HISPANIC SURNAME REGISTERED VOTERS: YAKIMA COUNTY
Under-identification in the 2024 General Election

Vickie Ybarra, PhD, MPH

Voting Rights in Yakima County

Yakima County, in central Washington state, is home to a large and growing Hispanic/Latino population, and since 1975 has been the subject of numerous federal and state Voting Rights actions. In 2004 the County entered into a federal consent decree with the U.S. Department of Justice requiring use of bilingual English/Spanish ballots and supports for Spanish-speaking voters. Although required federal reporting for the consent decree ended in 2006, the County Elections Office continues to regularly report Spanish-surname registration and voting summary data on its public website.[1]

The Yakima County Elections office relies on the Washington Office of the Secretary of State to identify Spanish surnames in the County voter file, using a reference surname list provided by the Yakima County Elections office. This list of surnames was provided to Yakima County by the U.S. Department of Justice with the 2004 consent decree.

Identifying Hispanic Surname[2] Registered Voters

Attempts by the Census Bureau to identify persons of Hispanic/Latino origin using surnames date back to the 1950s. In 1996 the Census Bureau published a 'short list' of the 639 most frequently occurring Hispanic surnames nationwide based on the 1990 Census.[3] Following publication, many jurisdictions and researchers began making use of that list to identify potential Hispanic/Latino populations in a variety of fields where self-identified race/ethnicity was not available. The DOJ Spanish-surname used by Yakima County is a list of 12,497 surnames, and appears to be derived from the 2000 Census.

In 2016, the Census Bureau published an updated comprehensive list of surnames appearing more than 100 times in the 2010 US Census, and their associated race/ethnicity frequency.[4] Some jurisdictions now make use of this updated nationwide reference list, denoting those 7,362 surnames with 85% or more of respondents self-identified as Hispanic/Latino as 'Hispanic surnames'.

More recently, scholars have developed and made public the statistical coding for a Bayesian Improved Surname Geocoding (BISG) method. This newer method updates the probability of surname race/ethnicity from the public Census surname list by incorporating public Census data on racial/ethnic makeup of the local community. This method addresses one long-observed concern with use of a surnames alone, that the accuracy of a surname list varies across geographies. Since spatial racial/ethnic segregation and concentration across states and communities impacts the utility of surname lists, incorporating that variation into a Bayesian probability calculation provides greater precision, especially at lower levels of geography. The method was originally published related to healthcare records to help healthcare systems identify and address racial/ethnic health disparities, and is now recommended by some voting rights scholars.[5]

Whatever method is chosen, scholars recognize that surname estimates are appropriate only when applied to population estimates, and cannot necessarily accurately identify the race/ethnicity of every individual in a population. There is always some error in any estimate, the goal is to find the estimation method that provides the most accurate population-level information for the task at hand.

Under-identification in 2024

Following the 2016 general election, the Yakima County Elections office publications reported 30,522 Spanish-surname registered voters countywide, or 26.7% of the county's total registered voters. The number reported increased to 37,978 Spanish-surname registered voters, or 29.7% of the county's total in the 2020 election. Then in the 2024 general election the reported number of Spanish-surname registered voters fell to 31,424, or 23.6% of the county's total.

Following the 2016, 2020, and 2024 general elections, I obtained from the Yakima County Elections office a complete voter file with their office's Spanish-surname designation. I compared methods of identifying Hispanic/Latino voters to try to find an explanation for the large unexpected drop in Hispanic/Latino registered voters in 2024. Table 1 below details the results of my analysis.[6]

Table 1. Hispanic Surname Registered Voter Undercount
Yakima County, 2016-2024

	2016	2020	2024
Total Yakima County Reg Voters	114,508	127,691	133,138
Hispanic Surname Reg Voters			
Yakima County published-N	30,522	37,978	31,424
Yakima County published-%	26.7%	29.7%	23.6%
1990 Census List	23.0%	27.5%	30.4%
2010 Census List	24.8%	28.9%	32.6%
BISG Method-N	30,001	38,946	46,031
BISG Method-%	26.2%	30.5%	34.6%
Estimated Over/Undercount-N	+521	-968	-14,618
Estimated Over/Undercount-%	+1.7%	-2.5%	-46.5%
Non-Hispanic Surname Reg Voters			
Yakima County published-N	83,986	89,713	101,714
Yakima County published-%	73.3%	70.3%	76.4%
1990 Census List	77.0%	72.5%	69.6%
2010 Census List	75.2%	71.1%	67.4%
BISG Method-N	84,490	88,745	87,096
BISG Method-%	73.8%	69.5%	65.4%
Estimated Over/Undercount-N	-504	+968	+14,618
Estimated Over/Undercount-%	-0.6%	+1.1%	+14.4%

Comparing my estimates using the BISG methodology to official Yakima County published counts, I find that the under-identification of Hispanic surname registered voters in Yakima County was less than 4% in the 2020 presidential general election. The under-identification then rose to an alarming 46.5% with the 2024 general election. Under-identification of Hispanic surname voters is necessarily associated with an over-identification of non-Hispanic surname voters – in the 2024 election I estimate this over-identification at 14.4%. It appears that in 2024 Yakima County reported over 14,000 Hispanic-surname voters as non-Hispanic surname, making it appear as though the number and percent of Hispanic/Latino registered voters had decreased between 2020 and 2024, when in fact they had increased substantially.

Examining more closely the undercounted records in 2024 (those identified as Hispanic surname by the BISG method but not identified as Spanish surname in the Yakima county file), I find that 30% have compound surnames, with spaces or hyphens. Of the remaining undercounted records, 88% have surnames that appear on the 2010 Census list, and 80% have surnames on the 1990 Census list.

As detailed in Table 1, the Yakima County published aggregate numbers are very close to the 2010 Census list in 2016 and 2020 – within 1.9 percentage points in 2016 and 0.8 points in 2020. Although they did not identify exactly the same individuals, the aggregate counts are very close. In 2016 and 2020 these aggregate numbers are also very close to what I find with the BISG method – within 0.4 points in 2016 and 1.0 point in 2020. It is only in 2024 that the results of the three methods diverge, with the Yakima County count performing much lower than either the 2010 Census surname list or the BISG method, suggesting that perhaps something went awry in the technical processing of the 2024 file.

This analysis also suggests that use of the 2010 Census Bureau reference list may be adequate to reasonably estimate the aggregate Hispanic surname population in Yakima County for purposes of routine public reporting. While there is additional precision gained by using the BISG method at smaller levels of geography, it places an additional burden for statistical processing and geocoding that may not be readily available to a public agency.

Citizen Voting Age Population

Since 2005, the Census Bureau has produced estimates of the Citizen Voting Age Population (CVAP) by race/ethnicity for multiple levels of geography. These estimates are drawn from the most recent American Community Survey 5-year population estimates, and serve to estimate the population potentially eligible to vote by race/ethnicity.[7]

Any identification of Hispanic surnames in the voter file, whether from the 1990 list, the 2010 list, or the BISG methodology, is only useful to the extent it permits jurisdictions to reasonably accurately estimate the actual number of Hispanic/Latino registered voters. Even in the absence of sophisticated statistical tests, we can examine the relationship between the registered voter count and the Census Bureau CVAP estimates as a minimal reasonableness check. When we do this for Yakima County, we see that the counts of Hispanic and non-Hispanic surname registered voters published by the County Elections office for 2016 and 2020 appear as though they could be reasonable, while the 2024 numbers do not.

Table 2. Hispanic/Latino & Non-Hispanic/Latino
Citizen Voting Age Population (CVAP)
Yakima County, 2016-2024

	2016	2020	2024
Hispanic/Latino			
Yakima County published Hispanic SN registered	30,522	37,978	31,424
Citizen Voting Age Population	40,150	48,250	53,713*
% of CVAP	76.0%	78.7%	58.5%
Non-Hispanic/Latino			
Yakima County published non-Hispanic SN registered	83,986	89,713	101,714
Citizen Voting Age Population	101,950	99,005	97,750*
% of CVAP	82.4%	90.6%	104.1%

Using the Yakima County published counts of Hispanic and non-Hispanic surname registered voters, we see that while the percent of Hispanic/Latino citizen adults registered ranges from what seems a reasonable 76%-79% in 2016-2020, it drops to about 58% in 2024. Yakima County published counts of non-Hispanic surname registered voters in 2024, at 101,714, represents 104% of the estimated 97,750 non-Hispanic/Latino citizen voting age population in 2024 in the entire county. This comparison provides further evidence that the 2024 Yakima County published counts are inaccurate.

Recommendations

After contacting both the Yakima County Elections office and the Secretary of State's office, I do *not* believe the 2024 under-identification was intentional. Rather, the undercount may simply represent lack of updated code, or a technical problem in processing the 2024 file. My hope is that this analysis points to the importance of updating methodology, and provides opportunity to incorporate best practices.

The science of racial/ethnic identification of surnames, in the absence of self-identified race/ethnicity, has advanced substantially in the past decade. Currently best practice points to the use of the most recent Census surname list, and applying whatever list is chosen to all parts of compound surnames. The BISG methodology provides the greatest precision at smaller levels of geography, but it may not be practical or necessary for routine public aggregate reporting at the county level.

To more accurately estimate Hispanic surname voters, I recommend that the Office of the Washington Secretary of State apply the most recent surname lists from the Census Bureau. While it may not be necessary to implement the BISG method for routine public reporting, use of the 2010 Census surname list, and future updated Census lists, will provide much better estimates than were produced in 2024.

Finally, I suggest that the Washington Office of the Secretary of State's and Yakima County Elections office work together to make the methods used to identify Hispanic surname voters more transparent to the public, so that affected communities may gain confidence in the published counts.

Endnotes

[1] See Yakima County Elections Office, Voting by Surname.

[2] The original Census Bureau term for surnames of people identifying as Hispanic/Latino was 'Spanish surname', reflecting the Spanish-language origins of the names of Americans with Latin American heritage. More recently the Census Bureau has adopted the term 'Hispanic surname' to reflect surnames associated with Census respondents who self-identify as Hispanic/Latino, so I adopt that usage here. Except when referring to Yakima County publications, where they still use the term 'Spanish surname'.

[3] Word, D.L. and Perkins, R.C., 1996. *Building a Spanish Surname List for the 1990's--: A New Approach to an Old Problem* (Vol. 13). Washington, DC: Population Division, US Bureau of the Census.

[4] Comenetz, J., 2016. *Frequently occurring surnames in the 2010 census*. United States Census Bureau.

[5] See Elliott, M.N., Morrison, P.A., Fremont, A., McCaffrey, D.F., Pantoja, P. and Lurie, N., 2009. Using the Census Bureau's surname list to improve estimates of race/ethnicity and associated disparities. *Health Services and Outcomes Research Methodology*, *9*, pp.69-83; and Barreto, M., Cohen, M., Collingwood, L., Dunn, C.W. and Waknin, S., 2022. A novel method for showing racially polarized voting: Bayesian improved surname geocoding. *NYU Rev. L. & Soc. Change*, *46*, p.1.

[6] For the 1990 Census I use the list of 639 surnames published by the Census Bureau in 1996. For the 2010 Census I use a list of 7,362 surnames published by the Census Bureau in 2016, applying the conventional threshold of 85% Hispanic/Latino self-identification. For the BISG Method I make use of methodology and Stata code published by the Consumer Finance Protection Bureau, applying the 85% threshold. For all three methods I applied surname lists to all versions of surnames – the original surname and all parts of compound surnames separately. All statistical coding was done in Stata/IC 15.1.

[7] Census Bureau Citizen Voting Age Population (CVAP) estimates: https://www.census.gov/programs-surveys/decennial-census/about/voting-rights/cvap.html

*Version 2 of this paper updates Table 1 2020 BISG estimated totals slightly based on re-analysis of 2020 data, with no substantive change in findings. Version 2 also updates the CVAP estimates in Table 2 to reflect 2023 Census data made available following original publication. The latest Census Bureau CVAP estimates available publicly are for 2023. To produce the 2024 estimates, I extend the published 2023 estimates by the last annual change trend one year for both populations (+1.9% Hispanic/Latino and -0.7% for non-Hispanic/Latino CVAP).