Using American Community Survey Headship Rates in Small Area Estimates

American Community Survey Data Users Conference
May 18, 2023

Rachel Cortes
Ken Hodges

claritas

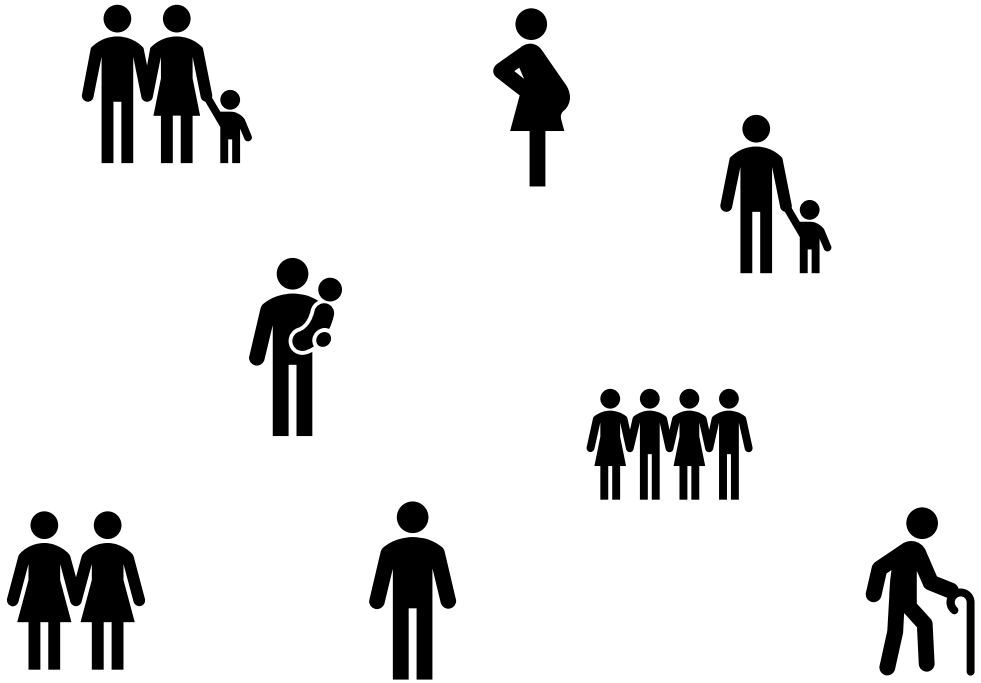
Claritas Products and Methods

- Build information products
  - For many businesses and use cases
- Annual demographic estimates
  - All block groups nationwide
  - Build from census bureau and other data
  - Develop headship rates by age, race/ethnicity, and sex

claritas

Differential Privacy and Headship Ratios

- 2020 Census products will use Differential Privacy (DP)
- Combining household and person records can cause impossible or improbable headship ratios
  - 103 persons and 124 households in a block group
  - 34 persons aged 24-29 and 52 household heads ages 24-29



Data and Methods

- Census Bureau provided versions of 2010 Summary File 1 files (SF1) using the DP privacy protection
  - iPUMS tabulations of the DP demo files and SF1 tables as published
- 2008-2012 ACS data, centered on 2010
- Created tables from the ACS, DP Demo files, and SF1:
  - Population by age and sex
  - Householders by age group

claritas

Data and Methods

- Calculate headship ratios
- Cases where pop = 0 but households > 0
  - Set the ratio equal to the number of households:
  - Pop = 0 and Households = 2, set headship ratio = 2
  - The higher the rate, the more “inconsistent” the data are

Descriptive Statistics for DP-based headship ratios

Age group	N	Min	Max	Mean
15-24	216,804	0.000	6.000	0.12750
25-34	216,794	0.000	19.000	0.44521
35-44	216,762	0.000	18.000	0.53846
45-54	216,727	0.000	14.000	0.57258
55-59	216,555	0.000	25.000	0.61414
60-64	216,561	0.000	53.000	0.63216
65-74	216,540	0.000	22.000	0.66214
75-84	216,333	0.000	34.000	0.73617
85 plus	215,856	0.000	31.000	0.83018
Valid N	215,644			

claritas

Large inconsistencies occur, but how often?

- Created a binary variable to categorize each headship rates as possible or impossible
- Block groups can have more than one inconsistent headship ratio

Sum the count of improbable rates to see extent of issues with DP data

Sum of improbable rates	Frequency	Percent	Cumulative	Total
0	143,714	66.0	66.0	0
1	55,726	25.6	91.6	55,726
2	13,940	6.4	98.0	27,880
3	3,414	1.6	99.6	10,242
4	722	0.3	99.9	2,888
5	176	0.0	100.0	880
6	40	0.0	100.0	240
7	7	0.0	100.0	49
8	1	0.0	100.0	8
Total	217,740	100.0	100.0	97,913

claritas

What to do about improbable headship ratios?

- In SF1 and in the ACS, there are no impossible values
- Can we substitute the out-of-bound rates with ACS data?
- If we do, do we substitute all the rates, or just the impossible ones?
- Even rates that are below 1.00 can be unrealistic

- Evaluate how different the DP-based headship ratios are from SF1-based rates
- Calculate the absolute difference
  - Ex: The DP-ratio = 0.519 and the SF1 ratio = 0.533
  - Absolute difference = 0.014
  - Sum these differences across categories
  - See where DP-induced differences were largest and compare to ACS and SF1 tables

claritas

Descriptive Statistics for Absolute Differences by Age Group: DP vs. SF1 and ACS vs. SF1

	DP vs. SF1		ACS vs. SF1	
Age group	**N**	**Mean**	**N**	**Mean**
15-24	216,633	0.03748	214,378	0.08483
25-34	216,653	0.07608	213,831	0.13315
35-44	216,608	0.07745	214,703	0.12480
45-54	216,577	0.07371	215,473	0.11192
55-59	216,344	0.12594	211,619	0.17426
60-64	216,269	0.13519	209,312	0.18376
65-74	216,186	0.13153	210,686	0.14863
75-84	215,742	0.19625	199,214	0.19147
85 plus	213,940	0.36104	157,509	0.26743
Average Absolute Difference	216,827	0.13545	216,648	0.15519

claritas

Limit comparison to only inconsistent ratios

	DP vs. SF1		ACS vs. SF1	
Age group	N	Mean	N	Mean
15-24	131	1.508	58	0.239
25-34	826	1.498	538	0.240
35-44	1,618	0.856	1,323	0.218
45-54	2,042	0.710	1,826	0.194
55-59	7,412	0.771	6,675	0.235
60-64	9,077	0.687	8,180	0.244
65-74	10,142	0.727	8,945	0.212
75-84	22,330	0.779	18,461	0.234
85 plus	42,845	1.022	28,662	0.276
Average Absolute Difference	73,658	0.834	61,141	0.248

claritas

ACS County Ratios

- Sometimes when the DP ratio is out-of-bounds, the ACS doesn’t have an alternative ratio
- Small ACS sample for that block group and age group and headship ratio cannot be calculated
- Could we use the ACS county-level ratio as a substitute?

- By substituting ACS county ratios:
  - The mean absolute difference is reduced slightly from 0.248 to 0.235
  - Provided valid headship ratios in more instances than where the DP ratio exceeded 1.00 and the mean difference was actually lower than before substitution

Data Source	N	Mean Absolute Difference
DP Data	73,658	0.834
ACS BG Only	61,141	0.248
ACS with County and BG	73,656	0.235

How often were the ACS county ratios closer to SF1 than DP when we substitute county ratios when ACS BG ratios are missing?

- ACS ratio improved on DP ratios in 88,886 out of 1,959,660 ratios (4.54 percent of the total)
- SF ratios with DP were closer to SF1 without DP than 5-year ACS data in most cases
- ACS can improve on SF1 with DP even when DP did not result in a ratio exceeding 1.00
  - Out of 97,913 ratios where DP > 1.00, the ACS ratio was closer in 88,886
  - In the other 8,827 the ratio > 1.00 but it was closer to SF1 than ACS
    - Ex: SF ratio = 0.950, DP ratio = 1.100, ACS ratio = 0.700

claritas

When the DP ratio exceeds 1.00, the corresponding ACS ratio always provided a valid ratio and was almost always closer to SF1

Age Group	Not Improved	Percent	Improved	Percent
15-24	34	21.0	128	79.0
25-34	55	6.3	815	93.7
35-44	129	7.8	1,534	92.2
45-54	135	6.5	1,942	93.5
55-59	532	7.1	6,954	92.9
60-64	756	8.2	8,415	91.8
65-74	660	6.4	9,624	93.6
75-84	1,790	7.9	20,821	92.1
85 plus	4,936	11.3	38,653	88.7
Total	97,913	9.2	88,886	90.8

claritas

DP County-level data vs. SF1

- What about summing the DP block group data to the county-level? Would it outperform the ACS block group + county-level data?
- Calculate headship ratios at the county-level by summing up the DP block group data
  - The county-level ratios exceeded 1.00 in 83 out of 1.9 million ratios

	DP vs. SF1		ACS vs. SF1		DP with County vs. SF1	
Age Group	N	Mean	N	Mean	N	Mean
15-24	216,633	0.03748	214,378	0.08483	216,688	0.07185
25-34	216,653	0.07608	213,831	0.13315	216,686	0.07747
35-44	216,608	0.07745	214,703	0.12480	216,661	0.05682
45-54	216,577	0.07371	215,473	0.11192	216,623	0.05096
55-59	216,344	0.12594	211,619	0.17426	216,425	0.06048
60-64	216,269	0.13519	209,312	0.18376	216,336	0.06384
65-74	216,186	0.13153	210,686	0.14863	216,251	0.06190
75-84	215,742	0.19625	199,214	0.19147	215,844	0.08573
85 plus	213,940	0.36104	157,509	0.26743	214,210	0.16291
Total	216,827	0.13545	216,648	0.15519	216,864	0.07739

claritas

How often do DP county-rates improve the DP block group rates?

- The max for DP-county does exceed 1.00 for age 25-34 and 85+. But only a few block groups.
- What about when the DP block group rate > 1.00?
- For both all rates and rates that exceed 1.00 the DP data summed to the county performed better than the ACS BG data
  - Improved 93.5 percent of ratios vs 90.8 percent of ratios
  - Better for all age groups except for 15-24 and 25-34 where ACS was slightly better

Conclusions

- Differential privacy gives impossible values for headship ratios at the block group-level
- This causes issues in our estimates process
- Decide what to do…
- Substituting the rates for ACS rates improved consistency with the SF1 data
- But ACS rates were not always available
- When DP rates were > 1.00, it was better to use DP based data at the county level than to use ACS rates…but only slightly

claritas

Thank You!

Rachel Cortes
rachel.cortes@claritas.com

claritas