Patterns of Uncertainty, Inaccuracy and Statistical Erasure: Where does the ACS-ED (2005-2019) Fail the Next Generation of Native American Scholars Most?

Jeffrey D. Burnette

Department of Sociology and Anthropology, Rochester Institute of Technology

PRELIMINARY DRAFT – DO NOT CITE WITHOUT PERMISSION

Abstract

This study examines patterns of both uncertainty and accuracy in the ACS – Education Tabulation (ACS-ED) from 2005 to 2019 by focusing on American Indian and Alaska Native (AI/AN) kindergarten through grade 12 (K-12) population totals for individual school districts across an array of different geographic definitions. It finds 80% of ACS-ED population estimates for AIs/ANs in the most recent data sample are either low or poor quality according to standard classifications using the coefficient of variation. Additionally, it finds troubling errors in the accuracy of ACS-ED estimates. More than half of all districts reporting AI/AN K-12 students in the Common Core of Data are estimated to contain no AI/AN students, thus the ACS-ED misses the existence of more than 29,000 AI/AN students attending more than 5,000 schools nationwide. The paper then identifies 15 school districts in or near AI/AN homeland area with the greatest need of additional resources for accurate estimates.

Key words: Indigenous Statistics, American Community Survey, Common Core of Data, ACS-ED, uncertainty, accuracy, margin of error

JEL Codes: J10, J15

Patterns of Uncertainty, Inaccuracy and Statistical Erasure: Where does the ACS-ED (2005-2019) Fail the Next Generation of Native American Scholars Most?

1. Introduction

The lack of quality data for American Indian and Alaska Native (AI/AN) individuals and communities has resulted in their frequent statistical erasure as small sample sizes and large margins of error have led researchers and government agencies to replace numbers describing their experiences with an asterisk (Shotton, Lowe & Waterman, 2013; Brayboy & Tachine, 2021). This practice has become so commonplace that AIs/ANs have been "described as the 'Asterisk Nation' because an asterisk, instead of a data point, is often used…" (National Congress of American Indian, https://www.ncai.org/policy-research-center/research-data/data). Though this issue is widely accepted as fact, there is a dearth of research in academic journals specifically documenting the amount of data meant to describe AIs/ANs, or its reliability. This study seeks to document one aspect of this issue by examining population estimates of the Kindergarten through Grade 12 (K-12) student population within the American Community Survey – Education Tabulation (ACS-ED). Consequently, it focuses on the uncertainty and accuracy of ACS-ED district level estimates along with the number of school districts estimated to have no AIs/ANs.

This study compares the coefficient of variation (CV) and the mean absolute percent error (MAPE) for AIs/ANs to that for the White student population in the same local education agencies across a number of different geographies. These well-established techniques have regularly been employed to analyze uncertainty and error within the American Community Survey (ACS). Here, an emphasis is placed on understanding how inaccuracies and uncertainty differs within and outside traditional AI/AN homeland areas as well as areas where AIs/ANs comprise the majority or secondary majority population. Overall, this analysis finds that the ACS-ED incorrectly estimates more than 5,000 school districts contain zero American Indian and Alaska Native students. Yet,

their existence is known because administrators report it to the Common Core of Data (CCD). As a result, the ACS-ED missed and thereby erased more than 29,000 American Indian and Alaska Native students from these districts. This is an especially disturbing result considering the increasingly dependence on ACS data to describe communities, distribute resources and develop policy meant to promote economic growth and reduce inequality (Jurjevich et. al, 2018; Jurjevich, 2019; Fuller et. al, 2022).

Lastly, it identifies 14 school districts in the greatest need of resources to improve ACS estimates. Overall, these 14 districts contain nearly 1,500 AI/AN students. Across these areas, AIs/ANs comprise 9.2% of the district's total student population and are either on or less than 25 miles from an American Indian homeland area. These last two points indicate that the issue of missing AI/AN student data in ACS surveys is not limited to sparsely populated by American Indian and Alaska Native communities that are hard to identify.

2. Literature Review

The ACS-ED is a relatively new product that began offering data in 2015, consequently there are no known studies investigating its reliability or error. This study uses two different techniques that have been employed to assess data quality. The first technique focuses on measuring the uncertainty of ACS estimates by examining patterns in the CV (Spielman, Folch & Nagle, 2014; Folch, Arribas-Bel, Koschinsky & Spielman, 2016; Jurjevich, 2019). The second concentrates on determining the accuracy of ACS estimates by comparing them to an established value and calculating the percent difference from that value (Swanson & Hough, 2012). While the uncertainty and accuracy of the ACS has received some attention, very little peer reviewed work has evaluated the quality of AI/AN estimates in the ACS. The most prominent is by Jordan and Beaghen (2013), which documented the coverage of the AI/AN population in 2010 ACS 1-year estimates.

2.1. Measuring Uncertainty within the ACS

Each of the studies utilizing CV to measure uncertainty within the ACS focus on income and use data prior to 2017. They also tend to determine the reliability of ACS estimates based upon either the measure referenced by the National Research Council (NRC) or that suggested by the Environmental Systems Research Institute (ESRI).[1]

While concentrating on estimates of household for African Americans at the tract level for the 150 largest Metropolitan Statistical Areas within the 2007 – 2011 ACS, Spielman, Folch, and Nagle (2014) find that the estimate reliability was lowest for groups with lower income and those living in lower income census tracts. Data for Folch et. al (2016) spans a similar time frame, 2006 – 2010, but their emphasis is on exploring geographic differences and how different characteristics impact the CV. They discovered that the correlation between the percentage of African Americans and the CV was positive and significant for census tracts in both northern and southern states. However, northern and southern differences emerged. The vast majority of statistically significant high-CV clusters were located in southern states while the low-CV clusters were predominately in northern states. Additionally, it was observed that increased racial diversity was associated with significantly lower CVs in southern states, but not in northern states.

Jurjevich (2019) explored a different angle of CV differences and focused on rural areas. Their study included data from 2006 – 2010 as well as that from 2013 – 2017 with the aim of determining if ACS data in rural communities was becoming more or less reliable. Ultimately, they found that statistical reliability (lower CV) tends to decrease for smaller tracts and while the ACS has become more reliable over time, the majority of areas with less than 20,000 people have income estimates classified as either "Somewhat reliable" or "Unreliable". Each of these studies shape

[1] The NRC considers a CV less than or equal to 12% a "reasonable standard of precision" (Citro and Kalton, 2007, p. 64). ESRI by comparison classifies a CV between 12% and 40% to have "medium reliability" (ESRI, 2018, p. 9).

expectations for the expected reliability of ACS data for AIs/ANs considering AIs/ANs tend to live in less racially diverse, rural areas, frequently associated with lower income.

2.2. Measuring the Accuracy Estimates and the ACS

The use of mean absolute percent error (MAPE) to assess the accuracy of population forecasts is extensive, but its application to accessing the accuracy of ACS data is more limited. Hough and Swanson (2006) compared preliminary findings from 1999 – 2001 ACS with long form estimates from the 2000 Census in Mulnomah County, the most populous county in Oregon. For most estimates they found ACS and Census data achieved similar results. However, they also discovered significant differences between these two data sets in estimates for disability and race along with population and housing estimates and Census 2000 counts. Later, Swanson and Hough (2012) utilized data from the decennial census to analyze the accuracy 2010 ACS 1 – year estimates. This time the analysis was extended to include 18 test site counties for the 1999 ACS, but focused on the persons per household variable, utilizing MAPE along with other methods of comparison to determine that the variance levels in ACS 1 – year estimates at the sub-county level are too high to be relied upon by demographers.

The latest and most complete discussion concerning the quality of ACS estimates is provided by Bazuin and Fraser (2013). Their analysis takes a two-pronged approach that includes both a case study and a nation-wide analysis. In both instances, they emphasize a comparison of population totals for different demographic groups based upon whether the ACS estimate and its 90% margin of error includes decennial census values for 2000 and 2010. At the national level, they found ACS population estimates were significantly different from those from the decennial census for all groups except those for single-race American Indians; all other racial groups were found to be more than 5% under and 5% over their estimates the confidence interval. It should be noted, that the lack of difference for American Indians could be the result of disproportionately large margins of error.

More recently, Graves and Gerney (2018) employed MAPE to determine the accuracy of demographic data from five different vendors. They compared income and population data from each vendor for 80 different census tracts in the fastest growing U.S. metropolitan areas against 2015 ACS data in those same areas. The goal of their study was not to assess the accuracy of ACS data. Instead, they used it as a benchmark to determine if the data across the different vendors produced consistent estimates.

3. Methods

This study applies frequently used techniques to analyze uncertainty and error in the American Community Survey (ACS) to document patterns within AI/AN student totals estimated by the American Community Survey – Education Tabulation from 2005 to 2019. First, data is separated into different geographic areas and their summary statistics are compared to determine the total AI/AN, White and the district student population vary inside and outside AI/AN homeland areas in addition to areas with higher concentrations of AIs/ANs. Next, the coefficient of variation (or relative standard error) for ACS-ED estimates of each of these populations for each local educational agency are combined to examine how estimate uncertainty for AIs/ANs varies across these areas and compares to that for other populations. Lastly, the level of accuracy in these estimates is observed by calculating its difference from the respective total in the CCD and then averaged for each geographic area to determine its mean absolute percentage error.

3.1. Data

The ACS-ED is part of the Education Demographic and Geographic Estimates (EDGE) program. It uses spatial data from National Center for Education Statistics boundaries and estimates from the American Community Survey data to generate estimates as granular as the school district level. The ACS-ED is "meant to help policymakers, program administrators, and the public understand relationships between educational institutions and the communities they serve"

(Geverdt, 2017, p. 1). Because the ACS - ED is so new, it was introduced in 2015, research on or using it is limited (Reardon, Kalogrides & Shores 2019; Xu, Solanki & Fink, 2021).

Studies relying upon ACS-ED have utilized the data it provides concerning socioeconomic status for different race and ethnicity groups, while obtaining enrollment totals from the National Center for Education Statistics (NCES). This approach plays to the intended strengths of each dataset. The ACS-ED collects its data from sampling a subset of the population, CCD totals are the result of public elementary and secondary local education agencies uploading their individual totals into a server managed by the NCES. Consequently, the CCD is a "comprehensive, annual, national database of all public elementary and secondary schools and school districts" (National Center for Education Statistics, https://nces.ed.gov/ccd/) One limitation of the CCD is its lack of variables describing socioeconomic status. In contrast, the ACS-ED provides estimates of population totals and a wider array of socioeconomic variables with the aim of providing "communities with reliable and timely demographic, social, economic, and housing data." (National Center for Education Statistics, https://nces.ed.gov/programs/edge/demographic/acs)

In order to offer race and ethnicity data at such a granular level, the ACS-ED pools five years of ACS data. Consequently, the data here is separated into the following intervals: 2005 – 2009, 2010 – 2014, and 2015 – 2019. The CCD is published on an annual basis. In order to account for this difference, values from the CCD are pooled into similar periods and the average for the period is used for each district.

3.2. Calculations and Geographic Areas of Interest

A particular benefit of the EDGE program is its work to connect NCES data with geographic boundaries of other types like: counties, Congressional Districts and American Indian/Alaska Native/Native Hawaiian areas. In each of these cases, there is a lack of direct correspondence; the shapes of local school district and agency boundaries don't perfectly align with

those of any of the other geographic types. School district geographic relationship files allow for school districts to be identified based upon whether a portion of the district overlaps with a shapefile that contains a legal or statistical American Indian/Alaska Native or Native Hawaiian entity (AIANNH homeland). These entities include: American Indian Reservation (AIR) and off-Reservation trust land; state recognized AIRs; Hawaiian Home lands; American Indian Tribal subdivisions; Alaska Native Regional Corporations; Alaska Native village statistical areas; Oklahoma Tribal statistical areas; Tribal designated statistical areas; and state designated Tribal statistical areas. For this analysis, the definitions from the 2020 AIANNH area shapefile are used to separate areas into whether or not they contain an AIANNH homeland area.

Geographic areas are also separated based upon whether or not AIs/ANs are the majority or second largest majority population. These areas are defined based upon CCD district enrollment data from 2015 – 2019. School districts where the AI/AN student population is larger than that of all other race or ethnicity groups are separated and combined to create majority AI/AN population areas. The same is done for secondary majority population areas, except these are school districts where the CCD indicates AIs/ANs are the second most populous group.

Mean values concerning the student population overall, as well as that for White and AI/AN students from ACS-ED estimates and CCD totals for each geographic area, are presented for each of the five-year periods in Table 1; the 90% margin of error for each estimate from the ACS-ED is also provided. The total number school districts is the same for each of the ACS-ED geographic areas from each five-year period from 2005 – 2019. However, due to intertemporal variations like districts merging or separating, and growth in the number of AI/AN student population, local educational agencies in the CCD are able to be mapped to a district in the ACS-ED in all time periods. This also makes it difficult to compare ACS-ED and CCD totals since the number of school districts included in the total is not the same.

Table 1. Mean Student Population Totals by Race, Geographic Area, Year and Dataset

	2005-2009			2010-2014			2015-2019		
	CCD	ACS-ED		CCD	ACS-ED		CCD	ACS-ED	
	Count	Estimate	90% MOE	Count	Estimate	90% MOE	Count	Estimate	90% MOE
All School Districts									
AI/AN	549,952	511,186	10,945	498,408	539,198	5,721	443,987	556,613	6,070
White	25,993,754	37,192,696	36,193	24,112,984	37,584,108	36,964	22,495,182	36,843,560	38,363
Total Population	46,563,008	54,153,120	39,330	46,871,908	55,087,168	39,105	46,002,252	54,988,592	41,020
# of School Districts	11,048	11,116		11,061	11,116		11,048	11,116	
AIANNH Homeland Areas									
AI/AN	346,038	298,386	4,168	310,713	312,689	3,605	286,858	320,599	3,549
White	2,549,573	4,074,802	11,727	2,319,222	4,114,149	11,689	2,161,933	4,077,182	12,228
Total Population	5,435,131	6,259,065	12,621	5,507,479	6,441,195	12,102	5,545,468	6,483,015	12,403
# of School Districts	1,193	1,203		1,199	1,203		1,198	1,203	
Outside AIANNH Homeland Areas									
AI/AN	203,913	212,800	10,120	187,695	226,509	4,442	157,129	236,014	4,925
White	23,444,182	33,117,896	34,240	21,793,764	33,469,958	35,067	20,333,250	32,766,380	36,362
Total Population	41,127,876	47,894,056	37,250	41,364,428	48,645,976	37,185	40,456,784	48,505,576	39,100
# of School Districts	9,855	9,913		9,862	9,913		9,850	9,913	
Majority AIAN Population Areas									
AI/AN	129,362	140,094	2,504	124,123	137,985	1,801	120,567	144,015	1,843
White	34,300	52,032	1,759	30,072	46,450	1,357	25,833	42,387	1,274
Total Population	184,636	221,825	2,947	182,106	216,775	2,218	176,365	217,280	2,140
# of School Districts	181	181		181	181		181	181	
AIAN Secondary Majority Population Areas									
AI/AN	90,154	56,394	1,769	87,299	64,888	1,533	82,988	67,175	1,525
White	326,350	401,580	3,662	307,637	398,270	3,287	292,734	383,260	3,134
Total Population	492,012	535,230	4,012	489,373	542,745	3,621	483,984	535,050	3,463
# of School Districts	440	440		444	444		444	444	

Overall, 10.8% of ACS-ED school districts contain an AIANNH homeland area. Meanwhile, AIs/ANs are the majority population in 1.6% and the second largest group in 4.0% of all school districts. It is worth noting that the average number of students in each district varies substantially across these different geographic areas, but is relatively consistent over time. For instance, the average number of students attending districts in outside AIANNH homeland areas tends to about 90% of the total in schools containing an AIANNH homeland area. In contrast, schools where AIs/ANs comprise the most or second most students tend to be much smaller than the average school. In both instances, the total number of students in the district is about one-fourth the size of the average school district.

The coefficient of variation provides a measure of relative amount of sampling error associate with an estimate. Calculated by dividing the standard error of an estimate by value of the estimate, it provides a measure of uncertainty, since it indicates the estimated percentage error of an

estimate. This calculation is performed for every school district able to be matched between the ACS-ED and CCD and essentially describes the Census' degree of confidence in the estimated value.

One particular issue with the CV is that it does not exist in areas where the estimated value is zero. Hence, this type of analysis excludes areas where the number of AI/AN students is estimated to be zero. Consequently, it becomes especially valuable to find an alternative measure of estimate quality. This is done by comparing the percentage difference between the ACS-ED estimate and the CCD. In this case, the total from the CCD is taken to be the true value and the percentage error is calculated by dividing the difference between two datasets by the value from the CCD for each district level total. The mean absolute percentage errors (MAPE) is then calculated for each geographic area and time period to assess the accuracy of AI/AN data within the ACS-ED.

4. Results

The mean and median values from for CV calculations for each geographic area and time period are displayed in Tables 2. A standard metric for interpreting CV comes from ESRI. They consider CVs: less than 0.12 to be high reliability; between 0.12 and 0.40 to be medium reliability; and those over 0.40 to be low reliability. According to this criterion, the average estimates for all school districts of the White and total population are highly reliable, whereas those for AIs/ANs have low reliability in all periods. In general, the estimates for the total student population exhibit relatively low uncertainty since they are classified as highly reliable, regardless of which geography or time period is chosen.

Disaggregating by race tends to increase the CV, but most of the estimates for the white student population tend to concentrate around the high and medium reliability border for at least the mean or median CV. There is a notable exception in areas where AIs/ANs are the majority population, the CV for white student totals is consistently in the medium reliability range and even approach the

low reliability category. This is nearly the opposite pattern exhibited for AIs/ANs as this is one of the few areas where either mean or median CVs approach the value for high reliability.

Table 2. Coefficient of Variation for ACS-ED Estimates by Race, Geographic Area, and Year

	2005-2009			2010-2014			2015-2019		
	Mean	Median	Number of School Districts	Mean	Median	Number of School Districts	Mean	Median	Number of School Districts
All School Districts									
AI/AN	0.74	0.72	3,719	0.73	0.70	3,895	0.72	0.70	3,648
White	0.11	0.09	11,048	0.11	0.09	11,049	0.11	0.09	11,047
Total Population	0.09	0.08	11,048	0.09	0.08	11,061	0.09	0.08	11,048
AIANNH Homeland Areas									
AI/AN	0.45	0.40	1,029	0.41	0.33	1,029	0.41	0.32	1,029
White	0.15	0.12	1,197	0.16	0.11	1,197	0.15	0.12	1,197
Total Population	0.11	0.10	1,193	0.10	0.09	1,199	0.10	0.09	1,198
Outside AIANNH Homeland Areas									
AI/AN	0.82	0.78	2,690	0.82	0.77	2,866	0.81	0.77	2,619
White	0.10	0.09	9,851	0.10	0.09	9,852	0.11	0.09	9,850
Total Population	0.09	0.08	9,855	0.09	0.08	9,862	0.09	0.08	9,850
Majority AIAN Population Areas									
AI/AN	0.23	0.20	181	0.17	0.15	181	0.16	0.14	181
White	0.36	0.30	157	0.40	0.27	156	0.37	0.27	156
Total Population	0.15	0.14	181	0.12	0.11	181	0.12	0.11	181
AIAN Secondary Majority Population Areas									
AI/AN	0.50	0.20	387	0.40	0.15	391	0.41	0.14	391
White	0.16	0.30	440	0.14	0.27	444	0.14	0.27	444
Total Population	0.14	0.14	440	0.12	0.11	444	0.12	0.11	444

Table 3 displays the mean and median absolute percentage error for each geographic area and five-year time period from 2005 to 2019. Overall, there is evidence that ACS-ED estimates from 2015-2019 are less accurate than those of prior years. MAPE values for this time period generally exceed their corresponding value in other periods in most areas for most groups. CV and MAPE offer similar information since CV measures the standard error of an estimate divided by the estimate, while MAPE measures the absolute difference between an estimate and its established value divided by its established value. One way to consider the relationship between the two is that CV is an estimate of the expected accuracy or MAPE. Consequently, ESRI definitions are also applied to assess the accuracy of MAPE values.

Again, estimates for the total population tend to straddle the high-medium border with values falling on both sides. Similarly, estimates concerning the White and AI/AN totals tend to be less accurate than that of the population as a whole. The main difference from CV values concerns the size of the difference between total and racially disaggregated values; AI/AN and White

estimates are less accurate than their CV estimates suggest. This is most prominent for AI/AN values in all districts and outside AIANNH homeland areas; the majority of mean MAPE values in these areas are greater than one, while their median values are equal to one. This implies that 50 percent of all estimates are off by the size of their estimated value with many estimates exceeding more than double their actual value. This can only happen if the ACS-ED estimates there are no AI/AN students in a district or if the estimate is more than twice the actual value. As a result, there is evidence that there may be a large number of school districts where the ACS-ED incorrectly estimates there are no AI/AN students and/or the ACS-ED provides an relatively large overcount in these geographic areas.

Table 3. Absolute Percentage Difference Between ACS-ED and CCD by Race, Geographic Area and Year

	2005-2009			2010-2014			2015-2019		
	Mean	Median	Number of School Districts	Mean	Median	Number of School Districts	Mean	Median	Number of School Districts
All School Districts									
AI/AN	1.71	1.00	10,052	1.65	1.00	10,099	1.77	1.00	9,895
White	0.39	0.11	11,045	0.28	0.11	11,060	0.54	0.13	11,048
Total Population	0.14	0.09	11,048	0.14	0.09	11,061	0.19	0.10	11,048
AIANNH Homeland Area									
AI/AN	0.72	0.58	1,192	1.13	0.50	1,193	0.78	0.48	1,186
White	0.97	0.15	1,191	0.56	0.15	1,198	0.44	0.16	1,197
Total Population	0.20	0.12	1,193	0.20	0.12	1,199	0.21	0.12	1,198
Outside AIANNH Homeland Areas									
AI/AN	1.85	1.00	8,860	1.73	1.00	8,906	1.90	1.00	8,709
White	0.32	0.10	9,854	0.24	0.11	9,862	0.55	0.12	9,850
Total Population	0.13	0.09	9,855	0.14	0.09	9,862	0.19	0.10	9,850
Majority AIAN Population Area									
AIAN	0.47	0.35	181	0.52	0.29	181	0.50	0.30	181
White	5.14	0.53	179	2.44	0.54	180	1.48	0.56	180
Total Population	0.39	0.21	181	0.38	0.20	181	0.43	0.21	181
AIAN Secondary Majority Population Areas									
AIAN	0.79	0.35	434	0.74	0.29	444	0.60	0.30	444
White	0.27	0.53	440	0.28	0.54	444	0.27	0.56	444
Total Population	0.22	0.21	440	0.22	0.20	444	0.23	0.21	444

Mean MAPE values for AIs/ANs in areas with an AIANNH homeland or where AIs/ANs are one of the majority populations are all less than one, but all fall within the low accuracy range since they are all more than 0.40. Median MAPE values offer evidence in areas where AIs/ANs are either the largest or second largest population large errors in a small number of districts may be a driving factor, since they fall in the medium accuracy range. It is also worth noting that estimates of

the White student population in these areas all fall in the low accuracy range and the mean MAPE value for White students in majority AI/AN areas is larger than one.

An examination of mean and median MAPE values for the total population in three geographic areas with a disproportionately high number of AIs/ANs shows that estimates are less accurate than those without an AIANNH homeland and in general. Median MAPE values for the total population fall within the high accuracy range for all school districts and outside AIANNH homeland areas. Meanwhile, these values are in the medium accuracy range in areas where AIs/ANs are the secondary majority population or containing an AIANNH homeland and in the low accuracy range in areas where AIs/ANs are the majority population.

This is broken down into additional detail in Table 4, it displays the distribution of district estimates based upon their level of accuracy for the period 2015 – 2019 along with the number of districts where the ACS-ED falsely estimates there are no AIs/ANs and the total number of AIs/ANs living in these districts. Overall, there are 6,299 districts with false zeros for the AI/AN student population and 29,811 AI/AN students living in these districts. The vast majority of these are in districts that are outside AIANNH homeland areas and is likely caused by the small number of AIs/ANs; the average total number of AIs/ANs in each district is only 16 students.

Unsurprisingly, there are no districts with false zeros where AIs/ANs make up the majority of students. However, other places where you would expect a sizable AI/AN population, those in an AIANNH homeland or where AIs/ANs are the secondary majority, the percentage of districts with false zeros makes up more than 11% of the total. Consequently, 2,552 AI/AN students living in districts with an AIANNH homeland where the ACS-ED estimates there are none. The corresponding total in AI/AN secondary majority areas is 1,191 AI/AN students.

Table 4. Distribution of Uncertainty, Accuracy and False Zeros by Race and Geographic Area (2015 – 2019)

	AI/AN		White		Total Population	
	Coefficient of Variation	Absolute Percentage Error	Coefficient of Variation	Absolute Percentage Error	Coefficient of Variation	Absolute Percentage Error
All School Districts						
# High Quality (0-12%)	114	369	5,859	5,326	7,626	6,251
# Moderate Quality (12% - 40%)	675	802	4,799	4,212	3,483	3,971
# Low Quality (40%-100%)	2,106	7,591	310	1,129	7	745
# Poor Quality (>100%)	701	1,133	24	380	0	81
# of School Districts with False Zeros	6,299		55		0	
# of AI/AN Students in False Zero Areas	29,811					
Mean Percentage of Group in District	0.96%		48.90%			
Average Student Group Total in District	40		2,036		4,164	
AIANNH Homeland Areas						
# High Quality (0-12%)	111	170	497	471	778	582
# Moderate Quality (12% - 40%)	485	326	604	467	417	473
# Low Quality (40%-100%)	375	593	59	189	3	122
# Poor Quality (>100%)	53	97	11	70	0	21
# of School Districts with False Zeros	162		26		0	
# of AI/AN Students in False Zero Areas	2,552					
Mean Percentage of Group in District	5.16%		38.99%			
Average Student Group Total in District	239		1,805		4,629	
Outside AIANNH Homeland Areas						
# High Quality (0-12%)	4	199	5,362	4,855	6,801	5,669
# Moderate Quality (12% - 40%)	195	476	4,195	3,745	3,045	3,498
# Low Quality (40%-100%)	1,732	6,998	251	940	4	623
# Poor Quality (>100%)	648	1,036	13	310	0	60
# of School Districts with False Zeros	6,130		29		0	
# of AI/AN Students in False Zero Areas	27,259					
Mean Percentage of Group in District	0.39%		50.26%			
Average Student Group Total in District	16		2,064		4,107	
Majority AI/AN Districts						
# High Quality (0-12%)	71	43	7	27	107	63
# Moderate Quality (12% - 40%)	106	64	100	41	73	68
# Low Quality (40%-100%)	4	59	40	68	1	37
# Poor Quality (>100%)	0	15	9	44	0	13
# of School Districts with False Zeros	0		24		0	
# of AI/AN Students in False Zero Areas	0					
Mean Percentage of Group in District	68.38%		14.68%			
Average Student Group Total in District	666		143		974	
Secondary Majority AI/AN Districts						
# High Quality (0-12%)	16	54	142	181	224	193
# Moderate Quality (12% - 40%)	226	131	296	190	220	198
# Low Quality (40%-100%)	136	238	5	60	0	47
# Poor Quality (>100%)	13	21	1	13	0	6
# of School Districts with False Zeros	53		0		0	
# of AI/AN Students in False Zero Areas	1,191					
Mean Percentage of Group in District	17.16%		60.46%			
Average Student Group Total in District	187		659		1,090	

Tables 5 and 6 concentrate on the districts where the ACS-ED mistakenly assigns a zero. Table 5 lists the ten false zero districts where AIs/ANs make up the highest percentage of the total district population. While Table 6 lists those with districts with the largest number of AI/AN students. Regardless of which approach is used to identify these ten districts more than 1,000 AI/AN students are complete and the average number of AIs/ANs missed in each district is more than 100.

Table 5. School Districts with False Zeros: 10 Worst Errors by Percentage of AI/AN Population (2015 – 2019)

School District	State	AIANNH Homeland Area	AI/AN Secondary Majority	AI/AN Total	District Total	Percent AI/AN
LaFayette Central School District	NY	Onondaga Nation Reservation	Yes	283	868	32.6
Ruidoso Municipal Schools	NM	Mescalero Reservation	No	316	2,026	15.6
Fountain Hills Unified District	AZ	No	Yes	199	1,429	13.9
Gregory School District 26-4	SD	Rosebud Off-Reservation Trust Land	Yes	51	373	13.7
Big Sandy K-12 Schools	MT	Turtle Mountain Off-Reservation Trust Land	Yes	17	153	11
Burke School District 26-2	SD	Rosebud Off-Reservation Trust Land	Yes	26	258	10.2
South Koochiching School District	MN	Red Lake Reservation	Yes	29	282	10.1
Flasher Public School District 39	ND	No	Yes	21	230	8.9
Woodruff Joint No. 1 School District	WI	No	Yes	47	525	8.9
Valentine Community Schools	NE	No	No	51	597	8.5
Total				1,040	6,741	15.4

The areas identified in Table 5, are all associated with an AIANNH homeland area or are districts where AIs/ANs are the secondary majority population. In addition, the four districts that are not in an AIANNH area are located relatively close to one. For instance, Fountain Hills Unified, Woodruff Joint No. 1, and Valentine Community school districts are all approximately 10 miles or less from an AIANNH homeland area.[2] This is not always the case for districts in Table 6. While Balsz Elementary and Greenwood school districts are approximately ten miles or less from Salt River Reservation and the Choctaw OTSA respectively, the other districts not in or near an AIANNH homeland area.

Table 6. School Districts with False Zeros: 10 Worst Errors by Total AI/AN Population (2015 – 2019)

School District	State	AIANNH Homeland Area	AIAN Secondary Majority	AI/AN Total	District Total	Percent AI/AN
Ruidoso Municipal Schools	NM	Mescalero Reservation	No	316	2,026	15.6
LaFayette Central School District	NY	Onondaga Nation Reservation	Yes	283	868	32.6
Meridian Joint School District 2	ID	No	No	208	39,179	0.5
Fountain Hills Unified District	AZ	No	Yes	199	1,429	13.9
Balsz Elementary District	AZ	No	No	124	2,378	5.2
Southampton Union Free School District	NY	Shinnecock Reservation	No	120	1,614	7.4
Archuleta County School District 50-JT	CO	Southern Ute Reservation	No	109	1,610	6.8
Greenwood School District	AR	No	No	99	3,805	2.6
Dallas School District 2	OR	No	No	91	3,170	2.9
Texas City Independent School District	TX	No	No	90	8,257	1.1
Total				1,639	64,336	2.5

5. Discussion and Conclusion

Until now, the only measure concerning the reliability of AI/AN estimates from the American Community Survey has been produced by the Census Bureau and there has been no research

[2] Salt River Reservation is less than 10 miles east of Fountain Hills, AZ. The Lac du Flambeau Reservation is less than 10 miles west of Woodruff, WI, and the Rosebud Indian Reservation is approximately 10 miles north of Valentine, NE.

examining how the uncertainty concerning its estimates vary. Similarly, the accuracy of this data has not been examined. The introduction of the ACS-ED in 2015 has created an opportunity to address the later point by comparing estimates from the ACS-ED to totals produced in the CCD for non-Hispanic, single-race AIs/ANs because it accounts for the lack of alignment between census tracts and school district boundaries.

The results of conducting this analysis raise a number of issues concerning the accuracy of AI/AN estimates in the ACS-ED and by extension ACS data in general. First, and foremost the mean absolute percentage error concerning the estimated number of AI/AN students in each district is 177% for the period 2015 - 2019. By comparison, this value is 54% for White students and 19% when estimated the total number of students in the district. It appears that much of this error comes from incorrectly estimating there are no students in a school district when they do in fact live there. However, in many of these cases the total number of AI/AN students in the district is actually quite small. For example, out of the 6,299 districts with false zeros 97.3% of them are in areas outside AIANNH homeland areas with an average of 16 AI/AN students.

The ACS-ED also appears to have issues with estimates in AIANNH homeland areas. The mean absolute percentage error in estimated student totals for AIs/ANs, Whites and the school district overall is 78%, 44% and 21% respectively. The area where the ACS-ED does best in estimating the total number of AI/AN students is where they are the majority population. In these areas, the MAPE is 50% and is should be viewed as low in quality. By comparison, the same values for White and district totals are 148% and 43%, respectively.

A particular problematic issue concerning ACS-ED estimates of AI/AN student totals is the frequency with which it falsely indicates that no AIs/ANs reside in an area where we should expect higher percentages or totals. Consequently, 17 different school districts have been identified as areas where ACS-ED estimates are extremely poor and the Census Bureau should investigate why its

estimates are so inaccurate and AI/AN responses were not included as part of the sample. Possible explanations for why this may have occurred include randomness; not drawing a representative sample in the area; and a conscious decision to not participate by AIs/ANs in the area. Regardless of which of these is the actual cause, there should be some indication that estimates in these areas and potentially others are known to be inaccurate to ensure that researchers exclude it from their analysis. In addition, there should be specific efforts to reach out to representatives of Indigenous nations in these areas to determine the best way to proceed in the 14 districts[3] located on or near an AIANNH homeland area.

References

Bazuin, J. T., & Fraser, J. C. (2013). How the ACS gets it wrong: The story of the American Community Survey and a small, inner city neighborhood. *Applied Geography*, *45*, 292-302.

Brayboy, B. M. J., & Tachine, A. R. (2021). Myths, Erasure, and Violence: The Immoral Triad of the Morrill Act. *Native American and Indigenous Studies*, *8*(1), 139-144.

Citro, C. F., & Kalton, G. (2007). Using the American Community Survey: Benefits and challenges. Washington, DC: National Academies Press. http://ndl.ethernet.edu.et/bitstream/123456789/29003/1/59.pdf.pdf. Retrieved December 19, 2021.

Environmental Systems Research Institute (ESRI). (2018). The American Community Survey: An ESRI white paper. https://www.esri.com/content/dam/esrisites/en-us/media/whitepaper/WP2012-j10020-american-community-survey-2018-rev.pdf. Retrieved December 19, 2021.

Folch, D. C., Arribas-Bel, D., Koschinsky, J., & Spielman, S. E. (2016). Spatial variation in the quality of American Community Survey estimates. *Demography*, *53*(5), 1535-1554.

Fuller, B., Bathia, S., Bridges, M., Kim, Y., Galindo, C., & Lagos, F. (2022). Variation in the Local Segregation of Latino Children—Role of Place, Poverty, and Culture. *American Journal of Education*, *128*(2), 000-000.

[3] These 14 districts are: Archuleta County School District 50-JT, Balsz Elementary District, Big Sandy K-12 Schools, Burke School District 26-2, Flasher Public School District 39, Fountain Hills Unified District, Greenwood School District, Gregory School District 26-4, LaFayette Central School District, Ruidoso Municipal Schools, South Koochiching School District, Southampton Union Free School District, Valentine Community Schools, and Woodruff Joint No. 1 School District.

Geverdt, D. E. (2015). Education Demographic and Geographic Estimates Program (EDGE): Locale Boundaries User's Manual. NCES 2016-012. *National Center for Education Statistics*.

Jurjevich, J. R. (2019). Confronting statistical uncertainty in rural America: toward more certain data-driven policymaking using American Community Survey (ACS) data. In *Population, Place, and Spatial Interaction* (pp. 115-133). Springer, Singapore.

Jurjevich, J. R., Griffin, A. L., Spielman, S. E., Folch, D. C., Merrick, M., & Nagle, N. N. (2018). Navigating statistical uncertainty: How urban and regional planners understand and work with American community survey (ACS) data for guiding policy. *Journal of the American Planning Association*, 84(2), 112-126.

National Congress of American Indian, Policy Research Center. *Data Disaggregation | NCAI*. Data Disaggregation | NCAI. https://www.ncai.org/policy-research-center/research-data/data. Retrieved December 19, 2021

Reardon, S. F., Kalogrides, D., & Shores, K. (2019). The geography of racial/ethnic test score gaps. *American Journal of Sociology*, *124*(4), 1164-1221.

Shotton, H., Lowe, S. C., & Waterman, S. J. (2012). Beyond the Asterisk: Understanding Native Students in Higher Education. In *Stylus Publishing, LLC*. Stylus Publishing, LLC.

Spielman, S. E., Folch, D., & Nagle, N. (2014). Patterns and causes of uncertainty in the American Community Survey. *Applied Geography*, *46*, 147–157. https://doi.org/10.1016/j.apgeog.2013.11.002

Swanson, D. A., & Hough, G. C. (2012). An evaluation of persons per household (PPH) estimates generated by the American Community Survey: A demographic perspective. *Population Research and Policy Review*, *31*(2), 235-266.

Xu, D., Solanki, S., & Fink, J. (2021). College acceleration for all? Mapping racial gaps in Advanced Placement and dual enrollment participation. *American Educational Research Journal*, 0002831221991138.