Guidelines for Using Racial and Ethnic Groupings in Data Analyses

Revision Date: April 30, 2010
Primary Contact: Juliet VanEenwyk, Ph.D., State Epidemiologist for Non-Infectious Conditions

Acronyms

ACS: American Community Survey
AIAN: American Indian or Alaska Native
API: Asian or Pacific Islander
BRFSS: Behavioral Risk Factor Surveillance System
CDC: Centers for Disease Control and Prevention
NCHS: National Center for Health Statistics
NHOPI: Native Hawaiian or Other Pacific Islander
OMB: Office of Management and Budget
WSCR: Washington State Cancer Registry

Purpose

The Assessment Operations Group in the Washington State Department of Health coordinates the development of guidelines related to data collection, analysis and dissemination in order to promote good professional practice among staff involved in assessment activities within the Washington State Department of Health and in Local Health Jurisdictions in Washington. While

the guidelines are intended for audiences of differing levels of training, they assume a basic knowledge of epidemiology and biostatistics. They are not intended to recreate basic texts and other sources of information related to the topics covered by the guidelines, but rather they focus on issues commonly encountered in public health practice and where applicable, on issues unique to Washington State.

Background

What Are Race and Ethnicity?

Concepts of race and ethnicity have changed considerably over time. Today, most scientists do not view race as a valid biological construct. Genetic changes tend to show gradual variation across geographic areas with social and cultural categories of race and ethnic group being only modest proxies for continental ancestry.[1,2,3] Researchers, such as Camara Jones, propose that "race is only a rough proxy for socioeconomic status, culture, and genes, but it precisely captures the social classification of people in a race-conscious society such as the United States. … That is, the variable 'race' is not a biological construct that reflects innate differences, but a social construct that precisely captures the impacts of racism."[4(p1212)] The meanings attributed to ethnicity also vary with some researchers emphasizing cultural heritage, while others emphasize social identity.[5] As with racism, to the extent that socio-cultural contexts maintain disadvantage among members of specific ethnic groups, ethnicity also captures the experience of discrimination.

Use of Terms in This Guideline

Because race and ethnicity are not precisely defined constructs, scientific writers use a variety of approaches for referring to these terms. For example, authors sometimes place quotation marks around race and ethnicity to remind readers of the imprecision and heterogeneity within categories. Others discuss race and ethnicity as one construct capturing the ambiguity of these terms, while emphasizing their social and cultural underpinnings. The imprecision of the constructs themselves can be magnified by data collection irregularities, such as when classification is based on appearance rather than self-report.

Race and ethnicity in this guideline refer to imprecise social and cultural categories with which individuals identify or as reported by next-of-kin in the case of death records. The term "race" indicates one of the five categories specified in the United States Office of Management and Budget (OMB) 1997 Standard[6] and "ethnicity" indicates Hispanic or non-Hispanic origin. "Subpopulation" indicates a grouping within a larger racial category, such as Chinese or Japanese within the Asian category. For the public health technical and professional audiences for whom this guideline is intended, we do not use quotation marks around these terms. We follow the conventions of the Chicago Manual of Style 15th Edition[7] in capitalizing designations based on national and ethnic groups, but not capitalizing those based loosely on color when we use these terms in the text. The conventions used in this guideline are not intended as a recommendation. The best terminology or ways of presenting these constructs will vary with the document's purpose and intended audience.

Why Include Analyses by Racial and Ethnic Groupings in Public Health Assessment?

In the United States and in Washington State, there are large differences in health status by self-identified racial and ethnic categories. Reducing these disparities is both a national and a state goal. We need to measure health status and associated risk factors by racial and ethnic groupings so that we understand the magnitude of the disparities and whether current gaps are increasing or decreasing. This knowledge can assist with developing interventions to decrease gaps, such as developing policies to reduce inequitable access to educational, economic and community resources that facilitate healthy ways of living; inequitable access to and quality of medical care; and inequitable exposure to environmental toxins. Additionally, health care providers and other service organizations sometimes serve people who primarily identify with one or a limited number of specific racial or ethnic categories. These providers and organizations

often want to know the health status of the populations they serve. Assessing data by racial and ethnic groupings is one way to obtain this perspective.

What Racial and Ethnic Groupings Are Used Nationally?

In the early 1990s, OMB reviewed Statistical Policy Directive No. 15, the federal guidelines for reporting race and ethnicity that had been in effect since 1977. Based on that review, OMB issued a revised standard in 1997. The standard included an explicit statement that the racial and ethnic categories serve social, cultural and political purposes and should not be interpreted as indicating primarily biological or genetic differences among people. The three major changes in the OMB 1997 Standard are

- People can identify ***more than one*** racial category.
- Pacific Islanders should ***not be classified*** with Asians.
- The question on Hispanic/Latino ethnicity should be asked ***before*** the race question.

The minimum categories established in the OMB 1997 Standard are

- **Race**
  - **American Indian or Alaska Native (AIAN):** A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.
  - **Asian:** A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent, including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand and Vietnam.
  - **Black or African American:** A person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American."
  - **Native Hawaiian or Other Pacific Islander (NHOPI):** A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
  - **White**: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.
- **Ethnic Group**
  - **Hispanic or Latino:** A person of Cuban, Mexican, Puerto Rican, South or Central American or other Spanish culture or origin, regardless of race. The term "Spanish origin" can be used in addition to "Hispanic or Latino."

Guidelines

General Considerations

These recommendations are based on articles by Kaplan MS and Bennett T[8] and Kahn J.[9]

- Consider the potential health or scientific benefit to providing data by racial and ethnic groupings; articulate the goal of such an analysis in reports and presentations.
- Specify how race and ethnicity were collected, including what categories were used; provide a rationale for the categories.
- Note the imprecision of the racial and ethnic categories, what race and ethnicity are intended to reflect in the specific analysis or document, and conventions for using specific terms. Include an explicit statement of the socio-cultural basis of racial and ethnic groupings in public health data.
- Assess potential bias by racial and ethnic groupings due to non-representative sampling, missing data, or other anomalies; use the bias analysis to determine how to best present and interpret data. Appendix 1 provides **an example** of such an analysis – as well as dataset-specific guidelines – for the Washington State Behavioral Risk Factor Surveillance System (BRFSS).

- In discussing differences among racial and ethnic groupings
  - Do not use racial and ethnic categories as proxies for genetic variation.
  - Avoid discussing differences as due to inherent underlying traits without clear evidence of such. Distinguish between racial and ethnic groupings as risk factors that cause disease and risk markers that are associated with disease but not causal. In most public health assessment, racial and ethnic groupings are risk markers. For example, one hypothesis for the persistently higher rates of preterm delivery among black compared to white women in the United States is that black women are exposed to stressors that affect birth outcomes. In this scenario, race is a marker for these stressors, but not, in itself, a risk factor for preterm delivery.[10]
  - Consider all conceptually relevant factors, such as socioeconomic factors and racism, specifically acknowledging contexts that maintain socioeconomic disadvantage or result in biological differences.
- Use caution when statistically controlling for socioeconomic factors in analyses by racial or ethnic groupings. To the extent that socioeconomic disadvantage is caused by racial or ethnic classification, socioeconomic factors mediate the relationship between these classifications and health outcomes. Statistical adjustment for intermediaries can result in biased estimates of the main effects.[11] As with socioeconomic factors, biological differences across racial categories can arise in response to environmental factors, such as exposure to harmful physical environments or experiences of racism.[12] To the extent that biological factors mediate relationships between racial or ethnic classification and health, statistically controlling for these factors is also likely to result in inaccurate estimates of the relationships between racial or ethnic category and health. Stratified analyses depicting the associations of racial and ethnic groupings, socioeconomic factors, and health might provide a more accurate picture of these relationships than analyses that "control" for one of these factors to understand the effect of the other. The Robert Wood Johnson brief, "Race and Socioeconomic Factors"[13] illustrates this approach. The small numbers of observations in some strata, however, make this approach difficult in Washington.

Data Collection

Unless otherwise indicated, the following recommendations are consistent with the OMB 1997 Standard and the Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity.[6,14]

- Collect data separately for race and Hispanic ethnicity, with Hispanic ethnicity collected first.
- For Hispanic ethnicity, allow the respondent to select only one option.
- To help provide more complete race data for Hispanic respondents, consider including, especially for self-administered questionnaires, an instruction to answer both the Hispanic ethnicity question and the race question.
- Allow respondents to select more than one racial grouping.
  - Allow selection of multiple response options within a single question rather than including a "multiracial" response option. The Provisional Guidance recommends that the question wording include "Mark one or more...," "Select one or more....," or "Choose one or more."
  - Consider the feasibility and usefulness of collecting information on a single racial grouping for respondents initially reporting more than one race. This approach might be useful for comparing current data to data collected under the old federal standard that allowed reporting of one racial category only. This information can also be used to assign respondents to a single category when data on multiple racial groupings cannot be used due to bias, small numbers, lack of denominators for developing rates, or other constraints. The OMB 1997 Standard does not recommend this approach. The Provisional Guidance suggests using this approach experimentally to provide survey-specific information for assessing changes over time. Appendix 2 provides information on barriers to using multiple race data in Washington, as well as

limitations to collecting data on a single racial grouping for those initially reporting more than one race.

- Carefully consider the terminology for specifying racial and ethnic categories and the order in which categories are presented. Unless there are compelling reasons for doing otherwise
  - Use the terminology in the OMB 1997 Standard.
  - Order the categories alphabetically.

  The Provisional Guidance explicitly states, "There is **not** 'one right way' to ask an individual to report his/her race and ethnicity. Rather, question wording and format should depend on the mode of administration as well as the context in which the questions are being asked. … Also OMB does not recommend a particular order of categories. There are advantages and disadvantages to various approaches, such as an alphabetical ordering versus the ordering of the most prevalent group followed by groups less prevalent."[14(p19)]
- Minimally, include the five racial categories specified by the OMB 1997 Standard and provide definitions for these categories when possible.
  - Include an "other" category with an open-ended request to specify. The OMB 1997 Standard does not recommend this category, but authorizes its use on the U.S. Census. Including an "other" category and asking people to specify their race can facilitate assigning records to one of the five standard categories and understanding the perspective of respondents who do not identify with any of those groupings.
- Collect subpopulation detail when there are sufficient numbers of events or participants to provide useable data. Subpopulation categories should be
  - Able to be aggregated into the minimum categories established in the OMB 1997 Standard.
  - Mutually exclusive. (Not explicitly stated by OMB.)
  - Consistent with available denominator data if one intends to calculate rates. (Not addressed by OMB.)
  - Meaningful to the populations about whom data are being collected. If possible, involve affected communities in developing categories. (Not addressed by OMB.)
  - Collected through check boxes, flashcards, or categories provided by an interviewer, allowing for an "other" option. Approaches included in the OMB Provisional Guidance include
    - Selecting from a list that includes subpopulations, but not the larger groups to which the subpopulations belong. With this method, the data analyst assigns individuals to the larger categories if needed for the analysis.
    - Follow-up questions for those who report a racial category for which additional detail is collected. The Provisional Guidance provides an example of this approach for Asians and Pacific Islanders. In the example, the subpopulation response options for Asians and those for Pacific Islanders are mutually exclusive. For example, Filipino is offered as an option for those reporting Asian, but not for those reporting Pacific Islander. The Provisional Guidance does not provide recommendations for classifying respondents who answer "other Asian" and then specify a subpopulation assigned to the Pacific Islander group and vice versa. For example, they do not recommend how to classify a respondent who first reports "Pacific Islander," then reports "other Pacific Islander" and then specifies "Filipino." (See Appendix 1, Asian and Pacific Islander Subpopulations, second bullet for Washington State Department of Health recommendations.)
- Consider the mode of administration when designing questions and instructions. The Provisional Guidelines suggest several approaches including
  - Showing respondents a flashcard with the categories for **face-to-face** surveys and having definitions available to the interviewer.
  - Using a check box format for **self-administered** forms and including definitions for the minimum racial categories if space allows.

    - Including the minimum number of response categories for **telephone** surveys and using follow-up questions to provide more detail. The manner in which the interviewer reads response options is important. To avoid confusion, the interviewer needs to pause between categories, e.g., white (*pause*) black or African American (*pause*) so that the respondents do not think they have to choose between black and African American. Standard definitions should be available to the interviewer.
- Use self-identification rather than observer identification. If self-identification is not possible (e.g., for a deceased person), obtain proxy responses from family or friends. Do not use observer identification.
    - OMB emphasizes that self-identification is the preferred method for collecting data on racial and ethnic categories, but allows observer identification if self or proxy reporting is not feasible. The Washington State Department of Health recommends recording race and ethnicity as missing if self or proxy reporting is not possible. This recommendation is based on large differences between how respondents classify themselves and how they report others usually classify them. In the 2004 Washington State BRFSS, only about half of respondents reporting AIAN also reported that others usually classified them as AIAN; among Hispanics, the comparable figure was about 70%.
- Use translated data collection forms to ensure inclusion of people from diverse backgrounds whenever possible. OMB does not address issues of language or provide advice on translation. Working with affected communities for translations that reflect local or regional dialects helps assure culturally appropriate translation. When collecting Washington State data by racial and ethnic groupings, forms in multiple languages can reduce bias that might result from variation in English-language proficiency across groupings. Appendix 1 discusses general potential bias by racial and ethnic groupings due to language barriers with additional detail for Hispanic Ethnicity and Asian and Pacific Islander Subpopulations .

Data Tabulation and Presentation

General Considerations

Assess the feasibility, reliability and validity of tabulating and presenting data by the racial and ethnic groupings available in the dataset. Do not present data that are not valid and reliable.

Appendix 3 provides detail on the numbers of respondents or events by racial and ethnic categories for selected Washington State Department of Health datasets. Although the **numbers** seem to be sufficient to support analysis for Hispanics and the OMB minimum racial groupings, as well as for selected multiple racial groupings and subpopulation detail, the ability to use these data may be limited due to

- Lack of denominators needed to calculate rates especially for records with more than one racial category and for Asian and NHOPI subpopulations.
- Small numbers of specific events, such as specific causes of death or specific types of cancer, affecting the reliability of estimates. (See Small Numbers Guideline.)
- Bias due to missing data, underreporting of some racial and ethnic groupings, non-representative samples, or other data anomalies that affect validity. For example,
    - Missing data on the birth certificate can vary by hospital, disproportionately affecting selected racial or ethnic categories.
    - Records with more than one racial grouping are underrepresented in some Washington State datasets potentially resulting in biased estimates. Appendix 2, More Than One Race in Washington State provides additional detail on underreporting of more than one race.
    - Washington residents who do not speak English or Spanish cannot participate in BRFSS, creating potential bias in estimates for groupings with large proportions of adults who are not proficient in these languages. (See, for example, Appendix 1, Asian and Pacific Islander Subpopulations.)

Recommended Racial and Ethnic Groupings

- Unless there are compelling reasons to do otherwise, **tabulate and display** data for Hispanics and the OMB minimum racial categories among non-Hispanics as follows:
  - American Indian or Alaska Native, non-Hispanic
  - Asian, non-Hispanic
  - Black or African American, non-Hispanic
  - Hispanic
  - Native Hawaiian or Other Pacific Islander, non-Hispanic
  - White, non-Hispanic

  The OMB Provisional Guidance states "an Hispanic or Latino respondent reporting one race should be reported both as Hispanic or Latino and as a member of that single race."[14(p12)] In 2005, however, users of Hispanic health data in Washington recommended that the Washington State Department of Health use the groupings specified above. The Centers for Disease Control and Prevention also reports data with Hispanics as one grouping and other groupings that include non-Hispanics only.

  NOTE: Data analysts in Washington State cannot currently use these recommended groupings if they need census-based population counts to develop rates. (See Alternative Racial and Ethnic Groupings for recommendations and a discussion of related issues.)
- To develop data for the categories listed above,
  - First, classify records as Hispanic ethnicity, non-Hispanic ethnicity, or unknown.
  - Second, classify records with non-Hispanic or unknown ethnicity into racial groupings.
    - This approach assumes that respondents who complete the race question without completing the Hispanic ethnicity question identify more strongly with their racial grouping than with their Hispanic or non-Hispanic ethnicity. If the proportion of records with unknown Hispanic ethnicity is relatively large, the data analyst needs to check the validity of this assumption. In 2007 Washington State data, Hispanic or non-Hispanic ethnicity was missing for the mother on 2% of birth certificates and for less than 1% of BRFSS records and death certificates. Thus, potential misclassification of these records is not likely to have a substantive impact on health statistics presented by racial and ethnic groupings.
    - Because the recommended groupings do not include categories for people who report more than one race, the data analyst must assign people who report more than one racial grouping into one category, a process referred to as bridging. Appendix 2 discusses issues related to and methods for bridging. If bridged data are not available and the data analyst cannot create a bridged dataset, use the alternative groupings described below.
- When displaying data for the recommended racial and ethnic groupings,
  - List groups in alphabetical order, unless there are compelling reasons to do otherwise.
  - Note that the racial groupings include those of non-Hispanic ethnicity only.
- Do not combine recommended racial and ethnic groupings, if such aggregations do not result in meaningful categories. For example, "nonwhite," "more than one race," and "other," are not usually meaningful for interpreting health-related data. (Appendix 2, More than One Race in Washington State provides information on why a multi-race category is not meaningful in Washington.) The data analyst might, however, include numbers, but not health statistics, for groupings that are not meaningful from a health perspective to account for the total numbers of respondents. Three such groups include:
  - **Other** for responses that do not match any of the standard categories.
  - **Not reported** for records that are missing information. If data are available, this category can be subdivided according to the reason that information was not obtained, e.g., refusal, don't know, and not ascertained.

    - **Not tabulated above** to aggregate responses for any categories that do not contain enough people to be published separately because of data quality or confidentiality concerns. (See Small Numbers Guideline.)
- Include as much additional detail on racial and ethnic groupings as possible without compromising data quality or confidentiality. More detail can usually be published for larger groups and population totals than for subgroups and attributes (e.g., income, education, or health outcomes). (See Data Tabulation and Presentation, General Considerations for a discussion of data quality and Small Numbers Guideline for a discussion of confidentiality issues.)
    - If some groups are excluded, explain why. Common reasons include small numbers, lack of denominator data needed to calculate rates, and bias.
    - Only aggregate groups to increase numbers if such an aggregation is meaningful.
    - If presenting data for multiple racial groupings, note that the other groupings include those who report one race only. (See Alternative Racial and Ethnic Groupings for an example of such labeling.)
    - If respondents selected from a list that included subpopulations, but not the larger racial groupings, assign records to the standard racial categories following the conventions of the U.S. Census. (See Appendix 1, Asian and Pacific Islander Subpopulations, second bullet for recommendations for classifying individuals in instances where race and subpopulation responses are not consistent with the U.S. Census designations.)

Alternative Racial and Ethnic Groupings

Use the following racial and ethnic groupings for **tabulating and displaying data** from datasets needing census-based population counts for denominators.

- American Indian or Alaska Native, single race only, non-Hispanic
- Asian, single race only, non-Hispanic
- Black or African American, single race only, non-Hispanic
- Hispanic
- Native Hawaiian or Other Pacific Islander, single race only, non-Hispanic
- White, single race only, non-Hispanic

This alternative is needed because census-based population estimates, such as those developed by the U.S. Census, the Washington State Office of Financial Management and the Washington State Department of Health, do not match the recommended racial and ethnic groupings. Census-based population estimates in Washington are available in two formats.

- One format provides bridged counts that classify all Washingtonians into one of four racial groupings that correspond to OMB's minimum racial categories **except** that estimates for Asians and NHOPIs are grouped into an Asian or Pacific Islander (API) category.
- The second format estimates population counts for Asians and NHOPIs separately, but does not bridge from multiple to single racial category and does not provide estimates for specific multiple racial groupings. Rather, this format estimates the total number of people who identify with more than one racial category irrespective of the specific combination of categories.

Our recommendation uses population data from the second of the two formats and omits records with more than one reported racial category. We selected this option, because

- This option provides rates separately for the Asian and NHOPI categories. As illustrated in Appendix 4, there tend to be large differences in health status for Washington residents reporting Asian compared to those reporting NHOPI. These differences are generally larger than differences in rates for those reporting a single race only and rates developed using bridging, especially in datasets, such as the Death Certificate System, where bridging uses a regression method. (See Appendix 2, Bridging Methods.)

- More than one race is underreported in most Washington State datasets. (See Appendix 2, More Than One Race in Washington State.) If few records contain information on multiple racial categories, omitting these records will not have a substantive impact on health statistics.
- Using the first format presented above requires assigning records with more than one racial category to a single category. In the absence of probabilistic bridging methods or self-report of a single category for those initially reporting more than one race, this assignment might introduce substantive error into estimates for specific racial groupings. (See Appendix 2, Bridging Methods and Appendix 4.)

Time Trends

Changes in data collection methods can create discontinuities for assessing trends over time. Data analysts need to identify and assess all such changes to determine the validity of treating rates for racial and ethnic categories over time as a continuous series. For example, as described in Appendix 1, Hispanic Ethnicity, adding a Spanish language option to the Washington State BRFSS created discontinuities in some trends for those reporting Hispanic ethnicity. Approaches for indicating a break in a time series due to discontinuities include separate tables or graphs, clearly demarcated breaks in trend lines, footnotes, technical notes, and explanations within the text.

In addition to dataset-specific changes that affect the continuity of data over time, two standards in the OMB 1997 Standard create potential discontinuities for many datasets commonly used for public health assessment. These include: 1) creating separate groupings for Asians and NHOPIs and 2) allowing reporting of more than one racial grouping. If assessing time trends that include data collected before and after the implementation of the OMB 1997 Standard (adopted by most Washington State Department of Health datasets in the early 2000s), data analysts need to consider the effect of those changes.

- It is not always possible to separate Asians and NHOPIs in data collected before implementing the OMB 1997 Standard. In these instances, the data analyst must choose between beginning a new time series or combining Asians and NHOPIs into a single API group comparable to the grouping used under the old standard. If there are large differences in rates for Asians and NHOPIs, we recommend beginning a new time series with Asians and NHOPIs as two separate categories.
- The approach to analyzing and displaying time trends for datasets affected by the reporting of more than one racial category varies with the completeness of reporting multiple races and the ability to assign records with more than one racial category to a single grouping.
    - If there are few records that include more than one racial grouping, reporting of more than one race is not likely to have a substantive impact on rates. In these instances, the data analyst might be able to treat data collected before and after the change as a continuous series, excluding records with more than one racial category. The data analyst should document the reason for the exclusion and note the percent of excluded records.
    - If records with more than one racial category can be assigned to a single category comparable to the one the respondent would have selected previously, the data analyst can make such an assignment and treat the years before and after the change as a continuous series. Assignment to a single racial category is best accomplished either by asking respondents directly which racial grouping they would select if they could select only one or through bridging with a regression method. (See Data Collection, 4th bullet, sub-bullet 2 for asking about a single racial category and Appendix 2, Bridging Methods for a discussion of bridging methods.)
    - If reporting of more than one racial category is likely to substantively affect estimates of rates or other health statistics and records with more than one racial grouping cannot be assigned accurately to a single grouping, clearly indicate a break in the data series between data collected before and after implementation of multiple race reporting. Approaches for indicating a break include separate tables or graphs,

clearly demarcated breaks in trend lines, footnotes, technical notes, and explanations within the text.

EndNotes

[1] Bamshad M. Genetic influences on health: Does race matter? JAMA 2005; 294(8): 937-46.

[2] Serre D, Paabo S. Evidence for gradients of human genetic diversity within and among continents. Genomic Research 2004; 14:1679-85)

[3] Additional detail in Genetics for the human race. Nat Genet 2004;36(suppl 11):S1-60.]

[4] Jones CP. Levels of racism: a theoretical framework and a gardener's tale. AJPH 2000; 90:1212-1215.

[5] Whaley A. Ethnicity/race, ethics, and epidemiology. J Natl Med Assoc 2003; 95:736-42.

[6] Office of Management and Budget, Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity Federal Register Notice October 30, 1997 retrieved September 2009 from http://www.whitehouse.gov/omb/fedreg_1997standards/

[7] University of Chicago. The Chicago Manual of Style, 15th Edition. Chicago: University of Chicago Press, 2004.

[8] Kaplan MS, Bennett T. Use of race and ethnicity in biomedical publication. JAMA 2003; 289:2709-16.

[9] Kahn J. Genes, race, and population: avoiding a collision of categories. AJPH 2006; 96:1965-70.

[10] Blackmore CA, Ferré CD, Rowley DL, Hogue CJ, Gaiter J, Atrash H. (1993). Is race a risk factor or risk marker for preterm delivery? Ethn Dis 3:372-7.

[11] Kaufman JS and Cooper RS. (2001). Commentary: Considerations for the use of racial/ethnic classification in etiological research. Am J Epidemiol 154:291-298.

[12] Kreiger N (2003) Does racism harm health? Did child abuse exist before 1962? On explicit questions, critical science, and current controversies: an ecosocial perspective. AJPH, 93:194-199.

[13] Robert Wood Johnson Foundation. Issue Brief 5: Race and Socioeconomic Factors. April 2009.

[14] Office of Management and Budget. Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity. December 15, 2000, retrieved September 2009 from http://www.whitehouse.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/re_guidance2000update.pdf.

Appendix 1: Racial and Ethnic Groupings in Washington State Behavioral Risk Factor Surveillance System

What is the Behavioral Risk Factor Surveillance System (BRFSS)?

The Washington State Behavioral Risk Surveillance System (WA-BRFSS) is a telephone survey of a random sample of non-institutionalized Washington residents ages 18 and older. From 1987-2002, WA-BRFSS included English-speaking residents only. Beginning in 2003, WA-BRFSS introduced a Spanish language questionnaire. The survey includes questions on health risk behaviors, preventive practices, health care access and use, prevalence of selected diseases, and health-related knowledge and attitudes. It is supported by the Centers for Disease Control and Prevention (CDC) and includes required CDC "core" questions – asked by BRFSS programs in all 50 states – as well as state-added questions.

From 1987-2007, WA-BRFSS included residents living in households with landline telephones only. In 2006, WA-BRFSS piloted a cell phone component and in 2008 added this component to the survey. The cell phone survey includes a random sample of non-institutionalized Washington residents ages 18 and older who have cell phones and live in households with no landline telephones. For the cell phone survey, WA-BRFSS interviews randomly selected cell phone owners whose cell phones are registered in Washington. If the cell phone owner now resides in another state, CDC transfers the respondent to the correct state when compiling and weighting the data. Likewise, CDC transfers cell phone respondents living in Washington and interviewed by other states to the Washington dataset. In 2008, 21 states included a BRFSS cell phone survey; in 2009, approximately 40 states participated. The cell phone survey asks a subset of questions, primarily those required by CDC. WA-BRFSS is awaiting CDC's initial weighting of 2008 cell phone data in order to integrate the cell phone and landline surveys. CDC anticipated providing these weights by August 2009, but has experienced delays. The cell phone survey will continue to the extent that resources allow. (See Living in households with no landline telephones for more information.)

Additional information on BRFSS is available at the WA-BRFSS and National BRFSS websites.

How are race and ethnicity collected in BRFSS?

Since its inception, the WA-BRFSS has used the CDC core questions to collect information on race and Hispanic ethnicity. The interviewer reads the questions and the racial or ethnic groupings that comprise the response options. The respondent can also reply "Don't know/not sure" or can refuse to answer the question, options that are not read by the interviewer. "Don't know/not sure" and refusals are generally treated as missing data.

The questions have changed over time, but CDC has maintained the original ordering of the response options, listing the largest group nationally first, followed by the next largest group. It is not clear why CDC originally selected this approach. One argument for maintaining the original order is for consistency over time. CDC requires WA-BRFSS to use their exact question and response options. The questions and time periods during which they were asked follow. From 1987-2000 one response was allowed for each question. From 2001 to the present, respondents are asked to name all racial groupings with which they identify.

- 1987-1992
  - What is your race, would you say
    1. White
    2. Black
    3. Asian, Pacific Islander
    4. Aleutian, Eskimo, Native American
    5. Some other (Specify)
  - Are you of Hispanic origin, such as Mexican American, Latin American, Puerto Rican, or Cuban?
    1. Yes
    2. No

- 1993-2000
  - What is your race? Would you say
    1. White
    2. Black
    3. Asian, Pacific Islander
    4. American Indian, Alaska Native
    5. Other (Specify)
  - Are you of Spanish or Hispanic origin?
    1. Yes
    2. No

- 2001-present:
  - Are you Hispanic or Latino (if female, Latina)?
    1. Yes
    2. No
  - Which one or more of the following would you say is your race? [CHECK ALL THAT APPLY.]
    1. White
    2. Black or African American
    3. Asian
    4. Native Hawaiian or Other Pacific Islander
    5. American Indian, Alaska Native
    6. Other (Specify)

The changes in 2001 reflect the implementation of the OMB 1997 Standard that asks the question on Hispanic ethnicity first and allows for reporting of more than one racial category. For respondents who report more than one race, the interviewer asks, “Which one of these groups would you say BEST represents your race?” The interviewer reads the same response options as in the original question. Beginning in 2004, WA-BRFSS included a state-added question that collects additional detail for respondents who report Asian or NHOPI either alone or in combination with another racial category. (See Asian and Pacific Islander Subpopulations.)

How well are Washington residents of different racial and ethnic groupings represented in the WA-BRFSS?

General Considerations

To assess whether WA-BRFSS collects a representative sample of Washington residents for every racial and ethnic group, we compared the percent distribution of WA-BRFSS respondents by race and Hispanic ethnicity to the percent distribution of adults in the American Communities Survey (ACS). (Table 1.1) Although similar response distributions in WA-BRFSS and the ACS do not guarantee a representative sample [1] dissimilar distributions are likely to indicate a non-representative sample especially for groups that are underrepresented.

Table 1.1. Percent of Washington residents ages 18 and older by racial and ethnic grouping based on the 2007 Behavioral Risk Factor Surveillance System (WA-BRFSS) and 2007 American Community Survey (ACS)

	WA-BRFSS design weights		WA-BRFSS design and post-stratification weights		ACS design weights	
Hispanic Ethnicity	Percent	Percent Margin of Error[a]	Percent	Percent Margin of Error[a]	Percent	Percent Margin of Error[a]
Hispanic	5.5	0.4	7.3	0.6	7.6	0.1
Non-Hispanic	94.5	0.4	92.7	0.6	92.4	0.1
Racial Category[b]						
AIAN[c]	1.4	0.2	1.6	0.2	1.3	0.1
Asian	2.8	0.3	3.0	0.3	6.9	0.1
Black	1.7	0.2	1.9	0.3	3.1	0.1
NHOPI[d]	0.4	0.1	0.6	0.2	0.4	0.03
White	88.5	0.5	86.7	0.7	82.6	0.2
Other single race	2.3	0.2	3.0	0.4	3.3	0.1
Two or more races	2.9	0.3	3.3	0.4	2.5	0.1

[a] Based on an alpha of .05; expressed as percent plus or minus.
[b] For comparison to ACS, racial categories include Hispanic and non-Hispanic respondents; except for the "two or more races" category, racial categories include respondents reporting a single race only.
[c] American Indian or Alaska Native
[d] Native Hawaiian or Other Pacific Islander

The U.S. Census Bureau conducts ACS under the authority of the United States code of law and response is mandatory. The Census Bureau conducts the survey in any language requested by the respondent and uses mail, telephone, and home visits to assure a representative sample. The Census Bureau reported a response rate of over 97% for 2005. With mandatory response and the subsequent high response rate, the ACS likely represents the most accurate picture of Washington's population distribution available for non-census years.

Table 1.1 provides percent distribution using both design and post-stratification weights for BRFSS and design weights (person and replicate) for ACS. Design weights adjust for unequal probabilities of being selected to participate in a survey. Post-stratification adjusts for over or underrepresentation of specific groups. Post-stratification does not eliminate bias if the respondents in a specific group differ from other members of that group. For example, as shown in Hispanic Ethnicity, before the 2003 introduction of the WA-BRFSS questionnaire in Spanish, Washington residents of Hispanic origin were substantially underrepresented in the survey. With the addition of the Spanish language option, it is apparent respondents of Hispanic origin before 2003 did not represent the larger group. That is, their patterns of risk factors and health status were different from those for the group as whole. Because Washingtonians of Hispanic origin who

took the survey prior to 2003 did not represent the larger group, post-stratification on Hispanic ethnicity prior to 2003 would have been likely to increase, rather than eliminate, bias.

Currently BRFSS post-stratification adjusts to the sex and age distribution of Washington residents. Because post-stratification cannot correct for bias, we compare design-weighted ACS to design-weighted BRFSS to assess over or underrepresentation. Due to the large sample sizes and subsequent very small margins of error, even small differences in Table 1.1 are statistically significant. Beyond the issue of statistical significance, however, is whether the proportions for specific racial and ethnic groupings are so low compared to those in ACS that the respondents are unlikely to represent the category as a whole. In this regard, the BRFSS proportions for the Asian, black and Hispanic categories are of concern with the proportion of respondents in the Asian grouping being particularly low. Within the Asian grouping, bias is likely to be exacerbated by a non-representative distribution of subpopulations. (See Asian and Pacific Islander Subpopulations.)

Underrepresentation in BRFSS most likely results from

- ***Cultural differences in willingness to participate.*** There is little information on which to assess cultural differences in willingness to respond to the BRFSS survey. Anecdotal information from public health personnel in Washington suggests that Washingtonians identifying with the AIAN grouping or with several of the Asian subpopulations might be less willing to share health and related information over the phone than are respondents who identify as white. While survey researchers express the need for culturally appropriate methods of obtaining data,[2,3] we were not able to locate systematic evidence identifying procedural barriers for specific cultures.
- ***Living in households with no landline telephones.*** Except for the NHOPI category, Table 1.2 shows a pattern of groupings that are underrepresented in the WA-BRFSS landline survey being better represented in the cell phone only survey and vice versa. (See What is the Behavioral Risk Factor Surveillance System for descriptions of the landline and cell phone surveys.) For example, the Asian, Hispanic, and black groupings, which are all are underrepresented in the landline WA-BRFSS, are better represented in the cell phone surveys.

Table 1.2. Percent of Washington Behavioral Risk Factor Surveillance System respondents in landline and cell phone surveys[4]

Racial[a] and Ethnic Grouping	Landline[b]	Cell Phone[b]
AIAN[c]	2%	2%
Asian	2%	6%
Black	1%	2%
Hispanic	8%	11%
NHOPI[d]	3%	1%
White	92%	82%

[a] Racial groupings include non-Hispanics only.
[b] Cell phone data from the 2006 cell phone pilot, 2008 cell phone survey and January-May 2009 cell phone survey combined, unweighted percentages, margins of error not available; landline data developed in the same manner using responses from the same time period.
[c] American Indian or Alaska Native
[d] Native Hawaiian or Other Pacific Islander

The percentages in Table 1.2 are not weighted for probability of selection into the sample and margins of error are not available. Thus, it is likely that some of the differences are not statistically significant. Nonetheless, these findings are similar to findings from national BRFSS data collected from 21 states, including Washington[5] and for respondents identifying as black or white in a survey conducted by the Pew Research Center for the

People and the Press.[6] These data suggest that disproportionate use of cell phones relative to landlines accounts for some of the underrepresentation of Washingtonians identifying as black, Hispanic, and especially Asian in the WA-BRFSS data that are currently available (i.e., the cell phone only data have not been integrated into the landline survey).

- ***Language barriers***. Combining U.S. Department of Labor estimates of English proficiency from the 2000 U.S. Census[8] with 2000 population counts, about 1.7% of Washington adults have limited proficiency in English or Spanish. Although the congruence of racial or ethnic category with language is not 100%, one can use the U.S. Census data to estimate the proportions of people by racial grouping who are not proficient in English under the assumption that most Washingtonians speaking an African language (not including Arabic) at home would identify with the black grouping, most speaking an Asian language with the Asian grouping, and so on. Using this assumption, between 17%-20% of Washington adults who identify as Asian are not proficient in English: they report speaking an Asian language[7] in the home and speaking English "not well" or "not at all." The range of the estimate depends on whether we include Asians reporting only Asian or Asian in combination with another racial category as the denominator for determining the percent. Based on those reporting one racial category only, the comparable figures are about 8% for Washington residents reporting NHOPI, less than 2% for residents reporting black, and less than 1% for residents reporting AIAN or white. Thus, at the state level, language poses larger barriers for those reporting Asian compared to other categories. Other groupings, however, may be affected at smaller geographic levels of analysis. For example, estimates for those identifying as black or white could be biased due to language barriers in small areas with large recent influxes of people from Africa or Eastern Europe, respectively. (See Hispanic Ethnicity for a discussion of language and bias that likely applies to all groups with language barriers.)

Hispanic Ethnicity

Since BRFSS' inception in 1987, there have been several changes to the Hispanic ethnicity question. (See How are race and ethnicity collected in BRFSS?) These changes have been relatively minor and likely did not have substantive impacts on the proportions of or findings for those reporting Hispanic ethnicity. For example, the proportion of respondents reporting Hispanic ethnicity remained stable before and after the 2001 change in the order of the race and Hispanic ethnicity questions. In each year of the three years before the change, about 5% (±1%) of WA-BRFSS respondents reported Hispanic ethnicity compared to 4%-5% (±1%) in each of the three years following the change.

In contrast, the addition of a Spanish language questionnaire in 2003 had a major impact on both the proportion of WA-BRFSS participants reporting Hispanic ethnicity and survey findings for this grouping. Following the introduction of the Spanish language questionnaire, the percent of respondents reporting Hispanic ethnicity increased by about 60%, from 5% (± 1%) in 2002 to 8% (± 1%) in 2003. About 3%-5% (±1%) of WA-BRFSS respondents reported Hispanic ethnicity each year from 1987-2002; this increased to 7%-8% (±1%) from 2003-2008.

In addition to the increase in the proportion of respondents reporting Hispanic ethnicity, rates for the Hispanic grouping changed for many variables after the introduction of the Spanish language questionnaire. These changes indicate that estimates before 2003 were not representative of the Hispanic ethnicity category as a whole. Some of these changes were large enough to affect the state rate.

- Health insurance coverage
    - Among those reporting Hispanic ethnicity, the percent reporting no health insurance increased from 20% (±6%) in 2002 to 42% (±4%) in 2003; these percentages ranged from 13%-29% (±6-8%) for each of the six years before the introduction of the Spanish language questionnaire and from 42%-47% (±4%) in the following six years.
    - For 2006-2008 combined, 68% (±3%) of WA-BRFSS respondents reporting Hispanic ethnicity who took the survey in Spanish reported not having health insurance

compared to 24% (±3%) who took the survey in English; the percent for those taking the English language survey is similar to the percent for the Hispanic grouping as a whole before the introduction of the Spanish language questionnaire.

- Including Spanish-speaking respondents who did not previously participate changed the statewide percent of those reporting no health insurance. In 2002, 12% (±1%) of all WA-BRFSS participants reported being uninsured compared to 14% (± 1%) in 2003; among those reporting non-Hispanic ethnicity, percentages did not change (12% ±1% in both 2002 and 2003).

- Smoking
  - From 2005-2008 combined, 2% (±1%) of females who reported Hispanic ethnicity and took the WA-BRFSS in Spanish reported current smoking compared to 15% (±3%) of females reporting Hispanic ethnicity and taking the survey in English. These differences were not apparent for males reporting Hispanic ethnicity: 19% (±4%) reported current smoking independent of language.
  - The very small proportion of smokers among females who reported Hispanic ethnicity and took the survey in Spanish lowered smoking rates for the Hispanic grouping as a whole. From 2005-2008 combined, 17% (±2) of respondents reporting Hispanic ethnicity who took the survey in English reported smoking compared to 14% (±2%) when including those who took the survey in Spanish.
  - Including a Spanish language questionnaire has not had a measurable effect on statewide smoking rates: from 2005-2008 combined, 17% (±0.4%) of Washington adults reported current smoking independent of whether those taking the survey in Spanish are included or excluded from the total.

The changes for the Hispanic grouping after the addition of the Spanish language questionnaire suggest that respondents reporting Hispanic ethnicity who participated in WA-BRFSS before the addition of this option did not represent the group as a whole. This is not surprising given 1) that 30% of adults who reported Hispanic ethnicity on the 2000 U.S. Census also reported speaking English not well or not at all[8] and 2) the relationship between English proficiency and many health factors. Although adding a Spanish language questionnaire does not guarantee representativeness of WA-BRFSS respondents reporting Hispanic ethnicity,[2] this option reduces one important source of bias.

Asian and Pacific Islander Subpopulations

Beginning in 2004, the Washington State Department of Health added additional detail for respondents who report Asian or NHOPI either alone or in combination with another race. Respondents are asked “Which one or more of the following best describes your Asian or Pacific Islander heritage?” The interviewer reads all response options irrespective of whether the participant reported Asian or NHOPI. Response options include: Native Hawaiian, Chinese, Japanese, Korean, Filipino, Vietnamese, Cambodian, Asian Indian, Samoan, Guamanian or Chamorro, Other. There are several issues to consider when using these data.

- Although WA-BRFSS allows reporting of more than one subpopulation, for some years, only one subpopulation is available in the database and it is not clear how that one subpopulation is selected. Beginning with 2007, up to three Asian or Pacific Islander subpopulations are available.

Table 1.3. Coding of Asian and Native Hawaiian or Other Pacific Islander (NHOPI) subpopulations in the Washington Behavioral Risk Factor Surveillance System

Year	Coding in Analytic File
2004	One Asian or NHOPI subpopulation.
2005	All options coded as yes/no
2006	One Asian or NHOPI subpopulation
2007 & later	Up to three Asian or NHOPI subpopulations.

The inconsistencies in coding would likely not substantively affect rates for subpopulations, because a relatively small proportion of respondents report more than one Asian or NHOPI subpopulation. For 2005 and 2007 combined, 4% (± 2%) of residents reporting Asian or NHOPI as their only race reported more than one subpopulation.

- With all response options available irrespective of whether a person reports Asian or NHOPI, not everyone classifies themselves in a manner consistent with state and federal conventions. For example, the U.S. Census and the Washington State birth and death certificate systems classify Filipinos as Asians. In the 2005-2007 WA-BRFSS, however, 30% (±11%) of Filipinos who reported one race only, reported NHOPI. Data analysts have several choices depending on their needs.
    - When looking only at the racial groupings and not subpopulations within those groupings, use the reported race irrespective of the subpopulation with which the respondent identifies. This method respects the preference of the respondent and provides consistency with national BRFSS data.
    - When comparing subpopulations to the larger groupings, classify respondents as Asian or NHOPI based on their subpopulation following the categories of the U.S. Census. This approach assigns all respondents reporting a specific subpopulation to the same racial grouping and assures that all respondents in a specific subpopulation are also in the larger comparison grouping.
    - When comparing subpopulations to each other rather than to the larger Asian or NHOPI groupings, use the reported subpopulation irrespective of reported race.

- The proportions of specific Asian subpopulations in WA-BRFSS are not the same as the proportions in the ACS. (See How well are Washington residents of different racial and ethnic groupings represented in the Washington BRFSS for a discussion of ACS and weighting.) As shown in Table 1.4, Asians reporting Filipino or Japanese are overrepresented, while those reporting Vietnamese or Korean are underrepresented.

  The issues discussed for underrepresentation in BRFSS including landline coverage, language barriers, and cultural willingness to participate, likely apply to Asian subpopulations. We have no information on willingness to participate or landline coverage by subpopulation. The 2000 U.S. Census provides information on English proficiency that can be used to evaluate potential language barriers.[8] (Table 1.5)

  Given the relatively large proportions of Washington residents reporting Vietnamese or Korean also reporting speaking English “not well” or “not at all,”, language barriers likely contribute to the underrepresentation of these groups in WA-BRFSS. In contrast, the high level of English proficiency among those reporting Filipino is consistent with an overrepresentation of this group. Large proportions of individuals who are not proficient in English can result in biased estimates, as illustrated by the addition of a Spanish language questionnaire in 2003. (See Hispanic Ethnicity.)

Table 1.4. Percent of Washington respondents ages 18 and older reporting specific Asian subpopulations: 2007 Behavioral Risk Factor Surveillance System (WA-BRFSS) and 2007 American Community Survey (ACS)[a]

Asian subpopulation	WA-BRFSS design weights		WA-BRFSS design and post-stratification weights		ACS design weights	
	Percent	Margin of Error[b]	Percent	Margin of Error[b]	Percent	Margin of Error[b]
Asian Indian	12.5	3.5	14.9	4.4	10.2	1.1
Cambodian	NP[c]		NP[c]		2.6	0.5
Chinese	23.3	4.4	23.7	4.8	22.0	1.5
Filipino	26.1	4.5	24.5	4.8	20.5	1.2
Japanese	14.6	3.4	11.1	2.8	8.2	0.6
Korean	8.6	2.8	9.3	3.5	14.4	1.2
Laotian[c]	NP[c]		NP[c]		2.6	0.7
Vietnamese	5.8	2.3	8.0	4.1	14.2	1.2
Other[d]	6.9	2.7	6.2	2.5	5.5	0.7

[a] Includes respondents reporting one race (Asian or NHOPI) only. For consistency with ACS, subpopulations include respondents independent of Hispanic ethnicity; fewer than 1% of Asians in Washington also report Hispanic ethnicity.
[b] Based on an alpha of .05; expressed as percent plus or minus.
[c] BRFSS data not presented due to fewer than 6 observations.
[d] People reporting Asian as their only race with other or unspecified Asian subpopulation or two or more Asian subpopulations.

Table 1.5. Estimated percent of Washington residents reporting speaking English "not well" or "not at all," ages 5 and older, 2000 U.S. Census[8]

Subpopulation	Estimated Percent
Asian Indian	1
Chinese	19
Filipino	5
Japanese	10
Korean	24
Vietnamese	31

Summary and Recommendations

All of the guidelines presented in Guidelines for Using Racial and Ethnic Groupings in Data Analysis apply to presenting WA-BRFSS data by racial and ethnic grouping. This appendix provides one approach for assessing the representativeness of these groupings. A number of findings suggest caution when presenting WA-BRFSS data by racial and ethnic categories. To help assure that those using WA-BRFSS information understand potential bias in the results, authors need to

- Clearly explain the ramifications of excluding Washington residents who do not
  - Speak English or Spanish. This is most important at the state level for the Asian and Asian subpopulations and might be important at smaller geographic levels for other groupings, as well.

- Live in households with landline telephones (although this might be changing; see What is the Behavioral Risk Factor Surveillance System). This exclusion might affect the Asian, black and Hispanic categories more than the other groupings.

- Assess other information to determine whether data from WA-BRFSS are consistent with information that is less subject to bias. Such information includes:
  - Data from other Washington State databases, such as the death certificate.
  - Published information from special studies.

  For example, relatively low death rates for major causes of death in the Asian grouping in Washington are consistent with low levels of health risk found on many factors measured in BRFSS. Thus, the magnitude of the bias in the WA-BRFSS estimates for the Asian category might not be large enough to obscure findings relative to other racial and ethnic groupings.

Understanding cultural differences in willingness to participate in government-sponsored telephone surveys would likely provide perspective on bias in WA-BRFSS by racial and ethnic grouping. This assessment would require special study developed and implemented in collaboration with stakeholders and leaders representing diverse racial and ethnic groupings. Such a study should include not only Washington residents identifying with the Asian, Hispanic and black groupings that we know are underrepresented in BRFSS, but might also include representatives from other groupings, such as AIAN, who have expressed concern about biased representation. Given the reliance of state and local agencies on BRFSS for policy development and program planning and the Washington State Department of Health's commitment to reducing disparities, developing this understanding and seeking alternative survey approaches, as indicated, could be extremely useful.

EndNotes

[1] Peytcheva E and Groves RM. Using variation in response rate of demographic subgroups as evidence of nonresponse bias in survey estimates. J Off Stat 2009; 25(2):193-201

[2] Link MW, Mokdad AH, Stackhouse HF, Flowers NT. Race, ethnicity and linguistic isolation as determinants of participation in public health surveillance surveys. [serial online] Jan 2006, available at http://www.cdc.gov/pcd/issues/2006/jan/05_0055.htm accessed January 2010.

[3] U.S. Department of Health and Human Services. Improving the collection and use of racial and ethnic data in HHS [Internet]. Washington DC: U.S. Department of Health and Human Services, HHS Data Council, December 1999, available at http://www.aspe.hhs.gov/datacncl/RaceRpt/index.htm accessed February 2010.

[4] Wynkoop Simmons K, Wintz JC. Conducting a Telephone Survey with "Cell Phone Only" Households. Washington State Department of Health and the Gilmore Research Group. Presented at the Puget Sound Research Forum June 18, 2009.

[5] Balluz L. Behavioral Risk Factor Surveillance System Challenges. Centers for Disease Control and Prevention. October 7, 2008, available at http://www.chronicdisease.org/files/public/AGM08_TUE_TheStateofBRFSS.pdf accessed February 5, 2010.

[6] Pew Research Center for the People and the Press. The Cell Phone Challenge to Survey Research. Summary of Findings. May 16, 2006, available at http://people-press.org/report/276/ accessed September 2009.

[7] Asian languages include Chinese, Gujarathi, Hindi, Japenese, Korean, Miao/Hmong, Mon-Khmer Cambodian, Other Asian, Other Indic, Tagalog, Tai, Urdu, Vietnamese.

[8] U.S. Department of Labor. The LEP Special Tabulation of Census 2000 Data on Limited English Proficient Adults, available at http://www.doleta.gov/reports/censusdata/introduction.cfm accessed July 2009

Appendix 2: Methods for Assigning Washington Residents Reporting More Than One Race to a Single Racial Category

More Than One Race in Washington State

In the 2000 U.S. Census, 3.6% of Washingtonians identified with two or more racial categories. Based on the 2007 American Community Survey (ACS), this percent has not changed. Younger Washingtonians are more likely to report (or have a proxy report) more than one racial category compared to older residents. (Table 2.1)

Table 2.1. Percent (point estimate) of Washingtonians reporting more than one racial category by age group, American Communities Survey 2005-2007

Age Group	Percent with more than one race	Age Group	Percent with more than one race
<5	8.3%	35 – 44	2.3%
5 – 9	7.0%	45 – 54	2.0%
10 – 14	6.3%	55 – 64	1.5%
15 – 19	5.7%	65 – 74	1.2%
20 – 24	4.2%	75 – 84	0.9%
25 – 29	3.3%	85+	0.5%
30 – 34	2.6%		

The percent reporting more than one racial category also varies by category. In the 2007 ACS, about 47% of people reporting American Indian or Alaska Native (AIAN) also reported one or more additional racial groupings; 40% of Native Hawaiians or other Pacific Islanders (NHOPI) also reported one or more additional categories, as did 24% of blacks, 17% of Asians, 9% of other, and 4% of whites. Of Washingtonians reporting more than one racial grouping, about one-quarter reported both AIAN and white, one-quarter reported Asian and white, and 20% reported black and white. No single combination predominates and it is unlikely that a “more than one race” category is meaningful in Washington.

Most Washington State Department of Health data systems currently allow respondents to select more than one racial grouping. (See Data Collection.) Data analysts face challenges, however, in providing information on multiple racial groupings when using these datasets for public health assessment. Common obstacles include:

- Small numbers in specific combinations of groupings. (See Appendix 3 for examples.)
- Lack of denominators for datasets in which rates – and especially age-adjusted rates – are needed. (See Alternative Racial and Ethnic Groupings.)
- Potential bias due to underreporting. For example, in contrast to the percentages in Table 2.1, for children under age 15, more than one race was recorded on fewer than 2% of 2004-2006 Washington State Cancer Registry (WSCR) records and fewer than 0.5% of 2005-2007 death certificates. Similarly, for ages 15-54, more than one racial category was included on fewer than 1% – and for most age groups fewer than 0.5% – of WSCR records and on fewer than 0.1% of death certificates.
- Discontinuities in assessing trend over time. (See Time Trends.)

These challenges often constrain data analysts to exclude records with more than one racial category or assign these records to a single racial grouping. The method for assigning to a single

race that best honors self-identification is to include a question on single race for those who initially report two or more categories. The analyst can use the self-reported single racial category if it is not feasible to use multiple categories. Collecting a single racial grouping for those initially reporting more than one category, however, is not always possible. Even when feasible, many people who initially report more than one racial category do not then report a single grouping.

Table 2.2. Sample sizes (n) and weighted percent distribution (with 95% confidence intervals) of single racial groupings selected by respondents initially reporting more than one race: 2006-2008 Washington State Behavioral Risk Factor Surveillance System (BRFSS) and 1997-2000 National Health Interview Survey

	2006-2008 Washington State BRFSS					
Multiple Racial Grouping		Single Racial Grouping				
	n	AIAN[a]	API[b]	Black	White	Don't Know/ Refused
AIAN/White	1358	15% (12-18)			75 %(71-78)	10% (8-13)
API/White	250		26% (19-34)		50% (42-58)	24% (18-32)
Black/White	87			45% (30-62)	34% (21-50)	21% (12-34)
	1997-2000 National Health Interview Survey[c]					
AIAN/White	1593	21%			74%	5%
API/White	1147		40%		41%	19%
Black/White	1138			45%	27%	28%

[a] American Indian or Alaska Native
[b] Asian and Pacific Islander, combined for comparability with Parker et al.
[c] Parker JD, Schenker N, Ingram DD, Weed JA, Heck KE, Madans JH. Bridging between two standards for collecting information on race and ethnicity: an application to Census 2000 and vital rates. Public Health Reports 2004; 119:192-204; confidence intervals not available

Bridging Methods

Bridging methods are sets of rules or statistical models that assign individuals reporting more than one race to one or more single racial categories. Bridging estimates are made only for records with more than one race under the assumption that records with a single race would have had the same race irrespective of whether respondents could select one or more than one racial category. Bridging methods can be **deterministic** (i.e., assigning records based on a set of rules external to the dataset) or **probabilistic** (i.e., assigning records according to a set of probabilities).

Commonly used **deterministic bridging** methods include:

- **Whole assignment:** assign an individual's responses to one racial grouping using the single racial grouping that
  - Has the smallest count (**smallest group**).
  - Has the largest count other than white (**largest group other than white**).
  - Has the largest count (**largest group**).
  - Is the grouping that other individuals having the same combination of races identify with most strongly (**plurality**).
- **Fractional assignment:** assign an individual's responses as fractions of multiple categories using
  - Equal fractions for each reported category (**equal fractions**).
  - The fractional proportion of individuals identifying most strongly with each racial category based on data (**preferred fractions**).

- **All inclusive:** assign an individual's responses to each racial category reported. In this case, the sum of the categories totals more than 100%.

Table 2.3 shows assignment of single racial categories using deterministic bridging methods for a Washington resident who reports both black and white.

Table 2.3. Deterministic assignment to a single racial category for a respondent reporting with both black and white racial groupings

Method	Number assigned to: Black	White
Whole Assignment		
Smallest group	1	
Largest group other than white	1	
Largest group		1
Plurality	1	
Fractional Assignment		
Equal fractions	0.5	0.5
Preferred fractions[a]	0.63	0.37
All Inclusive	1	1

[a] Based on national fractions in Parker et al.[1] reproduced in Table 2.4.

Probabilistic methods that assign some records with a specific combination of racial categories to a single category and other records with the same combination to a different category (or categories) include:

- **Hot deck imputation:** assign the racial grouping of the "nearest neighbor" within the dataset. The "nearest neighbor" has similar demographic characteristics to the record with more than one racial grouping and has a single racial grouping that is the same as one of the groupings in the multiple race record.
- **Regression:** assign an individual to a single racial category based on information about relationships between selected covariates and the category a respondent would have chosen under the system that allowed respondents to select one racial category only.

Because the regression method is based on data about preferences of people identifying with more than one racial grouping, it is likely the most accurate method for assigning a record to the racial category the respondent would have chosen under the system allowing selection of one category only. The regression method, however, is difficult to implement. Additionally, we know of no Washington data that are sufficiently robust for developing assignment algorithms. The Washington BRFSS collects data that, in theory, would enable development of regression algorithms, but barriers to using these data include:

- Large proportions of respondents who initially report more than one racial grouping do not then provide a single race. (See Table 2.2)
- Small numbers for most combinations of multiple racial categories make probability estimates for these groups unreliable. As shown in Table 2.2, the confidence intervals for the single racial groupings for participants reporting both black and white are large even though this category is the third largest multiple racial category in Washington.

The National Center for Health Statistics (NCHS) has developed regression algorithms based on the National Health Interview Survey. This survey asked people initially reporting more than one racial category to select a single grouping by asking, "Which one of these groups would you say BEST represents your race?" NCHS then developed models that best assigned people reporting

multiple races to the single category they had selected. Factors (covariates) in the regression models that are used in the current NCHS algorithms include age, Hispanic or non-Hispanic ethnicity, sex, region of the U.S., county-specific index of urbanicity, and county-specific racial composition.[1] The NCHS Procedures for Multiple-Race and Hispanic Origin Data: Collection, Coding, Editing, and Transmitting provides examples of implementing this model.[2] Of note is that in example 1, a person reporting Asian or Pacific Islander (API) and white categories has a 40% probability of being assigned to API and 60% to white; in example 3, the probabilities of a person reporting API and white are 88% and 12%, respectively. The large differences in probabilities in these examples illustrate the importance of the covariates. (NCHS does not separate Asians and NHOPIs for bridging, because the previous standard combined these groups.)

Recommendations for Assigning Multiple to Single Racial Groupings

- Use bridged data developed using a regression method if possible. In most instances this is only possible when these data are provided by an external organization, such as NCHS providing bridged birth and death certificate data.
- Use all other bridging methods with caution.
- If using fractional assignment, use the national fractions from Table 4 in Parker et al. reproduced in Table 2.4.[1] In the absence of local information, the national fractions are likely to be reasonable for Washington given the correspondence of the National Health Interview Survey and Washington BRFSS fractions shown in Table 2.2. As with the NCHS regression algorithms, these fractions assign Asians and NHOPIs to one API grouping because the API grouping was the standard under the old system of data collection.

Table 2.4. Percent distribution of single-race assignment after application of the NHIS[b]-regression method to bridge multiple-race counts to single-race categories: public-use Census Modified Race Summary file, United States, 2000[a]

	Single-race assignment			
Multiple-race response	*AIAN*[b]	*API*[c]	*Black*	*White*
AIAN/API	63.3	36.7	—	—
AIAN/black	15.9	—	84.1	—
AIAN/white	22.4	—	—	77.6
API/black	—	41.4	58.6	—
API/white	—	40.9	—	59.1
Black/white	—	—	62.9	37.1
AIAN/API/black	26.8	25.4	47.8	—
AIAN/API/white	2.2	8.7	—	89.1
AIAN/black/white	18.7	—	57.4	23.9
API/black/white	—	12.0	11.9	76.1
AIAN/API/black/white	0.9	1.0	2.1	95.9

[a] Reproduced from Parker JD, Schenker N, Ingram DD, Weed JA, Heck KE, Madans JH. Bridging between two standards for collecting information on race and ethnicity: an application to Census 2000 and vital rates. Public Health Reports 2004; 119:192-204. Available at http://wonder.cdc.gov/wonder/help/populations/bridged-race/PublicHealthReports119-2-p192.pdf; NHIS: National Health Interview Survey
[b] AIAN: American Indian or Alaska Native
[c] API: Asian or Pacific Islander

EndNotes

[1] Parker JD, Schenker N, Ingram DD, Weed JA, Heck KE, Madans JH. Bridging between two standards for collecting information on race and ethnicity: an application to Census 2000 and vital rates. Public Health Reports 2004; 119:192-204. Available at http://wonder.cdc.gov/wonder/help/populations/bridged-race/PublicHealthReports119-2-p192.pdf .

[2] Division of Vital Statistics, National Center for Health Statistics. Centers for Disease Control and Prevention. NCHS Procedures for Multiple-Race and Hispanic Origin Data: Collection, Coding, Editing, and Transmitting. May 7, 2004, retrieved March 9, 2010 from http://www.cdc.gov/nchs/data/dvs/Multiple_race_docu_5-10-04.pdf

Appendix 3: Numbers of Events or Respondents among Washington Residents by Race and Ethnicity in Selected Washington State Department of Health Datasets

	2007 Deaths	2007 Births[1]	2005-2007 BRFSS[2]
TOTAL	**47,115**	**88,921**	**72,943**
Hispanic	**1,098**	**16,839**	**3,954**
Single race, non-Hispanic			
AIAN[3]	**596**	**1,559**	**907**
Asian	**1,351[4]**	**7,225**	**1,415**
Asian Indian	122	1,358	175
Chinese	233	1,167	302
Filipino	282	1,244	333
Japanese	260	440	275
Korean	190	737	128
Cambodian	NA[5]	NA	19
Laotian	NA	NA	19
Vietnamese	120	1,140	71
Other Asian	152	1,139	93
Black	**1,220**	**3,527**	**935**
NHOPI[6]	**134[7]**	**900**	**127**
Hawaiian	23	78	55
Guamanian	31	173	17
Samoan	42	324	39
Other Pacific Islander	41	325	16
White	**42,444**	**55,833**	**62,956**
Other			178
Two races, non-Hispanic			
AIAN, Asian	4	26	13
AIAN, Black	7	60	40
AIAN, NHOPI	0	12	5
AIAN, White	116	744	1,370
Asian, Black	4	63	8
Asian, NHOPI	6	95	45
Asian, White	37	766	172
Black, NHOPI	0	23	4
Black, White	33	526	85
NHOPI, White	8	130	60
Other	0	0	28
Three or more races, non-Hispanic			
Total	7	235	96
White, Black, AIAN	5	89	
White, Black, Asian	0	25	
White, Black, NHOPI	0	2	
White, AIAN, Asian	2	24	
White, AIAN, NHOPI	0	12	
White, Asian, NHOPI	0	63	
Black, AIAN, Asian	0	4	
Black, AIAN, NHOPI	0	0	
Black, Asian, NHOPI	0	0	
AIAN, Asian, NHOPI	0	2	
Four or more races, non-Hispanic			
Total	0	14	
Unknown, non-Hispanic			
Total	50	358	545

[1] Births are classified according to mother's race and ethnic group.
[2] Behavioral Risk Factor Surveillance System
[3] American Indian/Alaska Native
[4] Subpopulations include 8 people selecting more than 1 Asian ethnicity.
[5] Not available
[6] Native Hawaiian or Other Pacific Islander
[7] Subpopulations include 3 people selecting more than 1 NHOPI ethnicity.

Appendix 4: Comparison of Rates Using Single Race Only and Bridged Files

Background

As described in Alternative Racial and Ethnic Groupings, it is not currently possible to use the recommended racial and ethnic groupings for rates of health events that require census-based denominators. Of the two options available for alternate groupings, we recommend using single race only and excluding records with more than one racial category.

The other approach, which we do not recommend, is to use a file that assigns records with more than one racial category to a single grouping using a process called "bridging." While this option uses all available records, the census-based population denominators in Washington State that are consistent with this approach currently do not separate the Asian and NHOPI categories. In addition to this limitation, the assignment of records with more than one racial category to a single category can be problematic. (See Appendix 2.)

Comparison of Rates by Racial and Ethnic Groupings Using Alternative Population Estimates

Figures 4.1 and 4.2 present age-adjusted mortality and cancer incidence rates, respectively. Each chart provides rates developed using groupings that correspond to those in the two currently available census-based population estimates. The numbers of events used to calculate some of the rates are relatively small, as indicated by a wide confidence interval. These rates would likely not be presented in a health assessment. They are presented here not as part of a health assessment, but rather to illustrate differences in alternative approaches to developing rates. The charts illustrate three phenomena.

- Rates for the combined Asian and NHOPI grouping (formerly called "Asian or Pacific Islander") are similar to the rates for the Asian category, but not the NHOPI category. In Washington about 17 times more people classify themselves as Asian than as NHOPI. The NHOPI category is effectively "lost" when combining the NHOPI and Asian groupings.
- For the most part, rates for the NHOPI grouping are higher than rates for the Asian grouping. The differences are particularly evident for the death data where rates for the NHOPI grouping are statistically significantly higher than those for the Asian category for four of the six causes of death presented. For cancer incidence, rates for the NHOPI grouping are statistically significantly higher than those for the Asian grouping for all causes of cancer and for prostate cancer.
- For the most part, rates for a specific racial grouping are similar irrespective of whether records with more than one racial category are assigned to a single category or excluded. The size of these differences, however, varies with bridging method. (See Appendix 2 for a discussion of bridging and recommended methods.)

  The National Center for Health Statistics uses a probabilistic regression method for bridging. Probabilistic regression is the method that is most likely to assign records with more than one racial category to the category the decedent's next of kin would have selected had they been asked to list one racial category only. With this method of bridging, the choice of groupings based on availability of population estimates does not affect interpretation of the findings.

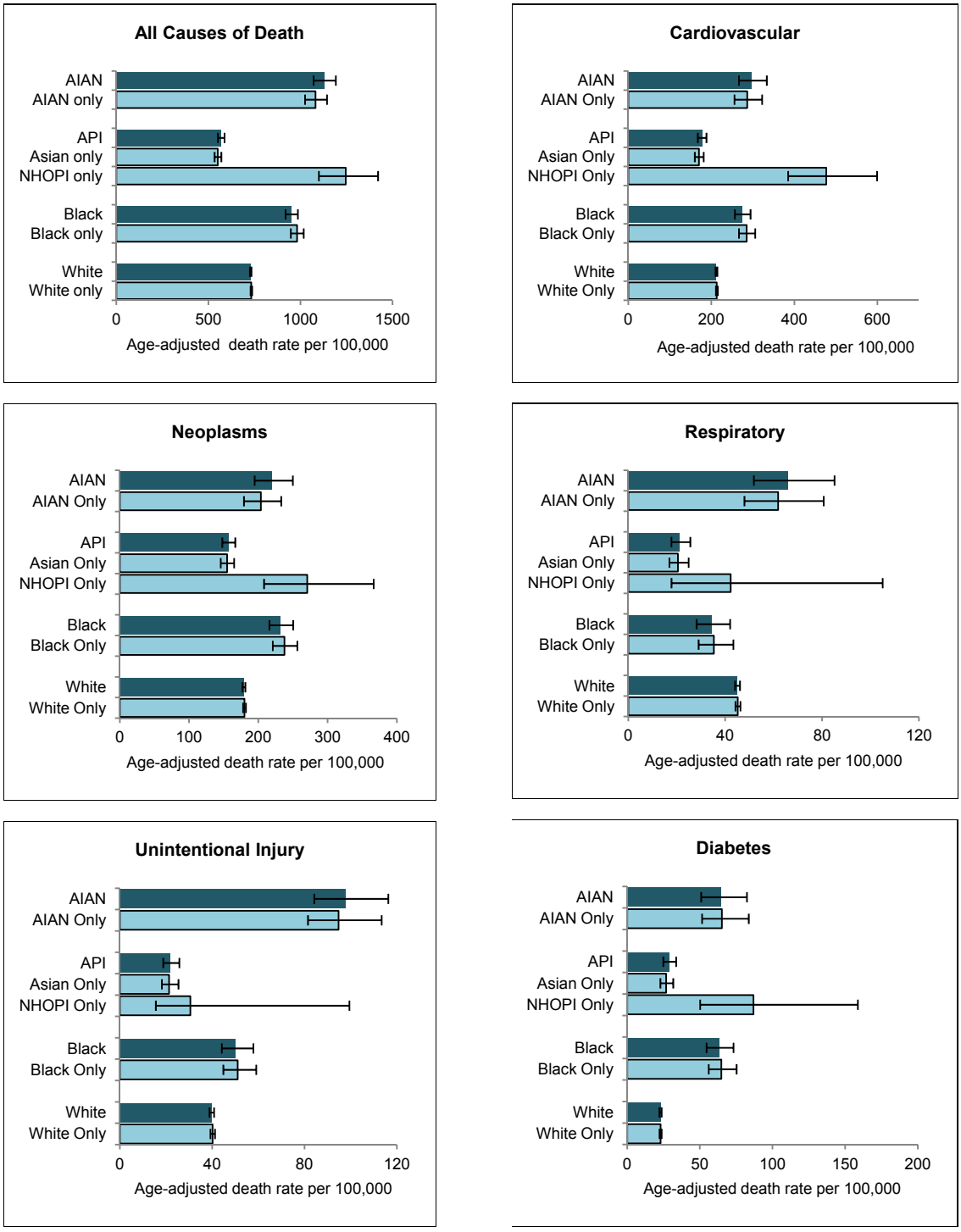
  The Washington State Cancer Registry (WSCR) follows the Surveillance, Epidemiology, and End Results (SEER) deterministic rules for assigning records with more than one race to a single category. The North American Association of Central Cancer Registries

supports these rules and the National Program of Cancer Registries (the primary funder of WSCR) requires their use. The SEER rules require assigning people who select white and another racial category to the non-white category. This rule likely over-assigns people to categories other than white, because some people reporting white and another category would classify themselves as white if they could select only one category.

In WSCR, the bridging rules likely have the largest impact on rates for the American Indian or Alaska Native (AIAN) grouping. As shown in Appendix 2, Table 2.2, only about one-quarter of Washington adults who report both AIAN and white racial categories, select AIAN when asked “Which of these groups would you say best represents your race?” Nonetheless, WSCR assigns all of the records with AIAN and white racial groupings to AIAN, likely representing over-assignment to that grouping.

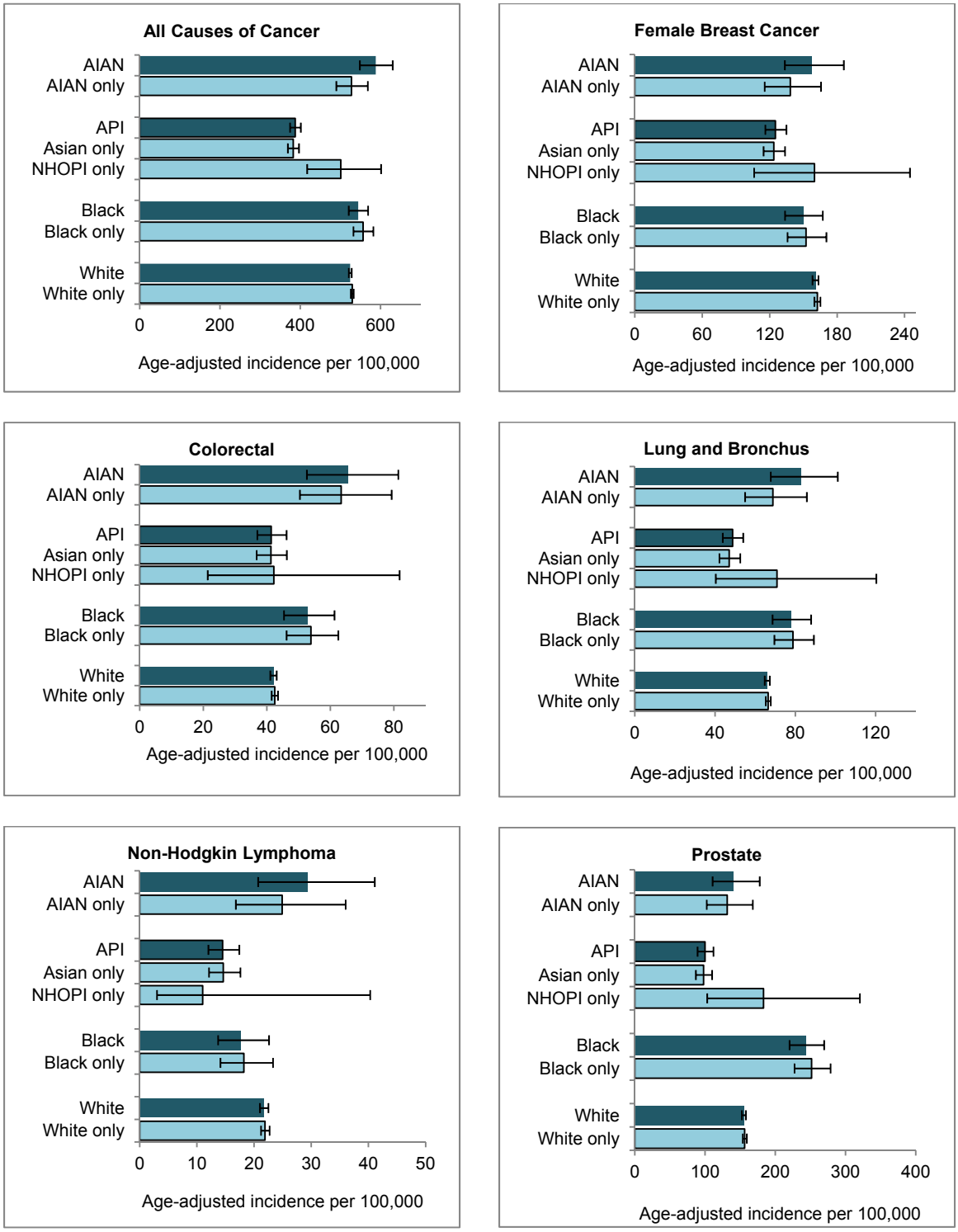
On the charts in Figure 4.2, rates for the AIAN only category are smaller than those for the AIAN category when records with more than one race are assigned to a single category. Most of these differences do not appear to be substantive, but can, nevertheless, affect the interpretation of the findings. For example, when assigning records with more than one race to a single category, the AIAN age-adjusted rates for all cancer types combined and for lung cancer are statistically significantly higher than the rates for the white category; when excluding records with more than one racial grouping, rates for the AIAN and white two groupings are similar. For the other four types of cancer presented, the interpretations are the same irrespective of whether records with more than one race are assigned to a single category or excluded.

Figure 4.1. Washington State Age-Adjusted Death Rates per 100,000, 2005-2007



- Data retrieved from CHAT, March 2010
- Rates are for non-Hispanics only and are age-adjusted to the 2000 U.S. standard population. using the 10-year age groups of the National Center for Health Statistics.
- Light bars ("only") include records with one reported racial category; dark bars include records with one reported category and those with two or more categories assigned to a single category using the probabilistic method of the National Center for Health Statistics. (See Appendix 2, Bridging Methods and Recommendations)
- AIAN: American Indian or Alaska Native; API: Asian or Pacific Islander; NHOPI: Native Hawaiian or other Pacific Islander

Figure 4.2. Washington State Age-Adjusted Cancer Incidence Rates per 100,000, 2005-2007



- Data developed by the Washington State Cancer Registry, February 2010.
- Rates are for non-Hispanics only and are age-adjusted to the 2000 U.S. standard population. using the 5-year age groups of the National Cancer Institute.
- Light bars ("only") include records with one reported racial category; dark bars include records with one reported category and those with two or more categories assigned to a single category using the deterministic rules of the Surveillance, Epidemiology and End Results, National Cancer Institute. This method assigns records with white and another racial category to the other category. (See Appendix 2, Bridging Methods and Recommendations)
- AIAN: American Indian or Alaska Native; API: Asian or Pacific Islander; NHOPI: Native Hawaiian or other Pacific Islander