



STATE HEALTH ACCESS  
DATA ASSISTANCE CENTER

## *The Impact of the American Community Survey on Small Area Estimation*

25th Annual Behavioral Risk Factor Surveillance System  
Conference

Orlando, FL  
March 17th, 2008

**Michael Davern, Ph.D.**

Assistant Professor, Research Director  
SHADAC, Health Policy & Management  
University of Minnesota

*Supported by a grant from The Robert Wood Johnson Foundation*

## Acknowledgements

- This presentation was developed along with Dr. Gestur Davidson, Senior Research Associate and Health Economist at SHADAC at the University of Minnesota
- Thanks to Dr. Mokdad for the invitation to speak as I am an enthusiastic supporter and data user of the BRFSS data.



## Introduction

- Small area estimation requires high quality auxiliary data on “small” geographies
- The goal is to develop educated guesses on the likely rates of characteristics of interest for places without much (or any) sample in the survey (e.g., BRFSS) using auxiliary information
- Auxiliary data often includes a combination of administrative data and census data developed on the areas/geographies of interest
- The source of the basic demographic auxiliary data used in these models is changing as the decennial census long form is replaced with the American Community Survey



3

## Outline of Talk

- This talk will address:
  - What is the American Community Survey?
  - How is it different from the long form decennial census data?
  - What are the strengths and challenges associated with working the ACS into small area estimates?
- Bottom Line: The ACS will provide critical data on a timely basis that will improve small area estimation across the US but it will also introduce new challenges

4

## What is the American Community Survey (ACS)

- The Decennial Census long form data is a fundamental component in most small area estimation efforts
- The American Community Survey (ACS) is the replacement for the Decennial Census long form survey
- The best way to introduce the ACS is by talking about the key differences/similarities from the long form



5

## Key Differences in Design

- ACS data are collected continuously and not just around census day once every ten years
- ACS data collected in one year is released in August of the following year (not 2 or 3 years later)
- ACS is a 3% sample of housing units every year versus a 17% sample once every 10 years
  - Over 10 years of ACS data 30% of households will have been in sample
- ACS samples non-respondents for follow-up by phone and in person
- ACS will have 3 year and 5 year data products
  - 65,000 or more in population, every year
  - 20,000-65,000 3 year average summary tables (first 3-year data products are being released in August of 2008)
  - Less than 20,000 5 year average summary tables (first 5-year data products coming in August 2009)

6

## Key Similarities in Design

- Conducted by the Census Bureau and are “mandatory”
- Two types of data products:
  - Summary table files by geography (down to census tract and even block group)
  - 1% Public Use Microdata Sample (PUMS)
- Most of the content is similar to long form
- Mixed mode data collection (mail, phone and in-person)
- Both use the Census Master Address File (MAF)
- Institutional group quarters included (certified nursing facilities, prisons, and dorms)
  - Homeless shelters are excluded from ACS
- Oversamples rural and less populated areas (sampling fraction varies from 10% to 1.6% of households)

7

## Key Changes to Content

- Income data is collected for reference period of last 12 months instead of last calendar year
- New data elements included in ACS
  - Food stamps, marital history, service-related disability, health insurance coverage
- Content changes we would like to see... Many... but two at the top of my list are:
  - A change in the telephone question to examine changes in telephone technology
    - Cell phone only households
  - Health status question

8

## Sample Size of ACS, CPS and BRFSS by Selected States

**Table X:** Household Sample Sizes by Selected States for the 2006 BRFSS, 2006 Current Population Survey's Annual Social and Economic Supplement (CPS), the 2006 ACS, and the 2006 ACS PUMS

Selected States	BRFSS	CPS-ASEC	ACS	ACS PUMS
District of Columbia	4,023	1,186	3,672	2,612
Wyoming	4,993	932	3,877	2,175
Alaska	2,113	1,013	5,835	2,237
Vermont	7,016	1,052	8,076	2,522
North Dakota	4,780	958	8,258	2,788
Connecticut	8,501	1,656	21,357	13,673
Oregon	4,866	1,020	23,785	15,001
Kentucky	6,174	1,059	28,658	17,486
Oklahoma	7,020	983	29,492	14,461
Minnesota	4,254	1,692	57,762	20,659
Florida	10,726	3,453	103,089	76,842
New York	5,928	3,309	121,011	72,476
Texas	6,854	3,959	129,186	84,474
California	5,707	6,519	178,666	125,071
<b>Total US</b>	<b>347,790</b>	<b>75,939</b>	<b>1,945,237</b>	<b>1,163,343</b>

Note: This table excludes institutionalized group quarters and US territories

9

## Strengths of ACS for Small Area Estimates

- The ACS will produce summary file data and microdata every year
  - Relying on decennial census data that is 10 years out of date is no longer needed
- Data processing is much faster (8 months after data collection is completed)
- Quality of data is improved
  - Professional interviewing staff versus temporary staff
- Will improve intercensal population estimates
- Census is producing variance estimates for summary tables which was missing from long form data
  - Will be able to be used in small area estimation variance to bias tradeoff decisions

10

## Weaknesses of ACS for Small Area Estimates

- In general the annual calendar year estimates will be available for those areas for which the BRFSS is already able to produce reliable estimates
- Areas where the BRFSS needs to borrow strength from auxiliary data the ACS will be producing 3 or 5 year estimates
  - Some small area estimates will be based on five year estimates and others on three year estimates
- Smallest area of identifiable geography on the microdata is the PUMA (which is a aggregation of counties)
  - PUMA is a city, county or group of counties within a state that generally represents about 100,000 people
  - Will have roughly 1,000 people in the 1% ACS PUMS
  - The PUMAs geographies do not match current BRFSS public use geography
  - Would be helpful for small area analysis if the BRFSS public use data also put PUMAs on their public use file

11

## Small Area Estimation Issues

- Researchers can combine the summary file data (especially the 3 and 5 year products) along with the ACS PUMS and BRFSS microdata to produce very good estimates using small area estimation approaches
  - Year can be taken into account and modeled using all the 3 and 5 years worth of data
- Traditional synthetic estimates that just rely on summary data will have issues of reference period to deal with

12

## Conclusions

- ACS will be a great improvement to auxiliary data available for small area estimation using BRFSS and other surveys
  - However it comes with its challenges
    - Most notably the reference period issue for small areas of geography varying between 3-5 years for summary data reference period
    - Getting to know a new data set and dealing with its growing pains

13

## Random Next Steps

- Would be helpful to get PUMA geography on the public use BRFSS as well
- As researchers begin to use the ACS data (both summary files and PUMS) there should be open dialogue to make sure problems with the these data for small area estimates are identified (and reported to Census and the data user community)
- Lobby census/Congress to change the telephone question and add health status

14

## SHADAC contact information

[www.shadac.org](http://www.shadac.org)

State Health Access Data Assistance Center

University of Minnesota

2221 University Avenue, Suite 345

Minneapolis Minnesota 55414

(612) 624-4802

shadac@umn.edu

