Understanding and Using NAMCS and NHAMCS Data

Data Tools and Basic Programming Techniques

2010 National Conference on Health Statistics
August 16, 2010

Chun-Ju (Janey) Hsiao
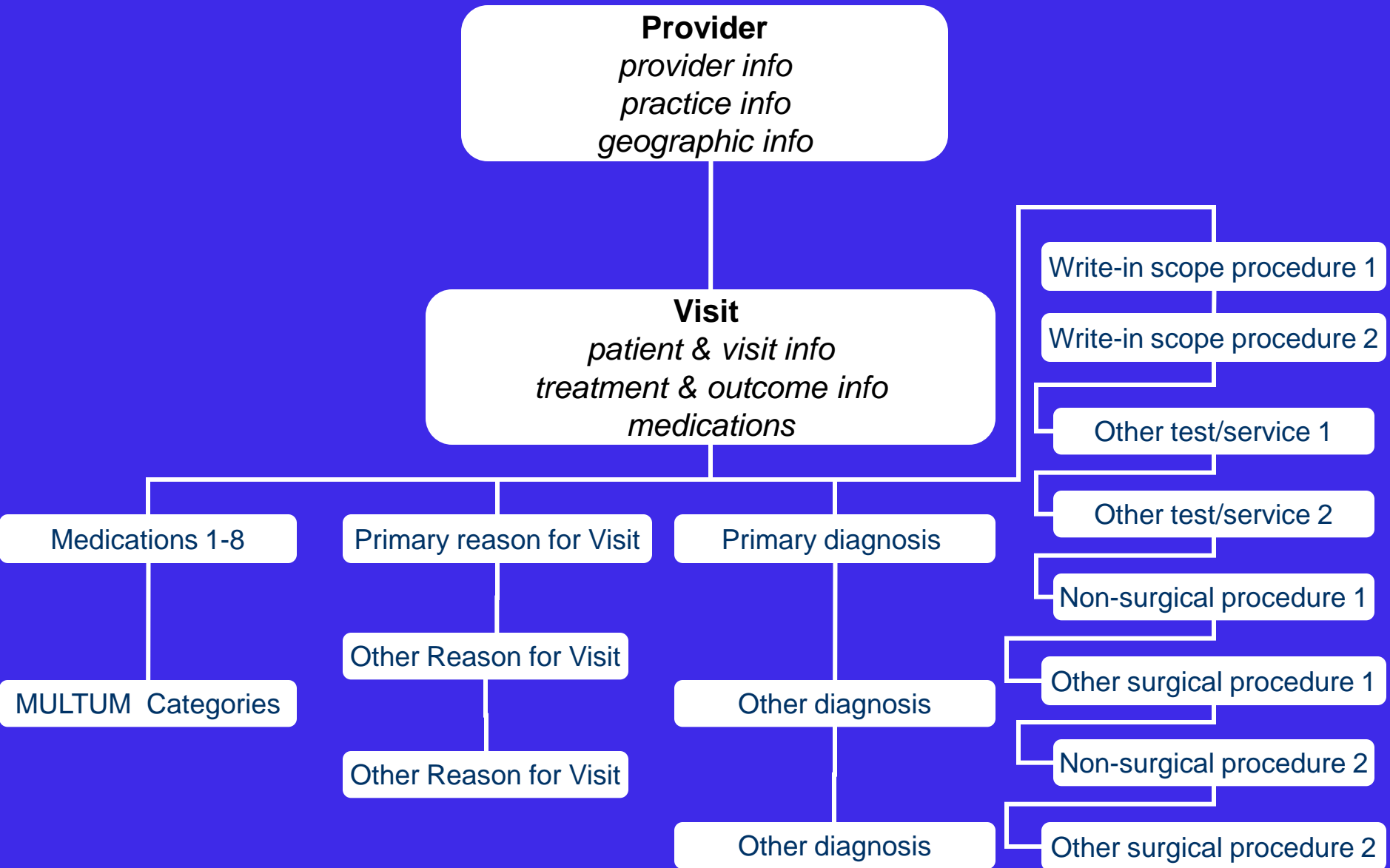
National Center for Health Statistics

Overview

- Some important features of NAMCS & NHAMCS

- File structure

- Exercises using SAS Proc Surveyfreq/Proc Surveymeans and STATA
  - Downloading data & creating a SAS/STATA dataset
  - Weighted and unweighted frequencies with/without standard errors
  - Creating a new variable-Asthma
  - Visit rates for asthma-male/female
  - Total number of digestive write-in procedures
  - Time spent with physician

- Considerations

- Summary

Organizational Structure-NAMCS Data

Provider
provider info
practice info
geographic info

Visit
patient & visit info
treatment & outcome info
medications

Write-in scope procedure 1

Write-in scope procedure 2

Other test/service 1

Other test/service 2

Medications 1-8

Primary reason for Visit

Primary diagnosis

Non-surgical procedure 1

MULTUM Categories

Other Reason for Visit

Other diagnosis

Other surgical procedure 1

Non-surgical procedure 2

Other Reason for Visit

Other diagnosis

Other surgical procedure 2

Data Items

- Patient characteristics
  - Age, sex, race, ethnicity

- Visit characteristics
  - Source of payment, continuity of care, reason for visit, diagnosis, treatment

- Provider characteristics
  - Physician specialty, hospital ownership

- MULTUM drug characteristics added in 2006

Sample Weight

- Each NAMCS record contains a single weight, which we call Patient Visit Weight.

  - Same is true for OPD records and ED records

- This weight is used for both visits and drug/procedure mentions.

2007 NCHS Coding Convention Changes

- Starting in 2007, missing data have consistent negative codes
  - Blank= -9
  - Unknown/Don't know= -8
  - Not applicable= -7

- Prior to 2007, missing data had positive codes
  - Blank code varied
  - Unknown/Don't know code varied
  - Not applicable=8

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: 1973-2007
        NHAMCS: 1992-2007
    - SAS input statements, variable labels, value labels, and format assignments for 1993-2007
    - SPSS syntax files for 2002-2006
    - STATA .do and .dct files for 2002-2006

Enhanced Public-Use Files (cont.)

- New survey items and facility level data

- Sample design variables
  - 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.

  - Starting in 2003, we only released masked variables for use in software using 1-stage.

Design Variables—Survey Years

2001

3- or 4-Stage design variables

2002

1-Stage design variables

3- or 4-Stage design variables

2003

1-Stage design variables only

Creating a Usable STATA Dataset

- Three options:

  1) Use the self-extracting file in the STATA folder to open a complete dataset for the 2005-2006 NAMCS, NHAMCS-ED, & NHAMCS-OPD.

  2) Use the DO file (*.do) and the dictionary file (*.dct) along with the flat data file (*.exe) to create a dataset.

  3) StatTransfer

Hands-on Exercises

STATA Users

- Double-click: C:\AHDATA\SAS
- Open STATA
- In the command window type:
  - Set **mem 100m**
  - Set **matsize 500**
- Under the "File" icon- double-click **namcs07.dta**
- Under "New Do File Editor"- double-click: **STATA 07exercises.do**

SAS/SUDAAN Users

- Double-click: C:\AHDATA\SAS
- Double-click: SAS07exercise

Visit Rate Estimates

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

Sample size=4   Visits=570   71.25 visits per 100 persons

*Note: Rate=est/pop=Σ patwt/pop=1/pop*Σ patwt.

Calculating Total Number of Write-in Procedures

Record	Proc1	Proc2	Proc3	Proc4	Proc5	Proc6	Proc7	Proc8	Totproc
1	1911	0000	0000	0000	0000	0000	0000	0000	1
2	2182	2186	0000	0000	0000	0000	0000	0000	2
3	5490	0000	0000	0000	0000	0000	0000	0000	1
4	0000	0000	0000	0000	0000	0000	0000	0000	0
5	8192	0000	0000	0000	0000	0000	0000	8200	2

Note: 0000=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 disproportionately larger shares of visits to EDs than physician offices and OPDs

Which Statistical Program?

Program	Categorical Variables	Continuous Variables
SAS	PROC SURVEYFREQ	PROC SURVEYMEANS
STATA	SVY: TAB	SVY: MEAN
SUDAAN	PROC CROSSTAB	PROC DESCRIPT

How Good are the Estimates?

- Depends … In general, OPD estimates tend to be somewhat less reliable than NAMCS and ED.

- Since 1999, our Advance Data Reports/National Health Statistics Reports include standard errors in every table so it is easy to compute confidence intervals around the estimates.

Reliability Criteria

- Estimates should be based on at least 30 sample records **AND**

- Estimates with a relative standard error (standard error divided by the estimate) greater than 30 percent are considered unreliable by NCHS standards.

- Both conditions should be met before considering estimates 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/*x*

- RSE for percent of visits by persons less than 21 years of age with diabetes
  - ✓ 1999 RSE = .08/.18 = .44 (44%)
  - ✓ 1998 & 1999 RSE = .06/.18 = .33 (33%)
 - ✓ 1998, 1999, & 2000 RSE = .05/.21 = .24 (24%)

Sampling Error

- NAMCS and NHAMCS are not simple random samples

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

- Must use generalized variance curve or special software (e.g., SUDAAN) to calculate SEs for all estimates, percents, and rates

# Calculating Variance with NAMCS/NHAMCS Estimates

- Old way (least accurate) = Generalized variance curves
- Better way (recommended) = Masked design variables
  - Multiple sampling stages for years
  - Single stage of sampling or ultimate cluster design
- Most accurate way (expensive) = Actual design variables

Comparisons of Relative Standard Errors (RSEs) for Patient Race
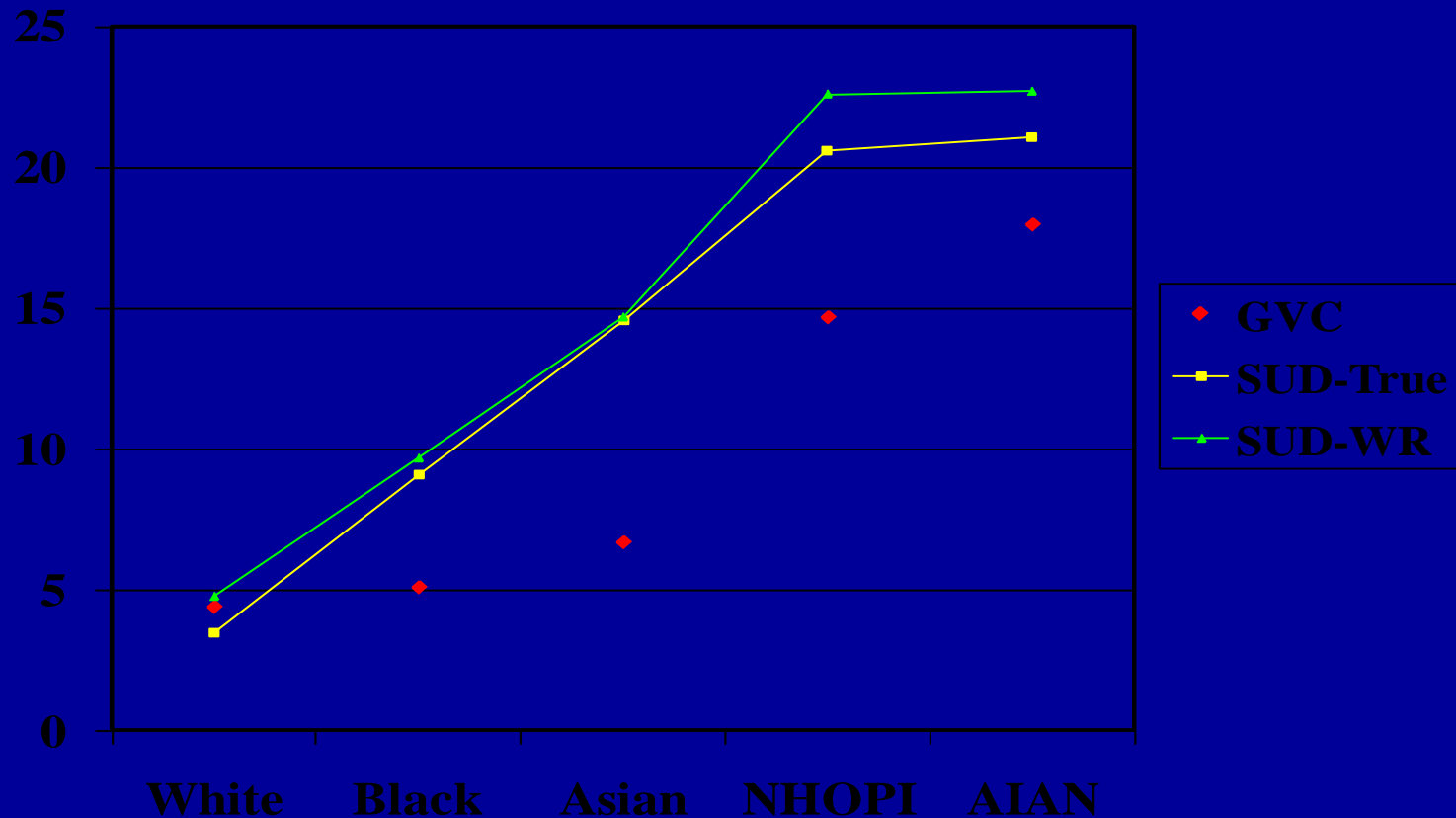
Variances for clustered items (like race, diagnosis, type of provider) are predicted less accurately using the GVC. If you use the GVC, use p = .01, not .05

Comparison of SEs Produced Using GVC, SUDAAN-True, and SUDAAN WR

24
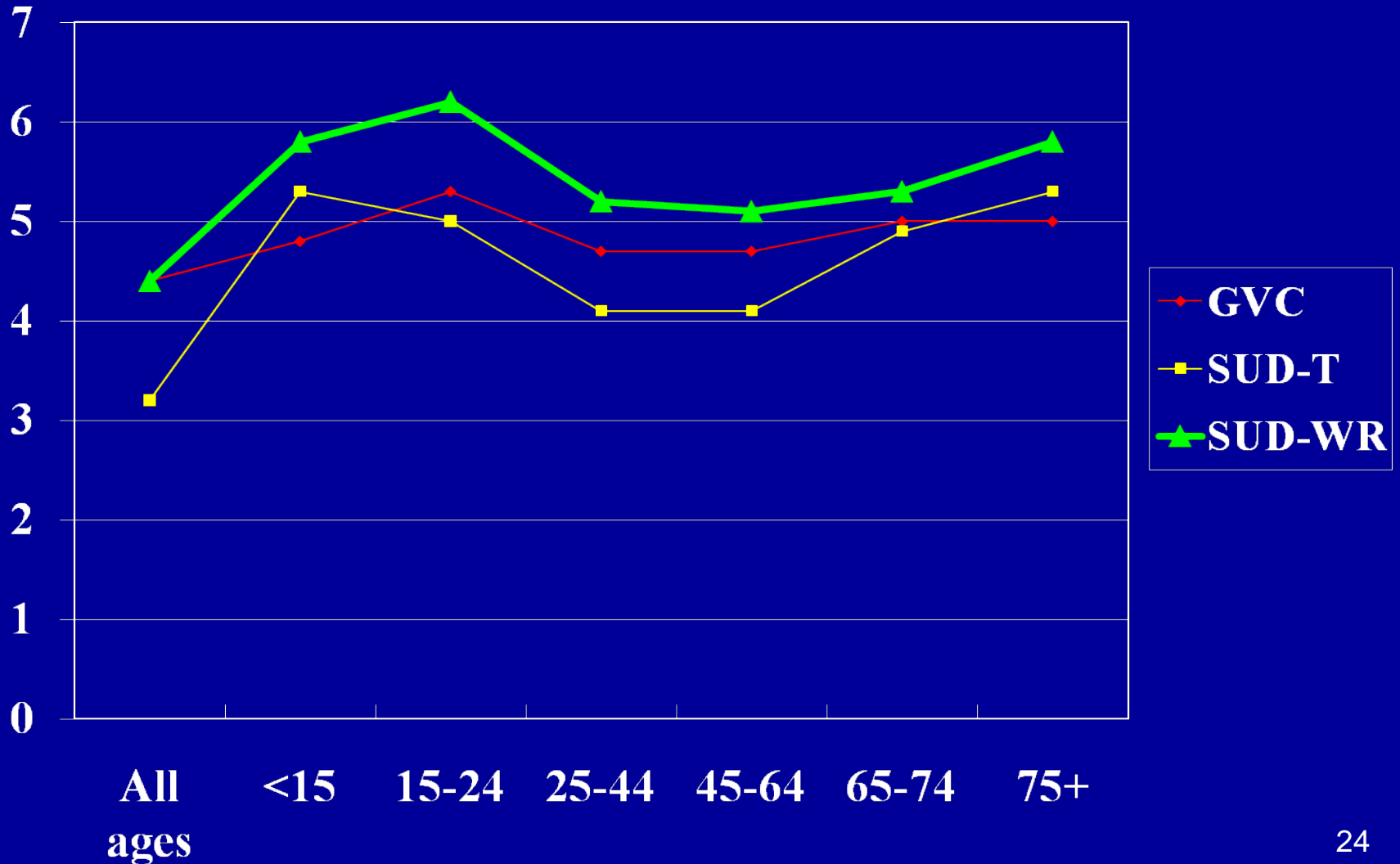
Some User Considerations

- NAMCS/NHAMCS sample visits, not patients

- No estimates of incidence or prevalence

- No state-level estimates

- May capture different types of care for solo vs. group practice physicians

- Data are only as good as what is documented in the medical record

Some User Considerations (cont.)

- High percentage of missing on some data items
  - 2007 NAMCS
    - Ethnicity (34.7%)
      - Imputed and unimputed data
    - Race (31.5%)
      - Imputed and unimputed data
    - Time spent with provider (26.2%)

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

- Each booklet includes:
    - A description of the survey
    - Record format
    - Marginal data (summaries)
    - Various definitions
    - Reason for Visit Classification codes
    - Medication & generic names
    - Therapeutic classes

Where to get more information?

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

- Call the Ambulatory and Hospital Care Statistics Branch at 301-458-4600

- Email jhsiao1@cdc.gov