American Community Survey Updates

California State Data Center Meeting

March 15, 2011

Mark E Asiala

U.S. CENSUS BUREAU

U.S. DEPARTMENT OF COMMERCE

Washington, DC 20233

Overview

- Sampling Update
- Working with Small Area Data
  - Tips for dealing with large variances
  - Aggregating Margins of Error
  - Tests of significance
- Population Estimates Used as Controls
- Methods Panels Update
- ACS Selected Population Tables

Sampling Update

- Housing Unit Address Sampling
- Group Quarters Sampling

Housing Unit Address Sampling

Sampling Frame

- Sample from the Master Address File (MAF)
  - Continually updated and maintained
  - Contains all Census 2010 addresses
  - Is updated with the delivery sequence file from the USPS twice a year
  - Is updated from current surveys including ACS

ACS Sample Design

- Main sample – select 99% of annual sample in summer of year prior to sample year
- Supplemental sample – select 1% of total sample in January of sample year
  - Selected from addresses new to the MAF since main sampling.
- Sampling for each phase is done in two stages

ACS Sample Design

- First Stage Sampling
  - Partitions universe into five groups – county level
    - Each is a representative sample of the nation
  - Approximately 20% of addresses are eligible for second stage sampling
  - Achieves unduplication across five year periods

ACS Sample Design

- Second Stage Sampling
  - Independent county sub-samples of first-stage sample
  - Base rate recalculated each year
  - 16 target sampling rates assigned to census tracts and small governmental units based on size (estimated occupied housing units)

ACS Sample Design

- Rate Definitions 2005 to 2010
- Sampling rates function of base rate (BR)
- One fixed rate stratum

Stratum	Block MOS Criteria	Sampling Rates
5	0 < GUMOS ≤ 200	10% (fixed)
2	200 < GUMOS ≤ 800	3×BR
3	800 < GUMOS ≤ 1,200	1.5×BR
1	TRACTMOS ≤ 2,000	BR
4	2,000 < TRACTMOS	0.735×BR

Reallocation of the HU Address Sample

- Five Year CVs for Typical Tracts, by Size Class

Tract Size Category	Average Tract Size	CV
0 – 400	291	66%
401 – 1,000	766	41%
1,001 – 2,000	1,485	29%
2,000 – 4,000	2,636	26%
4,000 – 6,000	4,684	19%
6,000 +	8,337	15%

Reallocation of the HU Address Sample - Improvement

- Increase the number of sampling strata
  - Smaller stratum intervals allows smoother transitions between rates

- Increase sampling rates for blocks in the very smallest governmental units
  - Increase reliability of the estimates

Reallocation of the HU Address Sample – 2011 Stratification

- New Stratification (small GUs)
  - increased number of fixed rate strata
  - increased the rates

Stratum	Block MOS Criteria	Sampling Rates
1	0 < GUMOS ≤ 200	15% (fixed)
2	200 < GUMOS ≤ 400	10% (fixed)
3	400 < GUMOS ≤ 800	7% (fixed)
4	800 < GUMOS ≤ 1,200	2.8 × BR

Reallocation of the HU Address Sample – 2011 Stratification

Stratum	Block MOS Criteria	Sampling Rates
5	0 < TRACTMOS ≤ 400	3.5×BR
6	0 < TRACTMOS ≤ 400 H.R.	0.92×3.5×BR
7	400 < TRACTMOS ≤ 1,000	2.8×BR
8	400 < TRACTMOS ≤ 1,000 H.R.	0.92×2.8×BR
9	1,000 < TRACTMOS ≤ 2,000	1.7×BR
10	1,000 < TRACTMOS ≤ 2,000 H.R.	0.92×1.7×BR
11	2,000 < TRACTMOS ≤ 4,000	BR
12	2,000 < TRACTMOS ≤ 4,000 H.R	0.92×BR
13	4,000 < TRACTMOS ≤ 6,000	0.6×BR
14	4,000 < TRACTMOS ≤ 6,000 H.R.	0.92×0.6×BR
15	6,000 < TRACTMOS	0.35×BR
16	6,000 < TRACTMOS H.R.	0.92×0.35×BR

Reallocation of the HU Address Sample – More equitable results

- Five Year CVs for Typical Tracts, by Size Class

Tract Size Category	Average Tract Size	Old Stratification CVs	New Stratification CVs
0 – 400	291	66%	41%
401 – 1,000	766	41%	30%
1,001 – 2,000	1,485	29%	29%
2,000 – 4,000	2,636	26%	29%
4,000 – 6,000	4,684	19%	29%
6,000 +	8,337	15%	28%

Sampling for Computer Assisted Personal Interview (CAPI)

- Increase Sampling Efficiency
- CAPI Eligible Universe
  - All addresses that do not respond by mail or Computer Assisted Telephone Interview
  - All non-mailable addresses

Sampling for Computer Assisted Personal Interview (CAPI)

- Sampling Rates
  - Every tract assigned to a sampling stratum
  - Four rates: 1-in-2, 1-in-3, 2-in-5, 2-in-3 (non-mailable only)
  - Rates based on historical mail/CATI response patterns

Group Quarters Sampling

Group Quarters Sampling

2008 ACS Definition (partial):

"A group quarters is a place where people live or stay, in a group living arrangement, that is owned or managed by an entity or organization providing housing and/or services for the residents. This is not a typical household-type living arrangement. …"

http://www.census.gov/acs/www/Downloads/2008_ACS_GQ_Definitions.pdf

Major GQ Type Groups (#s)

- Correctional institutions (1)
- Juvenile facilities (2)
- Nursing homes (3)
- Other long-term care facilities (4)
- College dorms (5)
- Military facilities (6)
- Other non-institutional GQs (7)

Why Do We Include GQ Residents in the ACS?

- ACS replaces the decennial census long form – total population
- ~ 7.8 million GQ residents in Census 2000. ~ 8.2 million from 2007 ACS.
- Can be a large component of total population for some small areas

Current Design

- Produce robust state level characteristic estimates of the GQ resident *population*
- Design could support state major GQ type group characteristic estimates also

Sample Design Overview

- Two phase sample
- Identify GQs to conduct interviews
- Select small GQs to conduct interviews
- Subsample (if needed) to identify people in sample
  - Automated in instrument

Sample Design Overview

- Independent samples by state
- Large and small GQ sample design
  - Small GQs eligible for sampling only once in a five-year period
  - Large GQs eligible for sampling each year
- Systematic selection
  - Sorted by state, GQ type, sub-state geography, address, special place, GQ

Sample Design Overview

- Stratification
  - Small GQs < 16 expected population
  - Large GQs ≥16 expected population
- Sampling rates
  - 2.5% in 2006 and 2007
  - In 2008 15 small states increased to > 2.5%

Challenges

- No systematic process to update frame
  - GQs deleted from frame faster than added
- Majority of data on frame from Census 2010
- Sparse samples
  - May miss entire GQ population in small areas

Challenges

- State-level design doesn't support current data release strategy
  - Total population for sub-state areas
- Incorporating the 2010 Census results
  - Refresh of ACS frame
  - Matching to existing frame or start new?

Working with Small Area Data

Issues of Variances and Margins of Error

- In general, the smaller the sample size the larger the sampling error
- Factors that can impact variance of the estimate of concern
  - Time specificity (1-year vs. 3-year est.)
  - Geographic specificity (tract vs. county)
  - Table specificity (0-4 vs. 0-17 age cat.)
- Sampling variability not unique to ACS

Methods to Address Issues of Higher than Desired Variances

- Solutions follow their sources
  - Use multiyear estimate in place of a single-year estimate
  - Aggregate areas: for example, tracts into neighborhoods
  - Aggregate detail: for example add the 0-4 and 5-17 cells to for 0-17 estimate of age

Aggregating Estimates and Calculating Margins of Errors

Overview of this Section

- How to aggregate estimates and approximate the associated margin of error (MOE).
- Two examples will be presented
- Discuss issues associated with derived MOE

Aggregating Estimates

- Published estimates may be aggregated to form additional estimates.
- With 5-year data, Tract/Block Group estimates can be aggregated to form user defined areas.
- Calculating the MOE for these defined areas is also important.

MOEs of Aggregated Estimates

- Estimate of sum is obtained by adding the published estimates
- Cannot simply sum the MOEs together
- The actual formula is

$$MOE\left(\hat{X} \pm \hat{Y}\right) = \sqrt{[MOE(\hat{X})]^2 + [MOE(\hat{Y})]^2 + \text{covariance}}$$

- Covariance is not published so the approximation is used covariance = 0.

Example 1: Total number of people with income below the poverty level

Characteristics	Estimate	Margin of Error	MOE squared
Males	42,945	4,653	21,650,409
Females	61,956	5,723	32,752,729

Total = 42,945 + 61,956 = 104,901

$$MOE(\text{Male} + \text{Female}) = \sqrt{21{,}650{,}409 + 32{,}752{,}729} \approx 7{,}376$$

So the total is 104,901 with an approximate MOE of 7,376.

Example 1: Total number of people with income below the poverty level

Characteristics	Estimate	Margin of Error (Published)	Margin of Error (Approximated)
Total	104,901	9,224	7,376

$$MOE\left(\hat{X} \pm \hat{Y}\right) = \sqrt{[MOE(\hat{X})]^2 + [MOE(\hat{Y})]^2 + \text{covariance}}$$

Covariance = 30,679,038

$$MOE(Male + Female) = \sqrt{21{,}650{,}409 + 32{,}752{,}729 + 30{,}679{,}038} = 9{,}224$$

Example 2: Total number of males with income below the poverty level

$$MOE(\text{Male}) = \sqrt{2{,}637{,}376 + 1{,}946{,}025 + 1{,}052{,}676 + 3{,}644{,}281}$$
$$\approx 3{,}046$$

- So the total is 23,001 with an approximate MOE of 3,046.

Characteristics	Estimate	Margin of Error	MOE squared
Wyoming	23,001	3,309	10,949,481
PUMA 00100	5,264	1,624	2,637,376
PUMA 00200	6,508	1,395	1,946,025
PUMA 00300	4,364	1,026	1,052,676
PUMA 00400	6,865	1,909	3,644,281

The Adjusted Covariance Matrix

- The values on the diagonal in blue are the squared MOEs. The off diagonals are the covariance.
- We can see that the covariance range in size and are non-trivial.

Covariance Matrix	Puma 00100	Puma 00200	Puma 00300	Puma 00400
Puma 00100	2,636,050	-89,304	302,634	223,141
Puma 00200	-89,304	1,945,210	-140,590	330,487
Puma 00300	302,634	-140,590	1,052,505	209,910
Puma 00400	223,141	330,487	209,910	3,644,845

What can be done?

- We have found that the approximation formula seriously breaks down when aggregating more than four estimates.
- So we suggest you aggregate the fewest number of estimates as possible.
- Other Options:
  - Calculate the estimates using the Public Use Microdata Sample (PUMS)
  - Request a special tabulation (fee based and certain criteria apply)

Statistical Testing

Case Study

Tracking Economic Well-Being in Washington, DC

- In 2005, city implements a series of job training initiatives to increase employment and reduce poverty rates
- In 2008, public officials want to assess changes in poverty rates in the city

Finding the Data

- Washington, DC has a population size greater than 65,000
- Comparable data for both 2006 and 2007 are available from the ACS
- Examine change in the percent of people living in poverty from 2006 to 2007

Finding the Data

2006 ACS data for Washington, DC

19.6% % of all people living in poverty

1.4% Margin of error

2007 ACS data for Washington, DC

16.4% % of all people living in poverty

1.4% Margin of error

Standard Error (SE)

- Definition
  - A measure of the variability of an estimate due to sampling
  - Depends on variability in the population and sample size
  - Foundational measure

Standard Error (SE)

- Formula
  - SE = MOE / 1.645
- 2007 ACS Data for Baltimore City:
  - 52.1% Percent of males who have never married
  - 1.7% Margin of Error
  - SE = 1.7% / 1.645
  - SE = 1.033%

Are the Estimates Reliable and Usable?

Check Coefficient of Variation for each estimate

2006: SE = 0.85% = (1.4% / 1.645)

CV = 4.3% = (0.85% / 19.6%) * 100

2007: SE = 0.85% = (1.4% / 1.645)

CV = 5.2% = (0.85 %/ 16.4%) * 100

Result = Both estimates are reliable

Comparing the Estimates

Compare Confidence Intervals:

2006: 18.2% - 21.0% (19.6 +/- 1.4)
2007: 15.0% - 17.8% (16.4 +/- 1.4)

- Is there a significant difference?

Test of Statistical Significance

Definition

A test to determine if it is unlikely that something has occurred by chance

A “statistically significant difference” means there is statistical evidence that there is a difference

Conducting Tests of Statistical Significance

Formula

$$\left| \frac{\hat{X}_1 - \hat{X}_2}{\sqrt{SE_1^2 + SE_2^2}} \right| > Z_{CL}$$

where Z is the critical value for the desired confidence level

for 90% confidence level = 1.645

Testing for Statistical Significance

Substituting the appropriate values:

$$\left| \frac{19.6 - 16.4}{\sqrt{(0.85)^2 + (0.85)^2}} \right| = 2.662$$

- 2.662 > 1.645
- Difference is statistically significant at the 90% confidence level

Drawing Appropriate Conclusions

- Short-term fluctuations versus real trends
- Increasing confidence level to 95% or 99%
  - For 95% conf. level compare to 1.960
  - For 99% conf. level compare to 2.576

Accuracy Documents

- The ACS Accuracy document contains:
  - All formulas used in this presentation.
  - More examples
- It is available at http://www.census.gov/acs/www/data_documentation/documentation_main/

Questions?

Population Estimates and Controls for the ACS

Overview

- Population estimates - what we produce and how
- Postcensal versus intercensal population estimates
- Population estimates as controls for the ACS
- Types of population controls for the ACS by year of release

Estimates Produced Annually

- Population
  - Nation by age, sex, race, and Hispanic origin
  - States by age, sex, race, and Hispanic origin
  - Counties by age, sex, race, and Hispanic origin
  - Incorporated places and minor civil divisions (total population only)
  - Puerto Rico Commonwealth and municipios by age and sex
- Housing units
  - States
  - Counties

Producing Population Estimates

- Estimates base is most recent Census with some modification (e.g., Some Other Race is recoded).
- From the last Census forward, population is estimated using a cohort-component method at the national, state, and county levels.
  - Nation:

    $Population_2 = Population_1 + Births - Deaths + NIM$

    NIM = Net international migration
  - States and counties:

    $Population_2 = Population_1 + Births - Deaths + NM$

    NM = Net domestic and international migration
- Subcounty estimates produced using a distributive housing unit method.

Postcensal Versus Intercensal Estimates

- Postcensal population estimates
  - Built off of the last census
  - “Vintage” identified by terminal year in the series
  - July 1 estimates, full series from last Census date forward (for Vintage 2009, series is April 1, 2000-July 1, 2009)
- Intercensal population estimates
  - Based on two consecutive censuses
  - Will be produced by age, sex, race, and Hispanic origin at the county level
  - Will be used to control 2010 ACS products

Population Estimates as Survey Controls for ACS

- Population estimates are the official estimates for the nation, states, counties, cities, and towns.
- Population estimates are used as survey controls for the ACS to reduce variance and coverage bias.

Population Controls Provided to ACS

- Population estimates provided as controls
  - County by age (single years 0-84,85+), sex (male, female), race (31 races), and Hispanic origin (Hispanic, non-Hispanic)
  - Puerto Rico municipios by age (single years 0-84, 85+) and sex
  - For ACS 2009 and beyond – subcounty total population estimates
- Group quarters population by the 7 major types at the state level and for Puerto Rico
- Housing units at the county level and subcounty level

ACS Controls

- ACS creates their set of controls from the population estimates for weighting areas which are counties or groups of counties
  - 13 age groups
  - 5 race alone categories (non-Hispanic)
  - Hispanic
- Group quarters controlled at the state level by type (7 major types)
- ACS uses the housing unit estimates to control the number of housing units in a weighting area

Population Controls by ACS Release Year: 2010

ACS Data	Year of Pop Control	Release Year
2009 ACS	July 1, 2009 controls from Vintage 2009 estimates (Census 2000 base)	2010
2007-2009 ACS	Average of July 1, 2007-2009 controls from Vintage 2009 estimates (Census 2000 base)	2010
2005-2009 ACS	Average of July 1, 2005-2009 controls based on Vintage 2009 estimates (Census 2000 base)	2010

Population Controls by ACS Release Year: 2011

ACS Data	Year of Pop Control	Release Year
2010 ACS	July 1, 2010 controls based on Intercensal estimates (based on Census 2000 and Census 2010)	2011
2008-2010 ACS	Average of July 1, 2008-2010 Intercensal Estimates	2011
2006-2010 ACS	Average of July 1, 2006-2010 Intercensal Estimates	2011

Note: No intercensal subcounty estimates for 2010

Population Controls by ACS Release Year: 2012

ACS Data	Year of Pop Control	Release Year
2011 ACS	July 1, 2011 controls from Vintage 2011 (Census 2010 base)	2012
2009-2011 ACS	Average of July 1, 2009-2011 controls based on Intercensal Estimates and Vintage 2011 (Census 2010 base)	2012
2007-2011 ACS	Average of July 1, 2007-2011 controls based on Intercensal Estimates and Vintage 2011 (Census 2010 base)	2012

Possible Issues with Population Controls 2010 and Beyond

- Timing of the production of intercensal estimates
  - Receive input files March 2011
  - Population controls need to be delivered to ACS May 2011
- For 2010 ACS data products, subcounty controls based on Vintage 2010 population estimates which are based on Census 2000

Questions?

Methods Panels Updates

ACS Content Policy

- New content for the ACS is determined by the OMB in consultation with Census
- Consider issues such as:
  - Frequency of data collection
  - Level of geographic detail needed
  - Other sources of data
- OMB's responsibility to ensure that respondent burden kept to minimum

ACS Content Testing History

- 2006 Content Test
  - Health insurance
  - Marital history
  - Veteran's service-connected disability rating
- 2007 Content Test: Field of Bachelor's degree
- 2010 Content Test

Timing for Testing Content Changes

- Fall/Winter 2008:
  - Worked with OMB Interagency Committee for the ACS to identify topics
- Spring/Summer 2009:
  - Conducted cognitive testing in English and Spanish
- Fall 2009 through Summer 2010:
  - Developed questionnaires/software for field test
- Fall 2010:
  - Conducted field test in mail, telephone and personal visit modes

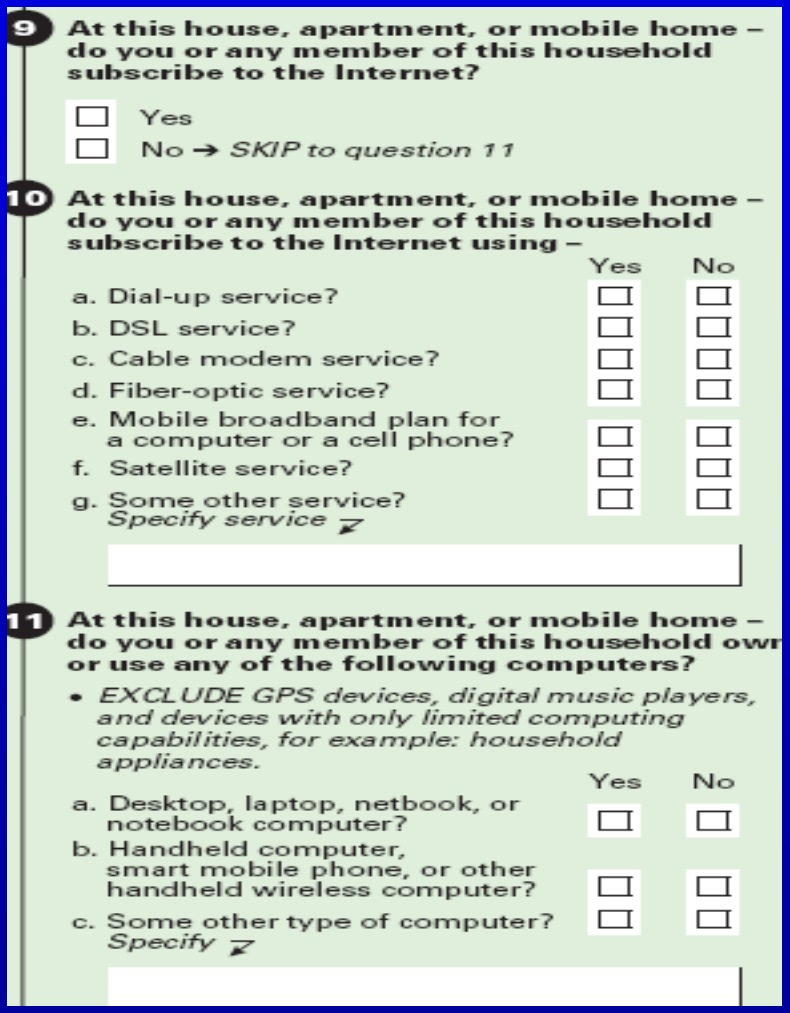
Timing for Testing Content Changes

- Winter 2010/Spring 2011:
  - Completed coding and analysis
- Summer/Fall 2011:
  - Prepare reports and brief stakeholders on results
- Winter 2011/Spring 2012:
  - OMB clearance process and approval of new and revised content
- January 2013 (or later):
  - Implement new and revised content into production

2010 ACS Content Test Questions

- New:
  - Household questions on Computer Ownership and Internet Access
  - Individual questions on Parental Place of Birth
- Revised:
  - Food Stamps - changing program name to SNAP
  - Veteran Identification and Period of Service
  - Public Assistance Income
  - Wages/Salary/etc. and Interest/Dividends/etc. income: changes to automated instruments only

Computer Ownership and Internet Access



9 **At this house, apartment, or mobile home – do you or any member of this household subscribe to the Internet?**

- ☐ Yes
- ☐ No → *SKIP to question 11*

10 **At this house, apartment, or mobile home – do you or any member of this household subscribe to the Internet using –**

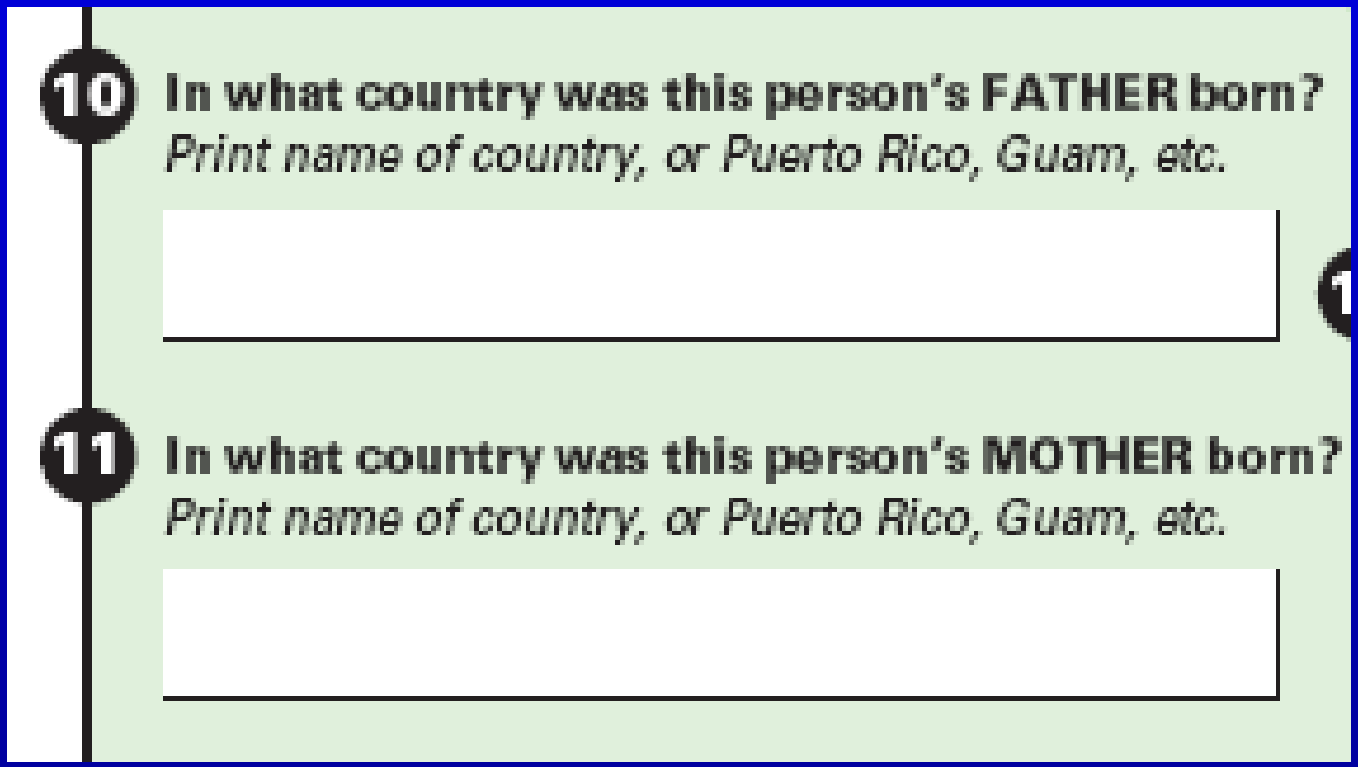
	Yes	No
a. Dial-up service?	☐	☐
b. DSL service?	☐	☐
c. Cable modem service?	☐	☐
d. Fiber-optic service?	☐	☐
e. Mobile broadband plan for a computer or a cell phone?	☐	☐
f. Satellite service?	☐	☐
g. Some other service? *Specify service*	☐	☐

11 **At this house, apartment, or mobile home – do you or any member of this household own or use any of the following computers?**

- *EXCLUDE GPS devices, digital music players, and devices with only limited computing capabilities, for example: household appliances.*

	Yes	No
a. Desktop, laptop, netbook, or notebook computer?	☐	☐
b. Handheld computer, smart mobile phone, or other handheld wireless computer?	☐	☐
c. Some other type of computer? *Specify*	☐	☐

Parental Place of Birth



2010 ACS Content Test - Revised Questions

- Veteran Identification and Period of Service
- Public Assistance Income
- Wages/Salary and Interest/Dividends income (CATI/CAPI only)
- Food Stamps
  - Change program name to SNAP

2010 ACS Content Test

- Field test wrapped up in December
- Content changes in ACS production in January 2013 or later

2011 Questionnaire Size Test

- Goal: to accommodate new content on mail form
- First step – small test studying impact on response rate from increasing form length to 36 pages (currently 28 pages)
- Scheduled for July 2011

ACS Content Reinterview Survey

- Study response error for ACS questions using reinterview for sample of production cases
  - Provide concrete estimates of response error
  - Help identify items for content testing
- Field January – December 2012

Testing an Internet Mode for the ACS

- Evaluating the feasibility of an ACS Internet response option in English and Spanish
- Experiment in 2011 will compare strategies for offering the Internet option in order to maximize combined response by Internet and mail
- Additional testing in 2012
- Possible implementation in 2013

Questions?

ACS Selected Population Products for ACS

Two Products

- Selected Population Tables (SF4-like)
- AIAN Tables
- Both include
  - Data profiles
  - Detailed tables
  - Imputation tables
  - Geographic comparison tables are proposed

Characteristics

- Full suite of characteristics including:
  - Demographic
  - Social
  - Economic
  - Housing
- All tables would be iterated by population group

Geography

- Selected Population Tables
  - Variety of geographies
  - Lowest level is census (and tribal) tract
- AIAN Tables
  - Fewer geographies included
  - Will include AI and AN areas as some of the lowest levels available

Population Thresholds and Iteration Groups

- Two thresholds to be aware of
  - Population group
  - Geographic area
- Population group
  - 7,000 nationally for selected population
  - 100 for AIAN tables
- Geography
  - 50 unweighted sample cases in the area

Frequency

- First release is planned for early 2012 based on 2006-2010 data
- Subsequent releases will be every 5 years after that for both products
- Available on AFF, DVD and download

Questions?

Thank you.

- Mark E. Asiala
  - mark.e.asiala@census.gov