Explaining Data Adjustments IHBG Negotiated Rulemaking

January 26, 2016

Agenda

- Overview - morning
  - Step-by-step – how did HUD come to this recommendation?
  - How the three adjustments work together
  - Q&A
- Adjustment 2: Aging the Decennial Census data - morning
  - The data accuracy problem we are looking to solve
  - The proposed solution and why
  - Q&A
- Adjustment 1: Decennial Census undercount adjustment - afternoon
  - The data accuracy problem we are looking to solve
  - The proposed solution and why
  - Q&A
- Adjustment 3: ACS reweighting adjustment - afternoon
  - The data accuracy problem we are looking to solve
  - The proposed solution and why
  - Q&A

Overview: HUD's process

- **Issues.** Study group had identified some issues about ACS and Census data and had offered up a consensus recommendation and several non-consensus options to mitigate those issues.

- **Research.** As promised, HUD sat down with Census staff to discuss some of the concerns raised about Census 2010 and ACS.

- **Findings.** Census staff and HUD discussed the Census 2010 undercount in some tribal areas and the ACS methodology for weighting data.

- **Policy.** With knowledge of both an undercount and the ACS methodology for weighting, HUD developed a mitigation strategy to improve the accuracy of the estimates in tribal areas.

- **Vetting.** After HUD developed an approach, HUD discussed the strategy with several statistical experts from other agencies. There was general agreement on the problem. No one raised any objections to the solution; and in the short time-frame, better solutions were not identified.

The three adjustments

Data Sources:

- Census 2010
- American Community Survey

Adjustments:

- **First**, correct for Decennial Census AIAN undercount in Reservation and Trust Land areas
- **Second**, age the corrected Decennial Census AIAN undercount using Census county level population estimates for AIAN
- **Third**, reweight the 6 needs variables from the ACS to match the AIAN aged decennial census population counts

Example

		Example Reservation/ Trust Land Area	Example non-Reservation/ Trust Land Area
A	Base: Census 2010 AIAN Alone Count	1,000	1,000
B	Adjust for Census 2010 Undercount	x 4.88%	
C=A*B	Subtotal	1,049	1,000
D	Adjust for county AIAN population growth to 2014 (latest year with pop estimates)	x 5.1%	x 5.1%
E=C*D	Subtotal	1,103	1,051
F	Base: ACS AIAN Count most recent 5-year data	900	900
G=E/F	Adjustment Ratio for Needs Data	1.23	1.17

Data Runs

- Explanation of the handouts

Overview of formula change impact on tribe allocations	ACS 2012 and Census 2010, No Adjustments	With census undercount adjustment and ACS reweighting (no aging)	With census undercount adjustment and aging, ACS reweighting	Added effect of the undercount and ACS reweight adjustment	Added effect of aging to 2014
Range of impacts for tribes with grants greater than minimum grant					
Min	-63%	-63%	-65%	-95%	-22%
10th pctile	-25%	-25%	-26%	-12%	-3%
1st quartile	-10%	-9%	-10%	-3%	-2%
Median	0%	0%	0%	0%	0%
3rd quartile	8%	8%	9%	4%	1%
90th pctile	23%	22%	21%	13%	2%
Max	186%	159%	157%	121%	20%
Number	495	495	495	495	495
Range of impacts for tribes with current grants over $250,000					
Min	-62%	-58%	-65%	-34%	-7%
10th pctile	-12%	-12%	-12%	-6%	-3%
1st quartile	-5%	-5%	-4%	-2%	-1%
Median	1%	2%	2%	0%	0%
3rd quartile	8%	8%	9%	2%	1%
90th pctile	18%	19%	17%	8%	2%
Max	86%	83%	81%	47%	20%
Number	310	310	310	310	310

Impact of updating ACS data from 06-10 to 08-12	ACS 06-10 to ACS 08-12 effect	ACS 10 v. ACS 12 with all adjustments
Range of impacts for tribes with larger than minimum grant		
Min	-62%	-89%
10th pctile	-13%	-13%
1st quartile	-5%	-5%
median	0%	0%
3rd quartile	5%	5%
90th pctile	14%	14%
Max	56%	114%
Number	495	495
Range of impacts for tribes with current grant over $250,000		
Min	-46%	-35%
10th pctile	-11%	-9%
1st quartile	-4%	-4%
median	0%	1%
3rd quartile	5%	5%
90th pctile	12%	10%
Max	42%	64%
Number	310	310

Overview Q & A

Adjustment 2: Aging the Decennial Census data

- The data accuracy problem we are looking to solve.
  - Decennial Census data are only collected every 10 years.
    - Just one of the seven IHBG formula variables are available from this source – AIAN population.
    - The 2010 Census data are "100% count" data. Everyone is expected to respond to the survey. That means no sample error. That is Great! (if there is no undercount)
  - If data only collected every 10 years, we do not account for annual population change.
- **Aging the data is our proposed solution to this data accuracy problem.**

Current method of aging the data

- The current formula ages the Census 2000 Decennial Census data using IHS Birth/Death data. Study Group identified these concerns with IHS data:
  - Using 2000 Census as base and not 2010 census
  - Underreporting concerns in tribal areas
  - Failure to address the migration of AIAN persons,
  - Does not distinguish between AIAN Alone and AIAN Alone or In Combination populations
  - Does not provide AIAN person data at the formula area, rather than county, level.

Census Population Estimates

- The Study Group evaluated an alternative to the IHS data for aging. That alternative is the Census Bureau's population estimate program:
  - Like IHS, data are only available at the county level and it likely has underreporting for tribal areas.
  - But it has three big improvements over IHS:
    - It uses Census 2010 as its base instead of Census 1990
    - In addition to births and deaths, it also adjusts for migration in and out of the county.
    - It provides separate estimates for both AIAN alone and AIAN alone with multi-race.

Pros and Cons

- Pros.
  - Aligns with the ACS. The ACS also uses the Census population estimates to create their county control totals on their weighting.
  - Publicly available, easy to incorporate.
- Cons.
  - County Level
  - Underreporting in Tribal Areas

Some data

- 2010 to 2014: AIAN population growth in 613 counties with IHBG formula areas:

	AIAN Alone	AIAN Multi
25% of counties have pop growth less than:	1.3%	2.0%
Median county	5.1%	5.5%
25% of counties have pop growth greater than	11.2%	10.9%

Data runs

- The data runs show that aging the data matters.
- Look at slide 7

Q & A

Aging Adjustment

Adjustment 1: Adjusting for undercount in Census 2010 Data

- The data accuracy problem we are looking to address:
  - 2010 Census Coverage Measurement Estimation Report
  - Key Finding: For Reservation/Trust Land only, a 4.88% undercount of AIAN persons.
  - Other tribal areas did not have the undercount.
  - Undercount in rural Alaska is unknown because it was not part of the study.
- **Increasing the AIAN 2010 Decennial Census population counts in Reservation/Trust Lands by 4.88% is the proposed solution to this data accuracy problem**.

Table 2: Components of Census Coverage by Race Alone-Or-In-Combination and Hispanic Origin

Race Alone-Or-In-Combination and Hispanic Origin	Census Count (Thousands)	Correct (%)	Erroneous (%)		Whole-Person Imputations (%)	Population Estimate (Thousands)	Correct (%)	Pct Undercount (%)	Omissions (%)
			Duplication	Other					
U.S. Total	300,703	94.7	2.8	0.5	2.0	300,667	94.7	-0.01	5.3
	(0)	(<0.1)	(<0.1)	(<0.1)	(0)	(429)	(0.1)	(0.14)	(0.1)
White	225,547	95.2	2.7	0.4	1.7	224,341	95.7	-0.54*	4.3
	(0)	(<0.1)	(<0.1)	(<0.1)	(0)	(319)	(0.1)	(0.14)	(0.1)
Non-Hispanic White Alone	191,997	95.4	2.6	0.4	1.6	190,413	96.2	-0.83*	3.8
	(0)	(<0.1)	(<0.1)	(<0.1)	(0)	(291)	(0.1)	(0.15)	(0.1)
Black	40,153	92.6	3.6	0.7	3.1	40,999	90.7	2.06*	9.3
	(0)	(0.2)	(0.2)	(<0.1)	(0)	(209)	(0.4)	(0.50)	(0.4)
American Indian and Alaskan Native	5,056	92.5	4.1	0.6	2.9	5,063	92.4	0.15	7.6
	(0)	(0.6)	(0.6)	(<0.1)	(0)	(36)	(0.6)	(0.71)	(0.6)
On Reservation	571	90.8	4.7	0.4	4.1	600	86.3	4.88*	13.7
	(0)	(0.6)	(0.6)	(<0.1)	(0)	(15)	(2.1)	(2.37)	(2.1)
American Indian Area Off Reservation	527	87.8	9.7	1.0	1.5	507	91.2	-3.86	8.8
	(0)	(4.1)	(3.9)	(0.5)	(0)	(15)	(2.6)	(2.99)	(2.6)
Balance of the U.S.	3,959	93.4	3.2	0.6	2.9	3,956	93.4	-0.05	6.6
	(0)	(0.4)	(0.4)	(<0.1)	(0)	(23)	(0.6)	(0.58)	(0.6)
Asian	16,969	94.7	2.4	0.9	2.1	16,969	94.7	0.00	5.3
	(0)	(0.2)	(0.2)	(<0.1)	(0)	(88)	(0.5)	(0.52)	(0.5)
Native Hawaiian or Pacific Islander	1,189	93.1	3.4	0.8	2.8	1,201	92.1	1.02	7.9
	(0)	(0.6)	(0.6)	(0.2)	(0)	(25)	(2.0)	(2.06)	(2.0)
Some Other Race	21,448	92.9	3.5	0.7	2.9	21,803	91.4	1.63*	8.6
	(0)	(0.3)	(0.3)	(<0.1)	(0)	(69)	(0.4)	(0.31)	(0.4)
Hispanic Origin	49,580	93.7	3.2	0.7	2.4	50,356	92.3	1.54*	7.7
	(0)	(0.2)	(0.2)	(<0.1)	(0)	(168)	(0.3)	(0.33)	(0.3)

A person can be included in multiple rows.
Standard Errors are shown in parentheses below the estimate. An asterisk (*) denotes a percent net undercount that is significantly different than zero.
The 2010 Census count excludes persons in group quarters and persons in Remote Alaska.

Data runs

- This adjustment applied just to the Decennial Census 2010 AIAN variable has a small effect because the AIAN variable allocates just 11% of the needs funding in the formula.
- It has a much larger impact is when it is used in conjunction with the reweighting of the ACS variables. That will be our next discussion.

Q & A

Decennial Census undercount adjustment of AIAN variable

Adjustment 3: reweighting the ACS

- The Data Accuracy problem we are looking to solve:
  - ACS is a sample survey that uses as population estimates (by race/ethnicity, age, gender) at the county level. For sub-areas within county – such as tribal areas - there is random variance in population counts that result in Census 2010 100% population counts not matching ACS population counts for the same area in the same time period.
  - This is different than for Census 2000, where small areas such as tribal areas used the Census 2000 100% total as the control totals for the Census 2000 sample data (that the need data are derived from).
- **Reweighting the ACS data so that they match population growth and undercount adjusted 2010 Census data improves the small area ACS estimates and all needs variables**

Census 2010 Compared to ACS 08-12

Census Defined Tribal Areas with 100 more AIAN alone or multi-race (N=411)			
	Census 2000 Sample Data Count AIAN Alone Compared to Census 2000 100% Data	ACS 2008-12Data Count AIAN Alone Compared to Census 2010 100% Data	ACS 2008-12 Data Count AIAN Multi Compared to Census 2010 100% Data
25% of areas have AIAN counts from sample data that are *less than* their 100% AIAN count by more than this amount:	-6%	-15%	-11%
Median	-1%	-3%	-1%
25% of areas have AIAN counts from sample data that are *greater than* their 100% AIAN count by more than this amount:	3%	11%	10%
Census Defined Tribal Areas with 1000 more AIAN alone or multi-race (N=124)			
25% of areas have AIAN counts from sample data that are *less than* their 100% AIAN count by more than this amount:	-4%	-13%	-7%
Median	-2%	-5%	-1%
25% of areas have AIAN counts from sample data that are *greater than* their 100% AIAN count by more than this amount:	0%	1%	3%

County with 10,000 people with 6 "areas" in county - 100% count

3000	2000	1000
2000	1000	1000

If 1-in-10 households are randomly selected, each household is weighted as "10". When the sample data are aggregated, we get this population count in each area. Still adds up to 10,000, but not an exact match to the 100% data because of sampling error. For non-incorporated places, this is what ACS uses.

2900	1900	1200
2100	900	1000

For Census 2000, they adjusted the sample weights so subcounty geographic populations for the weighted sample data matched the 100% data for areas as small as 400 people. By doing this, they improved the accuracy of the counts in the small areas.

3000	2000	1000
2000	1000	1000

Q & A

Reweighting ACS with Census 2010 AIAN counts