2015 CE Survey Microdata Users' Workshop

Sample Design and Weights

Brian T. Nix
Mathematical Statistician
Division of Price Statistical Methods
Bureau of Labor Statistics
U.S. Department of Labor

July 16, 2015

BLS
BUREAU OF LABOR STATISTICS
U.S DEPARTMENT OF LABOR

www.bls.gov

Overview

- History and Concepts
- Sample Selection
  - Define PSUs
  - Stratify and select a Sample of PSUs
  - Stratify and Select a Sample of Households
  - The Future (2010 Census-based Sample Design)
- Weighting the households (CUs)

History of Sample Redesigns

- New sample of geographic areas and addresses selected every decade (2010)
  - 1980 Census-Based Sample Design (1986–1995)
  - 1990 Census-Based Sample Design (1995–2004)
  - **2000 Census-Based Sample Design (2005–present)**
  - 2010 Census-Based Sample Design (2015–2024?)

BLS

Concepts

- Target Population=U.S. non-institutional civilian population
- Consumer Unit
  - person or a group of persons in a household related by blood, marriage, adoption, or other legal arrangements
    - OR are unrelated but pool their incomes to make joint expenditure decisions
    - Same as households approximately 98% of time

Concepts (cont.)

- Sampling Frame – List of Households from which we draw our sample
  - Unit Frame: Regular households (80%)
  - Area Frame: Rural households (10%)
  - Permit Frame: New construction (9%)
  - Group Quarters: (1%)
- Will Change to Census Bureau's Master Address File (MAF) in 2015 (2010 Census with updates twice per year by US Postal Service)

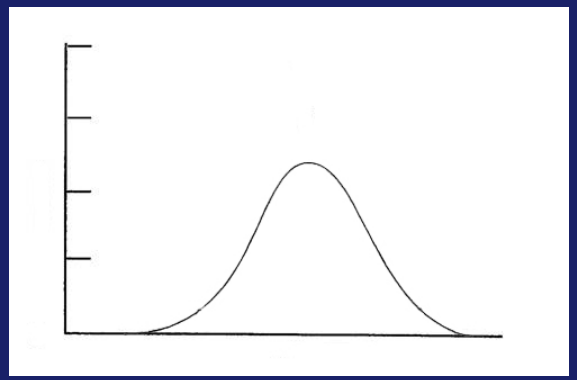
BLS

Sample Selection – Overview

- Geographic areas are randomly selected to represent the total U.S.
- Households are randomly selected to represent the geographic areas

Guiding principle:

"***Randomness ensures representativeness***."



Define PSU

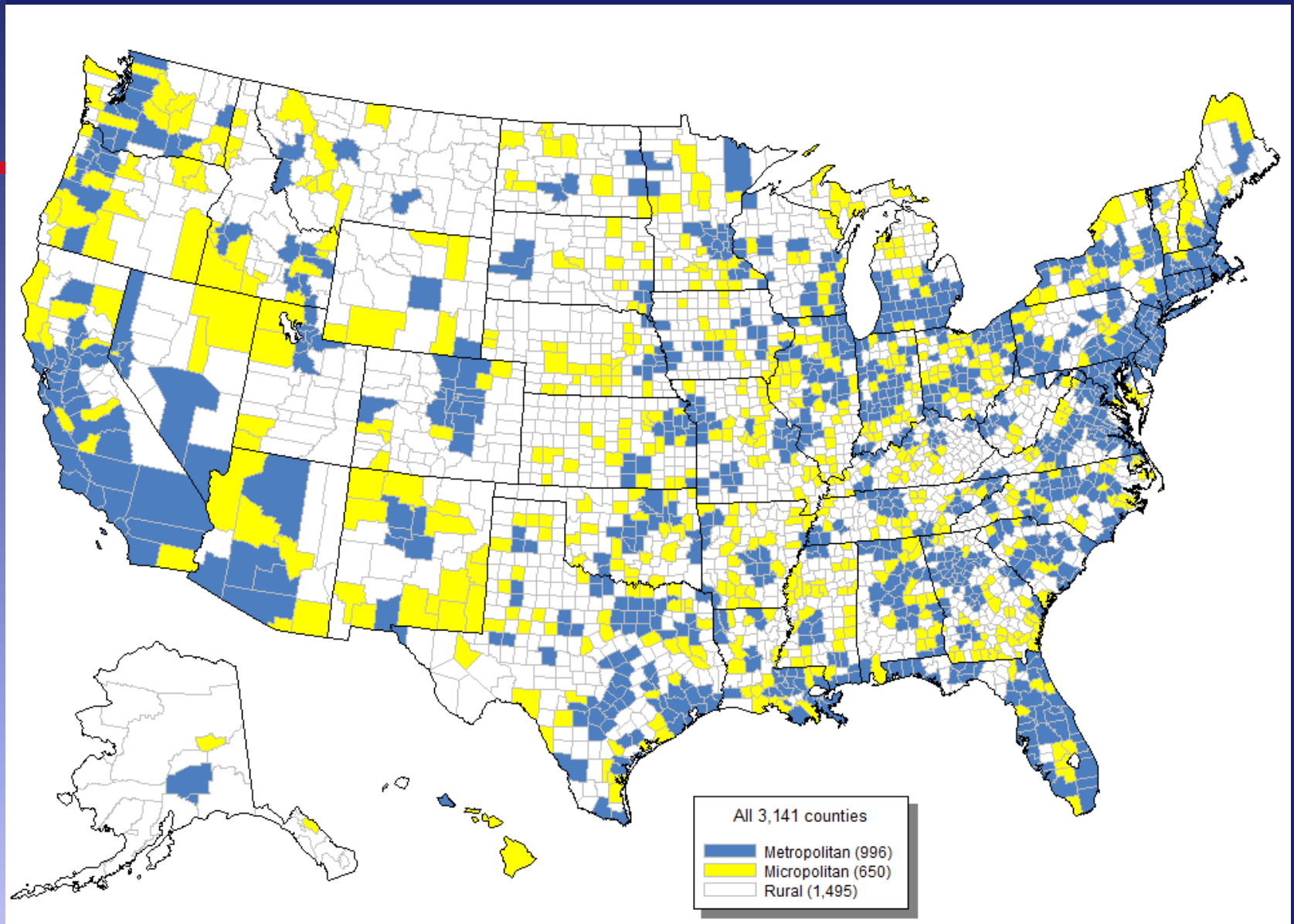
- PSU: Primary Sampling Unit
  - Counties are geographically grouped together to become units for sample selection
- CBSA: Core Based Statistical Areas (~old MSA)
  - Counties are grouped together into geographic entities called core based statistical areas (CBSA's) by Office of Management and Budget
  - **Metropolitan** – one or more counties centered around urban area of > 50,000 people

BLS

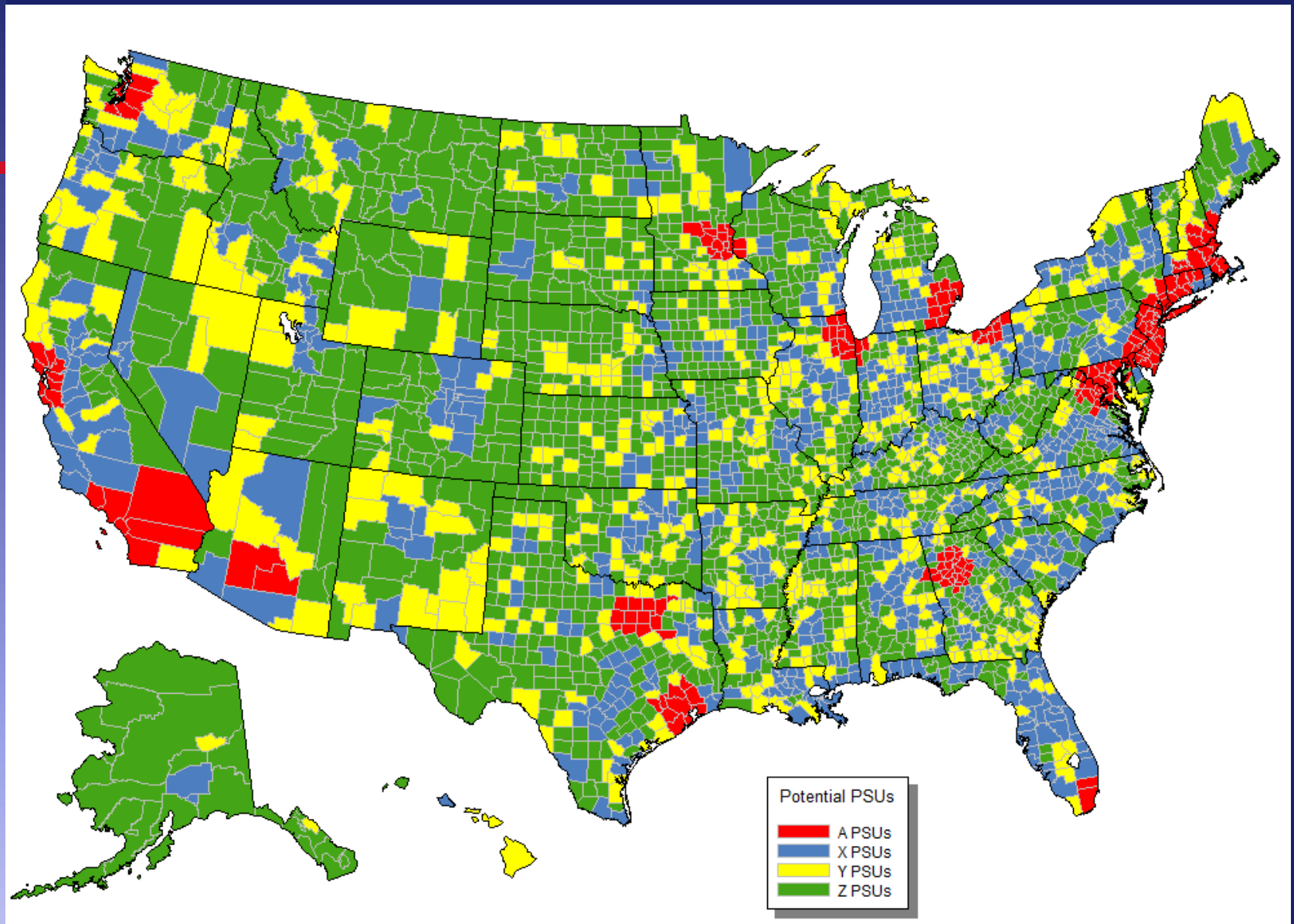
Define PSU (cont.)

- **Micropolitan** – one or more counties centered around urban area of 10,000 - 50,000 people

- Over 3,000 county and county equivalents in the U.S.
- Over 900 CBSAs defined by OMB


All 3,141 counties
Metropolitan (996)
Micropolitan (650)
Rural (1,495)

BLS



BLS

Selection of PSUs

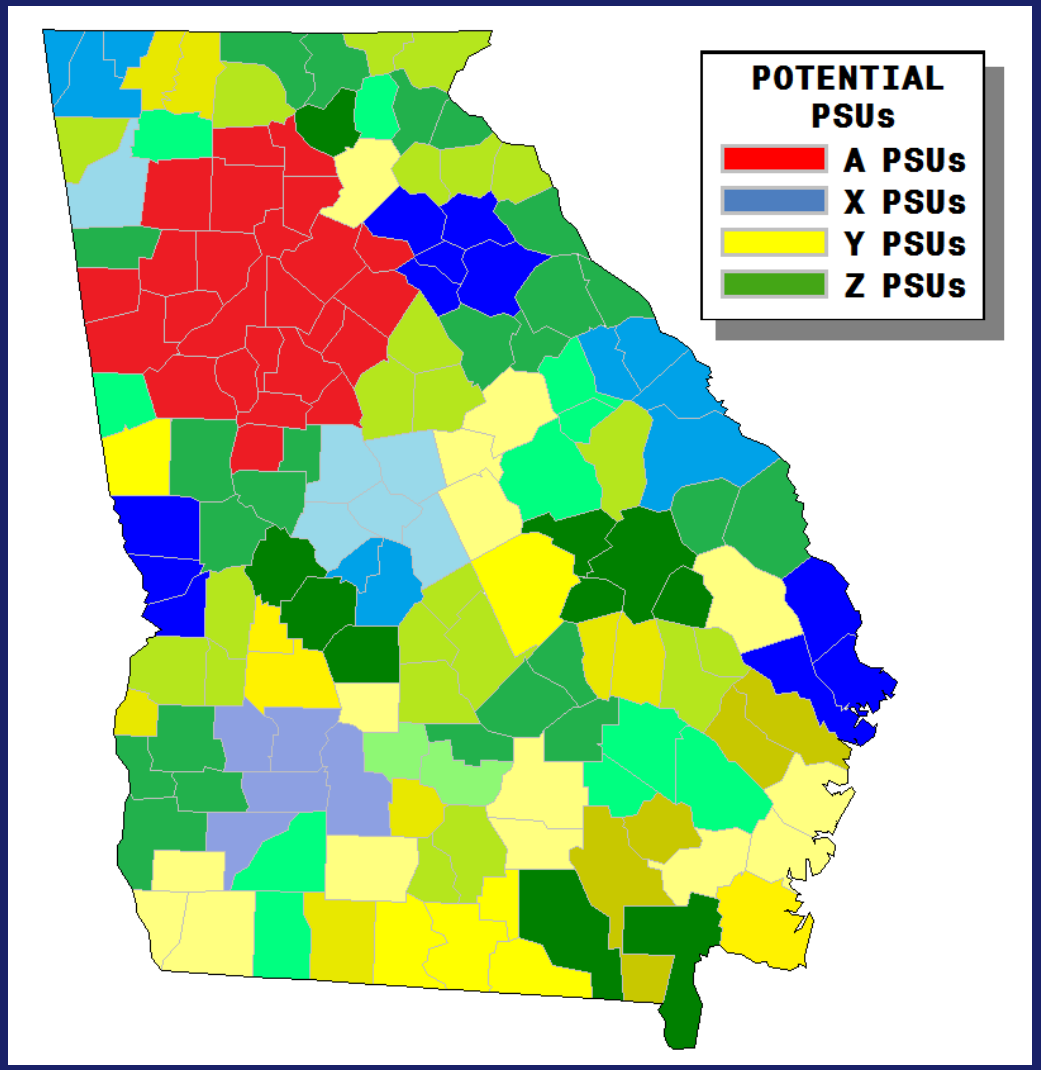
PSU size	SR/ NSR	CBSA/ Non-CBSA	Population Total	Examples
A	SR	Metropolitan (urban)	More Than 2,700,000	A103 Boston MA A210 Cleveland OH
X	NSR	Metropolitan (urban)	Less Than 2,700,000	X486 Denver CO X218 Cincinnati OH
Y	NSR	Micropolitan (urban)		Y104 Ithaca NY Y426 Newport OR
Z	NSR	Non CBSA (rural)		Z210 Holmes OH Z324 Montague TX

2000 Census-based Sample Selection
CPI – 75 PSUs; CE – 91 PSUs

PSU Size	Region				Total
	Northeast	Midwest	South	West	
A	5	4	6	6	21
X	4	10	16	8	38
Y	2	4	6	4	16
Z	2	4	6	4	16
Total	13	22	34	22	91

BLS

Georgia PSU Selection



BLS

Metropolitan CBSAs in Georgia

CBSA	2000 Population	Probability of Selection
Augusta-Aiken, GA-SC	523,162	0.54746
Columbus, GA-AL	274,624	0.28738
Albany, GA	157,833	0.16516
Total	**955,619**	**1.00000**

CBSA	2000 Population	Probability of Selection
Savannah, GA	293,000	0.30906
Macon, GA	222,368	0.23456
Athens, GA	207,668	0.21905
Warner Robins, GA	134,433	0.14180
Rome, GA	90,565	0.09553
Total	**948,034**	**1.00000**

BLS

Metropolitan CBSAs in Georgia

	CBSA	2000 Population	Probability of Selection
✓	**Augusta-Aiken, GA-SC**	**523,162**	**0.54746**
	Columbus, GA-AL	274,624	0.28738
	Albany, GA	157,833	0.16516
	Total	**955,619**	**1.00000**

	CBSA	2000 Population	Probability of Selection
	Savannah, GA	293,000	0.30906
	Macon, GA	222,368	0.23456
✓	**Athens, GA**	**207,668**	**0.21905**
	Warner Robins, GA	134,433	0.14180
	Rome, GA	90,565	0.09553
	Total	**948,034**	**1.00000**

BLS

Number of Households

- Allocate Target Sample to PSUs
  - Target size: ~7,000 interviewed households (**Based on Finite Budget)**
    - For Diary Survey per year
    - For Interview Survey per quarter
  - 6,600 to Households used jointly by CE and CPI (for CPI cost weight calculations)
    - 21 A-size PSUs
    - 54 X-size and Y-size PSUs
  - 400 to CE Households
    - 16 Z-Size PSUs

Number of Households (con't)

- **Target Sample Size**
  - 7,000 interviewed households per year (Diary)
  - 7,000 interviewed households per quarter (Interview, interviews 2-5 only)

- **Target Sample Yield**
  - 14,000 weekly diaries per year (=7,000×2)
  - 28,000 quarterly interviews per year (=7,000×4)

Number of Households (con't)

- **Local Target Sample Size**
  - Allocate 7,000 interviewed households to individual PSUs, proportional to each stratum's population
  - Minimizes CE's nationwide variance

Translate Addresses into Interviewed Households

- 80% “eligibility” rate: (most of the missing 20% are unoccupied)
- 75% response rate
- 60% “participation” rate (0.60 = 0.80×0.75)

Translate Interviewed Households into Addresses (con't)

PSU	Interviewed households	Addresses	%
A102 Philadelphia	169	322	52
A103 Boston	195	286	68
A109 New York City	220	420	52
A110 NY-Conn suburbs	212	335	63
A111 NJ suburbs	182	291	63
etc.	*etc.*	*etc.*	
Total	**7,000**	**11,750**	

Select a Random Sample of Households (Mechanics)

- Sort households from poor to rich based on information from Decennial Census and ACS:
  - Number of people in household
  - Tenure (owner, renter)
  - Market value of home (owners)
  - Monthly rent (renters)

Select a Random Sample of Households (Con't)

- Compute the sampling interval for each PSU
- Sampling interval = (# addresses in sampling frame) ÷ (# addresses in CE sample)
- Typical sampling intervals:
  - Every 1,000th address (X+Y, Z PSUs)
  - Every 5,000th address (A PSUs)

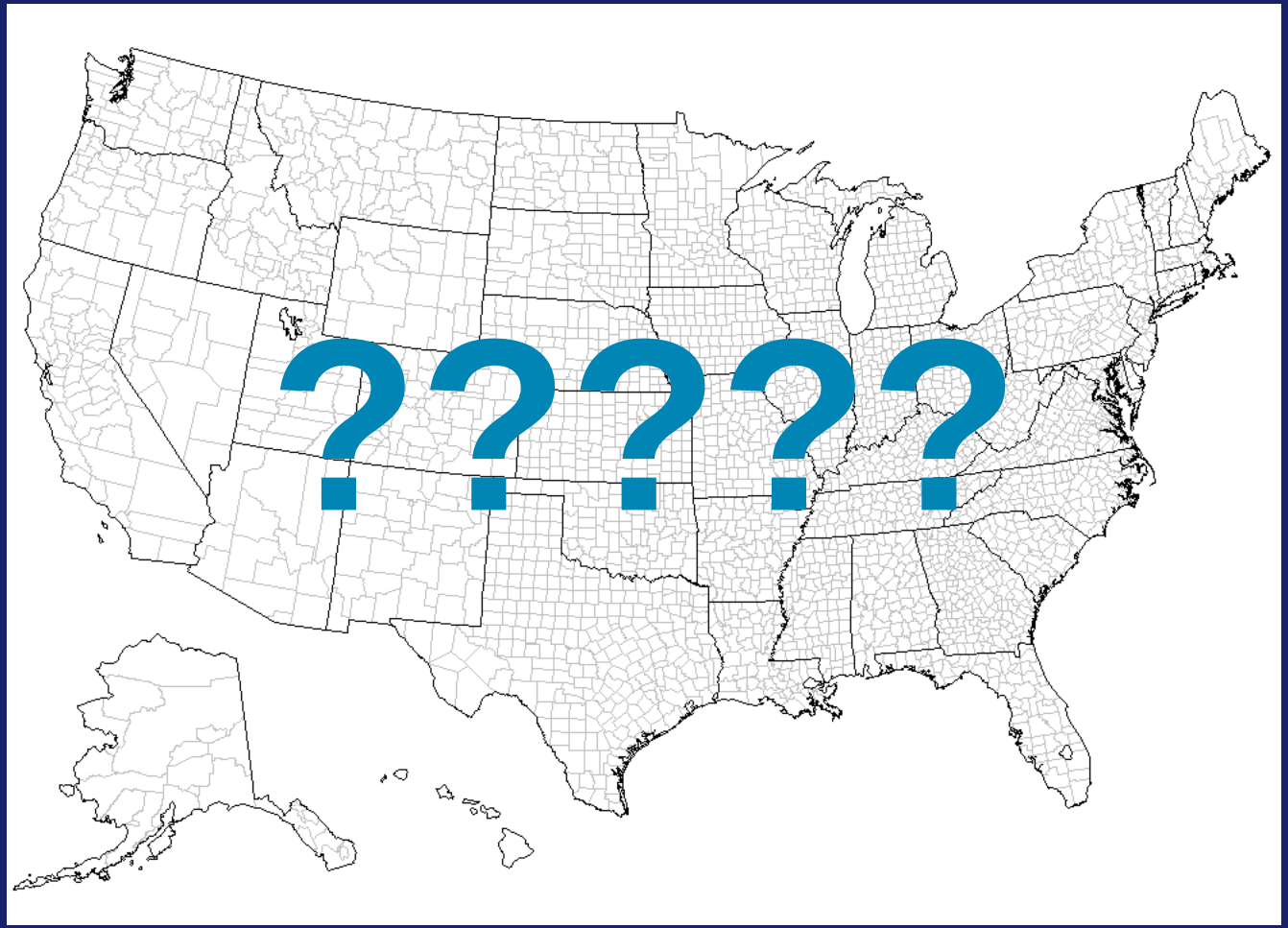
Select a Random Sample of Households (Cont.)

- -- D --- I --- D --- I --- D --- I --- D -- I -- -- D --- I --- D --- I --- *etc.*
- D=Diary, I=Interview
- Each “D” and “I” has enough sample to cover the next 10 years

The Future

- **New Sample Selection**
  - Based on 2010 Census
    - Updated Metropolitan/Micropolitan Statistical Area definitions from OMB
    - Updated populations used for selection probabilities
  - Currently Going Into Effect (2015)
  - No Currently Released Data

The Future



BLS

Weighting Process

- Usable Interviews
  - Diary Survey
    - Good interviews
      - Diary complete enough to be counted
  - Interview Survey
    - Good interviews
      - Interviews 2 – 5
      - Completed enough to be counted

Weighting Process (Con't)

- Base Weight (~10,000)
  - 9,999 CUs + Self
- Weighting Control Factor (~1.00)
  - Apartment Building instead of a House
- Non-interview Adjustment Factor (~1.3)
  - Type A: Refusal to Participate
- Calibration Adjustment Factor (~1.15)
  - Adjusts sample estimate to CPS Totals

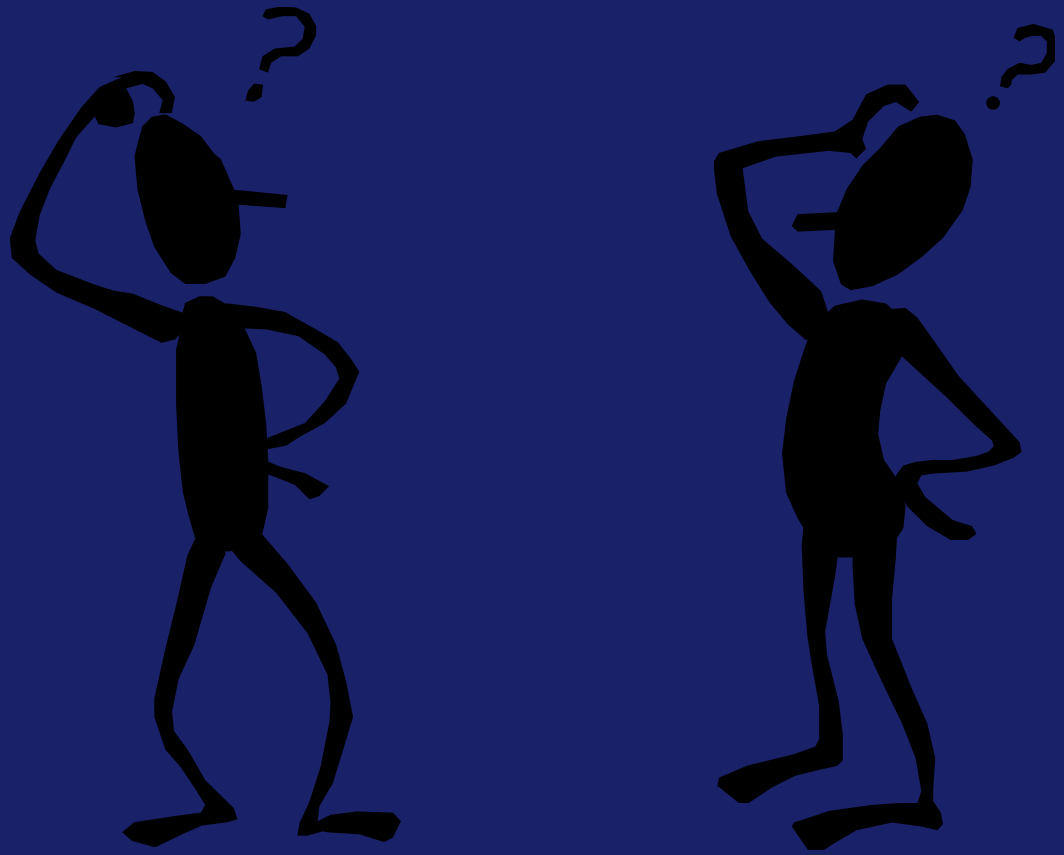
Weighting Process (Continued)

- Final Weight
  - Base Weight * Weighting Control Factor * Non-interview Adjustment Factor * Calibration Adjustment Factor
  - ~15,000 to 18,000

Conclusion

- Both Sample Design and Weighting Work Together to Produce:
  - Best Estimates of U.S. Expenditures
    - Subject to Allotted CE Budget

Any Questions?



Contact Information

Brian T. Nix

Mathematical Statistician

www.bls.gov/cex

202-691-6877

Nix.Brian@bls.gov

BLS

BUREAU OF LABOR STATISTICS

U.S DEPARTMENT OF LABOR