Understanding and Analyzing Ambulatory Health Care Data-The NAMCS & NHAMCS

Basic Data Analysis

Kelly L. Myrick, PhD, CPH

CDC

Overview

- **Important features of 2012 NAMCS**

- **File structure**

- **Exercises using SAS and Stata**
  - Downloading data & creating a SAS/Stata dataset
  - 2012 NAMCS: Weighted & unweighted frequencies with/without standard errors
  - 2012 NAMCS: Creating a new variable
  - 2012 NAMCS: Visit rates for asthma
  - 2012 NAMCS: Total number of write-in procedures
  - 2012 NAMCS: Analysis of a continuous variable

- **Data & User considerations**

- **Summary**

NAMCS File Structure

Sample Weight Pre-2012

- **Each visit record contains a Patient Visit Weight**
  - True across NAMCS, NHAMCS Outpatient Department (OPD) and NHAMCS Emergency Department (ED) records

- **This weight is used to generate all <u>visit</u>-related estimates, drug mentions & write-in procedures**

NAMCS 2012 Sample Weight

- **Each visit record contains two Patient Visit Weights**
  - National weight
  - State weight
- **National weight used for the overall total**
- **State weight used for state specific estimates**
- **Each weight is used to generate all <u>visit</u>-related estimates, drug mentions & write-in procedures**

Coding Systems Used

- **Reason for Visit Classification (NCHS)**

- **ICD-9-CM for diagnoses, causes of injury and procedures**

- **Two Drug Classification Systems**
  - In-house system
  - MULTUM

Survey Years 2007+ NCHS Coding Convention Changes

- **Missing data have consistent negative codes for years 2007 and beyond**
  - Blank = -9
  - Unknown/Don't know = -8
  - Not applicable = -7

- **Missing data had positive codes prior to 2007**
  - Blank code varied
  - Unknown/Don't know code varied
  - Not applicable code varied

Enhanced Public-Use Files

- **Download data and layout from website**

  http://www.cdc.gov/nchs/ahcd/ahcd_questionnaires.htm

  - Flat ASCII files for each setting and year:

    NAMCS: 1973-2010, 2012

    NHAMCS: 1992-2011

Enhanced Public-Use Files (cont. 2)

- **Flat ASCII files for each setting and year:**
  - SAS input statements, variable labels, value labels, and format assignments for 1993-2010, 2011 NHAMCS, 2012 NAMCS

  - SPSS syntax files for 2002-2010, 2011 NHAMCS, 2012 NAMCS

  - Stata .do and .dct files for 2002-2010, 2011 NHAMCS, 2012 NAMCS
    - Stata.exe files for 2005-2010, 2011 NHAMCS, 2012 NAMCS

Enhanced Public-Use Files (cont. 3)

- **New survey items and facility level data**

- **Sample design variables**
  - **Starting in 2003, we only released masked variables for use in software using 1-stage.**
  - In 2001 and prior years, masked variables for 3- or 4-stage sampling are available.
  - In 2002, NAMCS & NHAMCS masked variables have been available for use in software using multi-stage and 1-stage sampling.

Design Variables –Survey Years

2001

3-(NAMCS) or
4-(NHAMCS)
Stages

2002
1-Stage
(NAMCS &
NHAMCS

3-(NAMCS) or
4-(NHAMCS)
Stages

2003
1-Stage (NAMCS &
NHAMCS) only
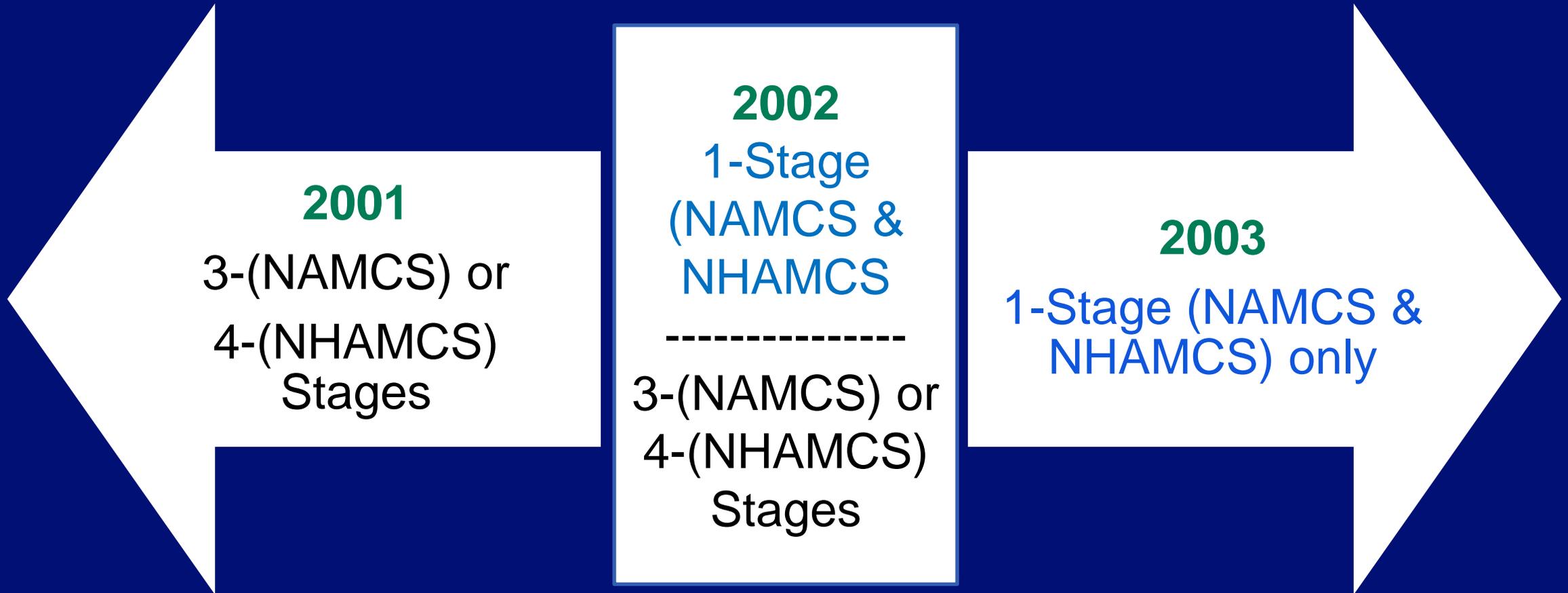
"Using Ultimate Cluster Models in NAMCS & NHAMCS Public Use Files"
http://www.cdc.gov/nchs/data/ahcd/ultimatecluster.pdf

11

Hands-on Exercises

SAS Examples

- **Double-click: My Computer\Local Disk C:\DATA**
- **Double-click: 2015 Basic SAS Exercises**

Stata Examples

- **Double-click: My Computer\Local Disk C:\DATA**
- **Open Stata**
- **In the command window type:**
  - Set **mem 200m**
  - Set **matsize 500**
- **Under the "File" icon-double-click NAMCS2015-stata.dta**
- **Under "New Do File Editor"-double-click:**
  2015 Basic Stata Exercises.do

Visit Rate Example

Female Population =		800	Calculation*	New Variable
Phycode	**Sex**	**Patwt**	**(Patwt/Pop)*100**	**Sexwt**
1401	1	100	(100/800)*100	12.5
1820	1	300	(300/800)*100	37.5
1001	1	50	(50/800)*100	6.25
500	1	120	(120/800)*100	15.0
	Sample size = 4	Visits=570		71.25 visits per 100 persons

*Note: Rate=estimate/population=Σ patwt/population=1/population*Σ patwt.

Patwt=patient weight, Phycode=physician code, Sexwt=sex weight

13

Write-in Procedures on the Patient Record Form (PRF)

- **The section of the Patient Record Form on the right shows the location of the write-in procedures that correspond to PROC1—PROC9.**

Calculating Total Number of Write-in Procedures

Record	Proc1	Proc2	Proc3	Proc4	Proc5	Proc6	Proc7	Proc8	Proc9	Tot proc
1	1911	-9	-9	-9	-9	-9	-9	-9	-9	1
2	2182	2186	-9	-9	-9	-9	-9	-9	-9	2
3	2121	3142	9547	-9	-9	-9	-9	-9	-9	3
4	-9	-9	-9	-9	-9	-9	-9	-9	-9	0
5	8192	8200	-9	-9	-9	-9	-9	-9	-9	2

NOTE: -9 = No procedure recorded

Data Considerations

NAMCS vs. NHAMCS

- **Consider what types of settings are best for a particular analysis**

  - Persons of color are more likely to visit OPDs and EDs than physician offices

  - Persons in some age groups make up disproportionately larger amounts of visits to EDs than physician offices and OPDs

Which Statistical Command?

- **SAS uses the Proc Surveyfreq command for the analysis of categorical variables and Proc Surveymeans command for the analysis of continuous variables**

- **Stata uses the svy:tab command for the analysis of categorical variables and the svy:mean command for the analysis of continuous variables**

- **SUDAAN uses the Proc Crosstab command for the analysis of categorical variables and the Proc Descript command for the analysis of continuous variables**

How Good are the Estimates?

- **Depends … In general, OPD estimates tend to be less reliable than NAMCS & ED estimates**

- **Since 1999 standard errors are included for calculation of confidence intervals around the estimates**
    - Advance Data Reports
    - National Health Statistics Reports
    - Web Tables

Reliability Criteria – The 30-30 Rule

- **Estimates should be based on at least <u>30 unweighted records</u> AND**

- **Estimates need to have a relative standard error or RSE (standard error divided by estimate) <u>less than 30%</u>**

- **Our standards dictate both conditions should be met before estimates are considered reliable**

Reliability of Estimates in NAMCS

Patient Race	Number of visits (millions)	Standard error (millions)	RSE for NAMCS (%)
White	843	42	5
Black	117	9	8

Reliability of Estimates in NHAMCS

Patient Race	RSE for OPD (%)	RSE for ED (%)
White	10	5
Black	12	7

- A higher RSE means that an estimate has a wider confidence interval and thus should be considered less reliable

Ways to Improve Reliability of Estimates

- **Combine NAMCS, ED and OPD data to produce ambulatory care visit estimates**

- **Combine multiple years of data**

- **Use multiple variables to define construct**

RSE Improves Incrementally with the Number of Years Combined

- RSE = SE of Estimate/Estimate (*x*)

- RSE for percent of office visits by persons less than 21 years with diabetes

  - ✓ 2010 RSE = .6/1.8 = .33 (33%)
  - ✓ 2009 & 2010 RSE = .3/1.3 = .23 (23%)
  - ✓ 2008, 2009, & 2010 RSE = .2/1.2 = .17 (17%)

Sampling Error

- **NAMCS and NHAMCS are not simple random samples**

- **Clustering effects:**
  - Providers within PSUs
  - Visits within physician practice or hospital

- **Must use special software (e.g., SAS / Stata survey procedures) to calculate standard errors for all estimates, percents, and rates**

Some User Considerations

- **High percentage of missing on some items**

  - 2012 NAMCS
    - Ethnicity (35.1%)
    - Imputed and unimputed data

    - Race (32.9%)
      - Imputed and unimputed data

Some User Considerations (cont.)

- **High percentage of missing on some items**

  - 2012 NAMCS

    - Time spent with provider (37.4%)
      - Imputed and unimputed data

    - Tobacco use (26.1%)
      - Unimputed data

If nothing else, remember…
The Public Use Data File Documentation is YOUR FRIEND!

- **Can be downloaded from NAMCS/NHAMCS Web site (http://www.cdc.gov/nchs/ahcd/ahcd_questionnaires.htm)**
- **Each booklet includes:**
  - A description of the survey
  - Record format
  - Marginal data (summaries)
  - Various definitions
  - Reason for Visit Classification codes
  - Medication codes & names
  - Medication therapeutic classes