Topics on American Community Survey

California Regional / Affiliate Data Center Meeting
June 1, 2012

Mark E Asiala
Chief, ACS Estimation Branch
Decennial Statistic Studies Division
US Census Bureau

United States Census Bureau
Measuring America

Overview

- Overview of ACS
- Internet Mode Update
- Integrating Results of 2010 Census
- Aggregating Margins of Error
- Sampling Update
- New GQ Estimation Procedures
- Special Data Products

Overview of ACS (High-level)

- Purpose
- Type of data collected
- Data collection
- Data products
- Availability of data

Purpose

- Part of plan for a redesigned decennial program
  - ACS collect all “long-form” sample characteristics on an annual basis
  - Decennial census focus on collecting critical 100% data items with improved quality
- ACS would provide higher quality, more frequent data

Type of Data Collected

- All census long-form style data
  - Demographic (age, race, sex, Hispanic, etc.)
  - Social (education, enrollment, etc.)
  - Economic (income, occupation, etc.)
  - Housing (tenure, home value, rent, etc.)

Data Collection

- Annual sample of 3.54 million housing unit addresses annually
- Everywhere, every month
- Annual sample of group quarters persons of 200,000

Data Collection (2)

- Multimode survey using primarily mail with telephone and personal visit nonreponse follow-up
- Monthly samples are interviewed over a 3 month data collection window
  - Mail (first month)
  - Telephone (second month)
  - Personal Visit (third month)

Data Products

- Annual processing
- Three primary data products
  - 1-year estimates for areas 65,000 plus
  - 3-year estimates for areas 20,000 plus
  - 5-year estimates for all areas including tracts
- 5-year data product is the source for long-form type estimates at all geographies
- Public Use Microdata Sets

Availability of Data

- All ACS data is available on AFF except block group data and a few large tables
- Also available as summary files for download

General Questions about the ACS?

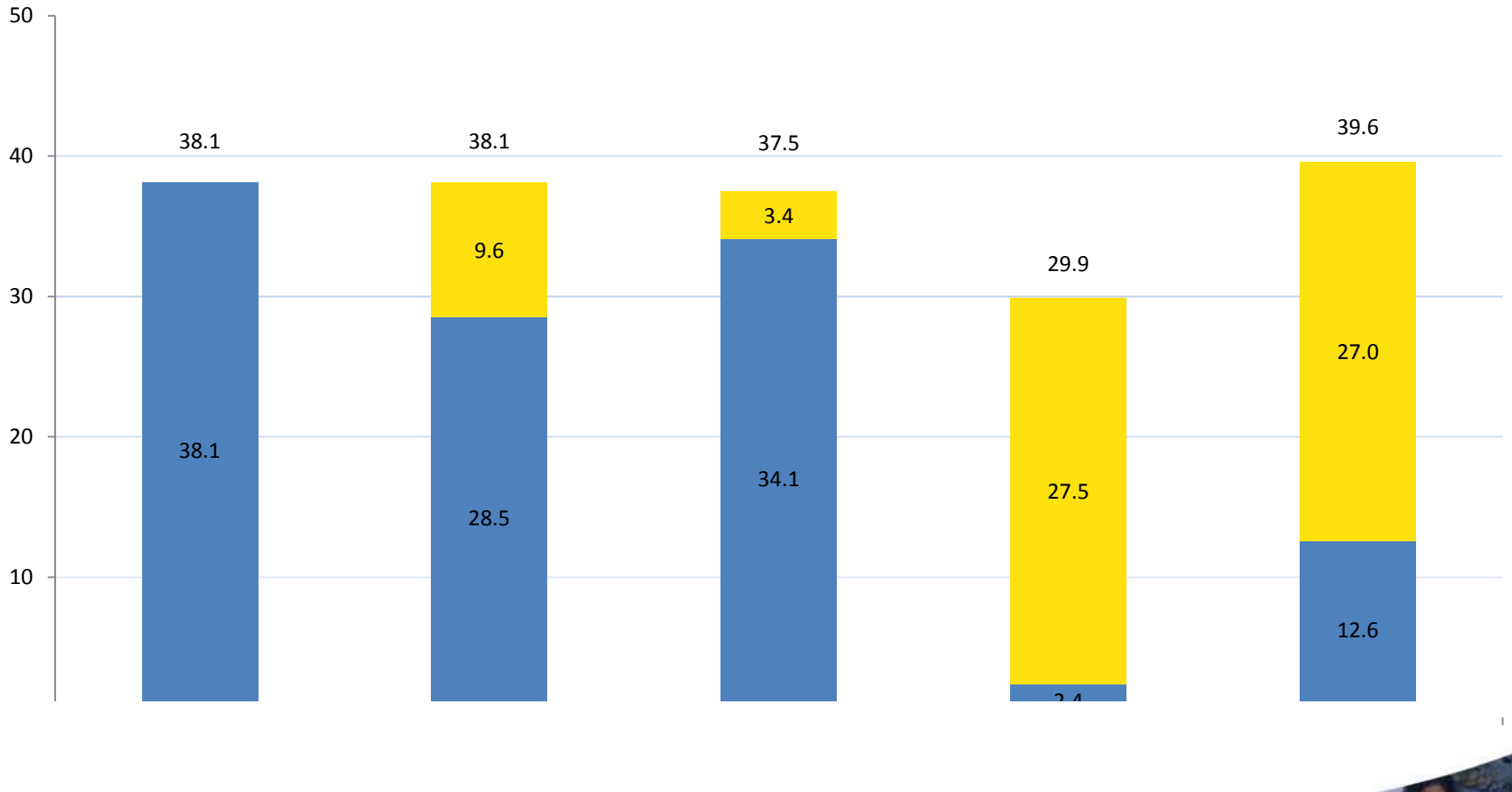
Internet Mode of Data Collection Update

- Results from
  - April 2011 ACS Internet Test
  - November 2011 ACS Internet Follow-up Test

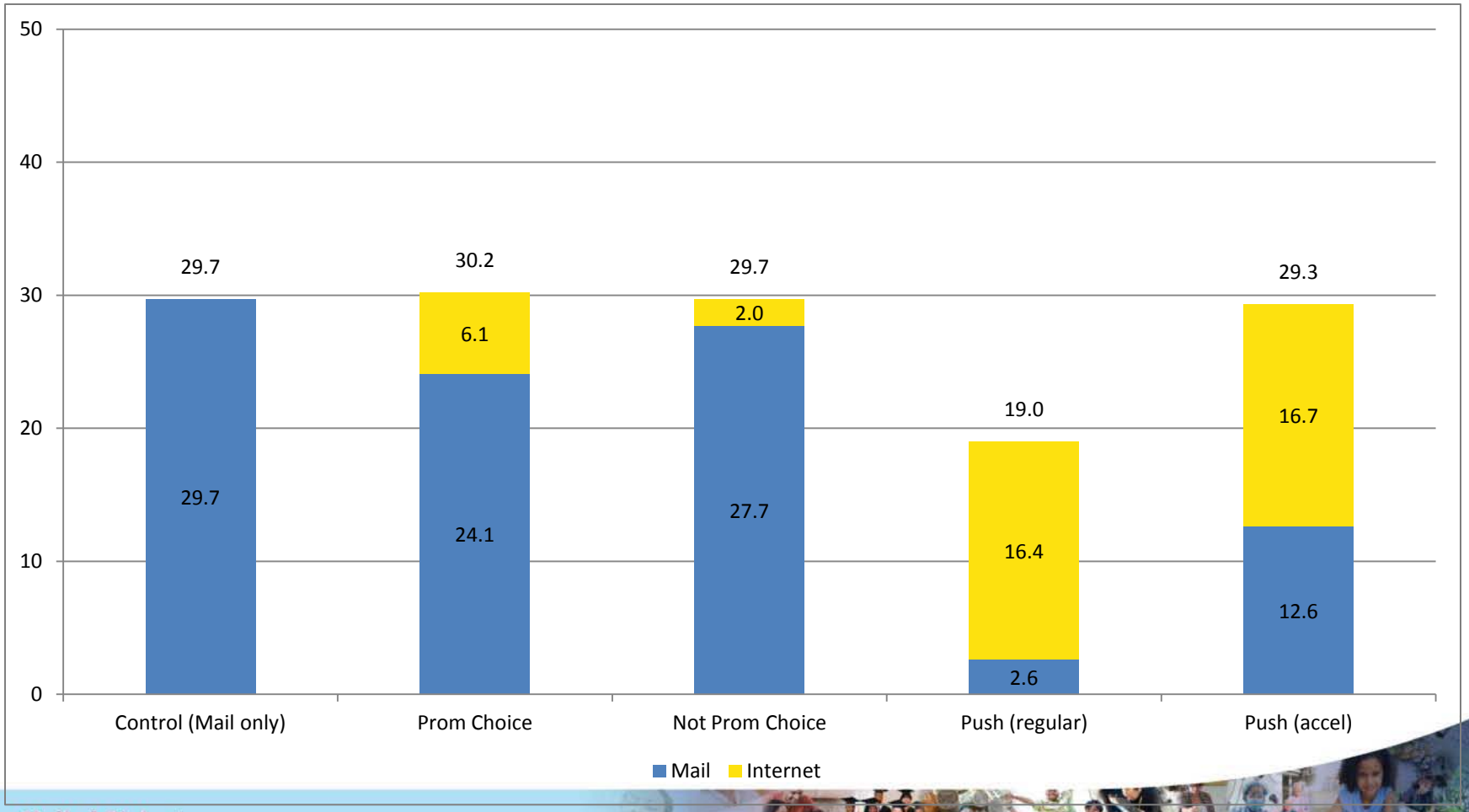
April 2011 ACS Internet Test

- Goal was to demonstrate feasibility and determine best method
- Four Treatments were tested
  - Prominent Choice
  - Not Prominent Choice
  - Push internet on regular mail schedule
  - Push Internet on accelerated mail schedule
- Two strata (targeted / non-targeted)

Self-Response Rates by Mode by Treatment in Targeted Stratum (April Test)



Self-Response Rates by Mode by Treatment in Nontargeted Stratum (April Test)

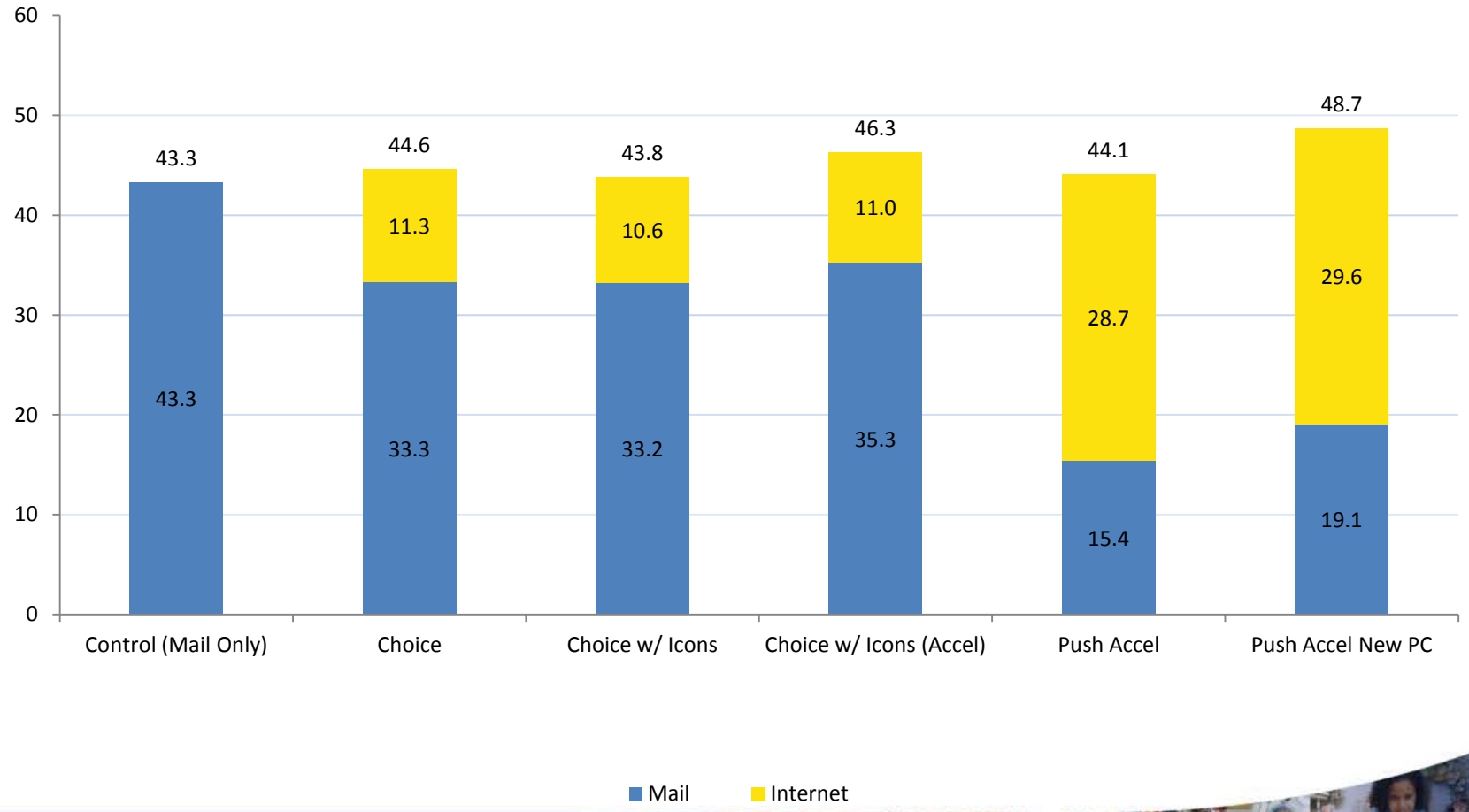


United States Census Bureau
Measuring America

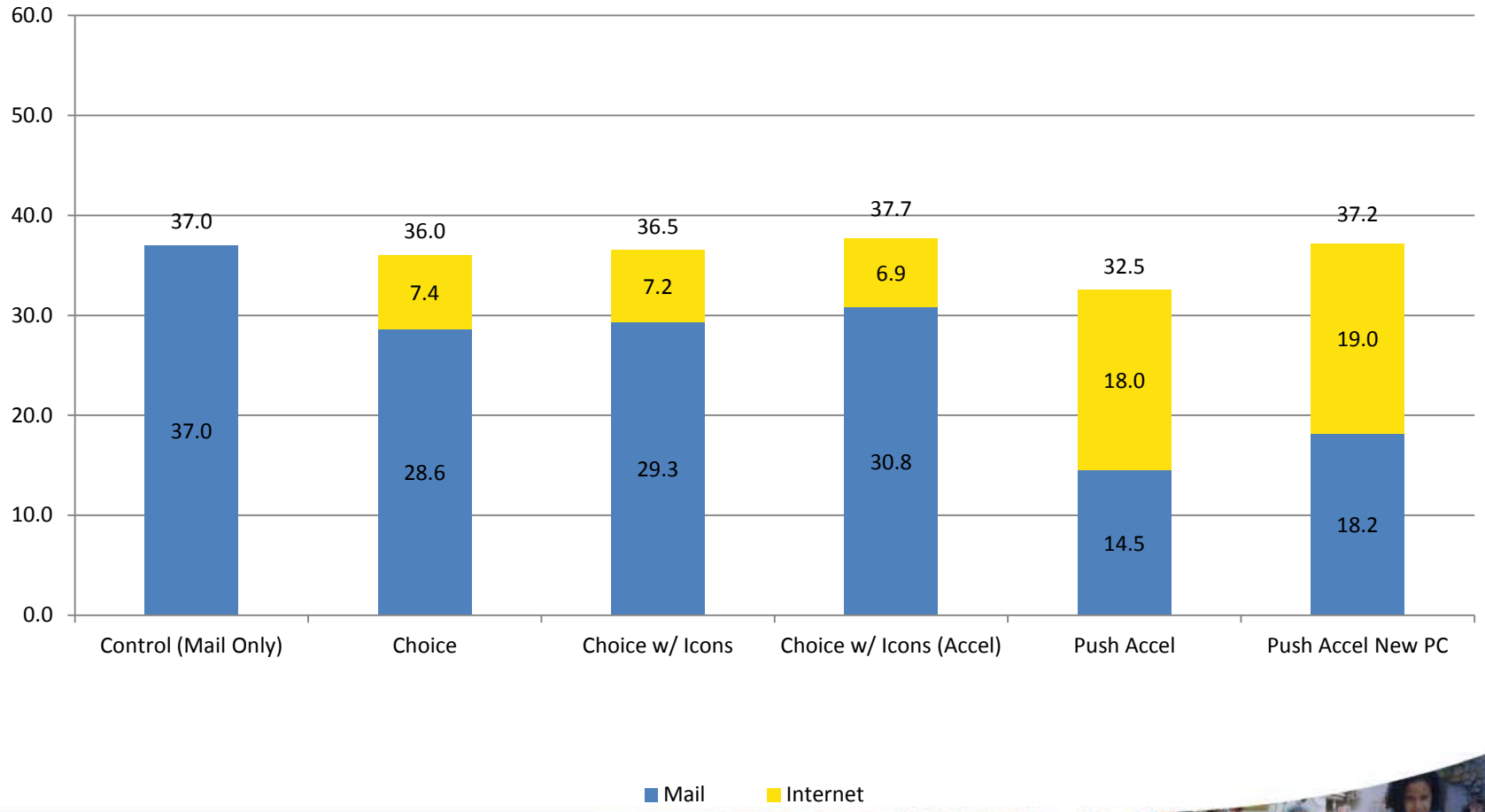
November 2011 ACS Internet Follow-up Test

- Goal was to test some refinements and tweaks to the first test
- Five treatments
  - Choice (same as prominent choice)
  - Choice with Icons
  - Choice with Icons, Accelerated Schedule
  - Push Internet, Accelerated Schedule
  - Same as above, plus reminder post card

Self-Response Rates by Mode by Treatment in Targeted Stratum (November Test)



Self-Response Rates by Mode by Treatment in Nontargeted Stratum (November Test)



Overall results

- Should compare ‘within’ NOT ‘between’ tests
- In the targeted strata
  - Push accelerated with post card had highest response rates
  - Statistically significant increase in self-response rate over control (~5.4%)
- In the non-targeted strata
  - Change in response rate not sig. different from control
  - Still have about ½ responses from internet

Plans / More Information

- Census Bureau plans to offer internet response option to the ACS in 2013
- Full reports are available on ACS website

April 2011 Test

http://www.census.gov/acs/www/Downloads/library/2012/2012_Tancreto_01.pdf

November 2011 Test

http://www.census.gov/acs/www/Downloads/library/2012/2012_Matthews_01.pdf

Integrating Results from the 2010 Census

- Housing unit sampling frame
- Group Quarters sampling frame
- Controls for ACS weighting

Housing Unit Sampling Frame

- Updates to the HU sampling frame during the decade primarily come from the postal service' Delivery Sequence File
- In some areas, this file cannot be used because of high risk of duplication
- Can be coverage issues with the file

Housing Unit Sampling Frame (2)

- The block canvass for decennial census is the largest field listing of addresses in the decade.
- ACS HU sampling began using the results of address canvassing in 2010 sample
- ACS will benefit from the final census cleanup of addresses in 2012 sample

Group Quarters Sampling Frame

- Group quarters frame is particularly dependent on the decennial update
- During the decade, most updates are for adult correctional facilities
- Some one-time updates were performed for military GQs and migrant worker camps

Group Quarters Sampling Frame (2)

- 2010 Census was first update to much of the rest of the frame
- For 2011
  - Frame was mostly 2010 Census GQ universe
  - A few GQs from 2010 ACS frame which we could not match to Census were kept on the frame
- For 2012
  - Updated the 2011 frame with final Census results
  - Removed duplicates, housing units, etc.

Controls for ACS Weighting

- 2010 Census results were used to create the intercensal estimates for 2000-2009
- Also were used to create simple extrapolated estimates for 2010
- Both of the above were used in processing the 2010 data products last year.
- 2011 will use the first post-censals based on 2010

Questions

Getting the Estimates You Need with Margins of Error

- Sometimes you may need to aggregate estimates because
  - You need an estimate for a less specific group than the detail it is published
  - You want a more robust estimate
  - You are creating custom geographies based on adding/subtracting published areas

First Steps

- Need a less specific estimate? --- Verify that it is not available elsewhere
- Check:
  - Profile tables (DP02 - DP05)
  - Subject tables (Sxxx)
  - (Collapsed) Detailed Tables (Cxxxxx or Dxxxxx)

Example: Unemployment rate 16+

- B23001: Sex by Age by Employment Status
  - Requires aggregating estimates unemployed and labor force over 26 cells
  - Still need to calculate the percentage
- This will require a lot of work calculating and the MOE will be less accurate
- However, S2301: Employment Status has it calculated already down to tracts

First Steps (2)

- So if you determine you can't find what you need elsewhere then try to assemble:
  - Source table with fewest cells that need to be combined
  - Source geographies that require the fewest areas to be combined
- Not only is it less work, your MOEs will be more accurate

Example: Unemployment rate 16-24 Males

- Suppose interested in unemployment rate for 16-24 males
- B23001 has ages 16-19, 20-21, 22-24
- C23001 has age ranges 16-19, 20-24
- Using the collapsed table will be easier and your calculated MOE generally will be more accurate

Example: Geographic Areas

- Suppose you are interested aggregating data for a set of 4 tracts
- Perhaps there is a larger published geography that includes these tracts plus one more
- Subtracting a single tract from the larger geography will be less work and more accurate than summing the four tracts

Aggregating Estimates and Calculating Margins of Errors

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

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

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

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

$$\approx 3{,}046$$

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

Example 2: Combining Fewest Areas

- Suppose you wanted the sum of 00100, 00200, and 00400.
- Combined estimate is 16,136.
- Approximated MOE using 3 cases = sqrt(1626^2 + 1395^2 + 1909^2) = 2,868
- If you subtract 00300 from Wyoming then MOE = sqrt(3309^2 + 1026^2) = 3,464
- Direct calculation = 3,025

What can be done?

- We have found that the approximation formula seriously breaks down when aggregating more than four estimates.
- So, in general, we suggest you aggregate the fewest number of estimates as possible.
- Try aggregating in different ways and see how sensitive the calculated MOE is to the method
- Other potential options:
  - Calculate the estimates using the Public Use Microdata Sample (PUMS)
  - Request a special tabulation (fee based and certain criteria apply)

Accuracy Documents

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

Questions

Sampling Update

- Sample Reallocation
- Sample Expansion
- Full non-response follow up in select areas

Sample Reallocation

- In update last year gave details of sample reallocation
- Highlights
  - Increase number of sampling rates to provide more equitable reliability of tract estimates
  - Increase sampling rate for blocks in the very smallest governmental units

Sample Reallocation

- Implemented with 2011 sample (January)
- Some 1-year areas with many small governmental areas may show improvements, i.e., Alaska
- Main impact is small area data, i.e. 5-year data
- Effects will be gradual improvement until 2011-2015 data is published

Sample Expansion

- Expansion of the HU sample from 2.9 million to 3.54 million
- Implemented starting in June 2011
- Partial benefit in 2011 1-year estimates, full benefit in 2012 1-year estimates
- Multiyear benefit phased in until 2011-2013 3-year and 2011-2015 5-year

Full Non-response Follow Up for Select Areas

- Select areas were selected to receive full follow up in CAPI mode of data collection
  - Remote Alaska (implemented Jan 2011)
  - Select American Indian Areas (implemented Aug 2011)
- Many of these areas were 1-in-2 sampling areas in Census 2000
- In combination with sample reallocation, can significantly improve reliability

Questions

New GQ Estimation Procedures

- Need to improve upon the existing methodology
- Overview of the new methodology

Problem Statement

- GQ sample is designed to produce state-level estimates of characteristics of the GQ population
- Estimates of GQ population are components of estimates of the total resident population
  - Published for small areas such as counties and tracts for 5-year estimates
- Incongruence between design and use of the GQ sample for estimates of total resident population
  - Lack of balanced representation of the GQ population across small areas

ACS GQ Sample in Tracts by Major GQ Type Group

(2006-2010 ACS Sample)

Major GQ Type Group	Tracts with ACS Sample	Tracts without ACS Sample	Total Tracts with Type Group
Correctional institutions	3,085	1,908	4,993
Juvenile facilities	1,343	1,582	2,925
Nursing homes	10,859	5,775	16,634
Other long-term care facilities	1,075	2,533	3,608
College dorms	2,538	827	3,365
Military facilities	304	276	580
Other non-institutional GQs	11,805	23,611	35,416

Solution

- Impute GQ person records to not-in-sample GQ facilities
- Donors are the sampled GQ person records
- Impute whole-person records
- Use expected population from ACS GQ sample frame

Objectives

GQ person representation (sample or imputed) in every combination of

- Tract by major GQ type for 5-year data
- County by major GQ type for 1- and 3-year data

Rules for Selecting Donors

- Period estimates - choose donors from same year that we impute into
- Expanding search: choose nearest donors
  - Nearness defined by geography and type of GQ facility

Rules for Selecting Donors (2)

- Take account of sex of donors when imputing to a single-sex GQ facility
- Made adjustments to the expected populations from ACS sample frame
- Limit the number of times a person can be used as a donor in a given tract

Evaluations

- Initial method was tested using simulated samples on Census 2000 data
- That work led to refinements and a broader evaluation using ACS data
- Analysts reviewed state and county estimates compared to existing ACS estimates and 2010 Census
- Current evaluation also comparing tract-level estimates to 2010 Census

Evaluations (2)

- Found that impact to state-level estimates is small
- Where there were state-level differences, many were found to be improvements
- County-level totals were improved significantly compared to 2010 Census
- Preliminary findings show similar pattern for tracts

Implementation

- Based on evaluations, the decision was made to implement this new methodology
- Will be implemented for all 2011 data products, i.e., 1-, 3-, and 5-year
- Expect improved estimates for characteristics for total population for substate areas
  - GQ contribution more in line with expectations
  - Better year-to-year stability

Questions

Special Data Products

- Select Population Tables
- American Indian Alaskan Native Tables
- Migration Data Products
- Commuting Flows
- Census Transportation Planning Package
- EEO File

Selected Population Tables a.k.a. SF4

- Released May 24, 2012 based on 2006-2010 ACS 5-year data
- Special data product to study detailed characteristics for 392 population groups based on
  - Race
  - Tribe
  - Hispanic
  - Ancestry
- Up to 300 tables on person and some housing characteristics
- Groups must have a population of 7,000 nationally
- The geography must have at least 50 group members in sample during the 5-year period (lowest level is census tract)

AIAN Tables

- Similar to SPT but only for American Indian and Alaskan Native groups
- Minimum national population is 100
- Same geographic restriction of 50 group members in sample during the 5-year period
- Fewer geographic types available
- Does include
  - Alaskan Native Regional Corporations
  - American Indian and Alaskan Native areas
  - Hawaiian Home Lands

Accessing Either Product

- Available on AFF
- Also available via FTP for direct download
  - One file per state
  - Can be large (CA is 1.6 GB)
  - Existing summary file tools do not work for these data products

Resources

- Use the URL below to access technical documenation, links to AFF, and links to FTP area.

Main Landing Page

http://www.census.gov/acs/www/data_documentation/special_data_release/

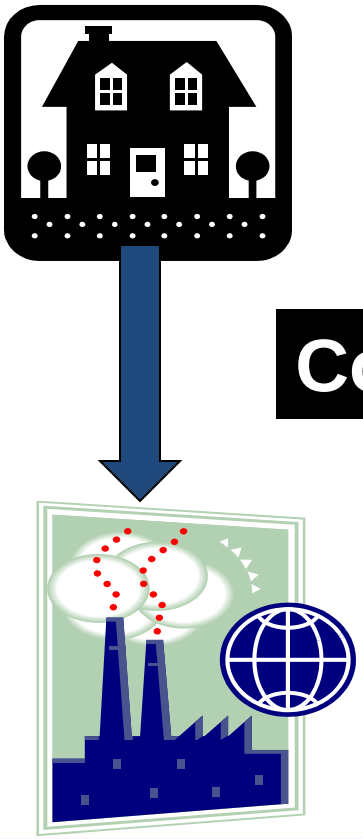
Migration Data Products from the American Community Survey

Outline

- County-to-county migration products from past censuses
- ACS migration flow products
- ACS county-to-county flow products
- Other proposed migration flow products
- Proposed commuting flow products


65

Geographically Dependent Characteristics: "Flows"



Commuting

Migration

Census County-to-County Flow Products

- 1970 Census – Universe Area Code-to-Universe Area Code issued on tape
  - UAC include all counties and some subcounty areas
- 1980 Census – Summary File Tape
  - 20 tables covering 11 characteristics
- 1990 Census – Migration CD
  - 21 tables covering 14 characteristics
- Census 2000 Migration Data DVD
  - 42 tables covering 19 characteristics

Census Bureau's Internal Flows Working Group

- New internal group formed to create new and modified flow products to become part of standard production

- Focus on data content and dissemination method

- Considerations for new product development: User needs, confidentiality, content of existing products

Standard Migration Products in AFF

- Tables with **In-Migration** by Characteristics
  - Since 2004 ACS
  - All publishable geographies

- Tables with **Out-Migration** by Characteristics
  - Since 2007 ACS
  - Limited geographies restricted to 50 or more cases

Good example of AFF limitations for presenting both ends of flow data

State-to-State Migration Flow Products

- ACS State-to-State Migration Flow Tables Published since 2005
- Shows state of current residence by state of residence 1 year ago (no characteristics)
- 3-year flow tables were added starting with 2007-2009 ACS
- ACS 2010 state-to-state single year was released in November 2011

County-to-County Migration Flow Products

- ACS 2005-2009 county-to-county and county/mcd-to-county/mcd flows (count estimates only) released in March 2012

- Working paper released in March 2012
  - Documentation for files
  - Issues related to creation of files
  - Some initial findings

(http://www.census.gov/hhes/migration/data/acs/county-to-county.html)

Proposed Approach for Presenting Flows: Alternate Characteristics across Years

- Release of initial 5-year ACS estimates not crossed by any characteristics (Done)
- Release of second 5-year ACS estimates crossed by Age, Sex, Race, Hispanic Origin (In progress)
- Each subsequent annual release within a 5-year period to include flows crossed with a small number of characteristics that vary by year
- Achieve a set of flows by multiple characteristics over the course of 5 years

Example Schedule

Year	2012	Late 2012	2013	2014	2015	2016	2017
ACS Data set	05-09	06-10	07-11	08-12	09-13	10-14	11-15
File:							
County-to-County	Released	In Works	X	X	X	X	X
County/MCD-to-County/MCD	Released	In Works	X	X	X	X	X
…by Age		In Works					X
…by Sex		In Works					X
…by Race		In Works					X
…by Hispanic Origin		In Works					X
…by Marital Status			?				
…by Place of Birth			?				
…by Nativity by Education				?			
…by Labor Force Status				?			
…by Industry Group					?		
…by Occupation Group					?		
…by Labor Force Status					?		
…by Poverty Status						?	
…by Tenure						?	
…by Individual Income						?	

Other Migration Flow Products

- Comparisons of ACS migration flows with administrative records
  - IRS county-to-county migration flows
  - National Change of Address files

- Graphic Visualization Products

Proposed New Commuting Flow Products

2006-2010 Commuting Flow Products

- Supports Office of Management and Budget's Metropolitan and Micropolitan Statistics Areas delineation program
- Will be released in 2013
- County and MCD level with no characteristics

Proposed New Commuting Flow Products

- Similar to Migration Flow Tables
- Released for counties and minor civil divisions
- Published for a set of basic social, demographic, and economic characteristics over the course of 5-years
- Characteristics will repeat every 5-year cycle.

Census Transportation Planning Products

What is the CTPP?

A set of special tabulations from the American Community Survey tailored for the data needs of transportation planners

Produced by the Census Bureau, sponsored and owned by American Association of State Highway Transportation Officials (AASHTO)

Key CTPP Contact: Penny Weinberger (AASHTO) 202-624-3556

2006-2010 5-Year CTPP Plan

- Similar tables to the 2006-2008 3-year CTPP
  - Residence tables
  - Workplace tables
  - Residence-to-workplace tables
- Similar geographic summary levels to the Census 2000 CTPP
  - Including Traffic Analysis Districts (TADs) and Traffic Analysis Zones (TAZs) – special user-defined geographies by members of the transportation community

2006-2010 5-Year CTPP Plans

- More detailed information at the TAZ level?
  - Current remedy – modeling a set of data to tabulate a subset of tables and flows at greatest risk to confidentiality
- Number of crosses for flows minimized to 5 variables to ensure respondent confidentiality.

79

We Want Your Feedback

- Which characteristics are most critical in understanding migration and commuting **FLOW** data

We need your help!

- Among existing ACS migration and place of work tables on AFF, which are most/least useful?

- How do you want to obtain the data?
  - Geo-IDs such as FIPS codes included in table?
  - What is the utility of maps to accompany data tables?
  - Other: Web applications (e.g. OnTheMap), other tools?

For More Information

Contact Information:

Alison Fields, Chief, Journey to Work and Migration Statistics Branch

alison.k.fields@census.gov or 301-763-2454

Migration Data including Flows
http://www.census.gov/hhes/migration/

Commuting Data
http://www.census.gov/hhes/commuting/

Census Transportation Planning Products
http://ctpp.transportation.org/Pages/default.aspx

EEO Tabulation

- External benchmark for conducting comparisons between the racial, ethnic, and sex composition of each employer's workforce to its available labor market

- Used by organizations to develop and update their affirmative action plans

What is the EEO Tabulation?

- Tabulation of the civilian workforce aged 16 and older
- Place of residence, place of work, and worksite commuting flows
- Occupation by sex by race/ethnicity by citizenship for detailed geography
- Other tables include industry, age, educational attainment, and median earnings

New This Time For the EEO Tab

- ACS 2006-2010 5-year file
- Margins of error
- 2010 SOC Occupation categories
- Census Occupation categories: 487 for Worksite tables and 488 for Residence tables
- Additional variable: Citizenship
- Data will be disseminated through American FactFinder

EEO Tab: 3 Geography Types

Residence geography

(where people live)

Workplace geography

(where people work)

Workflow geography

(where people work and where they commute from)

EEO Tab Geography Levels

- U.S. Total
- All states
- Core Based Statistical Areas (CBSAs) of 50,000 or more
- Counties and County Sets (Counties with populations of 50,000 or more or aggregations of counties that together have a population of 50,000 or more.)
- Places of 50,000 or more

EEO Data Products

- 12 sets of tables
- Residence/Worksite/Commuter flow tables
- Population threshold of 50,000
- EXCEPT Citizenship tables-
  - will include areas with a population threshold of 100,000 people or more
  - at least 3 unweighted cases per cell

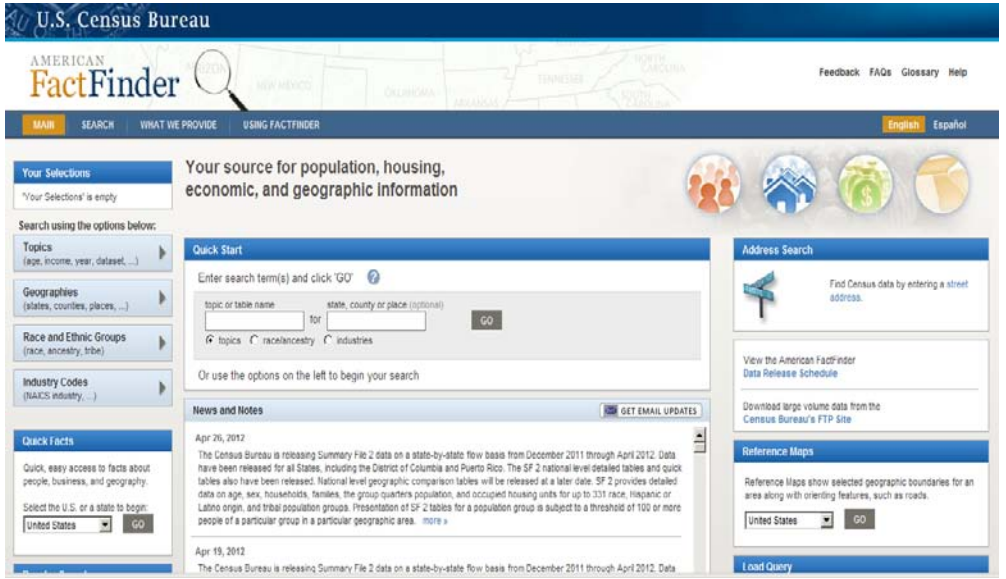
Table Descriptions

- By Race/Ethnicity, Sex, and Citizenship-
  - Detailed Census Occupation Categories*
  - EEO Occupational Groups
  - EEO-1 Job Categories
  - Federal Sector Job Categories
  - State and Local Government Job Groups
  - Educational Attainment and Younger Age groups
  - Detailed Census Occupation Categories and Educational Attainment*
  - Detailed Census Occupation Categories and Industry
  - Detailed Census Occupation Categories and Earnings
  - Detailed Census Occupation Categories and Older Age groups

*Due to Disclosure Review Board (DRB) geography restrictions, these two tables will also be published without the citizenship variable with a population threshold of 50,000 people.

EEO Data Dissemination Plans

- Summer 2012:
  - Occupational listings and crosswalks
  - County set listing
  - On http://www.census.gov/hhes/www/eeoindex/eeoindex.html
- End of 2012:
  - All EEO tables
  - On American FactFinder: http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml



Contact Information

Mark E Asiala

US Census Bureau

mark.e.asiala@census.gov