



## Principles for Collecting, Coding, and Reporting Social Identity Data

### ETHNICITY GUIDELINES

Interim February 8, 2011

#### Purpose

To promote transparency and consistency, so that the San Francisco Department of Public Health's (SFDPH) reports of health by ethnicity refer to the same populations and reflect meaningful differences in health outcomes, conditions that impact health, and delivery of health services.

#### Background

These guidelines were developed by SFDPH Community Programs epidemiologists, researchers, and analysts who share concerns regarding the collection, coding, reporting, interpretation, and use of social identity indicators. To monitor health outcomes and intervene on behaviors that are the underlying causes of disease and injuries, SFDPH must be able to incorporate changing definitions, relevance, and boundaries that individuals, communities, programs and/or institutions use to identify themselves and others.

These guidelines address the following key issues concerning race and ethnicity:

1. Desire for consistency in grouping or categorizing of race and ethnicity data across time and data regimes.
2. Need for flexibility to accommodate many different existing data collection practices.
3. Lack of clarity in the meaning and use of terms defining race and ethnicity.

The UCSF Family Health Outcomes Project addressed the first two of these concerns in a set of guidelines for the California Department of Health Services in 2003. SFDPH used their recommendations as a starting point for these guidelines, and extended them to accommodate continuing issues in race and ethnicity data collection methodologies (e.g., the use of separate Hispanic origin and "race" questions, selection of multiple "race" categories, and identification of primary "race" if more than one chosen) and also our understanding of the meaning and use of the terms "ethnicity" and "race." These guidelines also reflect the contributions of members of the community and SFDPH workgroups focused on reducing health disparities, including a peer review with members and guests of San Francisco State University's Health Equity Institute for Research, Practice and Policy.

#### General Guiding Principles

1. Ethnicity is an important marker of health differences but should not be assumed to be the source of health differences.

Collecting and using data by "ethnicity" (ethnicity, ethnic, so-called "race," or national identity/origin categories) is important for monitoring and addressing health inequities and differences, but should not substitute for, or detract from, continuing to understand and act on the conditions that more directly underlay and produce these disparities. These conditions may include distribution and opportunities for income, residence area, work and education, and culturally-based practices in areas such as diet, physical activity, and social relations.



## Principles for Collecting, Coding, and Reporting Social Identity Data

### ETHNICITY GUIDELINES

Interim February 8, 2011

2. Ethnicity is a social or cultural identity, not a biological or medical condition, as the term “race” may imply.

“It is important to remember that racial and ethnic data categories are social-political constructs and that they should not be interpreted as being genetic, biological, or anthropological in nature.<sup>1</sup>” Because use of the term “race” may contribute to historical and contemporary misinterpretations (i.e., that “race” is a biologically- or medically-based classification), the term itself should not be used unless it is required to be.<sup>2</sup> This concept should be referred to as “ethnicity” (or “ethnic identity”). If use of the term “race” is required, the Department of Public Health should put the term in quotes whenever possible. While we suggest avoiding the term “race,” we recognize the concept of (and use of the term) *racism* to characterize discriminatory social institutions and practices which influence life experiences and health.

3. Naming of ethnicities should allow for both consistency and relevancy.

Ethnicity is a dynamic concept, the meaning of which changes based on the ways individuals, communities and institutions identify themselves and others. The DPH approach to naming and categorizing ethnicities should maintain consistency across programs and over time, so that data are as comparable and interpretable as possible, while also striving to reflect the way people and communities construct and experience their own identities. Individuals need to be able to see themselves and their communities reflected. Reviewers of data need to be confident that any DPH data reported by ethnicity are using the same terms to refer to the same groups.

4. Categorization should allow for both compliance and comparability.

Ethnicity data should be collected, coded, and reported to allow:

- a. Programs to meet their specific reporting and funding requirements for data
- b. Programs to collect disaggregated (detailed) levels as they determine them to be useful and feasible
- c. Comparability with federal, state, and local data requirements
- d. Comparability across County departments and DPH programs

5. A single set of common mutually-exclusive core ethnicity categories that are aligned with state and federal minimum reporting categories should be used.

Ethnicity data should be minimally reported by these core categories and definitions:<sup>3</sup>

- a. **African American/ Black**  
A person having origins in any of the black ethnic groups of Africa
- b. **Asian**  
A person having origins in any of the original peoples of the Far East, Southeast Asia (including Philippines), or the Indian subcontinent
- c. **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
- d. **Native American<sup>4</sup>**  
A person having origins in any of the original peoples of North America, Central America, or South America



## Principles for Collecting, Coding, and Reporting Social Identity Data

### ETHNICITY GUIDELINES

Interim February 8, 2011

- e. **Latino/a**  
A person having origins in Mexico, Central America, South America, Puerto Rico, or Cuba
- f. **White**  
A person having origins in any of the original peoples of Europe, the Middle East, or North Africa
- g. **Multi-ethnic**  
A person having origins in more than one of the other core categories specified.

### Data Collection Principles

- 6. Ethnicity should be self-identified.  
Assigning of ethnicity should be self-identified and allow for as broad a range of identification as possible. If self-identification is not possible, declaration by next of kin or other close informant should be used. Ethnicity should not be ascribed by others based on appearance or language.
- 7. A single ethnicity question should be used.  
Federal and state practice is to use two questions, in this order: (1) “ethnic origins,” which is used to identify Latinos, and (2) “race,” which is used to identify all other groups. People are allowed to identify more than one category on the second “race” question. This format asks everyone, including Latinos, to answer both the Latino ethnicity question and the “race” question; e.g., a self-identified Latino may add that he or she is white. However, since the distinction between “ethnicity” and “race” as used in this format is neither clear nor meaningful to respondents, DPH should collect this information using a single ethnicity question which includes all the categories from both the Latino origin and “race” questions, and allows for multiple choices to be selected (see attached crosswalk for common rules of aggregation to be followed).
- 8. Ethnic-specific detail should be captured only to the degree it is useful for identifying differences in health, health-related conditions, or delivery of health information or services.  
Detail can be collected either by listing at least two checkbox options (see below) or by allowing a text fill-in for detail for any given checkbox option (e.g., “White (specify:)” or “Native American: Alaska Native (tribe:)”). If used, subcategory options under the core ethnicity categories should:
  - a. Allow for collapsing into the core categories (reference number 5 above; see attached crosswalk for common rules of aggregation to be followed, “Attachment CCSF DPH Ethnicity Codes by Core Categories and Subcategories Crosswalk”).
  - b. Include “Other” with a fill-in space option for text data under each core category that includes any detailed subcategory choices (e.g., if “Russian” is listed under “White”, there should also be an “Other White (specify:)”), so that anyone falling under that core ethnicity category will be able to check-off or fill-in a detailed subcategory choice that includes them.



## Principles for Collecting, Coding, and Reporting Social Identity Data

### ETHNICITY GUIDELINES

Interim February 8, 2011

9. Selections of multiple ethnicities by an individual should be allowed.
  - a. Allow persons to enter or check up to four ethnicity subcategories, including if feasible the possibility of writing in an ethnicity other than those specifically listed. For example, ask, “With which ethnicities do you identify?” and follow it by, “Check up to four that apply.”
  - b. Persons who select more than one ethnicity should be given the opportunity to also select their primary ethnicity, e.g., “Do you most identify with one ethnicity in particular?” If answered yes, follow by option of selecting which particular ethnicity they most identify with.
  - c. For individuals selecting “primary ethnicity,” the primary ethnicity choices should offer both “none primary” and “multi-ethnic” options.

### Coding and Reporting Principles

10. Core Ethnicity categories should be used to organize reporting, regardless of level of detail.

Report detailed ethnicity categories as subheads under the appropriate core ethnicity category using the attached crosswalk matrix and recommendations.
11. Each individual should be assigned to a single set of exhaustive, mutually exclusive core ethnicity categories.
  - a. Person selecting only one ethnicity is assigned to the core category selected, or the core category which includes the specific subcategory selected.
  - b. Person selecting more than one ethnicity and a “primary” ethnicity is assigned to the primary selection’s core and sub-category.
    - If ethnic-specific choices fall within one core category and person selects primary choice as multi-ethnic, person is classified under that core and within it, as [core] multi-ethnic subcategory. For example, person selecting Chinese, Japanese, and multi-ethnic, person would have “Asian” core ethnicity and “Asian multi-ethnic” specific ethnicity.
    - If specific choices cross core categories and person selects primary choice as multi-ethnic, person is classified into multi-ethnic core category.
  - c. Person selecting more than one ethnicity, but not specifying a “primary” ethnicity, is assigned as follows:
    - If all subcategory choices fall within a single core category, assign the person to that core category and under it, to the core category’s multi-ethnic category.
    - If their subcategory choices fall across more than one core category, assign them to the “Multi-ethnic” core category.
    - When compiling data from the two-question methodology datasets, further delineation must be used. As noted, Latinos are a special case because the federal two-question protocol asks Latinos to self-identify their Latino origins as well as to select one of the choices on the “race” question, in effect requiring those following this procedure to identify themselves as falling under multiple core ethnicity categories. Assigning Latinos may be done in several ways, but, however done, should be reported such that Latinos are not lost in



**Principles for Collecting, Coding, and Reporting Social Identity Data**

**ETHNICITY GUIDELINES**

Interim February 8, 2011

the “Multi-ethnic” category. Whatever way is used should be explained for data users.

- i. A common practice in California for crossing data from the two questions (Latino and “race”) into a single set of mutually-exclusive categories is to categorize anyone selecting a Latino choice as a Latino and then to categorize all non-Latinos among the other categories, which are then called “Non-Latino White”, “Non-Latino Black”, etc.
- ii. An alternative method is to apply the same rule to all categories and include a “Multi-ethnic” core category with subcategories allowing for all combinations including Latinos, i.e., “Latino and White”, “Latino and Black”, “Latino and Other.” This would allow the number of Latinos reported under “Multi-ethnic” in this method to be comparable to the number reported under “Latino” in method shown in section 11.c.i above.

Reference	If person selected Ethnicity(ies) as:	And if person selected Primary as:	Then Core Category of Person is:	And Subcategory of person is:
11.a.	✓ Chinese (Asian)	Not applicable	Asian	Chinese
11.b.	✓ Mexican (Latino)	Chinese	Asian	Chinese
	✓ Chinese (Asian)			
11.b, 11.c.1.	✓ Chinese (Asian)	None Selected, No primary, or multi-ethnic	Asian	Multi-Ethnic Asian
	✓ Japanese (Asian)			
11.c.2.	✓ Chinese (Asian)	None Selected, no primary	Multi-ethnic	Asian-Pac. Is.
	✓ Samoan (Pac.Isl.)			
11.c.3.i	✓ Mexican	None Selected, no primary	Latino	Latino- White
	✓ White			
11.c.3.ii	✓ Mexican	None Selected, no primary	Multi-ethnic	Latino- White
	✓ White			

- d. Where program and data needs will support it, other forms of reporting of multiple ethnicity choices in greater detail should be explored to supplement (but not replace) the one used above. When used, the basis for this additional classification should also be reported.

**12. Data reports by ethnicity should minimally represent all core ethnic categories.**

If data are not available to report reliably for any of the core categories, that fact should be reported. For example, if insufficient data are available to report on the health of Native American/ Alaska Natives in San Francisco, reports should note the reason for the lack of presentation of such data.



## Principles for Collecting, Coding, and Reporting Social Identity Data

### ETHNICITY GUIDELINES

Interim February 8, 2011

#### Supporting Data to Collect

13. Where possible, collect data on place of birth.

14. Where possible and useful, collect data on language.

For example, ask, “What is your primary/usual/principle language spoken at home?” or “What language do you feel best able to express yourself?”

#### Recommendations

- DPH to work with researchers inside and outside DPH to assess this proposed methodology with the aim of refining the guidelines and developing effective data collection tools.
- DPH Programs to incorporate these principles into their current data collection and reporting as feasible, and attempt to apply them further in any system, data collection, and reporting (re)designs.
- DPH to develop and make available implementation materials, including question and response options, best practices, community recommendations, and data coding and reporting programs that have been developed.
- DPH to develop a policy that each data system designed should articulate how questions will be asked, when they will be asked, and by whom; and also make public how they have been reported (coding and aggregation practices used).

---

<sup>1</sup> Office of Management and Budget, Executive Office of the President, *Provisional Guidance on the Implementation of the 1997 Standards of Federal Data on Race and Ethnicity*, Wash. D.C.: Dec. 15, 2000, p. 9., cited in Family Health Outcomes Project, UCSF, and Center for Health Statistics, CDHS: *Guidelines on Race/Ethnicity Data Collection, Coding and Reporting*. January, 2003 (<http://www.ucsf.edu/fhop/docs/pdf/prods/raceEth2003.pdf>). Many of the ideas in this set of guidelines are derived from this FHOP Guidelines document.

<sup>2</sup> “American Anthropological Association Response to OMB Directive 15: Race and Ethnic Standards for Federal Statistics and Administrative Reporting,” (Sept 1997), <http://www.aaanet.org/gvt/ombdraft.htm>

<sup>3</sup> These categories, along with “other single race” are used by the California Department of Finance to produce county population estimates by year, single year of age, sex, and ethnicity. The definitions are modifications of the federal minimal “race” categories, with “Latino” integrated from the federal “Hispanic origin” question. See <http://www.census.gov/prod/2001pubs/c2kbr01-1.pdf> for exact definitions. Thus, these categories also allow the calculation of city-wide rates requiring population estimates. Definitions are from OMB Standards and reflect the minimum categories for data on race and ethnicity for Federal statistics, program administrative reporting, and civil rights compliance reporting. See State of California, Department of Finance, *Race/Ethnic Population with Age and Sex Detail, 2000–2050*. Sacramento, CA, July 2007.



## Principles for Collecting, Coding, and Reporting Social Identity Data

### ETHNICITY GUIDELINES

Interim February 8, 2011

---

[http://www.dof.ca.gov/HTML/DEMOGRAP/Data/RaceEthnic/Population-00-50/RaceData\\_2000-2050.php](http://www.dof.ca.gov/HTML/DEMOGRAP/Data/RaceEthnic/Population-00-50/RaceData_2000-2050.php)

- <sup>4</sup> The OMB title used for this category is “American Indian / Alaska Native (AIAN),” a term that is used by other federal agencies strictly for identifying indigenous people whose origins are in the United States. However the same OMB category definition includes indigenous people of all the Americas. To encompass the OMB definition (for comparability) and to enable isolated tracking of AIAN individuals, SFDPH changed the core term to “Native American” and recommends minimally tracking the following subcategories, and optionally, further defined groups within subcategories.

**CORE CATEGORY: Native American**

A person having origins in any of the original peoples of North America, Central America, or South America

**SUBCATEGORIES:**

1. **AIAN-American Indian/Alaska Native** (United States)  
Provide checklist and/or provide space for “Optional: specify tribe: \_\_\_\_\_”
2. **First Nation** (Canada)  
Provide checklist and/or provide space for “Optional: specify tribe: \_\_\_\_\_”
3. **Indígena** (Mexico, Central and South America)  
Provide checklist and/or provide space for “Optional: specify group: \_\_\_\_\_”
4. Other Native American (specify \_\_\_\_\_):

**Principles for Collecting, Coding, and Reporting Social Identity Data ETHNICITY GUIDELINES, Attachment CCSF DPH Ethnicity Codes by Core Categories and Subcategories Crosswalk**

(CASPER INTERIM -020811- 415-255-3706)

CORE ETHNICITY Categories *	SUB ETHNICITY Categories (NOTE: most reflect current data collected for one or more databases. *** reflects subcategories proposed by community and academic consultations. CASPER is not recommending which subcategories to keep or delete.)	CURRENT CODES USED									CURRENT AGGREGATES USED		
		FHOP / Vital Records	OMB	2000 Census	Dept Finance	Health SF	LCR	MH BIS	SA BIS	FHOP Codes (California & Birth) Aggregates	OMB & Census Codes (Federal) Aggregates	DoF Codes (Population Est) Aggregates	
<b>African American/ Black</b>	Black/African American	20	Q2	X	X	X	2	B	B	2 Black/ African American Only (non-Latino)	2 Black/ African American Only (non-Latino)	5 Black/ African American Only	
A person having origins in any of the black ethnic groups of Africa	African American												
	***Caribbean African Descent												
	***African												
	***Other African American (specify):												
<b>Asian</b>	Asian	Y	Q2		X	X	5			5 Asian Only (non-Latino)	4 Asian only (non-Latino)	3 Asian only	
A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent	Asian Indian	40		X		X	I	S	E				
	Chinese	42		X		X	C	G	G				
	Japanese	44		X		X	J	K	K				
	Korean	45		X		X	K	U	L				
	***Mongolian												
	***Tibetan												
	Other Asian (specify):	49		X			O	M	P				
	Filipino	39		X		X	F	L	P				
	Cambodian	41				X	A	J	F				
	Hmong	43							2				
	Laotian	46				X	L	I	M				
	Thai	47											
	Vietnamese	48		X		X	V	H	O				
	Other Southeast Asian (specify):						W	P					
<b>Native Hawaiian or Other Pacific Islander (NHOP)</b>	Native Hawaiian or Pacific Islander	Y	Q2		X					7 Pacific Islander Only (non-Latino)	5 Pacific Islander Only (non-Latino)	4 Pacific Islander Only	
A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands	Native Hawaiian	50		X		X	H	T	J				
	Pacific Islander								7				
	Guamanian [or Chamorro]	51		X		X	G	Y	I				
	Samoaan	52		X		X	S	Z	N				
	Tongan					X							
	Other Pacific Islander (specify):	53		X					W				
<b>Native American</b>	AIAN-American Indian/Alaska Native (United States)	30	Q2	X	X		4	C	C	3 American Indian or Alaska Native Only (non-Latino)	3 American Indian or Alaska Native Only (non-Latino)	6 American Indian or Alaska Native Only	
A person having origins in any of the original peoples of North America, Central America, or South America	AIAN - American Indian					X							
	AIAN - Alaska Native					X							
	AIAN - Alaska Native - Eskimo	31					4		D				
	AIAN - Alaska Native - Aleut	32					4						
	Other AIAN (specify):	32											
	***First Nation (Canada)												
	***Indígena (Mexico, Central and South America)												
	***Indígena - Maya												
	***Other Indígena (specify):												
	***Other Native American (specify):												
<b>Latino/a</b>	Hispanic Yes/No	Y	Q1		X	X	3	Y/N	2	9 Latino/a (any "race")	6 Latino/a (any "race")	2 Latino/a	
A person having origins in Mexico, Central America, South America, Puerto Rico, or Cuba	Mexican	3		A			M						
	Central American	4					B						
	South American	5											
	Chicano/a/ Mexican American			A			N	Y					
	Puerto Rican	7		B			P		4				
	Cuban	6		C			D		3				
	Other Latino/a (specify):	8		D			U		5				
	Hispanic Unknown/ Not Reported							U					
<b>White</b>	White	10	Q2	X	X	X	1	A	A	1 White only (non-Latino)	1 White only (non-Latino)	1 White and Unknown only	
A person having origins in any of the original peoples of Europe, the Middle East, or North Africa	Russian						R	R					
	Russian Jew						X						
	Other European						T						
	Middle Eastern						E						
	Arab/ North African/ Mid East								V				
	***Other White (specify):												
<b>Multi-Ethnic</b>		Y			X				R	20 Two or more ethnicities	8 Two or more ethnicities	7 Multi-race	
<b>Options Not Recommended</b>	Other	60		X		X	6	I	Q	99 Other/ unknown ethnicity only (non-Latino)	9 Other/ unknown ethnicity only (non-Latino)	Listed under "White"	
	Other Non White						O						
	Unknown	99					8	O					

\* OMB Standards defines the minimum categories for data on race and ethnicity for Federal statistics, program administrative reporting, and civil rights compliance reporting. However, we changed "American Indian / Native Alaskan" to "Native American" so that we could incorporate indigenous peoples from all of the Americas.

\*\* Dept Fin: Department of Finance--data reporting only; not data collector. Produce county population figures (by year, sex and ethnicity); California state agencies required to use their population figures.