MAKE IT COUNT:

Native Hawaiian Population Estimates in Census 2000 and Implications for Other Small Racial Groups

Prepared by
Nolan J. Malone
Matthew Corry

May 2004

PASE REPORT 03-04: 30

AT A GLANCE

Despite the breakthrough of multiple-race responses in Census 2000, population estimates for Native Hawaiians and other small racial groups remain particularly tenuous for in-depth analyses. The PUMS 1-percent files, for example, account for only three-fourths of the actual Native Hawaiian population. The purpose of this report is to (1) outline the limitations of Census 2000 figures, (2) describe a proxy model that yields more accurate population estimates of the Native Hawaiian population, and (3) show that Census microdata products need to include detailed analyses for racial groups other than broad race categories (e.g., White and African American). To include multiple-race reporting without sufficient access to those data ultimately yields a similar result as not counting multiple-race responses in the first place.

CONTENT LINKS

PASE *Policy Analysis & System Evaluation*

KAMEHAMEHA SCHOOLS

Make It Count

Native Hawaiian Population Estimates in Census 2000 and Implications for Other Small Racial Groups

Nolan J. Malone
Matthew Corry

May 2004

Recommended Citation:
Malone, N.J., & Corry, M. (2004, May). *Make It Count: Native Hawaiian Population Estimates in Census 2000 and Implications for Other Small Racial Groups*. Honolulu, HI: Kamehameha Schools, PASE Report 03-04:30.

Inquiries should be addressed to:
Policy Analysis & System Evaluation (PASE)
Kamehameha Schools
567 So. King Street, Suite 400
Honolulu, Hawai'i 96813

Make It Count:

Native Hawaiian Population Estimates in Census 2000 and Implications for Other Small Racial Groups

EXECUTIVE SUMMARY

The year 2000 was a breakthrough for the U.S. Census Bureau, as the decennial census included for the first time the option for respondents to mark multiple races. The boon, however, has been somewhat blunted by inadequacies in public-use data files that would otherwise permit accurate population estimates for smaller racial groups. This report explains the inherent shortcomings of microdata products and discusses implications for Native Hawaiians and other small racial groups. Highlights of the report include the following.

- The reported Native Hawaiian population almost doubled from 1990 to 2000.
- Compared to other groups, Native Hawaiians are extremely diverse, with nearly two out of three reporting multiple races.
- Presently available Census data products do not permit in-depth analyses for Native Hawaiian population statistics.
- Based on the Census 2000 1-percent public use microdata sample (PUMS) files, about one-fourth of the Native Hawaiian population is "missing."
- A proxy method that uses PUMS 1-percent race data, along with ancestry variables, results in an estimate that is within 2 percentage points of the actual count for Native Hawaiians.
- The same proxy method is not necessarily useful for other smaller population groups such as Alaska Natives.
- More can be done to ensure that the benefits of multiple-race responses in Census 2000 be reflected in the actual reporting of those data.

INTRODUCTION

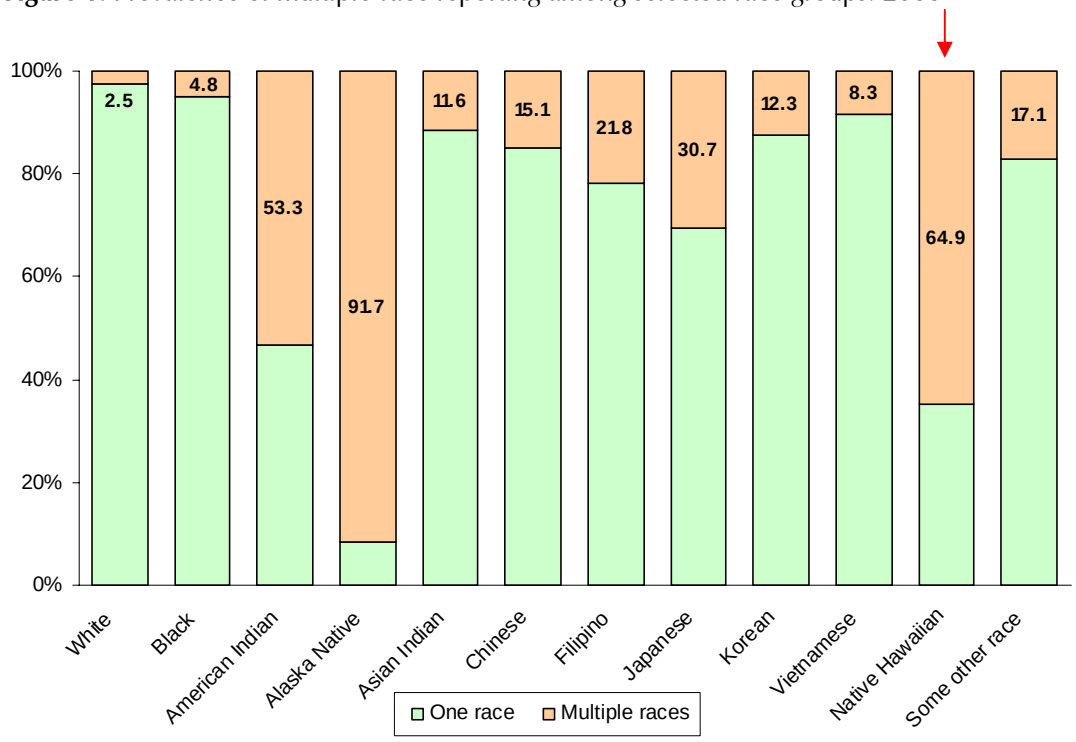
In 2000, the U.S. Census Bureau made great strides in accommodating multiple-race responses in the decennial census. This much-needed change acknowledged the nation's increasingly diverse population, providing population estimates for specific race groups that are believed to be the most accurate in the history of the census. However, this change did not solve inadequacies in reporting population estimates, especially those of smaller racial groups, owing to the inability of researchers to identify the certain race populations using Census Bureau public-use data sources. The purpose of this report is to (1) outline the limitations of Census 2000 figures, (2) describe a proxy model that yields more accurate population estimates of the Native Hawaiian population, and (3) show that Census microdata products need to include detailed analyses for racial groups other than broad race categories (e.g., White and African American).

THE BOON

Borne of Office of Management and Budget (OMB) Statistical Directive Number 15, the change in the Census 2000 questionnaire provided the American public with new race categories—Native Hawaiian and Other Pacific Islanders (NHOPI) among them—and permitted respondents to mark more than a single race group. For the first time in census history, individuals of multiple races were not required to choose only one race designation or opt for the "some other race" category.

The effort resulted in a noticeable increase in the Native Hawaiian population reported in the United States, rising from only 211,014 in 1990 to 401,162 in 2000. Furthermore, the Census 2000 data showed that roughly two-thirds of all Native Hawaiians are of multiple races, second only to Alaska Natives among the major race categories. The detail afforded by the Census 2000 questionnaire revealed the Native Hawaiian population to be one of the most ethnically diverse groups in the country, if not the world.

Figure 1. Prevalence of multiple-race reporting among selected race groups: 2000



The racial detail afforded by the newly designed census has been incorporated into several Census 2000 data products. Instead of a single race total, the new data products report statistics for both those who reported that race "alone" (i.e., the minimum) as well as for all those who reported that race whether "alone or in combination" with other races (i.e., the maximum). These efforts on behalf of the Census Bureau satisfy a wide array of data users' needs and make good use of the rich racial detail now available.

THE BUST

Despite the numerous Census briefs, data tables, and special reports that identify unique populations according to certain variables of interest, researchers inevitably encounter limitations when seeking to perform more in-depth analyses of specific race groups, such as Native Hawaiians. For example, the Census 2000 public use microdata sample (PUMS) files serve as the primary data source for those seeking to perform their own detailed analyses. Regrettably, in spite of the greater detail captured within the Census 2000 data, confidentiality protocols governing the release of data restrict the translation

of that rich detail to the public files, resulting in incomplete race codes that fail to fully capture certain populations. In essence, the boon of greater racial detail results in a bust for microdata researchers.

More specifically, the PUMS files limit the identification of all members of any specific race group (with the exception of White and African American). This limitation is even more pronounced for those race groups that are more likely to be of multiple races. Both the 1-percent and 5-percent PUMS files contain three categorical race variables[1] and six dichotomous race variables.[2] Identifying all White or African-American individuals is possible using the WHITE and BLACK dichotomous (recoded) variables, respectively. For other specific race groups, however, identification of corresponding individuals is limited to the three categorical variables mentioned above.

Because only the RACE3 categorical variable provides any detail regarding multiple-race combinations, it serves as the sole source for analyses of Native Hawaiians. However, the RACE3 coding scheme includes an incomplete list of race combinations that fails to fully capture the complete Native Hawaiian population. Because of Census Bureau confidentiality thresholds, a large number of race groupings are not assigned detailed codes but are instead collapsed into a residual category: "all other race combinations." This condition is even more pronounced for the 5-percent PUMS file. As an example, Table 1 presents a partial list of values for the RACE3 variable, highlighting those codes that directly identify Native Hawaiians in both the 1-percent and 5-percent PUMS files. As shown, the 5-percent file lacks four of the ten codes available in the 1-percent file that permit the identification of Native Hawaiians. Although the 5-percent file has traditionally served as a source for greater geographic detail and reliable race analyses, the finely detailed race categories reported in Census 2000 prove too threatening to respondent confidentiality and are, therefore, suppressed. Consequently, researchers of the Native Hawaiian population are resigned to using the 1-percent PUMS file for more accurate analyses.

[1] Categorical race variables include: RACE1, which codes individuals with eight broad single-race categories and one multiple-race category; RACE2, which offers 64 detailed single-race categories and one multiple-race coding option; and RACE3, which provides 71 detailed race and race-combination coding options.
[2] The six dichotomous race variables—WHITE, BLACK, AIAN, ASIAN, NHOPI and OTHER—can be used to identify anyone who reported a race within that group, whether alone or in combination with some other race.

Table 1. Limitations of PUMS 5-percent data for detailed race values (RACE3)

PUMS 1-percent	Code	PUMS 5-percent
Native Hawaiian alone	06	Native Hawaiian alone
Japanese; Native Hawaiian	14	Not used
Filipino; Native Hawaiian	18	Filipino; Native Hawaiian
Chinese; Native Hawaiian	22	Chinese; Native Hawaiian
Chinese; Filipino; Native Hawaiian	27	Not used
White; Native Hawaiian	52	White; Native Hawaiian
White; Japanese; Native Hawaiian	58	Not used
White; Filipino; Native Hawaiian	61	White; Filipino; Native Hawaiian
White; Chinese; Native Hawaiian	63	White; Chinese; Native Hawaiian
White; Chinese; Filipino; Native Hawaiian	64	Not used
All other race combinations	71	All other race combinations

Source: U.S. Census Bureau 2000 Census of Population and Housing, Public Use Microdata Sample, United States: Technical Documentation, 2003.

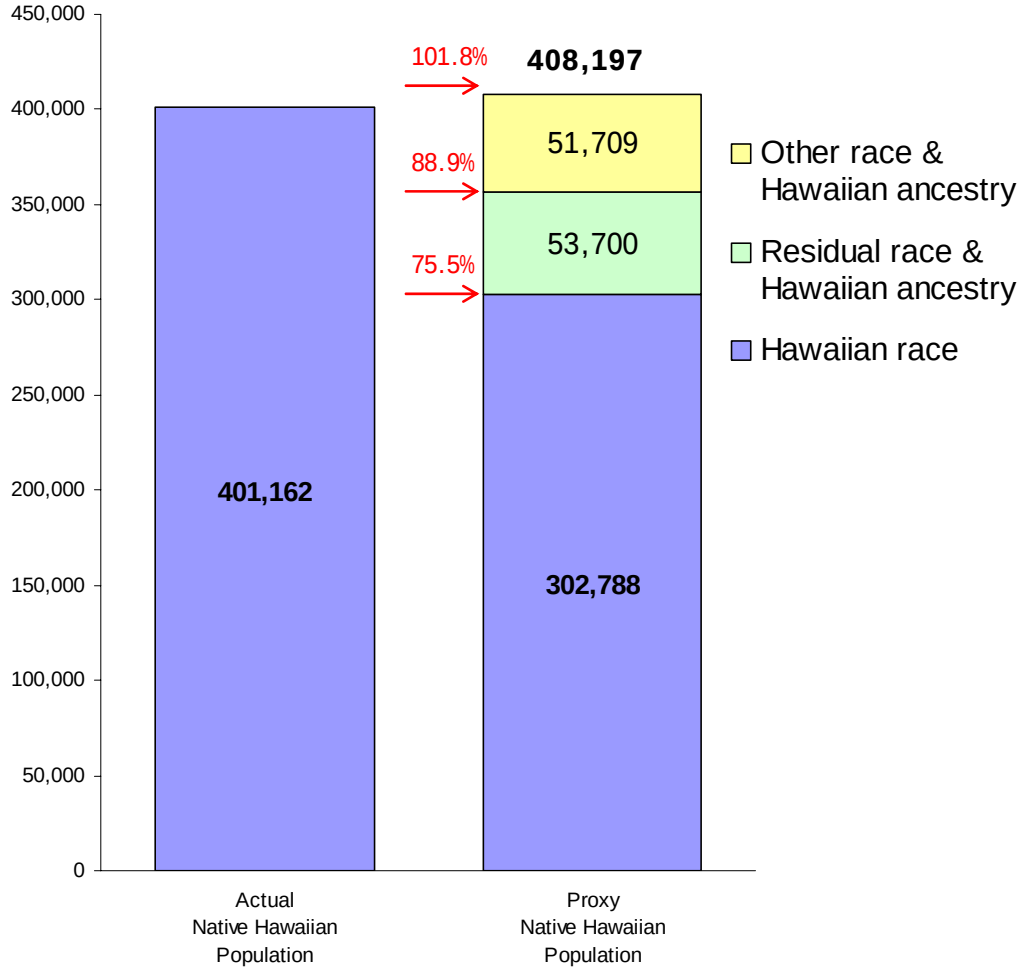
MAKESHIFT REMEDIES

Although the Census 2000 1-percent PUMS file identifies a relatively larger portion of the actual Native Hawaiian population, it only accounts for three out of four actual Hawaiians in the United States. The "missing" Hawaiians are likely those who reported less common race combinations and therefore did not meet the Census Bureau's reporting threshold, which resulted in their assignment to the residual race category, "all other race combinations." The reliance on a fraction of the actual population poses a serious threat to the representative nature of the PUMS data for this population group, casting all subsequent analyses into doubt.

To circumvent the race-variable limitations within the PUMS files, however, we have experimented with a rough *proxy* that relies on the PUMS file ancestry variables (ANCFRST1 and ANCSCND1). Specifically, we use reported Hawaiian (or part-Hawaiian) ancestry as a proxy for Hawaiian race, assuming that many multiple-race Hawaiians who reported rare race combinations—combinations that ultimately failed to reach the reporting threshold—were coded in the residual race category. Therefore, if one adds

those Native Hawaiians from the residual race category who recorded a Hawaiian ancestry on their census questionnaires, the total *proxy* Hawaiian sample rises to 88.9 percent of the actual count. The result is closer to the full count enumerated in Census 2000, but still falls short. In the case of Native Hawaiians, a possible solution is to add *all* PUMS respondents who reported Hawaiian ancestry—regardless of inclusion in the residual race category—to the Native Hawaiian proxy group. If we add individuals listed under other non-Hawaiian and non-residual race categories, but who cite Hawaiian ancestry, the final proxy Native Hawaiian sample comes within two percentage points of the true enumerated Native Hawaiian population.

Figure 2. Using a proxy method to estimate the "missing" 25% of the Native Hawaiian population



THE LIMITS OF PROXY

This data quandary is especially pronounced for race groups that include sizable multiple-race constituents. Consider, for example, Alaska Natives, for whom nine out of ten members are of mixed races. Using the RACE1 variable in the PUMS 1-percent file, one can determine an estimate of the total Alaska Native *alone* population: 97,589, which is 99.7 percent of the actual single-race count, but only 8.3 percent of the total Alaska Native enumerated population. Unfortunately, the RACE3 variable offers no specific race combinations for Alaska Natives owing to their small numbers. In fact, among the RACE3 code options, Alaska Natives and American Indians are combined as a single entity, thereby denying any opportunities to compile estimates of the full Alaska Native population. The ultimate discouragement is that ancestry variables offer little assistance in the formation of an Alaska Native proxy population: only 16,462 weighted PUMS cases identify Eskimo, Aleut or Inuit as ancestry responses exclusive of the "Alaska Native alone" race response, bringing the estimate of the entire Alaska Native proxy population to 114,051, only 9.7 percent of the actual total. For Alaska Natives, the proxy method is quite inadequate.

Table 2 illustrates the results of applying the proxy method to other race groups, resulting in varying degrees of success. The questionable results can be attributed to the basic assumption necessary for the method to succeed: specifically, respondents who have already identified their detailed race must repeat that response under the ancestry question to ensure that their race is acknowledged. However, for many respondents, the ancestry question (which is asked after race, marital status, enrollment status and educational attainment, respectively) serves to supplement their race responses, not to restate them. That is, for individuals who, for example, check race boxes for White, African-American and Filipino, the open-ended ancestry fields may be used to denote German and Jamaican heritages. The fact that only two ancestry responses are recorded, regardless of the number written, is another hurdle to using ancestry as a proxy for race in the PUMS files. Therefore, while seemingly reliable in the identification of Native Hawaiians, the nature of the ancestry question and the assumptions necessary for its implementation may produce highly unstable estimates for other race groups.

Table 2. The variableness of PUMS 1-percent proxy race populations, by selected race groups

	PUMS 1-percent estimates (proxy)				Census 2000 Actual
			Total		
	Race	Ancestry[1]	Number	% of actual	
NHOPI groups					
Native Hawaiian	302,788	105,409	408,197	101.8	401,162
Samoan	104,071	21,993	126,064	94.6	133,281
Tongan	44,542	12,281	56,823	154.2	36,840
Guamanian/Chamorro	63,741	18,898	82,639	89.2	92,611
Fijian[2]	0	10,521	10,521	77.5	13,581
AIAN groups					
American Indian	1,852,290	6,523,193	8,375,483	209.5	3,997,917
Alaska Native	97,589	16,462	114,051	9.7	1,179,517
Asian groups					
Asian Indian	1,779,005	177,654	1,956,659	103.0	1,899,599
Chinese	2,687,550	212,440	2,899,990	101.2	2,865,232
Filipino	2,309,400	193,179	2,502,579	105.8	2,364,815
Japanese	1,036,903	188,920	1,225,823	106.7	1,148,932
Korean	1,169,393	108,881	1,278,274	104.1	1,228,427

[1] Ancestry *exclusive* of identical race response.

[2] Fijians are not identifiable from any of the PUMS 1-percent race variables.

MAKE IT COUNT

While population estimates are, in fact, just estimates, figures for racial groups such as Native Hawaiians are particularly tenuous, even for experienced researchers. More can be done to ensure that the benefits of multiple-race responses in Census 2000 be reflected in the actual reporting of those data. Like Census briefs, data tables, and special reports, Census microdata products should be developed with the awareness that detailed analyses are needed for racial groups other than broad race categories (e.g., White and African American), especially those who are more likely to be multiracial, such as Native Hawaiians, American Indians and Alaska Natives. A simple dichotomous variable—such as those for the broader race categories—that denotes any Native Hawaiian response would suffice. To include multiple-race reporting without sufficient access to that detail ultimately yields a similar result as not counting multiple-race responses in the first place.