospital;
```

```
  IP_Fac_AmtPaid+IPFXP21X; *Total Facility AmtPaid;
```

```
  IP_Fac_Chgs+IPFTC21X; *Total Facility Charges;
```

```
  If Last.DUPERSID;
```

```
Run;
```

```
Data HOSP_IN_DEMO HINP_NOT_FYC FYC_NOT_HINP;
```

```
  Merge H229D_Agg(In=A) H233_DEMO(In=B);
```

```
  By DUPERSID;
```

```
  If A and B Then Output HOSP_IN_DEMO;
```

```
  Else If A and Not B Then Output HINP_NOT_FYC;
```

```
  Else If B and Not A Then Output FYC_NOT_HINP;
```

```
Run;
```

```
Proc Print Data=HOSP_IN_DEMO;
```

```
  Var DUPERSID NUMNIGHX IPXP21X IPTC21X
```

```
    AGE21X SEX RACEV1X HISPNCAT MARRY21X;
```

```
  Title1 'Merged Full Year-Hospital Inpatient Event File';
```

```
  Title2 'Select Demographic and Cost Variables';
```

```
Run;
```


Test your knowledge

Which of the following SAS procedures can be utilized to aggregate data in the proceeding exercise instead of the DATA STEP?

A. PROC SQL

B. PROC MEANS

C. PROC DESCRIPTIVE

D. BOTH A and B

SAS Output – Merging Data Files

Obs	DUPERSID	NUMNIGHX	IPXP21X	IPTC21X	AGE21X	SEX	RACEV1X	HISPNCAT	MARRY21X
1	2320005102	1	20933.16	101320.61	85	1	1	9	1
2	2320018102	2	6787.97	23507.86	73	1	1	9	1
3	2320045102	3	28972.60	53415.00	33	2	1	1	1
4	2320063101	3	15148.80	28241.68	-1	1	2	9	2
5	2320103101	28	15233.80	41815.23	-1	1	1	1	1
6	2320105103	3	9342.53	54082.11	-1	1	2	9	5
7	2320133101	4	20398.78	32755.81	64	2	1	9	2
8	2320141101	8	8969.28	49995.45	69	2	2	9	2
9	2320145102	4	12142.63	13071.40	77	1	1	3	2
10	2320145103	12	18453.85	166748.22	-1	2	1	3	1
11	2320160101	1	289.54	751.39	51	2	2	9	3
12	2320160102	1	6423.25	11162.47	23	2	2	9	5
13	2320206101	1	1741.95	10088.52	34	1	1	9	1
14	2320211102	3	2844.53	29277.55	84	2	1	9	1
15	2320222101	1	1969.88	4500.00	62	1	1	9	3
16	2320296101	15	18758.25	183009.88	78	2	1	9	3
17	2320320109	4	14270.23	30951.87	25	2	2	9	5
18	2320342102	2	4723.28	24128.60	85	2	1	9	1
19	2320357101	4	63757.69	307873.05	66	2	1	1	1
20	2320364101	9	16933.98	95060.29	81	1	1	9	1
21	2320368201	7	12302.37	68793.08	53	1	1	6	5
22	2320429101	3	39135.38	18427.25	53	2	1	9	2
23	2320441101	6	22364.92	52337.92	72	2	1	9	3
24	2320496101	2	3671.02	20117.36	85	1	1	9	2
25	2320504101	1	2805.30	20999.55	72	2	2	9	1

Unformatted variable values.

Apply formats to see descriptive labels for categorical variables, Sex, RaceV1X, etc.

Demonstration III

Demonstration III – SAS On-Demand for Academics

- I. Generating population estimates using PROC SURVEYMEANS
- II. Modeling and predicting healthcare expenditures using PROC SURVEYREG

SAS Code – Survey Data Analysis

```
PROC SURVEYMEANS DATA=PUFLIB.HOSP_IN_DEMO SUM MEAN STD MIN MAX STDERR;  
  VAR IPXP21X;  
  WEIGHT PERWT21F;  
  CLUSTER VARPSU;  
  STRATA VARSTR;  
RUN;
```

Specify the MEPS complex survey design properties with the CLUSTER and STRATA statements.

```
PROC SURVEYMEANS DATA=PUFLIB.HOSP_IN_DEMO SUM MEAN STD STDERR;  
  VAR IPXP21X;  
  WEIGHT PERWT21F;  
  CLUSTER VARPSU;  
  STRATA VARSTR;  
  DOMAIN RACEV1X;  
  FORMAT RACEV1X RACEV1XF.;  
RUN;
```

Use the DOMAIN statement to specify categorical variables for subpopulation analysis

SAS Output - PROC SURVEY MEANS

Data Summary	
Number of Strata	105
Number of Clusters	348
Number of Observations	1932
Number of Observations Used	1885
Number of Obs with Nonpositive Weights	47
Sum of Weights	19004115.8

Statistics							
Variable	Label	Minimum	Maximum	Mean	Std Error of Mean	Sum	Std Error of Sum
IPXP21X	TOTAL EXP FOR EVENT (IPFXP21X + IPDXP21X)	0	255297	16824	615.954608	319727601692	19209564981

Statistics for RACEV1X Domains						
RACEV1X	Variable	Label	Mean	Std Error of Mean	Sum	Std Error of Sum
1 WHITE - NO OTHER RACE REPORTED	IPXP21X	TOTAL EXP FOR EVENT (IPFXP21X + IPDXP21X)	17155	742.865684	255352942810	17780215078
2 BLACK - NO OTHER RACE REPORTED	IPXP21X	TOTAL EXP FOR EVENT (IPFXP21X + IPDXP21X)	14793	933.828475	40170016801	4666260618
3 AMER INDIAN/ALASKA NATIVE - NO OTHER RACE	IPXP21X	TOTAL EXP FOR EVENT (IPFXP21X + IPDXP21X)	10358	2207.241636	1825631162	612939977
4 ASIAN/NATV HAWAIIAN/PACFC ISL-NO OTH	IPXP21X	TOTAL EXP FOR EVENT (IPFXP21X + IPDXP21X)	16423	3288.328028	9665355696	3814315135
6 MULTIPLE RACES REPORTED	IPXP21X	TOTAL EXP FOR EVENT (IPFXP21X + IPDXP21X)	19890	3726.501728	12713655223	3027345101

SAS Code - PROC SURVEYREG

Model total expenditures for inpatient hospital visits on number of nights in hospital

Use race to subset to Afro-American respondents as our subpopulation

```
PROC SURVEYREG Data=PUFLIB.HOSP_IN_DEMO plot=fit(weight=heatmap shape=hex);  
    STRATA VARSTR;  
    CLUSTER VARPSU;  
    WEIGHT PERWT21F;  
    MODEL IPXP21X = NUMNIGHX;  
    DOMAIN RACEV1X ('2');  
RUN;
```

SAS Output - PROC SURVEYREG

Data Summary	
Number of Observations	1885
Sum of Weights	19004116
Weighted Mean of IPXP21X	16824.1
Weighted Sum of IPXP21X	3.19728E11

Fit Statistics	
R-Square	0.08284
Root MSE	20412
Denominator DF	243

Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
Model	1	7.83	0.0056
Intercept	1	118.38	<.0001
NUMNIGHX	1	7.83	0.0056

Estimated Regression Coefficients				
Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	13207.2689	1213.85114	10.88	<.0001
NUMNIGHX	731.1763	261.34099	2.80	0.0056

Design Summary	
Number of Strata	105
Number of Clusters	348

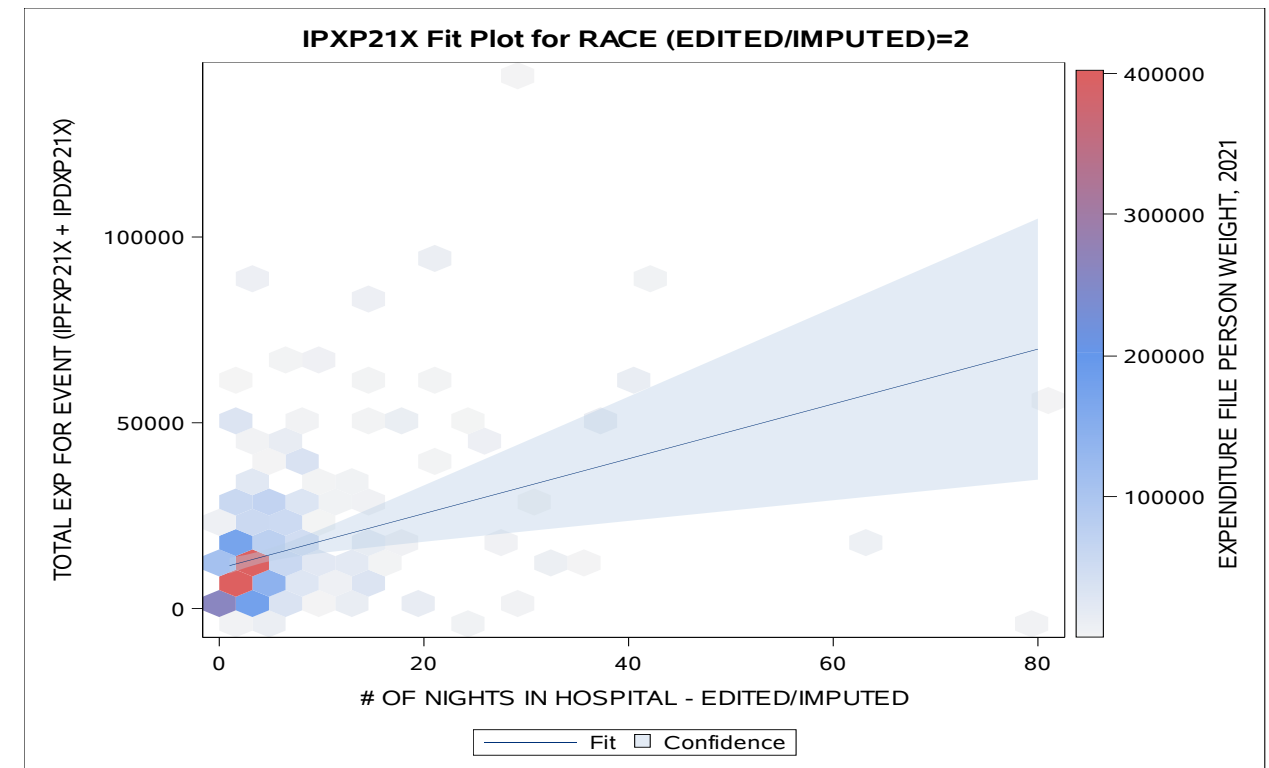
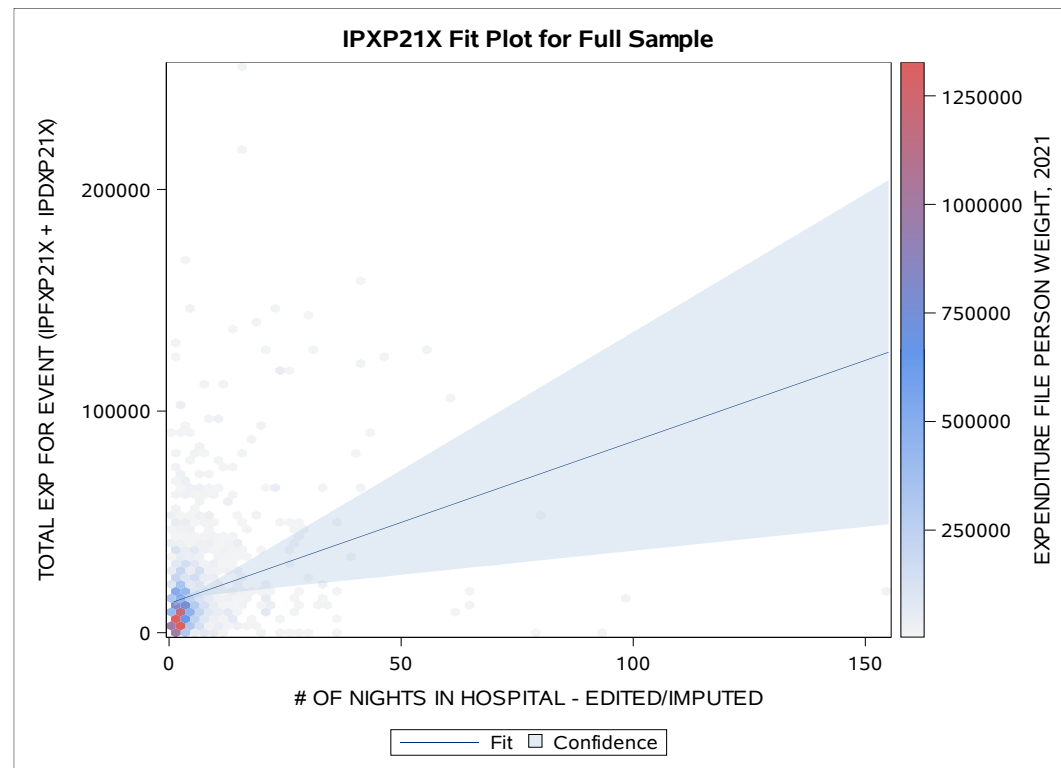
Domain Summary	
Number of Observations	1885
Number of Observations in Domain	369
Number of Observations Not in Domain	1516
Sum of Weights in Domain	2715481.2
Weighted Mean of IPXP21X	14793.0
Weighted Sum of IPXP21X	4.017E10

Fit Statistics	
R-Square	0.1405
Root MSE	14801
Denominator DF	243

Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
Model	1	9.92	0.0018
Intercept	1	76.80	<.0001
NUMNIGHX	1	9.92	0.0018

Estimated Regression Coefficients				
Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	10809.5793	1233.49801	8.76	<.0001
NUMNIGHX	737.6954	234.19772	3.15	0.0018

Medical Expenditure Panel Survey



Interpreting the SAS output and results

There's a clear positive relationship between number of nights spent in the hospital and expenditures for inpatient visits. This is a positive correlation.

The relationship between these variables is statistically significant at the .05 level.

Although the number of nights spent in the hospital only explains at most 14-15 % of the variation in expenditures for inpatient visits

The more nights you spend in the hospital, the larger the expenditures to cover that inpatient stay.

This is the case whether you're looking at the entire sample, or subsetting your data to include Afro-American patients only.

Course End / Contact Info

Thank you for Registering for WUSS Virtual Training.

Jay Iyengar

Director Data Systems Consultants LLC

datasyscon@gmail.com

<https://www.linkedin.com/in/datasysconsult/>