UNM-GPS Population Estimates Methodology
Estimates for 2020-2023

Jacqueline Miller, PhD
UNM Geospatial and Population Studies
Summer 2024

Key Changes to 2023 Estimates (released in 2024) and important considerations

- Race and ethnicity coding updated to reflect DOH's revised guidelines (see "*Two or More Races ('TOMR')*" section*)*
- Estimates of NM's Hispanic population continue to show a discontinuity in the years 2018 and 2019. This discontinuity originates in the data we use from the Census Bureau to distribute characteristics across the state and counties. It is unlikely that a true sudden shift in the Hispanic population occurred in these years. We expect this anomaly to be correct after the Census Bureau releases 2010-2019 intercensal estimates that include Hispanic origin.

Background

States, local agencies, and both public and private sector organizations need to know how many people live in their communities or catchment areas to be able to calculate rates of different events and plan for the population's needs. For example, during the COVID-19 pandemic, state Department of Health agencies needed to know how many people were in a community to calculated COVID-19 infection rates and vaccination rates, but they also could use population information to inform their outreach and target strategies so that they dedicated the necessary resources to each community.

Population estimates are produced for current or past populations, not future populations. This enables demographers to use data from the same period as the estimate period to make inferences about how the population is changing. This differs from population projections, which require demographers to make assumptions about how trends may change in the future. Though a census of the population that directly counts individuals within the community is considered the most reliable manner of determining the population size and its characteristics, complete censuses are extremely costly. Consequently, the United States only conducts a census every 10-years. Consequently, most states rely on population estimates as a cost-effective strategy to understand how their population is changing.

There are several methodologies for producing population estimates, each with their own strengths and weaknesses. For a thorough review of estimates methodologies in general, please see David Sawnson and Jeff Tayman's *Subnational Population Estimates*[1].

The Census Bureau produces estimates using the Cohort-Component Method (CCM) each year for the nation, states, counties, cities, and some towns[2]. These estimates are revised each year so that the entire series since the last Decennial Census is updated with new data. The equation below shows the CCM.

$$Pop_{2021} = Pop_{2020} + Births - Deaths + Net\ Migration$$

This method[3] starts with the most recent Census population count and then uses vital records to apply changes to the population that have occurred through births and deaths. They then incorporate migration, using data from the IRS, Medicare Enrollment, information on international migration, Armed Forces movement, and other sources. The census bureau first produces an estimate for the nation as a whole and then produces estimates for states and smaller geographies. Once a preliminary estimate is produced for each state, the totals are then proportionally adjusted so that the states sum to the original estimate for the whole nation. This process is repeated for each subsequent geography in order of relative size (i.e. counties, then cities, then towns).

The strengths of the Census Bureau's estimates are that they can incorporate data that are not accessible to people outside of the U.S. Government and they use a consistent methodology across all states which facilitates making comparisons between states. However, they cannot incorporate data that are not collected consistently across all states (such as school enrollment data) or adjust their estimates based on unique contextual factors and local conditions that may influence population dynamics. Furthermore, they only produce estimates at the nation, state, county, and city/town geographies. Consequently, most states produce their own population estimates that are more flexible, can incorporate local contextual knowledge and data, and are at the geographic-level necessary for state needs. The University of New Mexico's Geospatial and Population Studies (UNM-GPS) produces annual state and county population estimates for the state of New Mexico. Additionally, UNM-GPS produces annual tract-level population estimates, along with estimates of race, age, sex, and ethnicity, for the New Mexico Department of Health ('NM DOH'). This document will describe the procedures UNM-GPS uses to produce these estimates.

[1] Swanson, D.A. & Tayman, J. (2012). *Subnational Population Estimates*. New York: Springer.
[2] https://www.census.gov/programs-surveys/popest/technical-documentation/methodology.html
[3] https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2020-2022/methods-statement-v2022.pdf

State-Level Estimation Methodology

UNM-GPS produces population estimates using the Housing Unit Method and the Ratio Correlation Method for the total state population. These two methods are then averaged with Census Bureau's estimate to produce a total population estimate at the state-level. Estimates are produced for July 1 of each year. For our 2023 release, we produced estimates for July 2020, July 2021, and July 2022.

Each method is described below.

Housing Unit Method

Housing growth is used as an indicator of population growth. For this method, we begin with the number of housing units counted during the most recent census (April 2020) and add building permits issued to that count, using a multiplier of 0.95 to account for building permits that were issued but not completed. For most years, we use the prior calendar year building permit count to inform our measure of housing growth, as we assume a 6-month delay between the date of issue and the completion of the housing unit.

For 2020, we only needed to account for three months of building (April through June), thus we added a quarter of the 2019 building permits to the 2020 housing unit count. This may lead to an overcount of completed housing units as the COVID-19 pandemic made construction challenging, however, it is possible that many of these building permits were eventually completed and thus should be included in the housing stock for the decade.

Once we establish the number of housing units across the state, we calculate the number of units that are owner occupied, renter occupied, and vacant using occupancy rates from the American Community Survey (ACS) 1-year estimates. The average number of people who live in renter-occupied and owner-occupied units is also obtained from the ACS 1-year estimates. The estimated household population is then calculated using the number of occupied units and person-per-household estimates.

Currently, we assume the number of people living in group quarters remains stable over time. Consequently, we add the April 2020 Census group quarter population count to the household population estimate to produce a total housing unit method estimate for the state. Though the assumption of a stable group quarters population is likely not a valid assumption during the pandemic, we do not currently have data on how group quarters population changes. We are currently working to develop a tracking system to improve this component of our estimates.

Ratio Correlation Method

The ratio correlation method uses the ratio of people captured in a specific data source to the most recent census count of the population eligible for that data source. We adjust this method slightly by using the Census Bureau's April 2020 blended-base estimate, as this estimate adjusts for known over and undercounts in the original census data. For example, there were 368,932 New Mexican's over age 65 enrolled in Medicare in April 2020. The Census Blended-Base estimates that there were 385,343 New Mexican residents over the age of 65 on April 1, 2020. Using this information, we determine that 95.7% of New Mexico's population over the age of 65 is captured in the Medicare Enrollment data.

The ratio correlation method assumes that the population captured in a specific dataset remains consistent over time. Given that 95.7% of New Mexicans over age 65 were captured in the Medicare enrollment data, we assume that 95.7% of New Mexico's population over age 65 will be captured in the Medicare enrollment data from 2021-2029. To the extent that this assumption holds true, we can estimate the population age 65 and over in future years using the current Medicare enrollment data. Similarly, we can use MVD licensed drivers' data, birth counts, and school enrollment data to inform New Mexico's populations in other age groups.

While this method is both easy to use and provides a reasonably good estimate when the assumption of consistent data representation holds true, this assumption is not always true. Due to the COVID-19 pandemic, many families choose to homeschool their children or enroll them in private schools when private schools re-opened, leading to a large drop in the school enrollment data we obtain from the NM Public Education Department ('NM PED'). To account for this change, we obtained homeschool enrollment data from NM PED and private school enrollment data from the National Center for Education Statistics. Incorporating these two additional data sources prevented our estimates from showing a large and artificial population drop among school-aged children. Though we were fortunate to know that the PED enrollment data was no longer reliable as a sole source of school data, changes in the proportion of a population captured in a data source may change without our knowledge, thus introducing error into our estimates.

Furthermore, some data sources are more reliable than others. Though most children are enrolled in school (public, private, or home) and most people over the age of 65 are enrolled in Medicare, the proportion of people obtaining a license varies across time. Furthermore, the MVD is not always informed when a person moves, so a person moving out of state may remain on the list of licensed drivers until their license expires.

County-level Estimates

UNM-GPS produces population estimates for all New Mexico's 33 counties using the same method as described above for the state estimates, with a few notable exceptions. For counties that do not have complete building permit data, we may drop the housing unit method if the data obtained from the ACS suggests extreme changes that are inconsistent with other methods[4]. In a typical year, we use building permit data that is either publicly available or that is obtained by either UNM's Geospatial and Population Studies or UNM's Bureau of Business and Economic Research (UNM BBER). Many municipalities do not report or submit building permit data. In these instances, we either rely on data collected by the Census Bureau[5] or use estimates from the ACS. For the 2024 release, we relied solely on the Census Bureau Building Permit Survey and ACS as we were unable to sufficiently update our building permit database in time for the estimates production.

The estimates for each county are carefully reviewed. Counties showing large growth or declines are more carefully investigated, to ensure that data inputs are free from error and that trends from the individual sources (i.e. vital records, school enrollment, licensed drivers, and Medicare enrollment) support the existing estimates. In cases where our estimate is both lower than the Census Bureau's and difficult to justify with available data, we will replace our estimate with the Census Bureau's value for the same year.[6]

Raking

'Raking' is a process of proportionally adjusting the estimates in nested geographies so that their sum matches that of the parent geography. Once an initial estimate is produced for each county we rake the county-level estimates so that the county estimates sum to the previously established state estimate. To do this, we divide each initial county estimate by the sum of the estimate of all counties and then multiply that proportion by the estimated state total.

$$County_{raked} = County_{initial}/\sum(Allcounty_{inital})*StateEstimate$$

This process ensures that 1) each county retains their relative size compared to other counties, and 2) the counties sum to the established state totals.

---

[4] For small counties, the ACS estimates are averaged over a five-year period which can make it harder to notice change. Consequently, the housing unit method may show very small amounts of change while all other indicators suggest a large decline or growth. These differences could substantially skew the estimates if the housing unit method was retained.

[5] The Census Bureau collects data 'permit-issuing places', which excludes many regions of NM as smaller municipalities often do not issue their own permits. https://www.census.gov/construction/bps/methodology.html

[6] While our estimates benefit from a diverse set of data sources, there are known issues with each source that could impact the quality of the data and thus may increase error for some counties.

Tract-level Estimates

Because tracts are nested within counties, we have traditionally assumed that tracts follow patterns of change similar to the county in which they belong. This allows the tracts to sum to the county total and retain patterns that are indicated by the county-level data. In the last decade, we used methods that distributed the growth or removed the loss observed in the parent county to the tracts within that county either using indicators (such as building permits) or using methods for distributing the population proportionally. Consequently, all tracts in our estimates changed in the same direction as the county in which it belonged. However, a review of population change in New Mexico's tracts from 2010 to 2020 shows that the assumption that tracts change in a manner consistent with the parent county is not always valid. Thus, for the current estimates production, we used patterns of change between the 2010 and 2020 Censuses to inform the methods used to produce our 2020, 2021, and 2022 tract-level estimates.

Though several factors contribute to error in population estimates, one factor that contributes to error is the size of the population being estimated. Consequently, tract estimates typically have more errors than state or county estimates due to the nature of their size. In a review of small-area estimates, Wilson and colleagues[7] noted that averaging several methods or using a composite method could reduce the overall mean error more than any single method. The composite method in which methods for each geography are selected based on the population size and direction of change can reduce overall mean errors[8]. This method was found to reduce error slightly more than averaging method, perhaps because prior knowledge of the estimated geographies was taken into account when selecting the method. To improve our tract-level estimates for New Mexico, we use a composite method that selects either a *share-of-growth, exponential extrapolation, or linear extrapolation* method depending on characteristics of each tract and the county in which it resides.

The *share-of-growth* method estimates the tract population by calculating each tract's share of the county's growth[9] (or loss) across the prior decade and then applying this share to the estimated post-censal county population to determine the tract-level estimate. For our tract estimates, we used the *share-of-growth* method for tracts that changed in the same direction as its parent county from 2010 to 2020. Consequently, instead of applying the share-of-growth to the full county, we used the balance of the post-censal county estimates after removing the tract population for tracts that change in a different direction than the parent county. With the *share-of-growth* method, we assume that each tract's share of the county's growth is unchanged across the decade.

[7] Wilson, T., Grossman, I., Alexander, M., Rees, P., and Temple, J. (2022). Methods for small area population forecasts: State-of-the-art and Research needs. *Population Research and Policy Review, 41*, 865-898.

[8] Rayer, S. & Smith, S.K. (2010). Factors affecting the accuracy of subcounty population forecasts, *Journal of Planning Education and Research, 30*(2), 147-161. DOI: 10.1177/0739456X10380056

[9] The county census totals for these calculations excluded the tracts that had population change in the opposite direction of county, thus the balance of the county total after removing tracts estimated with linear extrapolation was used in the share of growth calculations.

When the tract population changes in a different direction than the parent county, the estimates produced by the *share-of-growth* method is likely to be unreasonable.[6] Consequently, for counties where the tract's population change is in the opposite direction than its county, we use either linear or exponential extrapolation methods to produce post-censal. Past research suggests that linear extrapolation performs better for growing populations while exponential extrapolation methods are preferred for declining populations.[4]**Error! Bookmark not defined.** Thus, we select the extrapolation method for each tract based on the direction of its change between the 2010 and 2020 censuses. The estimated population totals for these tracts are then removed from the county total to calculate the balance of the county population estimate for the remaining tracts (for which the *share-of-growth* method was used).

Raking

The raking process for the tract estimates is the same as that for the county estimates, with the exception that the county estimates are used as the control rather than the state total. Once we produce initial estimates for each tract, we adjust them so that their total matches the previously established county estimate. We do this by converting each tract's initial estimate into a share of the total across all tracts, then applying that share to the county total.

$$tract_{raked} = tract_{initial}/\sum(alltracts_{inital})*CountyEstimate$$

This process ensures that 1) the tracts retain their relative size compared to other tracts, and 2) the tracts sum to the established county totals.

Distribution of Characteristics

Once state, county, and tract totals are finalized we move to estimating the demographic characteristics (race, age, sex, and Hispanic-origin) of the people in those geographies using Multidimensional Iterative Proportional Fitting ('MIPF')[10],[11]. This method uses data from prior years and other sources to inform how characteristics are distributed across the population, while our previously determined tract, county, and state totals are maintained. This can essentially be thought of as a giant Suduko puzzle, in which we have some known values and need to use that known information to determine the unknown values. The known values in this case are: 1) the distribution of characteristics across tracts at the most recent Decennial Census ('seed'), 2) the Census Bureau's population estimates' distribution of demographic characteristics for the current year at the county level, and 3) our current year state, county, and tract totals. The output is the estimated population by demographic characteristics at the tract-level that sum to the previously estimated state, county, and tract totals. Once all characteristics are estimated, we combine the Asian and Native Hawaiian/Pacific-Islander categories, as requested by NM DOH. The larger combined size of the combined Asian-Pacific

---

[10] For those interested, this presentation provides a nice illustration of a simple 2x2 iterative proportional fitting: https://u.demog.berkeley.edu/~eddieh/IPFDescription/AKDOLWDIPFTWOD.pdf
[11] The process of fitting the population characteristics is done using the ipfn package for Python.

Islanders group also improves the reliability and usability of data as the Asian and Native Hawaiian/Pacific-Islander populations are very small in many of New Mexico's tracts.

The Census Bureau estimates data used for the fitting process requires some modifications to account for different racial groups in our final estimates compared to the Census Bureau's estimates. While the NM DOH requires estimates for single-race groups using the 1977 OMB standards, the Census Bureau uses the 1997 OMB standards and categorizes people who report multiple races in a "two-or-more" race category.

Two or More Races ('TOMR')

The Census Bureau's estimates can be used to proportionally distribute our population estimates in ways consistent with the Census Bureau's tabulation, however, some modifications are required to account for the inclusion of a "Two or More Races" ('TOMR') racial group in the Census Bureau's estimates which combines all people reporting multiple races into one group, irrespective of the specific racial groups (see *Appendix 1* for more information on why a modification was needed and how the decision was determined).[12] To address this, we redistribute people classified as TOMR based on the April 2024 guidance from NM DOH such that those reporting TOMR will be assigned in the following manner: 1) TOMR that include American Indian and Alaskan Native ('AIAN') will be classified as AIAN, and 2) TOMR groups that no not include AIAN as one of the selected races will be classified to the rarest race select based on NM's statewide prevalence (i.e. people who report race as White and Asian would be classified as Asian, because Asian is the rare racial group in New Mexico).

Seed File

A critical input for MIPF is the seed file (*Table 1*). While the Census Bureau's Population Estimates are not released at the tract-level, demographic characteristics are available from the Decennial Census at the tract-level. Consequently, using the 2020 Decennial Census Demographic and Housing Characteristics file ('2020 DHC'), we calculate the population in each tract for all possible combinations of age, sex, race, and ethnicity. For example, the file would indicate the population of each tract that is white (race), Hispanic (ethnicity), males (sex), who were age 10-years-old on April 1, 2020, and all other possible combinations. Age-specific mortality rates are then applied to this seed file and subsequently aged forward producing a new seed file for future years. The population under age 1 is produced by distributing county births across all tracts in the same proportion as was observed in the 2020 census.

---

[12] Please see Appendix 1 for an explanation of how this decision was made.

Table 1. *GPS's seed file variables and categories.*

Demographic Characteristic	Variable Categories
Census Tract	612 Census Tracts using 2020 Census Geographies
Sex	Male, Female
Age	Single-year of age
Race	White, Black, American Indian/Alaskan Native, Native Hawaiian/Pacific Islander, Asian, Two or More races
Hispanic-origin	Hispanic, non-Hispanic

The Decennial Census data presents one key challenge regarding demographic characteristics. While "some other race" is not a valid race categorization for the NM DOH, it is a valid response for the Decennial Census. Below is an explanation of how we handled this challenge when developing the seed file.

Some Other Race ("SOR")

New Mexico has a higher proportion of people reporting "some other race" than many other states.[13] This is because people who are of Hispanic-origin do not see Hispanic as their racial identity. Consequently, those who report *only* "some other race" and report being of Hispanic-origin would be given a racial categorization of white, as we expect those who are Hispanic-origin and a race other than white would indicate the other race on the Census along with "Some other race". The vast majority of New Mexican's who report "Some other race -alone" also report that they are Hispanic. Those who report "Some other race – alone" and are not Hispanic were distributed based on the proportion of the five race alone categories by sex and single-year age in each tract.[14] For those who reported SOR and a second race, we coded them as the second reported race group and distributed this population based on patterns in the Demographic and Housing Characteristics file.[15] For those who reported SOR and two or more additional races, we coded as TOMR.

> If Race1 + SOR, then Race = Race1.
> If Race1 + Race2+ + SOR, then Race = TOMR.

[13] In the past, the Census Bureau produced a modified-race file which included 31 race categories and recategorized people of "some other race." This file is not currently available but may become available for use in future years. We will consider revising our estimates methodology when a revised Modified Race File becomes available.
U.S. Census Bureau, Population Division. (2012 July 5). *Modified Race Summary File Methodology Statement.*
https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/modified-race-summary-file-method/mrsf2010.pdf
[14] If Hispanic and SOR-alone, then Race = White-alone.
If non-Hispanic and SOR-alone, then Race distributed across county proportionally based on counties age-sex specific distribution of non-Hispanic race.

[15] After reclassifying the population who were SOR and one of the five single race categories to their single-race category, the population was distributed across tracts based on the distributions found in the P8 and P9 tables. They were then distributed across age and sex groups in a manner consistent with the distribution for the same race group in the Demographic and Housing Characteristics file tables PCT12A through PCT12O.

Population Under Age 2

The Census Bureau's population estimates use vital records from the National Center for Health Statistics to estimate the child population born since the last decennial census. However, the Census projects expected births based on recent trends for children under age 2 because of delays in when this date become available. In recent years, these estimates have resulted in an overestimate of children under age 2 when compared to the birth counts recorded by NM DOH.

To improve the estimates for children under age 2, we use county-level birth counts provided by the DOH to estimate the population of children under age 2. The most recent period of births (i.e. January – July 2023), were preliminary at the time the population estimates were produced and thus may include some error. Estimates for this cohort will be updated using actual birth counts from the NM DOH in future releases.

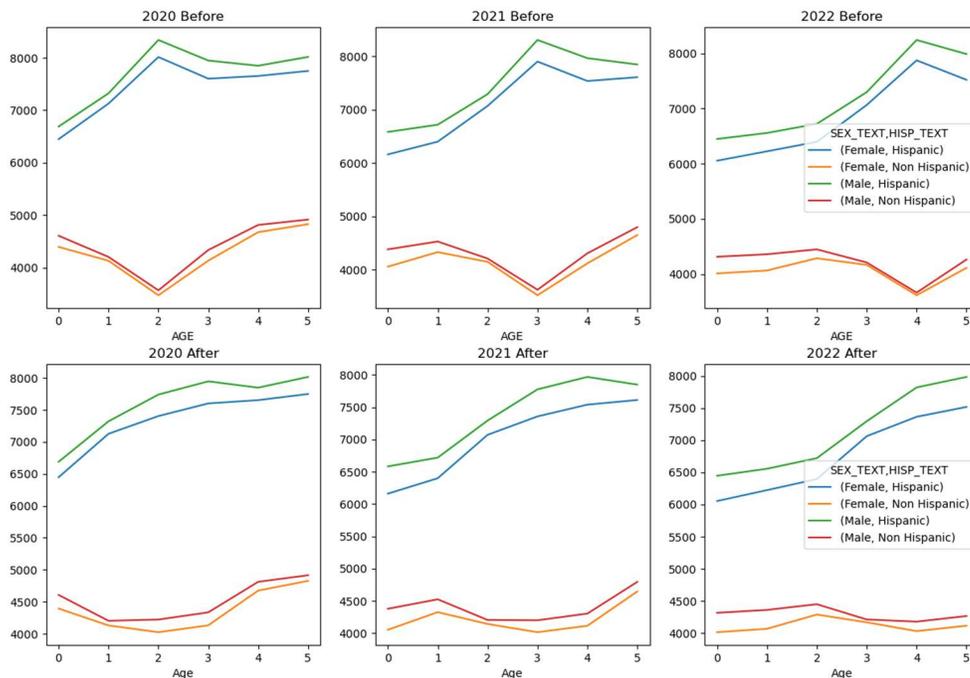**Evaluation and Adjustments v2024**

Figure 1. Population estimates for those under age 5 for years 2020-2022 before and after GPS's corrections.

After the complete set of estimates are produced, we carefully examine the estimates, investigate any areas of concern, and adjust the data accordingly. For example, in the current estimates, we noticed an unusual pattern showing a substantial drop in the number of non-Hispanic two-year-olds and a corresponding jump in the number of Hispanic two-year-olds in

2020. Upon further examination, we found that this pattern continued each year so that the same issue appeared for three-year-olds in 2021 and was introduced to our estimates through our fitting procedures using the Census Bureau's Special Tabulation file. An investigation into this issue revealed that this pattern is the result of an error in the 2017/2018 birth data that the Census Bureau obtains from the CDC and uses in their estimates.[16] Consequently, we adjusted our final estimates so that the Hispanic and sex distribution for this birth cohort used the same proportions as the birth cohort one year above them.

---

[16] The Census Bureau was aware of the issue and was in discussions with the CDC about fixing the issue.

Appendix 1

The Census Bureau currently tabulates statistics for their Population Estimates Program on the five race groups included in the 1997 OMB standards for people who report only one race ("White alone," "Black or African American alone," "American Indian or Alaska Native alone," "Asian alone," "Native Hawaiian or Other Pacific Islander alone").[17] Anyone reporting more than one race is tabulated in the "Two or More Races" category. In 2020, 10.2% of people in the United States identified as TOMR compared to 19.9% of people in New Mexico.[18] Consequently, decisions about how to distribute this group across other racial groups is more influential in New Mexico than in other states. The Census Bureau also produces population estimates for 31 race groups combined or in combination. These estimates count people in all racial groups that they identify so the totals exceed 100% of the population. The Census Bureau's estimates can be used to proportionally distribute our population estimates in ways consistent with the Census Bureau's tabulation, however, a decision about how to estimate the "Two or More Race" category was necessary.

In May 2023, discussions with staff at the NM DOH led to the decision that UNM-GPS would include a TOMR category in the population estimates to enable NMDOH to use their data to explore different approaches to distributing the TOMR category across the population. By lumping all individuals who are multi-racial into a single group, we may lose the ability to interpret statistics for this group in a meaningful way. The experiences and concerns of people in this group likely vary significantly as it would not be reasonable to expect that the concerns of someone who is both Black and White are more similar to someone who is Asian and Native American than one of the race-alone groups (i.e. Black or White).

Potential solutions[19] include assigning race using an algorithm based on rarest-race, most-popular race, or developing a bridging method using New Mexico-specific data from the Behavioral Risk Factor Surveillance System (BRFSS) using questions that ask about both multiple races and the person's self-identified most representative race.[20] When assigning a single race to a person who identifies as multiple races, the ideal scenario is to allow the individual to self-identify their primary race.[21] Consequently, using data from the BRFSS would be ideal, if the data are sufficient for drawing such conclusions at the population level. However, UNM-GPS does not have access to that data, and it was determined that NM DOH would further explore these options.

In the Spring of 2024, the NM DOH leadership, along with their Tribal Liaison and epidemiologists, determined that those reporting TOMR would be categorized using a statewide rarest race methodology with a special rule that ensures historically marginalized populations are not undercounted. Through this approach, people who are categorized as two-

[17] U.S. Census Bureau. "Measuring Racial and Ethnic Diversity for the 2020 Census." https://www.census.gov/newsroom/blogs/random-samplings/2021/08/measuring-racial-ethnic-diversity-2020-census.html

[18] U.S. Census Bureau, "Race and Ethnicity in the United States: 2010 Census and 2020 Census." https://www.census.gov/library/visualizations/interactive/race-and-ethnicity-in-the-united-state-2010-and-2020-census.html

[19] Miller., J.A. (2023). Race and Ethnicity Review and Options for the NM Department of Health. *Report produced for NM DOH with background and options to understand the concerns about the two-or-more-race grouping.*

[20] For a full list of BRFSS questions, please see https://data.cdc.gov/Behavioral-Risk-Factors/Behavioral-Risk-Factor-Surveillance-System-BRFSS-H/iuq5-y9ct.

[21] Mays, V.M., Ponce, N.A., Washington, D.L., and Cochran, S.D. (2003). Classification of race and ethnicity: implications for public health. *Annual Review of Public Health, 24,* 83-110. DOI: 10.1146/annurev.publhealth.24.100901.140927

or-more races and report being White and Black, would be categorized as Black as this is the rarer race in New Mexico. The special rule would have people who report that they are Native American and another race, would be categorized as Native American even if their other race is the rarer race in New Mexico.