FEDERAL INTERAGENCY TECHNICAL WORKING GROUP ON RACE AND ETHNICITY STANDARDS

ITWG Bridging Team Methods Report: Technical Documentation

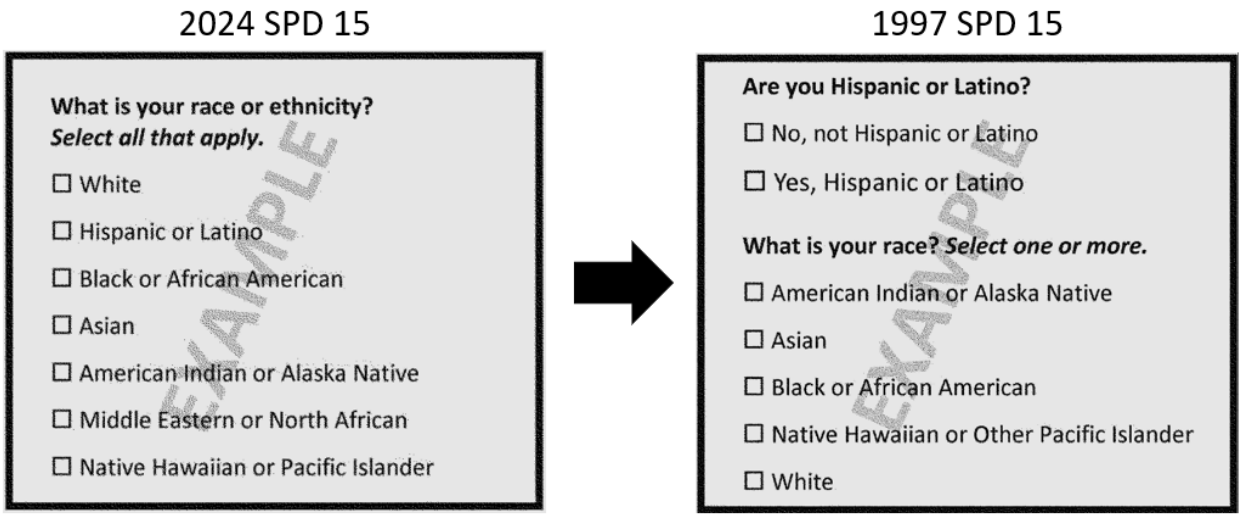
March 2024

Contents

1 Bridging from 2024 SPD 15 to 1997 SPD 15

Bridging from 2024 SPD 15 to 1997 SPD 15 refers to the process of making race/ethnicity data collected under the updated 2024 Statistical Policy Directive No. 15, *Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity* (SPD 15), consistent with the standards of data collection from the previous standards, which were codified in the 1997 SPD 15.

Figure 1. Bridging from 2024 SPD 15 Question Format to 1997 SPD 15 Question Format



To assist data users in the bridging of their data and to promote consistency across federal agencies, the Federal Interagency Technical Working Group (Working Group) has published a set of preliminary bridging factors and statistical programs for public use. Section 1 in the Technical Documentation provides detailed guidelines and instructions for data users who wish to implement the Working Group's bridging factors and statistical programs to bridge data collected under 2024 SPD 15 back to the 1997 SPD 15 standards.

1.1 Preparation

1.1.1 Software Requirements

To implement the Working Group's preliminary bridging methodology, data users must have access to one of two statistical programming software: SAS or Python. The Working Group has provided bridging programs in both SAS or Python statistical languages to provide flexibility to agencies and data users. Note that Python is available to download online at no cost. The Python program will also require the pandas and NumPy libraries.[1]

[1] NumPy and pandas are available to download at no cost and can be installed using Python's package manager `pip`. They are also available via the Anaconda distribution of Python: https://www.anaconda.com/download.

1.1.2 Preparing Input Data

The bridging programs require input data (i.e., the data to be bridged) that: (1) have been collected according to the 2024 SPD 15 standards and coded in a way that is consistent with the OMB Bridging Methods report; (2) retain all possible race/ethnicity combinations present in the data; (3) have been aggregated to count/distribution datasets with no stratification or stratification by age only; and (4) have been saved in a comma separated value (CSV) file format. We describe each of these criteria below.

Users can also reference a set of mock datasets provided with the bridging report and program. These datasets can be used as examples for how to structure agency-collected data for use with the Working Group's bridging programs. These datasets can also serve as test cases for agencies seeking to gain experience with the programs before inputting their own data.

1.1.2.1 Collecting and coding data according to 2024 SPD 15

Figure 2 shows an example of the proposed combined question with minimum categories based on 2024 SPD 15, where the race/ethnicity options have been listed in descending order by estimated population size[2]. Note that respondents are asked to select all options that apply. The seven race/ethnicity alone categories, along with all possible combinations of two or more race/ethnicity selections, results in a total of 127 race/ethnicity categories.

Since the details of the proposed question have not been finalized at the time of writing this report, we have used the ordering of the race/ethnicity groups in this example to determine the values of the categories for the race/ethnicity variable in our bridging program. Data users must also conform to this precedent in the assignment of values to the race/ethnicity categories for their input data.

Figure 2. 2024 SPD 15 with Groups Ordered by Estimated Population Size

Users should reference Section 3.2 to view the full list of race/ethnicity categories with their assigned variable values. Data users must code the race/ethnicity variable in their input dataset as specified in

[2] https://www.federalregister.gov/documents/2023/01/27/2023-01635/initial-proposals-for-updating-ombs-race-and-ethnicity-statistical-standards

Section 3.2, where the values for the race/ethnicity variable are numeric or string (text) numerals from 1 to 127. Including the value labels in the dataset is not necessary.

1.1.2.2 Retaining all race/ethnicity combinations

When presenting information on race/ethnicity in data products such as tables, figures, and other visualizations, federal agencies and researchers often present a collapsed, or simplified, set of race/ethnicity categories. Namely, agencies may show the minimum race alone categories with a single Two or More Races category that includes all multi-race combinations grouped together. This may be done for size or page limit considerations (e.g., a table with 7 categories is smaller than a table with 127 categories) and/or because many of the Two or More Races categories would otherwise show very small counts or even counts of zero. However, due to the nature of the changes involved in a shift from 1997 SPD 15 to 2024 SPD 15, the bridging of a collapsed race/ethnicity distribution is neither advised nor possible with the Working Group's current bridging programs. Namely, when the previously separate Hispanic or Latino question is combined with the race question in the 2024 SPD 15 format, the Hispanic or Latino category is now treated identically to the other major race groups and is no longer measured independently of race. This change leads to a new definition of the Two or More Races category, making the category incomparable between current and proposed formats.

For example, in the 1997 SPD 15 format, an individual who selects Hispanic or Latino in the ethnicity question and then selects Black or African American in the race question would be considered Hispanic, Black or African American alone (i.e., they would NOT be grouped in the Two or More Races category). In contrast, in the 2024 SPD 15 format, the same individual who selects Hispanic or Latino and Black or African American in the race question WOULD be grouped in the Two or More Races category. In other words, certain race/ethnicity combinations that are considered Two or More Races in the 2024 SPD 15 format would be considered single race categories in the previous format. Whether or not a race/ethnicity combination would be bridged to a single race or multiple race category depends on the race/ethnicity categories that comprise the Two or More Races response. Therefore, bridging requires that all possible race/ethnicity combinations present after data collection be retained throughout the bridging process BEFORE collapsing any categories. The individual race/ethnicity selections that comprise a Two or More Races combination are necessary inputs for the bridging program. Data users who wish to present a simplified, bridged race distribution must collapse categories only after the bridging process.

1.1.2.3 Structuring an input dataset

The bridging factors and programs are designed to work with count, or distribution, datasets, where each row in the dataset represents a population count associated with a race/ethnicity category. This data structure is analogous to a tabulation of the proposed race/ethnicity variable. This section provides details on how to convert a micro dataset into a count dataset, and on how to structure data that may or may not be stratified by additional characteristics.

1.1.2.3.1 Aggregating to a count dataset

If an agency or data user is starting with a *microdata* sample, where the observations or rows in the data are individual responses, the data must first be aggregated to a count/distribution structure. This procedure can be accomplished in any standard statistical software. Below is an example of how to restructure a micro dataset into a count dataset, where "user_microdata" is the name of a user's micro dataset; "RACE_ETHNICITY" is the name of a user's race/ethnicity variable; "DUMMY" is a dummy variable set to 1 for each individual; "count_data" is the name of the resulting count dataset to output; and "count" is the name of the resulting population count variable. SAS and Python users can replicate the code below, substituting data and variable names with those from their own dataset, to convert their data from individual-level to count-level. Users should refer to their software's user guide and technical documentation for additional instructions.

SAS code to tabulate microdata

```
proc summary data = user_microdata nway;
  class RACE_ETHNICITY;
  var DUMMY;
  output out=count_data(drop=_:) sum=count;
run;
```

Python code to tabulate microdata

```
count_data =
user_microdata.groupby(["RACE_ETHNICITY"]).size().reset_index(nam
e="count")
```

1.1.2.3.2 Using age stratifiers

The bridging program also permits the race/ethnicity variable to be stratified by a three-category age variable (ages 0-17, 18-64, 65 and up) or by single year of age. Regardless of whether the data are stratified by age, the columns in the dataset must include a numeric or character/string race/ethnicity variable ranging from 1 to 127 and a numeric, count variable indicating the number of people (or alternatively, a weighted estimate) in each corresponding category. See the tables below for examples of the three allowable data structures.

In Table 1, the RACE_ETHNICITY column indicates the numeric value, defined in Section 3.2, associated with each group. The RACE_ETHNICITY_LABEL is a character variable that provides a text label for the associated RACE_ETHNICITY value. Note that it is not necessary for a data user to include this *Label* column. The COUNT variable indicates the number of people from each race/ethnicity group in the dataset. Note that the values in the COUNT column in Tables 1-3 are mock values randomly generated for these examples. In a race/ethnicity count dataset with no additional stratification, there should be one row per RACE_ETHNICITY category, thus the dataset should have 127 rows of data (plus an additional row for column headers).

Table 1. Example 1 input data with no age stratification

ROW_NUMBER	RACE_ETHNICITIY	RACE_ETHNICITY_LABEL	COUNT
1	1	White alone	372
2	2	Hispanic or Latino alone	377
3	3	Black or African American alone	407
4	4	Asian alone	409
5	5	American Indian or Alaska Native alone (AIAN)	408
6	6	Middle Eastern or North African alone (MENA)	397
7	7	Native Hawaiian or Pacific Islander alone (NHPI)	400
8	8	White or Hispanic and Latino	337
9	9	White and Black or African American	373
…	…	…	…
127	127	White, Hispanic or Latino, Black or African American, Asian, AIAN, MENA, NHPI	403

Table 2 shows a distribution dataset where the race/ethnicity variable is stratified by a three-category age variable. Each row represents a race/ethnicity and age combination with an associated population or sample count. For example, the first row of data shows a count of 85 for the White alone race/ethnicity group (RACE_ETHNICITY=1), ages 0 to 17 (AGE_CATEGORY= 1). Since there are three age categories per race/ethnicity group, there are three rows for each of the 127 groups, totaling 381 rows of data (plus an additional row for column headers) .

Table 2. Example 2 input data stratified by three-category age variable

ROW_NUMBER	RACE_ETHNICITY	AGE_CATEGORY	COUNT
1	1	1	85
2	1	2	195
3	1	3	92
4	2	1	89
5	2	2	205
6	2	3	83
7	3	1	87
8	3	2	228
9	3	3	92
…	…	…	…
381	127	3	104

Table 3 shows a distribution dataset where the race/ethnicity variable is stratified by a single year of age variable top-coded at 85. Each row represents a race/ethnicity and age combination with an associated population or sample count. For example, the first row shows a COUNT of 6 for the White alone

race/ethnicity group (RACE_ETHNICITY=1), age 0 (SINGLE_YEAR_AGE = 0). Since there are 86 age values per race/ethnicity group, there are 86 rows for each of the 127 groups, totaling 10,800 rows of data.

Table 3. Example of input data stratified by single year of age

ROW_NUMBER	RACE_ETHNICITY	SINGLE_YEAR_AGE	COUNT
1	1	0	6
2	1	1	9
3	1	2	7
4	1	3	6
5	1	4	6
6	1	5	3
7	1	6	2
8	1	7	5
9	1	8	7
…	…	…	…
10,800	127	85	4

1.1.2.3.3 Using stratifiers besides age

It is important to note that the bridging programs do not permit race/ethnicity to be stratified by additional characteristics other than the age variables described above. For example, if data are broken out by a two-category sex variable, where there are separate counts by male and female for each race/ethnicity category, a user must run the bridging program separately for each stratum. In the example of a binary sex variable, this would entail running the bridging program twice: once on a dataset for females and once on a dataset for males. Likewise, if a user's data include race/ethnicity distributions separately by state, the user must run the bridging program once for each state dataset.

1.1.2.4 Using provided "mock" datasets

To help users set up their data, we have provided three example datasets with simulated, or mock, population counts using 2024 SPD 15. These datasets were designed for testing the Working Group's bridging programs and to allow data users to practice use of the programs prior to collecting their own data under 2024 SPD 15. The three example datasets correspond to the data structures outlined in Tables 1, 2, and 3: *race_ethnicity_data_example_1.csv* is a count dataset with no age stratification; *race_ethnicity_data_example_2.csv* is a count dataset stratified by the three-category age variable; and *race_ethnicity_data_example_3.csv* is a count dataset stratified by single year of age. Users can follow instructions in the bridging program to use any of the three example datasets.

It is also important to note that the example datasets were randomly generated and do not represent demographically reasonable race/ethnicity distributions. Once bridged from 2024 SPD 15 to 1997 SPD 15, example datasets 1-3 are not designed to be used with the 1997 SPD 15 to 2024 SPD 15 program. Doing so may yield unexpected results.

1.1.3 Understanding the Bridging Factors

The bridging factor dataset (2024 SPD 15 to 1997 SPD 15) contains three sets of bridging factors appended into a single comma separated value (CSV) file. The first set of factors are designed for data with no age stratification; the second set of factors are stratified by the three-category age variable; and the third set of factors are stratified by single year of age. Users will specify the structure of their input data in the bridging program (i.e., whether age is being used), and the program will automatically select the appropriate set of factors for their data. The provided bridging factor dataset requires no setup prior to use with the bridging program.

All race/ethnicity categories other than Hispanic or Latino alone have a single direct bridging factor equal to 1, regardless of age detail. This is also true for each race/ethnicity by age combination, excluding Hispanic or Latino alone. The Hispanic or Latino alone category instead is bridged proportionally to each of the 31 possible race values from current format. This means there are 31 bridging factors for each Hispanic or Latino by age combination. Proportional bridging factors for this group are decimals between 0 and 1 that have values out to ten decimal places (10^{-10}).

1.2 Running the Program

Both SAS and Python bridging programs are designed to require minimal input from data users. This section provides details on each step of the "USER INPUT" section of either program.

User Input Step 1: Specify the file path where the data and bridging factors are located.

In the first step, indicate the folder where your input data (i.e., the race/ethnicity distribution in 2024 SPD 15 format) and bridging factor dataset have been saved.

User Input Step 2: Specify the file path where the final, bridged dataset should be output.

Next, indicate a folder where you would like the bridged data to be output. This can be the same file path specified in Step 1.

User Input Step 3: Specify the name of the input dataset.

Next, indicate the name of your input dataset. Remember that the datafile should be in csv format and thus the file name should include a .csv suffix. You may use your own data, or for testing purposes, one of the three provided "mock" datasets included with the Working Group's preliminary bridging report.

User Input Step 4: Specify name of race/ethnicity variable in input dataset.

Indicate the name of the race/ethnicity variable in your input dataset. The name specified here must match the name of the variable in your dataset. The values of this variable should range from 1 to 127 and can be found in Section 3.2.

User Input Step 5: Specify the name of the count variable in the input dataset.

Indicate the name of the count variable in your input dataset. This should be a numeric variable containing the population total or estimate of each race/ethnicity group in the dataset.

User Input Step 6: Specify whether or how the data are stratified by age.

This step determines which set of bridging factors are merged with your data—the factors with no age stratification, the factors stratified by a three-category age variable, or the factors stratified by single year of age.

See below for example code in SAS and Python that corresponds to the three options for age stratification.

SAS code to set up no age stratification	**Python code to set up no age stratification**
`%let categorical_age= no;`	`CATEGORICAL_AGE = False`
`%let single_year_age= no;`	`SINGLE_YEAR_AGE = False`

SAS code to set up three-category age stratification	**Python code to set up three-category age stratification**
`%let categorical_age= yes;`	`CATEGORICAL_AGE = True`
`%let single_year_age= no;`	`SINGLE_YEAR_AGE = False`

SAS code to set up single year of age stratification	**Python code to set up single year of age stratification**
`%let categorical_age= no;`	`CATEGORICAL_AGE = False`
`%let single_year_age= yes;`	`SINGLE_YEAR_AGE = True`

User Input Step 7: Specify name of age variable in input dataset.

Indicate the name of the age variable in your input dataset. If your data are not stratified by age (e.g., if you are using the provided example dataset *race_ethnicity_data_example_1.csv)*, then specify the following:

SAS (no age variable)	**Python (no age variable)**
`%let age= ;`	or `AGEVAR = ""`

Otherwise, ensure that the variable name you specify matches the age variable name in your data. If using the provided example dataset *race_ethnicity_data_example_2.csv,* the age variable you would specify is "agecat". If using the third example dataset, *race_ethnicity_data_example_3.csv,* you would specify "sya".

OPTIONAL User Input Step 8: Specify whether the final population counts should be rounded to nearest integer.

The final step in the user input section determines whether the program rounds the final bridged population counts to integers. Additional notes on the rounding step can be found in the next section. The default selection is to use controlled rounding (see example code below).

SAS (Yes to controlled rounding)	Python (Yes to controlled rounding)
`%let controlled_rounding=yes;`	CONTROLLED_ROUND= `True`

If you do not wish to use controlled rounding, see the code below.

SAS (No to controlled rounding)	Python (No to controlled rounding)
`%let controlled_rounding=no;`	CONTROLLED_ROUND= `False`

1.2.1 Optional rounding step

The use of proportional bridging factors leads to bridged population counts that may include decimals. To allow data users to output bridged population distributions with whole numbers, thus avoiding the presentation of population totals with decimals or proportions of people, we have included an optional rounding step. This step also ensures that the population total of the bridged output dataset matches exactly the population total of the input dataset. Rounding to the nearest integer, on the other hand, could result in the total population count in the bridged data differing slightly from the count in the input data.

Specifically, we use a greatest-mantissa rounding algorithm—a type of "controlled" rounding procedure commonly used in population research. This algorithm rounds a list of numbers according to the relative size of each number's mantissa—the value to the right of the decimal point. The algorithm sequentially rounds up the numbers with the largest mantissas until the sum of the rounded numbers is the same as the sum of the unrounded numbers. The remaining numbers in the sequence with relatively lower mantissas are rounded down.

1.3 Interpreting the Results

Upon running the bridging program, a final bridged dataset in CSV format will be output to a folder specified by the user. The data output will be a race/ethnicity distribution in the format defined by 1997 SPD 15, which uses the separate race and Hispanic ethnicity questions. If no age stratification is used, the final dataset will include five columns: a column that corresponds to the race variable, called IMPRACE; a column that corresponds to the Hispanic or Latino origin variable, called CENHISP; a column with labels for the race variable called IMPRACE_LABEL, a column with labels for the Hispanic origin variable called CENHISP_LABEL, and a column named FINAL_COUNT with the bridged population totals for the race/ethnicity categories. If data were stratified by age, there will be an additional column that corresponds to the age variable used: either AGECAT (categorical age) or SYA (single-year of age).

The IMPRACE and CENHISP variables have been named and coded to match the Census Bureau's race and ethnicity variables that conform to 1997 SPD 15. The variable values and category labels are defined in Section 3.1.

There are 31 race combinations and 2 Hispanic origin categories, resulting in a total of 62 total possible race/ethnicity combinations. Table 4 below shows an example of a truncated bridged dataset with no age stratification. It also includes a ROW_NUMBER column for illustrative purposes that will not appear in a bridged dataset. The first row of data shows a FINAL_COUNT of 1,137 people in the non-Hispanic (CENHISP=1), white alone (IMPRACE=1) category. The last row shows a count of 1,236 people in the Hispanic (CENHISP=2), White, Black or African American, AIAN, Asian, NHPI (IMPRACE=31) category.

Table 4. Example of bridged distribution (truncated)

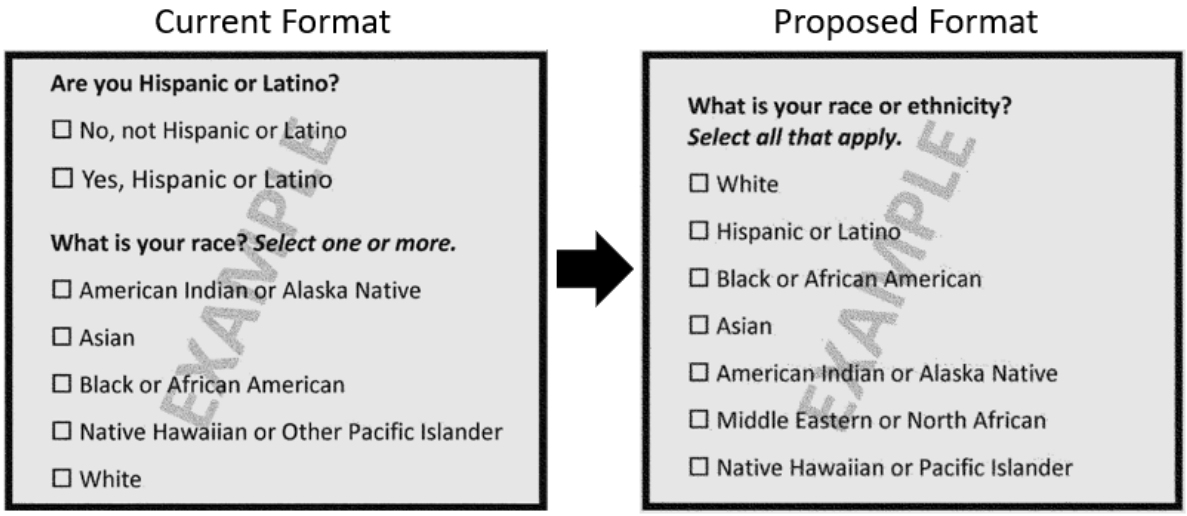
ROW_NUMBER	CENHISP	IMPRACE	FINAL_COUNT	IMPRACE_LABEL	CENHISP_LABEL
1	1	1	1,137	White alone	Not Hispanic or Latino
2	1	2	407	Black alone	Not Hispanic or Latino
3	1	3	408	AIAN alone	Not Hispanic or Latino
4	1	4	409	Asian alone	Not Hispanic or Latino
5	1	5	400	NHPI alone	Not Hispanic or Latino
6	1	6	1,167	White and Black or African American	Not Hispanic or Latino
…	…	…	…		
62	2	31	1,236	White and Black and AIAN and Asian and NHPI	Hispanic or Latino

Upon outputting a bridged dataset, users may then collapse race/ethnicity categories, if desired, to show a fewer number of race/ethnicity categories. For example, the categories that include two or more races may be collapsed or aggregated into a single Two or More Races category. To do this, a user could sum all of the FINAL_COUNT values for the categories that include two or more races using a PROC SUMMARY procedure in SAS (see Section 1.1.2.3.1 for a similar example).

2 Bridging from 1997 SPD 15 to 2024 SPD 15

Bridging from 1997 SPD 15 to 2024 SPD 15 refers to the process of making race/ethnicity data collected under the previous 1997 Statistical Policy Directive No. 15, *Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity* (SPD 15), consistent with the standards of data collection from the updated 2024 SPD 15.

Figure 3. Bridging from 1997 SPD 15 Question Format to 2024 SPD 15 Question Format



To assist data users in the bridging of their data and to promote consistency across federal agencies, the Federal Interagency Technical Working Group (Working Group) has published a set of preliminary bridging factors and statistical programs for public use. Section 2 in the Technical Documentation provides detailed guidelines and instructions for data users who wish to implement the Working Group's bridging factors and statistical programs to bridge data collected under 1997 SPD 15 to 2024 SPD 15.

2.1 Preparation

2.1.1 Software Requirements

To implement the Working Group's preliminary bridging methodology, data users must have access to one of two statistical programming software: SAS or Python. The Working Group has provided bridging programs in both SAS and Python statistical languages to provide flexibility to agencies and data users. Note that Python is available to download online at no cost. The Python program will also require the pandas and NumPy libraries.

2.1.2 Preparing Input Data

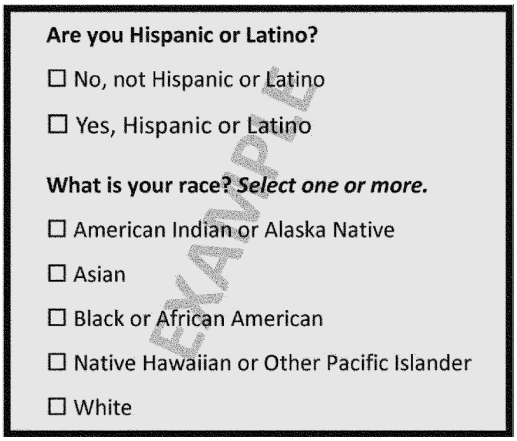
The bridging program requires input data (i.e., the data to be bridged) that: (1) have been collected according to 1997 SPD 15 standards and coded in a way that is consistent with this report; (2) retain all possible race/ethnicity combinations present in the data; (3) have been aggregated to count/distribution datasets with no stratification; and (4) have been saved in a comma separated value (CSV) file format. We describe each of these criteria below.

2.1.2.1 Collecting and coding data according to 1997 SPD 15

Figure 4 shows an example of the 1997 SPD 15 separate question format with minimum categories, where the race categories have been listed in alphabetic order. Note that respondents are first asked to select whether they are Hispanic or Latino or not Hispanic or Latino, and then are asked to select all race options that apply. The five race alone categories, along with all possible combinations of two or more

race selections, results in 31 race categories. Combined with the two Hispanic or Latino ethnicity options, there are a total of 62 possible ethnicity and race combinations.

Figure 4. 1997 SPD 15 with Race Groups Ordered Alphabetically



While the ordering of race categories in the example above is alphabetical, we have coded the categories in the bridging factor datasets to conform to the coding scheme used by the Census Bureau. The coding of both variables, including the values and category labels, can be found in Section 3.1.

Data users must code the race and ethnicity variables in their input dataset to match the coding scheme in Section 3.1, where the values for the ethnicity variable are numeric or string (text) numerals 1 or 2 for the Hispanic origin variable and the values for the race variable are numeric or string numerals from 1 to 31. Including the value labels in the dataset is not necessary.

2.1.2.2 Retaining all race/ethnicity combinations

While race and ethnicity data are often presented in collapsed or simplified formats (e.g., all two or more race combinations are combined into a single category), the bridging programs require that all possible race/ethnicity combinations present after data collection be retained throughout the bridging process BEFORE collapsing any categories. The individual selections that comprise a Two or More Races combination are necessary inputs for the bridging program. Data users who wish to present a simplified, bridged race distribution must collapse categories only after the bridging process. See Section 1.1.2.2. for additional details.

2.1.2.3 Structuring an input dataset

The bridging factors and programs are designed to work with count, or distribution, datasets, where each row in the dataset represents a population count associated with a race/ethnicity category. This data structure is analogous to a tabulation of the proposed race/ethnicity variable. This section provides details on how to convert a micro dataset into a count dataset, and on how to structure data that may or may not be stratified by additional characteristics.

2.1.2.3.1 Aggregating to a count dataset

If an agency or data user is starting with a *microdata* sample, where the observations or rows in the data are individual responses, the data must first be aggregated to a count/distribution structure. See Section 1.1.2.3.1 for more instructions.

2.1.2.3.2 Using stratifiers

The program for bridging from 1997 SPD 15 to 2024 SPD 15 does not include age-based stratification, unlike the 2024 SPD 15 to 1997 SPD 15 bridging found in Section 1. Input data should be structured such that there is one observation (i.e., one row in the dataset) for each of the 62 ethnicity and race categories. If data are broken out by additional variables where there are multiple observations for each race/ethnicity combination (e.g., race by sex distribution), a user must run the bridging program separately for each stratum of the additional variable. See Section 1.1.2.3.3 for additional discussion on stratification.

2.1.3 Understanding the Bridging Factors

The bridging factor dataset for converting from 1997 SPD 15 to 2024 SPD 15 contains one set of factors in a comma separated value (CSV) file. The provided bridging factor dataset requires no setup prior to use with the bridging program.

Each race/ethnicity category that can be bridged directly (i.e., with a bridging factor equal to 1) represents one row in the dataset. Each race/ethnicity category that must be bridged proportionally represents multiple rows in the dataset, where the number of rows for a given category is equal to the number of all possible race/ethnicity combinations to which that group could be bridged. Proportional bridging factors are decimals between 0 and 1 that have values out to ten decimal places (10^{-10}).

2.2 Running the Program

Both SAS and Python bridging programs are designed to require minimal input from data users. This section provides details on each step of the "USER INPUT" section of either program.

User Input Step 1: Specify the file path where the data and bridging factors are located.

In the first step, indicate the folder where your input data (i.e., the race/ethnicity distribution in the current SPD 15 format) and bridging factor dataset have been saved.

User Input Step 2: Specify the file path where the final, bridged dataset should be output.

Next, indicate a folder where you would like the bridged data to be output. This can be the same file path specified in Step 1.

User Input Step 3: Specify the name of the input dataset.

Next, indicate the name of your input dataset. Remember that the datafile should be in csv format and thus the file name should include a .csv suffix.

User Input Step 4: Specify name of ethnicity variable and race variable in input dataset.

Indicate the name of the ethnicity variable and the name of the race variable in your input dataset. The names specified here must match the name of the variables in your dataset. The values of the ethnicity variable should range from 1 to 2, and the values of the race variable should range from 1 to 31, following the ordering convention shown in Section 3.1.

User Input Step 5: Specify the name of the count variable in the input dataset.

Indicate the name of the count variable in your input dataset. This should be a numeric variable containing the population total or estimate of each race/ethnicity group in the dataset.

OPTIONAL User Input Step 6: Specify whether the final population counts should be rounded to nearest integer.

The final step in the user input section determines whether the program rounds the final bridged population counts to integers. Leave "yes" to round the counts or indicate "no" or any other response to leave the counts unrounded. See Section 1.2.1 for details on the option rounding step (same procedure used in standard bridging).

2.3 Interpreting the Results

Upon running the bridging program, a final bridged dataset in CSV format will be output to a folder specified by the user. The data output will be a race/ethnicity distribution in the format defined by 2024 SPD 15, which is based on the combined race and Hispanic ethnicity question. The dataset will include three columns: a column that corresponds to the race/ethnicity variable, called NEW_RACE_ETHNICITY, a column named FINAL_COUNT with the bridged population totals for the race/ethnicity categories, and a label column called NEW_RACE_ETHNICITY_LABEL. The value labels for the 2024 SPD 15 race/ethnicity variable "NEW_RACE_ETHNICITY" can also be found in Section 3.2.

There are a total of 127 race/ethnicity combinations, resulting in a total of 127 total rows of data. Table 5 below shows an example of a truncated, bridged dataset with mock values for FINAL_COUNT. Users can refer to the variable definitions for NEW_RACE_ETHNICITY in Section 3.2 to interpret and label the values in the resulting distribution. For example, the first row of data shows a FINAL_COUNT of 5,797 people in the White alone (NEW_RACE_ETHNICITY=1) category. The last row shows a count of 250 people in White, Hispanic or Latino, Black or African American, MENA, AIAN, Asian, and NHPI (NEW_RACE_ETHNICITY=127) category.

Table 5. Example of bridged distribution (truncated)

ROW_NUMBER	NEW_RACE_ETHNICITY	FINAL_COUNT	NEW_RACE_ETHNICITY_LABEL
1	1	5,797	White alone
2	2	2,003	Hispanic alone
3	3	843	Black alone
4	4	490	Asian alone
5	5	301	AIAN alone
6	6	300	MENA alone
…	…	…	…
127	127	250	White and Hispanic and Black and Asian and AIAN and MENA and NHPI

Upon outputting a bridged dataset, users may then collapse race/ethnicity categories, if desired, to show a fewer number of race/ethnicity categories. For example, the categories that include two or more races may be collapsed or aggregated into a single Two or More Races category. To do this, a user could sum all of the FINAL_COUNT values for the categories that include two or more races using a PROC SUMMARY procedure in SAS (see Section 1.1.2.3.1 for a similar example).

3 Variable Codebook

3.1 1997 SPD 15

Hispanic or Latino ethnicity (CENHISP):

Value	Category Label
1	Not Hispanic or Latino
2	Hispanic or Latino

Race (IMPRACE):

Value	Category Label
1	White alone
2	Black or African American alone
3	American Indian or Alaska Native (AIAN) alone
4	Asian alone
5	Native Hawaiian or Other Pacific Islander (NHPI) alone
6	White, Black or African American
7	White, AIAN
8	White, Asian
9	White, NHPI
10	Black or African American, AIAN
11	Black or African American, Asian
12	Black or African American, NHPI

13	AIAN, Asian
14	AIAN, NHPI
15	Asian, NHPI
16	White, Black or African American, AIAN
17	White, Black or African American, Asian
18	White, Black or African American, NHPI
19	White, AIAN, Asian
20	White, AIAN, NHPI
21	White, Asian, NHPI
22	Black or African American, AIAN, Asian
23	Black or African American, AIAN, NHPI
24	Black or African American, Asian, NHPI
25	AIAN, Asian, NHPI
26	White, Black or African American, AIAN, Asian
27	White, Black or African American, AIAN, NHPI
28	White, Black or African American, Asian, NHPI
29	White, AIAN, Asian, NHPI
30	Black or African American, AIAN, Asian, NHPI
31	White, Black or African American, AIAN, Asian, NHPI

3.2 2024 SPD 15

Race/ethnicity (NEW_RACE_ETHNICITY)

Value	Category Label
1	White alone
2	Hispanic or Latino alone
3	Black or African American alone
4	Asian alone
5	American Indian or Alaska Native (AIAN) alone
6	Middle Eastern or North African (MENA) alone
7	Native Hawaiian or Other Pacific Islander (NHPI) alone
8	White, Hispanic or Latino
9	White, Black
10	White, Asian
11	White, AIAN
12	White, MENA
13	White, NHPI
14	Hispanic or Latino, Black or African American
15	Hispanic or Latino, Asian
16	Hispanic or Latino, AIAN
17	Hispanic or Latino, MENA
18	Hispanic or Latino, NHPI

19 Black or African American, Asian
20 Black or African American, AIAN
21 Black or African American, MENA
22 Black or African American, NHPI
23 Asian, AIAN
24 Asian, MENA
25 Asian, NHPI
26 AIAN, MENA
27 AIAN, NHPI
28 MENA, NHPI
29 White, Hispanic or Latino, Black or African American
30 White, Hispanic or Latino, Asian
31 White, Hispanic or Latino, AIAN
32 White, Hispanic or Latino, MENA
33 White, Hispanic or Latino, NHPI
34 White, Black or African American, Asian
35 White, Black or African American, AIAN
36 White, Black or African American, MENA
37 White, Black or African American, NHPI
38 White, Asian, AIAN
39 White, Asian, MENA
40 White, Asian, NHPI
41 White, AIAN, MENA
42 White, AIAN, NHPI
43 White, MENA, NHPI
44 Hispanic or Latino, Black or African American, Asian
45 Hispanic or Latino, Black or African American, AIAN
46 Hispanic or Latino, Black or African American, MENA
47 Hispanic or Latino, Black or African American, NHPI
48 Hispanic or Latino, Asian, AIAN
49 Hispanic or Latino, Asian, MENA
50 Hispanic or Latino, Asian, NHPI
51 Hispanic or Latino, AIAN, MENA
52 Hispanic or Latino, AIAN, NHPI
53 Hispanic or Latino, MENA, NHPI
54 Black or African American, Asian, AIAN
55 Black or African American, Asian, MENA
56 Black or African American, Asian, NHPI
57 Black or African American, AIAN, MENA
58 Black or African American, AIAN, NHPI
59 Black or African American, MENA, NHPI
60 Asian, AIAN, MENA

61 Asian, AIAN, NHPI
62 Asian, MENA, NHPI
63 AIAN, MENA, NHPI
64 White, Hispanic or Latino, Black or African American, Asian
65 White, Hispanic or Latino, Black or African American, AIAN
66 White, Hispanic or Latino, Black or African American, MENA
67 White, Hispanic or Latino, Black or African American, NHPI
68 White, Hispanic or Latino, Asian, AIAN
69 White, Hispanic or Latino, Asian, MENA
70 White, Hispanic or Latino, Asian, NHPI
71 White, Hispanic or Latino, AIAN, MENA
72 White, Hispanic or Latino, AIAN, NHPI
73 White, Hispanic or Latino, MENA, NHPI
74 White, Black or African American, Asian, AIAN
75 White, Black or African American, Asian, MENA
76 White, Black or African American, Asian, NHPI
77 White, Black or African American, AIAN, MENA
78 White, Black or African American, AIAN, NHPI
79 White, Black or African American, MENA, NHPI
80 White, Asian, AIAN, MENA
81 White, Asian, AIAN, NHPI
82 White, Asian, MENA, NHPI
83 White, AIAN, MENA, NHPI
84 Hispanic or Latino, Black or African American, Asian, AIAN
85 Hispanic or Latino, Black or African American, Asian, MENA
86 Hispanic or Latino, Black or African American, Asian, NHPI
87 Hispanic or Latino, Black or African American, AIAN, MENA
88 Hispanic or Latino, Black or African American, AIAN, NHPI
89 Hispanic or Latino, Black or African American, MENA, NHPI
90 Hispanic or Latino, Asian, AIAN, MENA
91 Hispanic or Latino, Asian, AIAN, NHPI
92 Hispanic or Latino, Asian, MENA, NHPI
93 Hispanic or Latino, AIAN, MENA, NHPI
94 Black or African American, Asian, AIAN, MENA
95 Black or African American, Asian, AIAN, NHPI
96 Black or African American, Asian, MENA, NHPI
97 Black or African American, AIAN, MENA, NHPI
98 Asian, AIAN, MENA, NHPI
99 White, Hispanic or Latino, Black or African American, Asian, AIAN
100 White, Hispanic or Latino, Black or African American, Asian, MENA
101 White, Hispanic or Latino, Black or African American, Asian, NHPI
102 White, Hispanic or Latino, Black or African American, AIAN, MENA

103 White, Hispanic or Latino, Black or African American, AIAN, NHPI
104 White, Hispanic or Latino, Black or African American, MENA, NHPI
105 White, Hispanic or Latino, Asian, AIAN, MENA
106 White, Hispanic or Latino, Asian, AIAN, NHPI
107 White, Hispanic or Latino, Asian, MENA, NHPI
108 White, Hispanic or Latino, AIAN, MENA, NHPI
109 White, Black or African American, Asian, AIAN, MENA
110 White, Black or African American, Asian, AIAN, NHPI
111 White, Black or African American, Asian, MENA, NHPI
112 White, Black or African American, AIAN, MENA, NHPI
113 White, Asian, AIAN, MENA, NHPI
114 Hispanic or Latino, Black or African American, Asian, AIAN, MENA
115 Hispanic or Latino, Black or African American, Asian, AIAN, NHPI
116 Hispanic or Latino, Black or African American, Asian, MENA, NHPI
117 Hispanic or Latino, Black or African American, AIAN, MENA, NHPI
118 Hispanic or Latino, Asian, AIAN, MENA, NHPI
119 Black or African American, Asian, AIAN, MENA, NHPI
120 White, Hispanic or Latino, Black or African American, Asian, AIAN, MENA
121 White, Hispanic or Latino, Black or African American, Asian, AIAN, NHPI
122 White, Hispanic or Latino, Black or African American, Asian, MENA, NHPI
123 White, Hispanic or Latino, Black or African American, AIAN, MENA, NHPI
124 White, Hispanic or Latino, Asian, AIAN, MENA, NHPI
125 White, Black or African American, Asian, AIAN, MENA, NHPI
126 Hispanic or Latino, Black or African American, Asian, AIAN, MENA, NHPI
127 White, Hispanic or Latino, Black or African American, Asian, AIAN, MENA, NHPI