# Overview

The National Center for Health Statistics has conducted the National Health and Nutrition Examination Survey (NHANES) – and its precursor, the National Health Examination Survey (NHES) – since 1959. Continuous NHANES began in 1999 and released data in two-year cycles (1999-2000, 2001-2002, etc.)

Oral health items were included in many, but not all of the NHANES and NHES surveys. Oral health and other items were collected differently at different points in the survey. Variable names and coding may differ from one data cycle to the next, and variables may be located in different data sets. Public use data sets for a specific cycle may be changed after their initial release.

This tutorial focuses on oral health items included in the Continuous NHANES from 1999-2004 and describes some strategies for discovering and handling the differences across multiple years. The programming examples are presented using SAS and SAS-callable SUDAAN. The examples focus on use of the demographics, oral health questionnaire and oral health examination components, which contain one record per respondent. Laboratory and dietary data files can contain more than one record per respondent. We’ll explore these as time permits.

For additional information on NHANES, see <http://www.cdc.gov/nchs/nhanes>, especially the Web tutorial at <http://www.cdc.gov/nchs/tutorials/>. Continuing education credit is available for completing the Web tutorial. It contains crosswalks of variables over data release cycles.

# File and folder set up

Before beginning analyses of NHANES data, I find it useful to set up file folders as follows. The examples will use this type of structure. If you prefer something different some changes to file paths may be needed.

NHANES

* Original data sets (for data sets downloaded from NCHS Web site)
* Documentation (for data set documentation and analytic guidelines, downloaded or created)
* Analytic files
* Analysis data files (data files created from the original data using programs in folders below)
* Data management (programs to merge data files)
* Variable Creation (programs to recode variables or create indices, DMFT for example)
* Analysis programs (programs for conducting descriptive and multivariate analyses)

A note about programming style:

Programming examples avoid macros and “include” files, for the purpose of having examples that are easy to follow by relatively novice users of the software, perhaps epidemiology graduate students or residents. I’m certain that you can create much more efficient programs to do the same thing.

# Find data from NCHS NHANES Web site

1. Navigate to <http://www.cdc.gov/nchs>
2. Follow link in left menu to National Health and Nutrition Examination Survey
3. Follow link in left menu to Questionnaires, Datasets, and Related Documentation
	1. We’ll download data from years 1999-2000, 2001-2002 and 2003-2004.



Figure . Screen shot of NHANES Web page showing links to data release cycles and other NHANES documentation

# Download demographic data from 1999-2004

1. Follow the link to 1999-2000.
2. Under “Contents at a Glance”, note the Historic NHANES component matrix. This is a high-level view of the topics collected from NHANES I through Continuous NHANES, updated through the 1999-2006 data releases.
3. Under “Data, Documentation, Codebooks, SAS Code”, follow the link to “Demographics”.
	1. Download the “Merge Code example” (right-click, save target as)
	2. Download the SAS transport data set (DEMO.xpt)
	3. View the documentation (“Docs”) and begin an analysis-specific codebook.
	4. Convert the data file from the transport format to the SAS data format (or other format)
4. Continue to download demographic data for 2001-2002 and 2003-2004 and fill in the codebook.

Table 1. Example analysis specific codebook for demographic variables created from NHANES data set documentation

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1999-2000 | 2001-2002 | 2003-2004 |
|  | DEMO | DEMO\_B | DEMO\_C |
| Respondent ID | SEQN | SEQN | SEQN |
| Survey cycle | SDDSRVYR |  |  |
| Interview/Exam status | RIDSTATR |  |  |
| Sex | RIAGENDR |  |  |
| Age in years | RIDAGEYR |  |  |
| Family Poverty Income Ratio | INDFAMPIR |  |  |
| Education (Head of Household) | DMDHREDU |  |  |
| Weight – interview 2 year | WTINT2YR |  |  |
| Weight – exam 2 year | WTMEC2YR |  |  |
| Weight – interview 4 year | WTINT4YR | WTINT4YR | *Not applicable* |
| Weight – exam 4 year | WTMEC4YR | WTMEC4YR | *Not applicable* |
| Primary Sampling Unit | SDMVPSU |  |  |
| Strata | SDMVSTRA |  |  |

Note that the 4-year weights are only used when data from 1999-2002 are included in the analysis plan (1999-2004 will also involve use of the 4-year weights).

|  |
| --- |
| LIBNAME L XPORT 'C:\My Files\Temp\OHQ.xpt';LIBNAME D XPORT 'C:\My Files\Temp\DEMO.xpt';LIBNAME P 'C:\My Files\Temp\NHANES';DATA P.OHQA; MERGE D.DEMO (KEEP=SEQN RIDAGEYR RIAGENDR)  L.OHQ; BY SEQN;RUN; |

SAS Code . Example code to merge Demographics data file with a questionnaire file

# Download oral health questionnaire data from 1999-2004

1. Follow the link to 1999-2000.
2. Under “Data, Documentation, Codebooks, SAS Code”, follow the link to “Questionnaire”
	1. Download the SAS transport data set for the Oral Health Questionnaire (OHX.xpt)
	2. View the documentation (“Docs”) and begin an analysis-specific codebook.
	3. This codebook example includes variable coding. Note that the response options changed, so that a code of “1” has different meaning in the first and second cycles.
3. Continue to download Oral Health Questionnaire data for 2001-2002 and 2003-2004 and fill in the codebook.

Table 2. Example codebook for Oral Health Questionnaire

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1999-2000 | 2001-2002 | 2003-2004 |
|  | OHQ | OHQ\_B | **OHXREF\_C** |
| Respondent ID | SEQN | SEQN | SEQN |
| General condition of mouth and teeth | OHQ010 | OHQ010 | **OHQ011** |
| Target population | Aged 2-120 yrs | Aged 2-150 yrs | Aged **16** -150 yrs |
| Question | Now I have some questions about {your/SP's} mouth and teeth. How would you describe the condition of {your/SP’s} mouth and teeth? Would you say . . . | Now I have some questions about {your/SP's} mouth and teeth. How would you describe the condition of {your/SP's} mouth and teeth? Would you say . . . | How would you describe the condition of your teeth? Would you say...... |
| Excellent | Not available | Not available | 1 |
| Very Good | 1 | 1 | 2 |
| Good | 2 | 2 | 3 |
| Fair | 3 | 3 | 4 |
| Poor | 4 | 4 | 5 |
| Refused | 7 | 7 | 7 |
| Don’t Know | 9 | 9 | 9 |
| Missing | . | . | . |
| Other Notes | INCLUDE FALSE TEETH AND DENTURES | INCLUDE FALSE TEETH AND DENTURES | No other instructions |

# Example Analysis of Oral Health Questionnaire data

## Converting transport files (old way)

libname in XPORT "C:\NHANES\Original data sets\DEMO.xpt";

libname out "C:\NHANES\Analytic files\Analysis data files";

**data** out.demo;

set in.demo;

**run**;

1 libname in XPORT "C:\Documents and Settings\lub2\My Documents\NHANES\NHANES Workshop

1 ! 20100506\NHANES\Original data sets\DEMO.xpt";

NOTE: Libref IN was successfully assigned as follows:

 Engine: XPORT

 Physical Name: C:\Documents and Settings\lub2\My Documents\NHANES\NHANES Workshop

 20100506\NHANES\Original data sets\DEMO.xpt

2 libname out "C:\Documents and Settings\lub2\My Documents\NHANES\NHANES Workshop

2 ! 20100506\NHANES\Analytic files\Analysis data files";

NOTE: Libref OUT was successfully assigned as follows:

 Engine: V9

 Physical Name: C:\Documents and Settings\lub2\My Documents\NHANES\NHANES Workshop

 20100506\NHANES\Analytic files\Analysis data files

3

4 data out.demo;

5 set in.demo;

6 run;

NOTE: There were 9965 observations read from the data set IN.DEMO.

NOTE: The data set OUT.DEMO has 9965 observations and 144 variables.

NOTE: DATA statement used (Total process time):

 real time 1.17 seconds

 cpu time 0.20 seconds

## Merging and concatenating

|  |  |  |
| --- | --- | --- |
|  |  | *🡨-----------------------------------------------Merging------------------------------------------------🡪* |
|  |  | Demographics | Oral Health Questionnaire | Oral Health Examination - Dentition |
| *Concatenating* | 1999-2000 | DEMOSEQN 1-10000 | OHX SEQN 1-10000 | OHXDENTSEQN 1-10000 |
| 2001-2002 | DEMO\_BSEQN 10001-20000 | OHX\_BSEQN 10001-20000 | OHXDEN\_BSEQN 10001-20000 |
| 2003-2004 | DEMO\_CSEQN 20001-30000 | OHX\_CSEQN 20001-30000 | OHXDEN\_CSEQN 20001-30000 |

Note: SEQN numbers shown here are just for illustration of the concept. Actual SEQN numbers in the file may differ. Within a data cycle (1999-2000) the SEQNs will be the same across each data file, or will be subsets of the SEQNs in the Demographics file for that data cycle. SEQNs from 1999-2000 will not appear in 2001-2002 or 2003-2004, or vice versa.

## Descriptive analysis of 1999-2000 demographic and oral health questionnaire data (working from transport files directly)

libname demoa XPORT "C:\NHANES\Original data sets\DEMO.xpt";

libname ohqa XPORT "C:\NHANES\Original data sets\OHQ.xpt";

libname out "C:\NHANES\Analytic files\Analysis data files";

/\*read data in from transport files and merge, keeping only variables needed\*/

**data** out.demohqa;

merge demoa.demo(KEEP=SEQN RIDSTATR RIAGENDR RIDAGEYR WTINT2YR SDMVPSU SDMVSTRA)

 ohqa.ohq (KEEP=SEQN OHQ010);

by SEQN;

**run**;

/\*Variable Creation

Recode self-reported oral health to treat Don’t Know and Refused as missing data.

In practice, suggest more exploration of the refused and don't know responses before recoding as missing.

\*/

**data** out.demohqa;

set out.demohqa;

if OHQ010 in (**7**,**9**,**.**) then SROH=**.**; /\*set refused and don't know to missing\*/

else SROH=OHQ010;

**run**;

/\*Sort by Strata, and PSU within Strata\*/

**proc** **sort** data=out.demohqa;

by SDMVSTRA SDMVPSU;

**run**;

/\*SAS-callable SUDAAN - Descript\*/

**proc** **descript** data=out.demohqa design=wr DEFT;/\*D1. Sealant prevalence - not age adjusted\*/

weight WTINT2YR;

nest SDMVSTRA SDMVPSU/strlev=**1** psulev=**2** missunit;

var SROH ;

catlevel **1**;

subpopn ridageyr >=**16** and ridageyr<=**19** and RIDSTATR>=**1** and RIDSTATR<=**2**;

subgroup RIAGENDR ;

level **2** ;

tables RIAGENDR;

print nsum wsum percent sepercent DEFFPCT/

 style =nchs wsumfmt=f15.1 percentfmt=f6.2 sepercentfmt=f6.2 deffpctfmt=f6.2;

rtitle "SUDAAN Descript - Prevalence of Very Good oral health among ages 16-19 yrs by sex - NHANES 1999-2000";

**run**;

/\* SAS-callable SUDAAN - Crosstab (adapted from Donna Brogan's class) \*/

**proc** **crosstab** data = out.demohqa design = wr notot nocol DEFT;

nest SDMVSTRA SDMVPSU ;

weight WTINT2YR ;

tables RIAGENDR \* SROH ;

class RIAGENDR SROH ;

test chisq llchisq / waldchi waldf adjwaldf ;

title "SUDAAN Crosstab - Self-reported oral health among ages 16-19 yrs by sex - NHANES 1999-2000" ;

print nsum wsum sewgt rowper serow lowrow uprow deffrow/

 wsumfmt = f9.0 sewgtfmt = f9.0 rowperfmt = f7.3

 serowfmt = f7.3 lowrowfmt = f7.3 uprowfmt = f7.3 deffrowfmt=f6.2 style = nchs ;

**run** ;

/\* SAS SURVEYFREQ (adapted from Donna Brogan's class) \*/

**proc** **surveyfreq** data = out.demohqa nomcar ;

strata SDMVSTRA ;

cluster SDMVPSU ;

weight WTINT2YR ;

tables RIAGENDR \* SROH / nocellpercent row cl deff

 chisq chisq1 lrchisq lrchisq1 wchisq wllchisq ;

title "SAS Surveyfreq - Self-reported oral health among ages 16-19 yrs by sex - NHANES 1999-2000";

**run** ;

# Download oral health examination Dentition data from 1999-2004

1. Follow the link to 1999-2000.
2. Under “Data, Documentation, Codebooks, SAS Code”, follow the link to “Examination”
	1. Download the SAS transport data set for the Oral Health Dentition (OHXDENT.xpt)
	2. View the documentation (“Docs”) and begin an analysis-specific codebook.
	3. This codebook example includes variable coding.
3. Continue to download Oral Health Dentition data for 2001-2002 and 2003-2004 and fill in the codebook.
4. Refer to Data Resource Center draft materials for example variable creation programs

Table . Example codebook for Oral Health Examination - Dentition

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1999-2000 | 2001-2002 | 2003-2004 |
|  | OHXDENT | OHXDEN\_B | OHXDEN\_C |
| Respondent ID | SEQN | SEQN | SEQN |
| Oral Health Exam Status | OHAEXSTS |  |  |
| Dentition Exam Status | OHASTSC3 |  |  |
| Tooth Count | OHX01TC-OHX32TC |  |  |
| Coronal Caries tooth | OHX02CTC- OHX15CTCOHX18CTC- OHX31CTC |  |  |
| Coronal Caries surface | OHX02CSC- OHX15CSCOHX18CSC- OHX31CSC |  |  |

Table 4. Variable names by tooth position

|  |  |  |  |
| --- | --- | --- | --- |
|  | Tooth Count | Caries Tooth level | Caries Surface Level |
| Upper right 3rd molar | OHX01TC | Not examined | Not examined |
| Upper right 2nd molar | OHX02TC | OHX02CTC | OHX02CSC |
| Upper right 1st molar | OHX03TC | OHX03CTC | OHX03CSC |
| Upper right 2nd bicuspid/2nd primary molar | OHX04TC | OHX04CTC | OHX04CSC |
| … |  |  |  |
| Upper left 2nd bicuspid/2nd primary molar  | OHX13TC | OHX13CTC | OHX13CSC |
| Upper left 1st molar | OHX14TC | OHX14CTC | OHX14CSC |
| Upper left 2nd molar | OHX15TC | OHX15CTC | OHX15CSC |
| Upper left 3rd molar | OHX16TC | Not examined | Not examined |
| Lower right 3rd molar | OHX17TC | Not examined | Not examined |
| Lower right 2nd molar | OHX18TC | OHX18CTC | OHX18CSC |
| Lower right 1st molar | OHX19TC | OHX19CTC | OHX19CSC |
| Lower right 2nd bicuspid/2nd primary molar | OHX20TC | OHX20CTC | OHX20CSC |
| … |  |  |  |
| Lower left 2nd bicuspid/2nd primary molar  | OHX29TC | OHX29CTC | OHX29CSC |
| Lower left 1st molar | OHX30TC | OHX30CTC | OHX30CSC |
| Lower left 2nd molar | OHX31TC | OHX31CTC | OHX31CSC |
| Lower left 3rd molar | OHX32TC | Not examined | Not examined |

Table 5. Tooth Count variable coding

|  |  |  |  |
| --- | --- | --- | --- |
| OHX01TC-OHX32TC | 1999-2000 | 2001-2002 | 2003-2004 |
| Data set | OHXDENT | OHXDEN\_B | OHXDEN\_C |
| Primary tooth | 1 | 1 | 1 |
| Permanent tooth | 2 | 2 | 2 |
| Implant | 3 | 3 | 3 |
| Not Present | 4 | 4 | 4 |
| Permanent root tip is present | Not recorded | Not recorded | 5 |
| Missing | . | . |  |

Table 6. Coronal Caries Tooth Level variable coding

|  |  |  |  |
| --- | --- | --- | --- |
| OHX02CTC-OHX15CTC, OHX18CTC-OHX31CTC | 1999-2000 | 2001-2002 | 2003-2004 |
| Data set | OHXDENT | OHXDEN\_B | OHXDEN\_C |
| Sound primary tooth | D | D | D |
| Missing due to dental disease | E | E | E |
| Permanent root tip present, no replacement present | Not recorded | Not recorded | J |
| Primary tooth with surface conditions | K | K | K |
| Missing due to other causes | M | M | M |
| Missing due to dental disease, removable restoration | Not recorded | Not recorded | P |
| Missing due to other causes, removable restoration | Not recorded | Not recorded | Q |
| Missing due to dental disease, a fixed restoration | R | R | R |
| Sound permanent tooth | S | S | S |
| Permanent root tip is present, restorative replacement  | Not recorded | Not recorded | T |
| Unerupted | U | U | U |
| Missing due to other causes, fixed restoration | X | X | X |
| Tooth present, condition cannot be assessed | Y | Y | Y |
| Permanent tooth with surface condition (s) | Z | Z | Z |
| Missing | < blank > | < blank > | < blank > |

Table 7. Coronal Caries Surface Level variable coding

|  |  |  |  |
| --- | --- | --- | --- |
| OHX02CTC-OHX15CTC, OHX18CTC-OHX31CTC | 1999-2000 | 2001-2002 | 2003-2004 |
| Data set | OHXDENT | OHXDEN\_B | OHXDEN\_C |
| Multiple surface conditions found | 01234, 0628 | 01234, 0628 | 01234, 0628 |
| Lingual surface caries | 0 | 0 | 0 |
| Occlusal/incisal caries | 1 | 1 | 1 |
| Facial surface caries | 2 | 2 | 2 |
| Mesial caries | 3 | 3 | 3 |
| Distal caries | 4 | 4 | 4 |
| Lingual surface restoration | 5 | 5 | 5 |
| Occlusal/incisal restoration | 6 | 6 | 6 |
| Facial surface restoration | 7 | 7 | 7 |
| Mesial restoration | 8 | 8 | 8 |
| Distal restoration | 9 | 9 | 9 |
| Missing |  |  | < blank > |

# Download oral health examination Periodontal data from 1999-2004

1. Follow the link to 1999-2000.
2. Under “Data, Documentation, Codebooks, SAS Code”, follow the link to “Examination”
	1. Download the SAS transport data set for the Oral Health Periodontal (OHXDENT.xpt)
	2. View the documentation (“Docs”) and begin an analysis-specific codebook.
	3. This codebook example includes variable coding.
3. Continue to download Oral Health Dentition data for 2001-2002 and 2003-2004 and fill in the codebook.
4. Refer to Data Resource Center draft materials for example variable creation programs

Table . Periodontal variables

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1999-2000 | 2001-2002 | 2003-2004 |
|  | OHXPERIO | OHXPRU\_BOHXPRL\_B | OHXPRU\_COHXPRL\_C |
| Respondent ID | SEQN | SEQN | SEQN |
| Oral Health Exam Status | OHAEXSTS | OHAEXSTS | OHAEXSTS |
| Perio Exam Status | OHASTSC4 | OHASTSC4 | OHASTSC4 |
| Free gingival margin to cemento-enamel junction | ## tooth positions 02-15 and 18-31CJ can be negative99 is a special code for cannot be assessed |
| Mesial | OHD##CJS | OHD##CJS | OHD##CJS |
| Mid-facial | OHD##CJM | OHD##CJM | OHD##CJM |
| Distal | Not measured | OHD##CJD | OHD##CJD |
| Pocket Depth | ## tooth positions 02-15 and 18-3199 is a special code for cannot be assessed |
| Mesial | OHD##PCS | OHD##PCS | OHD##PCS |
| Mid-facial | OHD##PCM | OHD##PCM | OHD##PCM |
| Distal | Not measured | OHD##PCD | OHD##PCD |
| Calculated Loss of Attachment LA=PC-CJ | ## tooth positions 02-15 and 18-3199 is a special code for cannot be assessed |
| Mesial | OHD##LAS | OHD##LAS | OHD##LAS |
| Mid-facial | OHD##LAM | OHD##LAM | OHD##LAM |
| Distal | Not measured | OHD##LAD | OHD##LAD |

Note: 1999-2004 periodontal exam randomly selected one upper quadrant and one lower quadrant for examination. Each tooth (with the exception of 3rd molars) was probed at mesial and mid-facial sites on the buccal (facial, labial) side of the tooth in 1999-2000. In 2001-2004, the distal site was also probed. No lingual sites were probed. Oral Health examiners were dentists. The total number of measurements per participant is 14 teeth \* 2 probing sites per tooth = 28 measurements in 1999-2002. In 2003-2004 the total number of measurements is 14 teeth \* 3 sites per tooth = 42 sites.

The 2009-2010 periodontal exam was a full-mouth exam with 6 probing sites per each of 28 teeth, for a total of 168 measurements per participant. Public release data are not yet available, but expected later this year. Documentation is available under the 2009-2010 data cycle.