A Review of ACS Weighting and Estimation Methods and Associated Issues

Alfredo Navarro, U.S. Census Bureau

Instituto Brasileiro de Geografia e Estatistica Rio de Janeiro, Brazil May 29, 2006

USCENSUSBUREAU

## **Overview of the Session**

- Data Collection
- Sample Design
- Weighting and Estimation
  - Major Components
  - Multi-year Estimates
- ACS Research Weighting and Estimation

# **Data Collection**

## **Data Collection**

- Methodology based on best practices from decennial census and demographic surveys
- Monthly samples using overlapping multi-mode data collection methods
  - Mail
  - Telephone
  - Personal Visit

# **Data Collection**

Month of Data Collection 2005		2006				
ACS Sample Panel	November	December	January	February	March	April
November, 2005	Mail	Phone	Personal Visit			
December, 2005		Mail	Phone	Personal Visit		
January, 2006			Mail	Phone	Personal Visit	
February, 2006				Mail	Phone	Personal Visit
March, 2006					Mail	Phone
April, 2006						Mail

# **Sample Design**

# Sample Design

- Survey designed to include

   U.S. Stateside and Puerto Rico
   Population in both housing units and group quarters (group quarters started in 2006)
- Survey designed to produce annually updated single-year and multi-year estimates

### Sample Design Frame

 Sample cases selected from an updated Master Address File (MAF)

MAF updated through the use of...
 – Postal Service updates in most areas
 – Special field updating in more rural areas

# Sample Design

- Unclustered one-stage systematic sample selected as initial sample each month
- Sub-sample of nonrespondents selected after mail and phone attempts for personal visit follow-up

# **Initial Sampling Rates**

Governmental Unit Size: Estimate of Occupied Housing Units	ACS 1-year Sampling Rates
0 - 200	10.0%
201 - 800	~7.0%
801 - 1200	~3.5%
Census Tract Size	
2000 or less	~2.4%
Over 2000	~1.7%

# Sub-sampling Rates Nonresponse Follow-up

Address and Tract Characteristics	Sub-sampling Rate
Unmailable Addresses	2 – in - 3
Mailable addresses with the lowest mail/CATI rates	1 — in — 2
Mailable addresses in tracts with average mail/CATI rates	2 — in — 5
Other mailable addresses	1 — in — 3

#### USCENSUSBUREAU

# **Weighting and Estimation**

## Annual Weighting Process 3 Major Components

- Initial weights to reflect the probability of selection
- Adjust weights of interviewed households to account for noninterviews
- Adjust weights to independent housing unit and population estimates (controls)

### Initial Weight Probabilities of Selection

- Initial probability of selection is assigned as a function of the sample design
- Nonresponse follow-up (Personal Visit CAPI) sample design

# Initial Weight Variation in Monthly Response Factor

- Seasonal variations in response patterns
- Smooth out the total weight for all sample months

 Makes tabulated HUs in a month = sample HUs in a month

# Variation in Monthly Response Factors

1 <sup>st</sup> Quartile	Median	3 <sup>rd</sup> Quartile
0.875	1.014	1.096

## Nonresponse Adjustment

- The weight of the nonrespondents is transferred to the respondents
- Nonresponse adjustment is carried out at the census tract level for groups of households with characteristics correlated with nonresponse:
  - Census tract
  - Type of building (single vs. multi-unit)
  - Month of data collection

# **Nonresponse Adjustment Factors**

1 <sup>st</sup> Quartile	Median	3 <sup>rd</sup> Quartile
1.000	1.000	1.010

# Ratio Adjustments to Housing Unit and Population Controls

- Intercensal estimates are produced by updating the previous census results using various administrative records data
- In a multi-stage process, housing unit and population adjustment ratios are applied to the weights
- Applied at the county (or group of counties) level by race/ethnicity and age/sex groups

Ratio Adjustments to Controls - Why?

- Reduce variability of the estimates
- Reduce bias
  - Undercoverage of housing units
  - Undercoverage of people within housing units

## Housing and Population Control Factors

	1 <sup>st</sup> Quartile	Median	3 <sup>rd</sup> Quartile
Housing	1.000	1.021	1.052
Population	0.844	1.079	1.397

#### USCENSUSBUREAU

## Weighting and Estimation Single-Year Estimation

### **Estimates include:**

- population estimates
- rates
- means, medians

### Weighting and Estimation Single-Year Estimation

Percent foreign born population in year 1:

 $P_1$  = Percent Foreign Born =

 $\frac{\text{Number Foreign Born}}{\text{Total Population}} = \frac{N_1}{T_1}$ 

## **Multi-Year Estimates**

- Combining or "pooling"
- Population controls
- Tabulation geography
- Inflation adjustments

# Both Single- and Multi-Year Estimates are Period Estimates

- 2005 single-year estimates are based on Jan 2005 – Dec 2005 interviews (12 months)
- 2005-2007 three-year estimates are based on Jan 2005 – Dec 2007 interviews (36 months)
- 2005-2009 five-year estimates are based on Jan 2005 – Dec 2009 interviews (60 months)

## Multi-Year Estimates Pooling Advantages

- Improved accuracy of estimates taking advantage of increased number of sample cases
- More up-to-date controls
- Flexibility of developing weighting procedures
- Production of multi-year data products mirror the 1-year data products

#### USCENSUSBUREAU

### Multi-Year Estimates Population Controls

- Simple average of the set of population controls for the years comprising the multi-year estimate
- For example, for the 2005-2009 five-year estimates, sum the controls released in 2010 for 2005, 2006,...,2009 divided by 5
- Use the most recently released estimates for each year

## Multi-Year Estimates Tabulation Geography

- Boundary changes can occur through annexations during the multi-year period
- Plan is to tabulate using the geography of the most recent year in the multi-year estimate
- For 2005-2009 estimates, tabulate using all interviews for the period of 2005-2009 that were conducted in blocks that define the area in 2009

### Multi-Year Estimates Inflation Adjustments

- The Consumer Price Index is used to compute inflation factors
- Dollar valued data items are inflation adjusted to the most recent year of the period
- For example, for the 2005-2009 estimates, appropriate inflation factors are applied to reported income values for 2005, 2006, ..., 2008 to adjust to 2009 constant dollars

### Multi-Year Estimates Medians

- Medians are produced using combined data records from all years
- A 3-year median household income estimate is determined by combining the household records from the 3 years into one data set and determining the median from this combined distribution

# ACS Research Weighting and Estimation

## ACS Research Weighting and Estimation

- Multi-year estimates
- Family equalization
- Use of alternative population estimates as ACS controls
- Small area estimates

### ACS Research Multi-Year Estimates

- Production of a series of multi-year estimates for the ACS test sites (1999 – 2005)
- The objective of this research is to assess issues of stability and comparability
- A secondary objective is gaining insights to develop and improve user materials

# ACS Research Family Equalization

Assess impact of modifying weighting procedures to ensure consistency of the following estimates:

- Households and householders
- Householders and spouses (and unmarried partners) in married (unmarried) couple (partners) households
- Subfamily husbands and wives in married-couple subfamilies

## ACS Research Use of Alternative Population Controls

- The quality of national and state level population estimates is generally accepted
- The benefits of using of these controls by several current surveys is well documented and understood
- The quality of county level estimates, in particular estimates of detailed demographic groups, has not yet been determined

### ACS Research Use of Alternative Population Controls

- Define alternative population controls based on geographic and demographic detail
- Control to state level estimates by age and sex, and race/ethnicity and county population total only
- Control to county population estimates by sex and age and to place level total population

### ACS Research Small Area Estimates

- Research on 1999-2001 ACS data indicated tract-level standard errors were higher than anticipated
- Lack of tract-level controls identified as likely primary cause

### ACS Research Small Area Estimates

- Research into variance reduction is ongoing
  - Administrative records-based tract-level controls
  - Model-assisted weighting adjustment also incorporating administrative records
- Tract data not published until 2010 lots of time for research

## **Contact Information**



U.S. DEPARTMENT OF COMMERCE U.S. Census Bureau Washington, DC 20233 **Alfredo Navarro** Room: 2023-2 Phone: 301-763-3600 Email: Alfredo.Navarro@census.gov

USCENSUSBUREAU

#### USCENSUSBUREAU