Changes to Income Processing in the CPS ASEC

Jonathan Rothbaum

U.S. Census Bureau

Population Association of America Annual Meetings
April 9, 2019
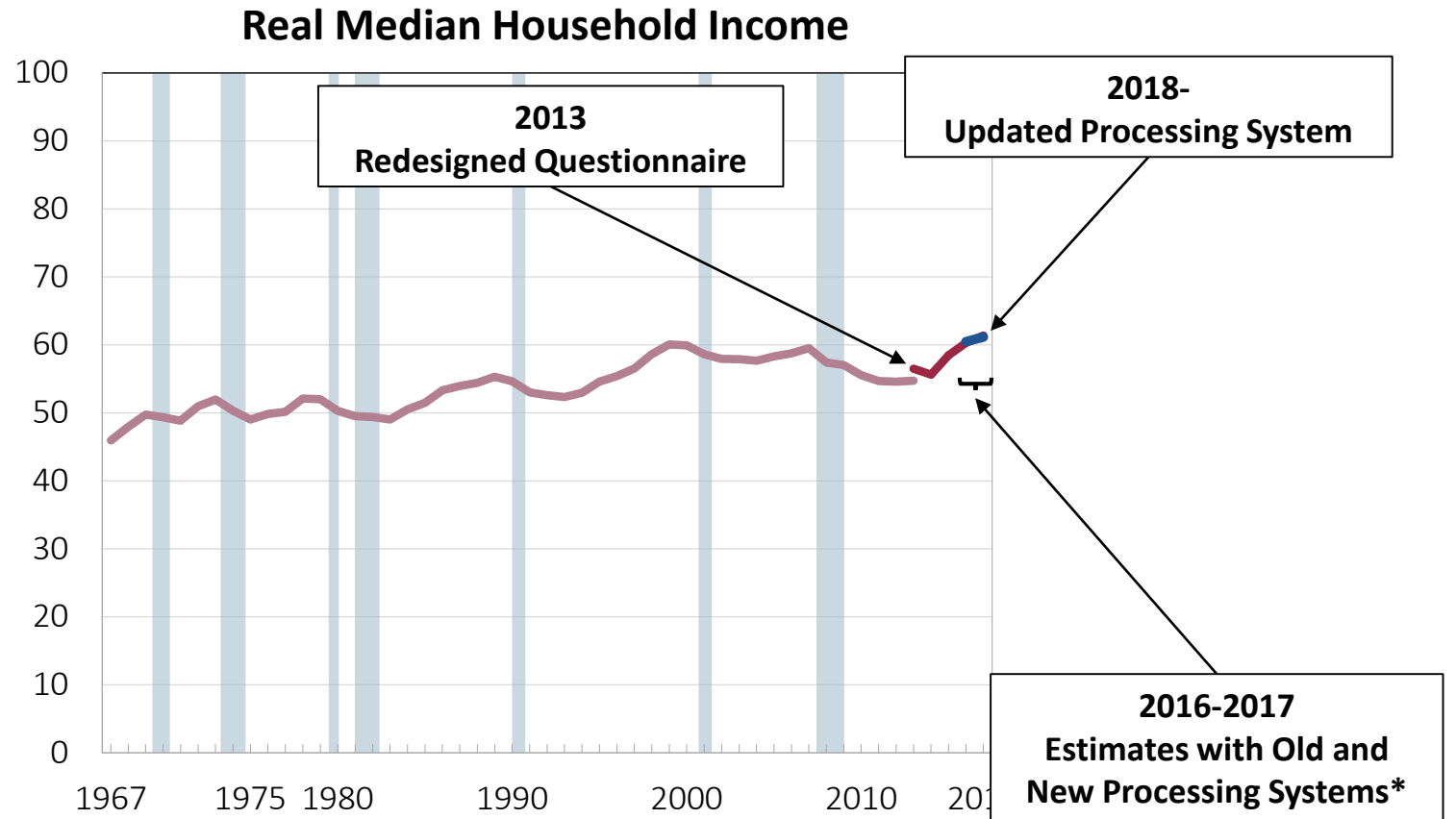
United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Email: jonathan.l.rothbaum@census.gov

1

Redesign of CPS ASEC: Income

- Redesigned Questionnaire in 2014
  - Income in 2013

- Updating processing system in production in 2019
  - Income in 2018

- Re-release estimates for 2017 with new system
  - 2016 data available in with new processing system in 2017 Research File

Real Median Household Income

2013
Redesigned Questionnaire

2018-
Updated Processing System

2016-2017
Estimates with Old and
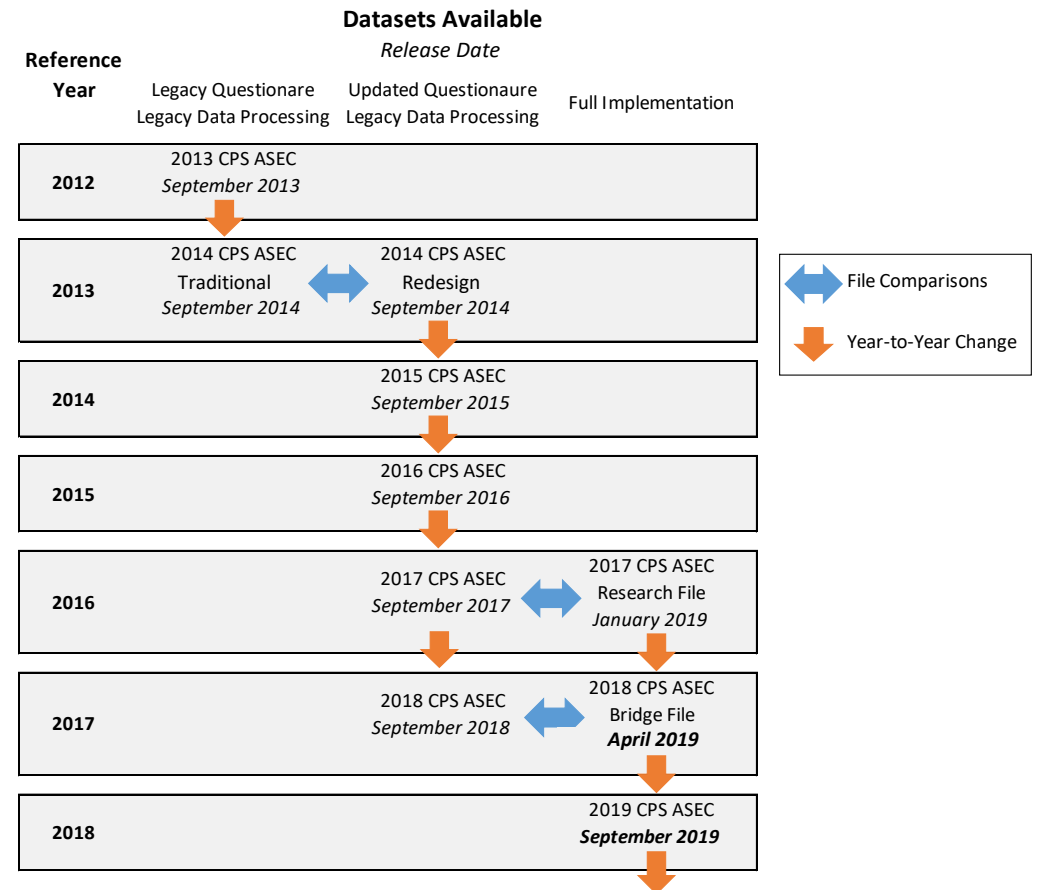New Processing Systems*

Source: CPS ASEC

* 2016 estimates with new processing system from 2017 Research File, 2017 estimates with new processing system from 2018 Bridge File.

Implementation – Income and Poverty

- Research
  - Over a decade of research on income and health insurance underestimates.
  - WESTAT contract for income questions
  - Field test in March 2013.
- 2014
  - Split Panel Test
    - 69% Traditional instrument, 31% Redesigned
- 2015+
  - Redesign implemented for full sample
  - Processed using old system
- 2019+
  - Updated processing system implemented in production

Datasets Available

Release Date

Reference Year	Legacy Questionare Legacy Data Processing	Updated Questionaure Legacy Data Processing	Full Implementation
2012	2013 CPS ASEC *September 2013*		
2013	2014 CPS ASEC Traditional *September 2014*	2014 CPS ASEC Redesign *September 2014*	
2014	2015 CPS ASEC *September 2015*		
2015	2016 CPS ASEC *September 2016*		
2016	2017 CPS ASEC *September 2017*	2017 CPS ASEC Research File *January 2019*	
2017	2018 CPS ASEC *September 2018*	2018 CPS ASEC Bridge File ***April 2019***	
2018			2019 CPS ASEC ***September 2019***

File Comparisons

Year-to-Year Change

Source: Creamer and Edwards, 2019

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

3

Processing Changes

- Earnings ranges used in imputation

- Other income edit and imputation overhaul

- Mortgage imputation change

- Means-tested program benefit caps removed

- Income allocation flags changed

- Increased top codes for some income types

- Various other small fixes

United States™
Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Earnings ranges used in imputation

- Impute income within ranges given
- 73% of item non-respondents had income imputed using a range

Could you tell me if you earned
- less than $45,000,
- between $45,000 and $60,000, or
- $60,000 or more

for the TOTAL yearly amount from this employer before taxes and other deductions during 2016?

Did you earn
- less than $15,000,
- between $15,000 and $30,000, or
- $30,000 or more

from this employer during 2016?

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Other income edit and imputation overhaul

- Without processing changes, redesigned questions were run through the old processing system
  - For example, multiple questions on interest and retirement income combined into one income amount for each
  - Effectively converts non-response to some questions to zero income

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Other income edit and imputation overhaul

- Use ranges provided in responses for income imputation

Allocation System Improvements

- Imputations using ranges (for select variables)

Variable		Share of Item Non-Respondents with Ranges
Earnings (Longest Job)		0.73
Unemployment Insurance		0.41
Worker's Compensation		0.40
Social Security		0.39
SSI		0.37
Interest		0.52
Dividends		0.36
Overall	**(share of income items missing due to item non-response)**	**0.50**

Other income edit and imputation overhaul

- Use ranges provided in responses for income imputation
- Update hot deck models to address critiques on match bias (to the extent possible with hot deck approach)

Hot Deck Imputation

- Match non-respondents to "similar" respondents along a set of characteristics in the model

- Donate response as imputation from respondent to non-respondent

- Example: 2 variables, 2 categories each – 4 cells
  1. Race: White/non-White
  2. Gender: male/female
     - Two non-respondents (A and B)
       - **Person A**: white, female – randomly select a white, female respondent and use her response as the imputed value
       - **Person B**: non-white, male – randomly select a non-White, male respondent and use his response as the imputed value

Other income edit and imputation overhaul

- Use ranges provided in responses for income imputation
- Update hot deck models to address critiques on match bias (to the extent possible with hot deck approach)
  - Increase the number of variables used in models
  - Multiple levels of matches – match people on as many variables as possible

United States **Census** Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Other income edit and imputation overhaul

- Use ranges provided in responses for income imputation
- Update hot deck models to address critiques on match bias (to the extent possible with hot deck approach)
  - Increase the number of variables used in models
  - Multiple levels of matches – match people on as many variables as possible
- Which variables to include?
  - Use RandomForest to select variables that best predict each income item
    - Supervised machine learning method that allows variables to be ranked based on how much they reduce variance when included in model

Allocation System Improvements

- Example – Dividends

Not Possible Before

	Old	New
Recipiency (Yes/No)	Gender Age Relation to Householder Individual/Household Earnings	Age Presence of Child Retirement Interest Money Market Interest Education Savings Interest Marital Status Checking Interest Relation to Householder Spouse Labor Force Participation Census Region
Value	Individual/Household Income Race Age Education Presence of Child	Age Presence of Child Retirement Interest Money Market Interest Education Savings Interest Marital Status Checking Interest Relation to Householder

Allocation System Improvements

- Example – Rental Income

Recipiency (Yes/No)	Relation to Householder Age Gender Education Race	Age Presence of Child Relation to Householder Savings Interest Census Region Earnings Marital Status Dividends Household Income (to this point)
Value	Earnings Age Race Gender Relation to Householder	Age Presence of Child Relation to Householder Savings Interest Census Region Earnings Marital Status Dividends Household Income (to this point)

Mortgage Imputation

- Pre-update
  - Mortgage non-response imputed from lagged American Housing Survey (AHS)
    - AHS conducted every two years with slight lag after collection for processing before data is available for use in CPS ASEC imputation
    - Imputed some very infrequently used variables from AHS data
    - Essentially mixing distributions as responses are not identical on AHS and CPS ASEC
- Update
  - Impute missing mortgage information directly from CPS ASEC respondents as we do for all other missing data

Means-Tested Benefit Caps Removed

- Pre-update
  - Energy assistance – if household income above a set nominal threshold, edit energy assistance receipt to $0, even if respondent reported receipt
  - Other means-tested benefits – no imputation of means-tested benefits to respondents above a set nominal threshold
    - SNAP, energy assistance, free or reduced price lunches, public housing assistance
- Update
  - Removed all of these caps
    - Research on benefits has consistently shown receipt for benefits for households with income above these caps, including in ACS responses and administrative program data

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Income Allocation Flags

- Pre-update
  - Need to check two flags for each income variable to assess if value imputed
    - Imputation flag for that income variable
      - Earnings from the longest job – i_ernval > 0
      - Social security – i_ssval > 0
      - SSI – i_ssival > 0...
    - Imputation flag for supplement non-response
      - ALL income items imputed if supplement non-respondent
      - Check if FL_665 > 1
      - Anecdotally - not well understood by even some experienced CPS ASEC researchers

Income Allocation Flags

- Update
  - Include a code for supplement non-response in each variable
    - If i_ernval = 9, then earnings imputed due to supplement non-response
  - More detail in imputation codes
    - Values 1-3: range bracket used in imputation
    - Values 4+: no range used
    - Lower values within these groups (1-3, 4+) indicate better matches (more variables in imputation model)
  - Composite variable imputation information
    - Example – interest income is a composite of several questions about various interest types
    - Y/N variables – imputation code indicates if all/some/or none of the variables were imputed
    - Value variables – imputation code indicates share of income imputed (0-25%, 25-50%,…)

Increased Top Codes for Some Income Types

- Pre-update
  - Nominal top codes for many income types had not been changed in many years
  - Small, but increasing share of responses above top codes

- Update
  - Increased top codes for:
    - Property income (interest, dividends, rental income)
    - Retirement income (annuities, pensions, withdrawals from defined-contribution plans)
    - Financial assistance
    - Other income
  - Changed process for updating top codes so that they are revisited annually and easier to update
    - Prevent fixed nominal values from declining in real terms over decades

Various Other Small Fixes

- In updating/editing code, various bugs/errors fixed and minor improvements made
  - Examples
    - Changed a small number of variable names to be consistent with established CPS ASEC naming conventions (such as adding an "h" prefix for household variables that did not have them)
    - Fixed an error in the assignment of the family market value of SNAP benefits (F_MV_FS)
    - Fixed an error when calculating family income for subfamilies where a member of the primary family had a loss in self-employment or rental income

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Various Other Small Fixes

- Examples, cont.
  - Fixed an error where the redesign instrument caused interview/respondent confusion.
    - This confusion resulted in a multiple pension sources being recorded when the respondent had only one source
      - Discovered in the course of processing updates
    - The second pension sources was recorded as having exactly the same income amount as the first
    - Pre-redesign, having multiple pension sources with the same amounts occurred ~10 times per year, post redesign, this occurred ~500 times per year
      - Minimal impact on distribution statistics for the 65+ population, but affects aggregate pension income in the survey

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Personal Income by Type
Percent Difference between Files

Type	People with Income	Percentile 25	Percentile 50	Percentile 90	Average	Total
Total	0.0	1.3***	0.3	0.8	1.4***	1.4***
Earnings	0.0	1.8***	0.3	-0.1	0.4	0.4
Social Security	0.0	-2.0***	-0.4	0.4	-1.3***	-1.3***
SSI	0.5	-17.7***	-1.3***	0.1	-2.6***	-2.2*
Public Assistance	0.1	-7.3***	-11.0***	-4.8	-6.5***	-6.4**
Interest	-0.1				15.1***	15.0***
Dividends	1.4***				13.0***	14.6***

Source: 2017 CPS ASEC Production and Research Files. Positive numbers indicate more recipients or income in the Research file. All percentiles are calculated using linear interpolation. Interest and dividend interpolations not shown as the width of the bins used for interpolation are too large for meaningful estimates. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels respectively.

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Personal Income by Type
Percent Difference between Files

Type	People with Income	Percentile			Average	Total
		25	50	90		
Retirement Income	4.2***	-4.3**	-5.6***	-2.4	1.7	6.0***
Company/Union Pension	-4.2***	-0.9	1.3	2.5	6.4***	1.9
Federal Gov't Pension	-5.8***	-20.4***	-9.1***	-9.1*	-9.8***	-15.0***
Military	-18.1***	-28.2***	-11.9***	4.5*	-1.1	-19.0***
State/Local Gov't Pension	-11.6***	6.6	2.5	3.6	4.6***	-7.5***
Defined Contribution (IRA, 401(k), etc.)	73.7***	-2.9	0.9	-7.7	8.5**	88.4***

Source: 2017 CPS ASEC Production and Research Files. Positive numbers indicate more recipients or income in the Research file. All percentiles are calculated using linear interpolation. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels respectively.

Household Income for Select Subgroups Median (in dollars)

Characteristic	Production	Research	Percent Difference
All Households	59,040	59,210	0.3
Family households	75,060	75,050	0.0
Nonfamily households	35,760	35,770	0.0
Race			
White	61,860	61,950	0.1
White, not Hispanic	65,040	65,440	0.6
Black	39,490	39,750	0.7
Asian	81,430	80,880	-0.7
Hispanic (any race)	47,680	46,930	*-1.6
Age			
25 to 34 years	60,930	60,020	**-1.5
35 to 44 years	74,480	73,880	-0.8
45 to 54 years	77,210	77,700	0.6
55 to 64 years	65,240	65,710	0.7
65 years and older	39,820	40,530	**1.8

Source: 2017 CPS ASEC Production and Research Files. All percentiles are calculated using linear interpolation. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels respectively.

Household Income for Select Subgroups 10th Percentile (in dollars)

Characteristic	Production	Research	Percent Difference
All Households	13,650	13,720	0.5
Family households	21,470	21,430	-0.2
Nonfamily households	9,227	9,138	-1.0
Race			
White	15,370	15,430	0.4
White, not Hispanic	15,950	16,030	0.5
Black	8,452	8,366	-1.0
Asian	16,830	17,190	2.2
Hispanic (any race)	12,290	12,160	-1.1
Age			
25 to 34 years	16,350	16,220	-0.8
35 to 44 years	19,280	19,450	0.9
45 to 54 years	17,350	17,290	-0.3
55 to 64 years	12,470	12,880	*3.3
65 years and older	11,850	11,580	**-2.3

Source: 2017 CPS ASEC Production and Research Files. All percentiles are calculated using linear interpolation. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels respectively.

Household Income for Select Subgroups 25th Percentile (in dollars)

Characteristic	Production	Research	Percent Difference
All Households	29,390	29,330	-0.2
Family households	40,700	40,820	0.3
Nonfamily households	17,100	17,090	-0.1
Race			
White	31,450	31,500	0.2
White, not Hispanic	32,680	33,060	1.2
Black	18,300	18,030	-1.4
Asian	40,100	41,090	2.5
Hispanic (any race)	25,520	25,190	*-1.3
Age			
25 to 34 years	32,850	31,870	**-3.0
35 to 44 years	39,520	39,900	1.0
45 to 54 years	40,320	40,710	1.0
55 to 64 years	30,930	31,480	*1.8
65 years and older	20,840	20,700	-0.7

Source: 2017 CPS ASEC Production and Research Files. All percentiles are calculated using linear interpolation. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels respectively.

United States™
Census
Bureau

Household Income for Select Subgroups 75th Percentile (in dollars)

Characteristic	Production	Research	Percent Difference
All Households	106,200	106,200	0.0
Family households	124,700	125,600	*0.7
Nonfamily households	66,060	65,480	-0.9
Race			
White	110,400	110,200	-0.2
White, not Hispanic	114,900	115,400	0.4
Black	73,370	74,000	0.8
Asian	145,500	142,700	-2.0
Hispanic (any race)	84,140	82,560	**-1.9
Age			
25 to 34 years	99,950	99,030	-0.9
35 to 44 years	122,900	122,100	-0.6
45 to 54 years	132,300	134,000	1.3
55 to 64 years	117,000	118,800	1.6
65 years and older	76,330	77,050	0.9

Source: 2017 CPS ASEC Production and Research Files. All percentiles are calculated using linear interpolation. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels respectively.

Household Income for Select Subgroups 90th Percentile (in dollars)

Characteristic	Production	Research	Percent Difference
All Households	170,900	173,600	***1.6
Family households	193,900	197,800	***2.0
Nonfamily households	110,500	108,300	**-2.0
Race			
White	175,400	178,700	***1.9
White, not Hispanic	182,100	186,100	***2.2
Black	120,700	121,700	0.8
Asian	217,300	226,000	*4.0
Hispanic (any race)	131,800	130,600	-0.9
Age			
25 to 34 years	147,000	148,000	0.7
35 to 44 years	189,900	191,300	0.7
45 to 54 years	204,200	208,100	**1.9
55 to 64 years	191,200	194,700	1.8
65 years and older	134,100	134,400	0.2

Source: 2017 CPS ASEC Production and Research Files. All percentiles are calculated using linear interpolation. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels respectively.

Household Income for Select Subgroups 95th Percentile (in dollars)

Characteristic	Production	Research	Percent Difference
All Households	226,000	233,300	***3.2
Family households	254,000	264,100	***4.0
Nonfamily households	151,200	147,600	**-2.4
Race			
White	231,300	240,200	***3.8
White, not Hispanic	240,200	250,600	***4.3
Black	158,000	159,800	1.2
Asian	276,000	289,600	*4.9
Hispanic (any race)	170,500	168,400	-1.2
Age			
25 to 34 years	189,500	190,400	0.5
35 to 44 years	250,500	259,000	***3.4
45 to 54 years	273,600	280,700	*2.6
55 to 64 years	251,700	262,200	***4.2
65 years and older	181,700	190,300	***4.7

Source: 2017 CPS ASEC Production and Research Files. All percentiles are calculated using linear interpolation. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent levels respectively.

Contact Information

Jonathan Rothbaum

Chief, Income Statistics Branch

Social, Economic, and Housing Statistics Division

jonathan.l.rothbaum@census.gov

(301) 763-9681

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov