

## Race & Ethnicity Data: Developing a Common Language for Public Health Surveillance in Hawaii

Catherine A. Sorensen, Betty Wood, Edward W. Prince

*Hawai'i Department of Health*

### Abstract

Race and ethnicity are key variables used in the field of public health surveillance for monitoring and tracking health status and outcomes of populations. However, over the last decade the collection and use of these traditional variables has come under scrutiny. Central to the arguments are the manner in which race and ethnicity are conceptualized and the lack of standards in terms of how the data elements are defined. This paper describes some of the challenges to collecting and reporting race and ethnicity data in a state that has a large and growing multi-racial and multi-ethnic population base. A model is presented that describes a methodology for incrementally clustering very discrete ethnic sub-populations to ethnic subgroups and eventually to the Office of Management and Budget five race classifications.

© 2003 Californian Journal of Health Promotion. All rights reserved.

*Keywords: race, ethnicity, data standards, informatics*

### Introduction – The Importance of Ethnicity & Race Data for Public Health

Race and ethnicity data have been collected for more than 200 years in the U.S. in order to provide denominators for social and demographic analysis. Prior to 1930 “race” was ascribed to national groups that were more often than not immigrants (e.g., Hebrews, Italians, Celts). Such groups along with the American Indians and Blacks were regarded as quite apart from the White founders. Racial distinctions were used to support “scientific” evidence that “morals, physical and intellectual capacity were inherited” (Oppenheimer, 2001).

Today ethnicity and race data continue to be collected to frame and describe public health issues. In one of the first reports on health disparities, Du Bois described differences between African-American and White populations in the early 1900’s (Du Bois, 1906). Because disease, injury, access to prevention and treatment services, cost of care, quality of life, and other public health concerns persist disproportionately affecting some ethnic and race populations over others, the Healthy People 2010 Initiative adopted as one of its three

overarching goals the elimination of health disparities (US DHHS, 2000). Ethnicity and race are important determinants in health patterns, whether representing true biological and genetic differences or a set of factors that affect health and health status.

The last 15 years have seen a rise in the criticism, particularly in the U.S., of the collection and use of race and ethnicity data. U.S. experts in the natural and social sciences agree that the biological concept of race has no scientific basis (CDC & P, 1993; NCI, 2000). The Institute of Medicine argues that “in all instances race is a social and cultural construct” with its base in “perceived differences in biology, physical appearance, and behavior (Haynes & Smedley, 1999, p.38). In fact, data show there are substantially more genetic variation within “races” than between them (NCI, 2000).

### Challenges Using Race and Ethnicity

There are two key issues regarding race and ethnicity. First is the manner in which the terms are conceptualized and implemented, and second is the range of definitions used to cluster

population groups. The heart of the challenge is that race and ethnicity are not simply terms but are concepts. They are complex and personal concepts that have context and, as noted, that context is a “product of social and political history” (CDC & P, 1993). The shape of one’s eyes or nose, the texture of one’s hair are not accurate or legitimate methods for classifying people into meaningful racial subgroups. As a biological or genetic measure race may suggest genetic homogeneity across population groups. In truth, these populations are heterogeneous. Asian, Black, White, Native American are examples of genetic admixtures of geographic stock from all over the world. Vietnamese, Pakistani, and Filipino are all Asian however each is derived from a different geographic genetic stock. The same fact applies to persons classified White. The last two decades in particular have seen a heightened sensitivity to differences among minority ethnic groups. On the other hand, White is such a ubiquitous and familiar term the diversity within the White race is not even regarded. White is inclusive of such diverse groups as Scottish, Greek, Spanish, Canadian, Iranian, and Moroccan. Clustering or aggregating diverse population groups to a single race category denies or ignores genetic variability.

Use of race and ethnicity categories also implies cultural homogeneity for such matters as health beliefs, dietary practices, and physical activity. If in fact culture and health are linked then the presumption might be made that health risks and outcomes would be homogenous. Data indicate otherwise. For example, breast cancer incidence rates for Asian-American ethnic groups are not the same (Deapen, Liu, Perkins, Bernstein, & Ross, 2002). In the U.S. breast cancer risk among women of Japanese and Filipino ancestry is twice that of Chinese and Korean women. These four individual ethnic groups are all Asian, yet each has its own distinct culture.

Classifying by race and ethnicity suggests that racial or ethnic identities are static and mutually exclusive. Research shows that these data are fluid as people can and do change how they self-report their race or ethnicity for political and economic reasons. Parents may report detail

about their children’s racial and ethnic background, particularly for multi-racial births. As these children age and, in particularly those who leave home, may simplify how they describe their racial and ethnic background (Walters, 2000).

These challenges have given rise to arguments that the collection and use of race and ethnicity should be discontinued. The proposals have suggested emphasizing socioeconomic data and other “life factors” including health insurance status, geography (e.g., place of residency and length of residency in U.S.), personal/family income status, religion, personal health beliefs and practices (Fullilove, 1998; Krieger, 2000; NCI, 2000). While acknowledging the salience of such arguments others contend that race and ethnicity information should be expanded. The increase in such collection would be done in concert with a broad awareness of the implications of using race and ethnicity (La Veist, 1996; Thomas, 2001; Williams, Laviszzo-Mourney, & Warren, 1994). Other recommendations include providing detailed guidelines to address issues such as multiple response sets, information on the collection method (self-report or observation), and study design (Jones, 2001; Kaufman & Cooper, 2001).

Despite the widespread practice of asking a person about his/her race and/or ethnicity the charge to collect these data is not universal. Many European countries are prohibited constitutionally or legislatively to collect these data though actual practices do vary among the European nations. In the U.S. public health organizations and private entities voiced major concerns with a California ballot initiative (Proposition 54) which would have restricted state and local government offices from collecting information on a person’s race, ethnicity, color, or national origin for certain purposes. The initiative was defeated.

Because the use of race and ethnicity is rooted in public health surveillance, the likelihood the collection and use of these descriptors will be abandoned anytime soon is remote. As one epidemiologist expressed – “This is what we are taught to do --” (Jones, 2001). A critical reason

for continuing to collect these data is health disparities do exist between different racial and ethnic groups. The challenge is differential patterns of disease, risk factors, health beliefs and practices, and access to services across racial and ethnic populations are poorly understood and what is known is incomplete.

### **Federal Reporting Standards**

The U.S. Census is an essential data source for information identifying trends and changes in economic, social, and health characteristics by race. Unfortunately, the method used to gather Census data has not been stable with almost each decade experiencing changes in how the data are collected. It was not until the mid 1970's that the federal government began a concerted effort to develop and implement a common language for reporting ethnicity and race. The impetus to develop a standard was the need for comparable data to monitor equality issues including the availability of and access to health care services, employment opportunities, education, and housing for population groups that experienced discrimination.

Since 1977, the Office of Management and Budget's (US OMB, 1997) Statistical Policy Directive 15 provided the common language promoting uniformity and comparability for collecting and reporting of ethnicity and race data. The standard recognized four categories for race data (American Indian or Alaskan Native, Asian and Pacific Islander, Black, and White) and two categories of ethnicity data ("Hispanic origin" and "Not of Hispanic origin"). In 1997 the OMB announced revisions to Directive 15 to address the need for more refined data that would take into account demographic changes including growth in immigration and increase in interracial marriages (US OMB, 1997). The revised standards established a new fifth race category - Native Hawaiian and Pacific Islander.

The current race category standards are:

- American Indian/Alaskan Native: 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/African-American: A person having origins in any of the black racial groups of Africa.
- Native Hawaiian/Other Pacific Islander: 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.

The 1997 standards retained the two ethnic categories: (1) Hispanic or Latino; and (2) Not Hispanic or Latino with Hispanic defined as: A person of Cuban, Mexican, Puerto Rican, South or Central American or other Spanish culture or origin, regardless of race.

Undoubtedly the most significant change was the federal guideline that required individuals be given the opportunity to select one or more races. For the first time in the history of the U.S., citizens were allowed to check more than one box to identify their race. Agencies are encouraged to not only collect multiple responses but even more detailed information on specific racial combinations as long as the data reliability and confidentiality concerns can be met.

The preferred federal data collection method uses a two-question format that asks a person to first declare their ethnicity then race. In instances where the data will be obtained from observer-collected methods, a combined format may be employed. The combined format makes use of six minimum categories:

- American Indian or Alaska Native
- Asian
- Black or African-American
- Native Hawaiian or Other Pacific Islander

- White
- Hispanic

Regardless of the collection method there are no criteria or qualifications such as blood quantum to “prove” race affiliation. The responses are based on self-perception and are by definition accurate. Unless a person intentionally misreports their ethnicity and/or race, their response is not “wrong” even if clinical tests would indicate otherwise. The Office of Management and Budget acknowledges that these categories are “neither anthropologically nor scientifically-based, but rather represent a social-political construct designed for collecting data on race and ethnicity of broad population groups in the U.S.” (US OMB, 1997).

It is important to note that OMB does not mandate the collection of race and ethnicity data, but rather provides the standards or guidelines by which these data are collected and reported. The OMB standards address the minimum amount of information that should be obtained thereby assuring the data will be collected in a prescribed and consistent manner. These standards do not apply for states and private industry. The exception is state agencies that receive federal funds. Agencies and organizations that are not required to use the standards have a reason to adopt them. For example, public health and social service organizations are dependent on data to describe their population(s) and empirically demonstrate a need. Often local data are used in comparison to national data to substantiate a problem or need. These data are usually expressed as rates. The Census is the source of population denominators for rate calculation used by federal offices and other state systems allowing for national comparisons.

### **Race and Ethnicity Data Standards**

Race and ethnicity are central to public health surveillance activities and programs and the demand for detailed data continues to grow locally and nationally. How are these terms understood within the professional community? Are they considered synonymous and thus used interchangeably? Are race and ethnicity in fact one and the same? Does it really make a

difference if the terms are transposed? What is the gain if these terms are standardized?

The Merriam-Webster dictionary defines “race” as: “a family, tribe, people, or nation belonging to the same stock; a division of mankind possessing traits that are transmissible by descent and sufficient to characterize it as a distinct human type.” This definition suggests there is a genetic or biological link to phenotypic traits such as skin color and facial features. “Ethnic” is defined as: “a member of a minority group who retains the customs, language, or social views of the group.” “Ethnicity” is thus associated with cultures, behavioral attitudes, beliefs, lifestyle patterns, diet, and environmental living conditions. The dictionary definitions indicate that “race” and “ethnicity” are not synonymous. The reality is that often in the day-to-day of public health these terms are likely to be used more casually.

The Hawaii Department of Health (DOH) does not employ a standardized method to define, collect or report race and ethnicity data resulting in a number of special challenges. [Appendix A](#) provides a dramatic visual representation of the problems. A sample of 11 data sets identified approximately 100 separate race and ethnic categories highlighting such issues as inconsistency in spelling (e.g., Belauan/Palauan), inconsistent combining of groups (e.g., Samoan and Samoan/ Tongan), and formatting differences (American Indian/ Alaskan Native and American Indian or Alaskan Native). The disparity across the descriptors and collecting practices are largely based on past practices and are often linked to funding agency guidelines (e.g., federal). While many DOH programs collect ethnic-specific data these data are reported at an aggregate level. The chance to conduct in depth analyses on specific ethnic populations is lost. The lack of data conformity results in lost opportunities to mine the data to provide new information. An important goal is to have processes in place whereby race and ethnic data are conformed (standardized) in terms of how they are defined, collected, and aggregated ensuring the accuracy and usefulness of the information.

A related challenge is the practice of collecting race and ethnicity information that is limited in its usefulness. The 11 DOH data sets lists over 20 such categories including: Other Asian, Mixed Asian Asian, Melanesian Unspecified, Other East Asian, Other South East Asian, South Asian Unspecified, South East Asian Unspecified, Other Asian/Pacific Islander, Micronesia Unspecified, Other Micronesian, Native American Unspecified, Other Melanesian, Other Polynesian, Polynesian Unspecified, Pacific Islander Unspecified, All Other, Mixed Race, Other Race, Not Specified, Mixed Non-Hawaiian, East Asian, and Eurasian. It would be beneficial to establish guidelines to minimize the use of general categories and improve the initial data collection process.

To nurture the critical issue of standardizing key data elements and to ensure clarity and continuity for readers this paper has adopted the dictionary definitions of race and ethnicity. The terms “race” incorporates: American Indian or Alaska Native, Asian, Black or African-American, Native Hawaiian or Other Pacific Islander, and White. “Ethnicity” references discrete population groups associated by geography, culture or language (e.g., Japanese, Chuukese, Fijian).

### **Hawaii Department of Health – Vital Statistics**

The Office of Health Status Monitoring (OHSM) is one of the key offices within the DOH. Much of the data it collects is used across many Divisions. OHSM has statutory authority and the responsibility under *Chapter 38, Hawaii Revised Statute* to collect and report all birth and death events. The birth certificate generates a wealth of information (e.g., maternal and infant medical risk factors, presence of congenital abnormalities, obstetrical procedures). It is also an important source for race and ethnicity information. Decision rules established by the DOH in the 1940’s are used to assign infant race and ethnicity as reported by the parent(s). Infant ethnicity is derived from the ethnicity of the father. If that is unknown or not reported, then the ethnicity of the mother is used to determine the infant’s classification. The decision rules allow parents to report multiple ethnicities on

the birth certificate form; however, only one ethnicity is captured electronically from the certificate. The following rules apply if more than one ethnicity is listed:

- If Hawaiian is one of the multiple ethnicities listed, Part-Hawaiian is coded.
- If a non-Caucasian\* ethnicity is listed with Caucasian, the non-Caucasian ethnicity is coded.
- If there is more than one non-Caucasian ethnicity listed, the first one listed is coded.
- If there is more than one Caucasian ethnicity listed, the first one noted is coded.

[\* “Caucasian” is used as an ethnicity. The dictionary definition of Caucasian is “originating in the Caucasus region; widely used as equal to “White”.]

Death certificates also generate indispensable information including: underlying cause of death, place of death, date of birth, occupation, and ethnicity and race. Death certificate ethnicity information is substantially more problematic. Whereas birth ethnicity information is reported by a parent, death certificate ethnicity information is based on the observation of the funeral director, physician or mortician. It is not unusual that birth and death records do not match for ethnicity. A Center for Disease Control and Prevention study found the differential racial classification on birth and death certificates was more than 31 times as likely to occur with different race than with same-race parents (Hahn, 1999). This situation is particularly problematic for persons of multi-ethnicities.

A major consideration in using birth and death data is these data are collected for administrative purposes. As such their application in public health surveillance, research, policy development, and program evaluation can be challenging as the administrative goals may not parallel these other needs.

### **Hawaii Race and Ethnicity Model**

In 1999, the DOH leadership envisioned a data warehouse that would serve the Department’s

data needs. The goals of the Hawaii Health Data Warehouse were to: coordinate resources across Administrations and Divisions, ensure consistency of the public health data with national recommendations, and increase public access to health data. Foremost in the design vision was integrating data from disparate data sets to enrich the information.

Integrating data requires that critical demographic data elements (i.e., race, ethnicity, age, gender, and geography) be available in a standard or common format. As the warehouse has moved from concept to reality, technological requirements have heightened awareness and driven the need to address a number of data issues. While the OMB standards provide a base for defining race, collecting and reporting ethnicity is a more complex matter.

A DOH working group was formed to research methodologies and practices for collecting and reporting race and ethnicity data. Five important criteria were identified. The model must:

- Provide for the continued collection of “program-level” data to support community-based planning and decision-making, and identify health disparities;
- Identify a process whereby ethnicity data could be clustered using a set of standards ensuring that individual ethnic groups are exclusive to a single aggregated racial group;
- Ensure federal reporting standards are met; and
- Increase the capacity to provide population-based (rate) information as long as sufficient numerator and denominator data are available.

The Centers for Disease Control and Prevention and the Agency for Toxic Substances and Disease Registry (CDC/ATSDR) produced a document titled, “Common Data Elements – Implementation Guide” (version 2.4) which proposed standard data elements for use in health information and surveillance systems (US DHHS, 2000). While incorporating the 1997 OMB race standards, the guide also

disaggregates each racial group into discrete ethnic sub-categories. This design forms a base for the Hawaii model (see [Appendix B](#)).

The Hawaii model distinguishes between race and ethnicity. Aggregated at the highest level are the five OMB categories. These race groups are disaggregated into different levels of specific ethnic populations based on the size of each discrete group in Hawaii. These selected ethnic groups are those most commonly reported by the DOH for public health surveillance. The most discrete ethnic groups listed, including the “unspecified” and “other” categories, are currently in use by different DOH programs.

The Hawaii model is a living or dynamic model. As new information becomes available or data needs change the model will be revised. Because there is no DOH requirement that data elements be standardized, adoption and application of the model will be implemented through consensus.

### **A Special Challenge**

The collecting and reporting of multi-race and mixed ethnicity data has long been an issue in Hawaii because of diverse population groups. In 1995, 15,407 births occurred in Hawaii with 7,843 (58.1%) reported as mixed ancestry. In 2000, the percent jumped to 61% with 7,990 of 13,120 babies identified as mixed ancestry (State of Hawaii, 2001). Nationally the percent of interracial births for the period 1991-1995 was just under 4% (Atkinson, Maccorman, Parker, 2001). In California, a state also known for a diverse population mix, less than 2% of mothers reported more than one race on their child’s birth certificate for 2000 (Heck et al., 2001).

Despite having a large population reporting multiple ethnicities, technical and programmatic issues preclude many DOH divisions and programs from reporting this information, even though it is collected. Programs that do attempt to report their multi-race or mixed ethnicity data use an assortment of methodologies, including:

- Assigning the person to a single race group using the OMB race categories;
- Requiring (forcing) individuals to self-select a single race or ethnicity; or

- Assigning respondents to a single race group using a set of program-specific guidelines.

There are a few cases where persons are assigned to a single racial group based on observation and/or the derivation of the person's last name.

The 1997 OMB Directive which permits individuals to select more than one race has had major implications for public health research and practice, especially for programs engaged in monitoring patterns, trends, and changes in health, social and economic characteristics by race or ethnicity. Data collected under the 1977 standards likely are not comparable to data collected under the 1997 standards. To help understand the race data collected under the new guidelines it may be necessary to bridge new data with the old. Bridging provides a means to measure "real or true" change vs. change produced by the new methodology. Bridging also offers a way to present multi-race data and provides for consistent numerators and denominators for the transition period. Bridging is not necessarily required for every situation. If some discontinuity in a data series is acceptable, or if substitute or surrogate data are available then bridging may not be necessary. Presumably bridging is of most interest when analyzing multiracial responses, as single race responses are assumed to be similar to past responses.

A number of bridging methods have been proposed including: Whole Assignment, Fractional Assignment, All Inclusive, and Hot Deck Imputation. Different bridging methods can change the outcome of reporting health data. Whole assignment and all inclusive methods were applied to analyze data from the Hawaii Health Survey (Baker et al., 1999). Analyses showed that "the reporting of health outcomes may differ drastically depending on the method of OMB coding used for multiple race." "Health outcomes, such as asthma and hypertension, may differ significantly depending if individuals were coded full or part race." The prevalence of asthma among Filipinos varied from 6% to over 16% depending on which method was applied.

## **Recommendations**

Public health surveillance is the cornerstone of public health practice. Surveillance data, of which race and ethnicity variables are a part, facilitate identifying patterns of health, disease, and personal/group health behaviors. While acknowledging this as true, we find ourselves dealing with an opposite truth. We complain about the dearth of available ethnic-specific data noting that without it the true health status among subpopulations is inaccurate, inadequate, and ill-defined. Deficiently detailed race data also limit our ability to report rates, following trends, and identify patterns in health problems. We struggle with the lack of consensus how race and ethnicity are defined and measured. We acknowledge that differences in terminology and data collection procedures may affect the reliability and validity of the analyses. We are affected by the changing demographics of the U.S. and Hawaii's population (i.e., increase in multi-race births) and look to methods of accurately using these data.

On the other hand we debate if by continuing to report out racial and ethnic difference in health outcomes we are maintaining stereotypes, or worse, racist notions. The debate similarly questions whether science supports a relationship between race and ethnicity variables to health outcomes. We ask what role these factors should play in influencing policy, program development, planning, and evaluation. Outside of reporting requirements and simple descriptive statistics, how are these data applied within DOH programs? Anecdotal reports indicate that some DOH programs do make use of their race data by ensuring that education materials and contact strategies are culturally sensitive and appropriate. Should such practices be applied throughout the DOH or are there special circumstances where it would be appropriate? If a reason for collecting these variables is all races have unique experiences then it follows that single race and mixed-race groups must be treated as distinct. Considering that Hawaii's population has a large percent of persons that self-identify more than one race begs the question, "Which race should be preeminent when designing prevention and intervention strategies to reduce health

disparities”? What is the importance of culture, geography, nativity, language competency...?

As the public health profession considers these issues a number of recommendations are proposed for Hawaii:

1. Continue to collect race and ethnicity data as long as differences in health outcomes are a reality with the aim of eliminating the disparities.
2. View race and ethnic data as clues to be mined, not ends in themselves.
3. Consider using a scale that measures factors such as income, residency, education, nativity, language proficiency in conjunction with race and ethnicity. As scale development and validation is an academic activity a partnership with the University of Hawaii might be considered.
4. Be precise when using race and ethnicity by not using the terms interchangeably.
5. Review current ethnic categories and their usefulness in generating information for programs.
6. Review mixed ethnicity data collection method and definition measures.
7. Review method(s) employed to count or classify mixed ethnicity data.

8. Document if the information is based on self-report or observation, and work to eliminate (when feasible and appropriate) observer-derived measures.
9. Document if multiple responses are allowed.
10. Eliminate the use of vague and imprecise descriptors (Other, Mixed, Not Sure, Unknown, etc.).
11. Standardize ethnic group names and spelling (e.g., Guamanian/Chamarro and Pohnpeiian/Ponapean).
12. Standardize rules for handling unknown, not specified, etc., type cases.

These recommendations will help establish DOH-wide standards for collecting, defining, measuring and reporting race and ethnicity data. When using data, care must be taken to not imply there is an association between an individual's or group's health status and their race or ethnicity. Health outcomes are the result of a complex relationship of diverse factors including but not limited to genetic/inherited traits, personal health behaviors, access to care, quality of available health care, and socio-economic status. Careful and complete data interpretations are critical as the application can be used to redirect resources, close or start programs, and affect health policy.

## References

- Atkinson, J. O., MacDorman, M. F., Parker, J. D. (2001). Trends in births to parents of two different race in the United States, 1971-1995. *Ethnicity and Disease*, 11, 273-278.
- Baker, K. K., Onaka, A. T., Reyes-Salvail, F., Horiuchi, B., Dannemiller, J., Tom, T., et al. (1999). Multi-race health statistics : A state perspective Hawaii health survey (HHS) 1998. Presentation at the National Conference for Health Statistics, August 1999.
- Centers for Disease Control and Prevention. (1993). Use of race and ethnicity in public health surveillance. Summary of the CDC/ATSDR workshop. *MMWR*, 42 (No.RR-10), 5.
- Deapen, D., Liu, L., Perkins, C., Bernstein, L., Ross, R. K. (2002). Rapidly rising breast cancer incidence rates among Asian-American women. *International Journal of Cancer*; 99, 747-50.
- Du Bois, W. E. B. (1906). *The health and physique of the Negro American*. Atlanta, GA: Atlanta University Press.
- Fullilove, M. T. (1998). Comments: abandoning “race” as a variable in public health research – an idea whose time has come. *American Journal of Public Health*, 88, 1297-1300.
- Hahn, R. A. (1999). Why race is differentially classified on U.S. birth and infant death certificates: An examination of two hypotheses. *Epidemiology*, 10, 108-111.
- Haynes, M. A., Smedley, B. D. (Eds.). (1999). *The unequal burden of cancer: an assessment of NIH research and programs for ethnic minorities and the medically underserved*. Washington, D.C.: National Academy Press.
- Heck, K., Parker, J., McKendry, C. J., Schoendorf, K. (2001). Multiple race mothers on the California

- birth certificates 2000. *Ethnicity and Disease*, 11, 303-316.
- Jones, C. P. (2001). Invited commentary: "Race", "racism", and the practice of epidemiology. *American Journal of Epidemiology*, 154, 299-304.
- Kaufman, J. S., and Cooper, R. S. (2001). Commentary: considerations for use of racial/ethnic classification in etiologic research. *American Journal of Epidemiology*, 154, 291-298.
- Krieger, N. (2000). Counting accountably: implications of the new approaches to classifying race/ethnicity in the 2000 Census. *American Journal of Public Health*, 90, 1687-1698.
- La Veist, T. A. (1996). Why we should continue to study race...but do a better job: an essay on race, racism, and health. *Ethnicity and Disease*, 6, 21-29.
- National Cancer Institute. (1997). Statement from the president's panel, 2000. The meaning of race in science – considerations for cancer research. Retrieved November 29, 2003, from <http://deainfo.nci.nih.gov/ADVISORY/pcp/archive/pcp0497/minutes.htm>
- Negrin, K. (1992). Collecting ethnic data: an old dilemma, the new challenge. Retrieved November 29, 2003, from <http://www.eumap.org/articles/content/92/924>.
- Oppenheimer G. M. (2001). Paradigm lost: race, ethnicity, and the search for a new population Taxonomy. *American Journal of Public Health*, 91, 1049-1055.
- State of Hawaii. (2001). The State of Hawaii data book 2000. Retrieved November 29, 2003, from <http://www.state.hi.us/dbedt/db00/db2000.pdf>
- Thomas, S. B. (2001). The color line: Race matters in the elimination of health disparities. *American Journal of Public Health*, 91, 1046-1048.
- U.S. Department of Health and Human Services. (2000). Common data elements implementation guide, version 2.4. Retrieved November 29, 2003, from <http://www.cdc.gov/data/impv24.pdf>
- U.S. Department. of Health and Human Services. (2000). Healthy people 2010: Conference edition.
- US Office of Management and Budget. (1997). Revisions of the standards for the classification of federal data on race and ethnicity. Retrieved November 29, 2003, from <http://www.whitehouse.gov/omb/fedreg/ombdir15.html>
- Walters, M. (2000). Immigration, intermarriage, and the challenges of measuring race and ethnic identities. *American Journal of Public Health*, 90, 1735-37.
- Williams, D. R., Lavizzo-Mourney, R., Warren, R. C. (1994). The concept of race and health status in America. *Public Health Reports*, 109, 26-41.

### **Acknowledgements**

This project was made possible by funding from the Hawaii Tobacco Settlement Funds and the Hawaii Department of Health Preventive Health and Health Services Block Grant.

#### Author Information

Catherine A. Sorensen, Dr. P.H., M.P.H.\*  
Betty Wood, PhD  
Edward W. Prince, MBA  
Hawai`i Department of Health  
Special Informatics Project  
1250 Punchbowl St., Room 422  
Honolulu, HI 96813  
Ph. 808-586-4497  
E-Mail: [casorens@mail.health.state.hi.us](mailto:casorens@mail.health.state.hi.us)

\* corresponding author

## Appendix A

### Hawaii Department of Health Race and Ethnicity Descriptors (Represented in 11 Data sets)

\* Codes

- |   |   |
|---|---|
| 1= Youth Alcohol & Drug Survey                      | 7= TB Rpt Verified                      |
| 2= AIDS Surveillance                                | 8= Hansen's Disease                     |
| 3= Behavioral Risk Factor Survey                    | 9= Vital Statistics                     |
| 4= National Electronic Tracking Surveillance System | 10= STD Surveillance                    |
| 5= Hepatitis  | 11= Youth Risk Behavioral Factor Survey |
| 6= TB Client  |   |

Race/Ethnicity	1	2	3	4	5	6	7	8	9	10	11
Aleute				X							
All Other									X		
All races, Hispanic							X				
American Indian				X	X			X		X	
American Indian/Alaska Native						X					
American Indian or Alaska Native	X	X					X				
America Samoa					X						
Asian			X								
Asian Indian		X	X	X			X		X		
Asian Pacific Islander		X				X	X			X	
Asian Polynesian				X							
Asian Unspecified				X							
Belauan					X						
Black			X	X	X	X		X	X	X	X
Black non-Hispanic		X					X				
Black or African-American	X										
Burmese					X						
Cambodian		X		X	X		X			X	
Caucasian				X	X			X	X		X
Chinese	X	X	X	X	X		X	X	X	X	X
Chuukese							X	X			
Cook Islander					X			X			
Cosovo										X	
Cuban									X		
East Asian				X							
Eskimo				X	X						
Eurasian				X							
Fijian		X		X	X			X			
Filipino	X	X	X	X	X		X	X	X	X	X
French Polynesian					X			X			
Guamanian		X	X	X	X		X	X		X	X
Hawaiian	X				X		X	X	X	X	
Hawaiian Part-Hawaiian		X	X	X							X
Hispanic		X					X	X		X	

Race/Ethnicity	1	2	3	4	5	6	7	8	9	10	11
Hispanic or Latino	X										
Indian					X				X		
Indo-Chinese	X										
Indonesian		X			X		X			X	
Iranian										X	
Japanese	X	X	X	X	X		X	X	X	X	X
Kiribatian								X			
Korean	X	X	X	X	X		X	X	X	X	X
Kosrean								X			
Laotian		X		X	X		X	X			
Malay					X						
Maori					X						
Marshallese		X		X	X		X	X			
Melanesian Unspecified				X							
Mexican			X		X					X	
Micronesian		X			X			X		X	
Micronesian Unspecified				X							
Mid East Arabian				X							
Mixed Asian Asian				X							
Mixed Non-Hawaiian		X	X								
Mixed Race				X							
Native American Unspecified				X							
Nauruan								X			
New Caledonian								X			
Niuean								X			
Non-Hispanic							X				X
Northern Mariana Islander		X						X			
Not Specified		X									
Okinawan					X						
Other			X		X		X	X		X	X
Other Asian	X	X	X	X				X	X		
Other East Asian				X							
Other Melanesian				X							
Other Micronesian				X							
Other Native American				X							
Other Pacific Islander	X	X	X	X			X	X			
Other Polynesian			X								
Other Race				X							
Other South East Asian				X							
Pacific Islander Unspecified				X							
Palauan		X				X	X				
Papua New Guinean								X			
Part Hawaiian								X		X	
Pohnpeian					X		X				
Polynesian Unspecified				X							
Ponapean								X			
Portuguese	X		X	X	X				X		
Puerto Rican			X		X				X		

<b>Race/Ethnicity</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>
Saipanese							X				
Samoan	X	X		X	X		X	X	X	X	X
Samoan/Tongan			X								
Solomon Islander								X			
South Asian Unspecified				X							
South East Asian Unspecified				X							
Tahitian		X									
Taiwanese					X						
Thai				X	X						
Thailand		X						X		X	
Tokelauan								X			
Tongan		X		X	X			X		X	
Tuvaluan								X			
Unknown				X		X	X	X	X	X	
Unknown/Unsure			X								
Unstated									X		
Vanuatuan								X			
Vietnamese	X	X	X	X	X		X	X	X	X	
Wallisan and Futunian								X			
West Samoan					X						
White	X		X			X				X	
White non Hispanic		X						X			
Yapese							X				
<b>Total # Race or Ethnic Descriptors by Program</b>	<b>15</b>	<b>29</b>	<b>21</b>	<b>44</b>	<b>36</b>	<b>6</b>	<b>27</b>	<b>40</b>	<b>18</b>	<b>24</b>	<b>11</b>

\* Codes

1= Youth Alcohol & Drug Survey  
 2= AIDS Surveillance  
 3= Behavioral Risk Factor Survey  
 4= National Electronic Tracking Surveillance System  
 5= Hepatitis  
 6= TB Client

7= TB Rpt Verified  
 8= Hansen's Disease  
 9= Vital Statistics  
 10= STD Surveillance  
 11= Youth Risk Behavioral Factor Survey

**Appendix B**  
**Hawaii Model**  
**Standard for Clustering Race/Ethnicity Categories**

RACE	ETHNIC GROUPS			
	Aggregated Ethnic Groups		Discrete Ethnic Groups	
White (non-Hispanic)	European		Portuguese	
	Middle Eastern or North African			
	Arab			
Black (non-Hispanic)	African American	Haitian		
	African	Jamaican		
	Barbadian	Tabagoan		
	Dominican	Trinidadian		
		West Indian		
American Indian/Alaska Native	American Indian			
	Alaska Native			
Asian	Japanese (includes Okinawan, Iwo Jiman)			
	Filipino			
	Chinese (includes Taiwanese)			
			Korean	Indian
			Vietnamese	Sri Lankan
			Laotian	Maldavian
			Cambodian	Asian
			Indonesian	Eurasian
			Madagascar	East Asian
			Nepalese	Mixed Asian Asian
		Asian Other	Burmese	Other Asian
			Asian Indian	Asian Unspecified
			Bangladeshi	Asian Polynesian
			Bhutanese	Other East Asian
			Thai	South Asian Unspec.
			Hmong	Asian Pacific Islander
			Indo-Chinese	Paskistani
			Other Asian Pacific Islander	
			Southeast Asian Unspecified	
			Other South East Asian	
Native Hawaiian or Other Pacific Islander	Native Hawaiian		Hawaiian	Part-Hawaiian
			Hawaiian/Part Hawaiian	
		Polynesian	Samoan	Samoan/Tongan
			Tahitian	American Samoan
			Tongan	West Samoan
			Tokulauan	Cook Islander
			Other Polynesian	Niuean
			Polynesian Unspecified	

RACE	ETHNIC GROUPS			
	Aggregated Ethnic Groups		Discrete Ethnic Groups	
	Other Pacific Islander		Guamanian or Chamorro	
			Chamorro	Belauan
			Palauan	Kiribati
			Chuukese	Kosraen
			Marshallese	Guamanian
			Tuvaluan	Maori
			Wallisan & Fortuna	French Polynesian
		Micronesian	Pohnpeian	Carolinian
			Yapese	Saipanese
			Mariana Islander	
			Northern Mariana Islander	
			Micronesian Unspecified	
			Other Micronesian	
		Melanesian	Papua New Guinean	Fijian
			Solomon Islander	New Hebrides
			New Caledonian	Vanuatan
			Other Melanesian	Nauruan
			Melanesian Unspecified	
		South Pacific Islander	Other Asian Pacific Islander	
			Other Pacific Islander	
			Pacific Islander Unspecified	
	Hispanic or Latino		Spaniard	Central American
			Mexican	South American
			Latin American	Puerto Rican
			Cuban	Dominican
Other	Mixed		Mixed non-Hawaiian	Other Race
			Mixed	All Others
			Non-Hispanic	Other
	Unknown		Unknown/Unsure	Unspecified
			Unstated	Not specified
			Refused	Unsure
			Unknown	