Some Thoughts on Health Surveillance Data, Race, and Population Categorization

Otis W. Brawley, MD

In this issue of CA: A Cancer Journal for Clinicians, Torre and colleagues have published “Cancer Statistics for Asian Americans, Native Hawaiians, and Pacific Islanders, 2016: Converging Incidence in Males and Females.”1 This is one in a series of articles published regularly by the American Cancer Society Surveillance and Health Services Research group on the cancer burden suffered by various racial and ethnic groups. It is done because the American Cancer Society takes seriously its obligation to serve all Americans and to point out suffering from cancer wherever it might be. These data are gathered from state and local cancer registries; from the US Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute; and from national vital registration. It is important to study and compare data grouped by race/ethnicity to shed light on inequalities in the cancer burden that can be addressed through targeted interventions.

However, those of us who study cancer demographics group Americans into racial categories with hesitance.2 Racial data must be interpreted carefully. When misunderstood, they can truly do an injustice to those who suffer unnecessary disparate health care outcomes by distracting from real issues and taking away from support for efforts to achieve health equity.

We worry that the presentation of data by race perpetuates the mistaken beliefs that:

1. Racial differences are due to biologic or genetic differences among populations.
2. The defined populations are homogenous.
3. Racial categories tell the whole story about health care differences.

Racial categorization of populations is more an art than a science and, at times, the motive has been nefarious. Early 18th century categories of Negroid, Caucasoid, and Mongoloid were used to justify European superiority and slavery.2 Today, the US Office of Management and Budget in the White House is charged with defining racial categories to be used for the collection of federal population data concerning employment, economic, and health status.3 The categories are published 2 to 3 years before every decennial census.

The racial categories used by the federal government have changed over the years, sometimes dramatically. Using these definitions, a 65-year-old person from India who has lived in the United States from early childhood has been 4 different races over that period. Americans from India have been categorized as Caucasian, Indian (Asian Subcontinent), Asian Pacific Islander (API), and now Asian American (AA).

The US government currently collects data in 6 broad categories that very loosely parallel with area of origin.3 They are:

- **American Indian or Alaska Native:** A person having origins in any of the original peoples of North, Central, or South America and who maintains tribal affiliations 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 (note that individuals from the Philippine Islands have been recorded as Pacific Islanders in previous data collection strategies).
- **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.

Chief Medical and Scientific Officer, American Cancer Society, and Emory University, Atlanta, GA

**Corresponding author:** Otis W. Brawley, MD, Chief Medical and Scientific Officer, American Cancer Society, 250 Williams Street NW, Atlanta, GA 30303; otis.brawley@cancer.org

**DISCLOSURES:** Otis W. Brawley reports no specific funding and no conflicts of interest.

I acknowledge JoAnn Tsark, MPH, of the ‘Imi Hale Native Hawaiian Cancer Network (U54CA153459), a National Cancer Institute-funded Community Network Program Center.

doi: 10.3322/caac.21346. Available online at cacancerjournal.com
There is one ethnic category: Hispanic or Latino. A Hispanic or Latino is a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race.

In the 2000 and 2010 US Census, people were encouraged to choose from all racial and ethnic categories that applied. This is recognition that some people are multiracial.

The US cancer registries use these categories to collect data on cancer diagnosis and death. They must also rely on the US Census to provide a count of the population in a given geographic area. This is crucial for determining the population denominator as incidence and mortality rates are calculated.

**Mistaken Belief 1**

**Racial Differences Are Due to Biologic or Genetic Differences Among Populations**

The observation that black rates are higher than white rates leads many to the rather naive and simplistic conclusion that it is because blacks are biologically different from whites.4 The Office of Management and Budget has consistently said that its racial/ethnic categories are sociopolitical and not based on biology. The anthropologic community has reiterated the point.4 Ironically, race is used to assess biological outcomes although these categories are not based in biology. Since 1993, the National Institutes of Health have been legally required to compel their grantees to do subset analyses of results by race within clinical trials.6,7

Such legislated pseudoscience can be harmful. Subset analyses are underpowered by definition and are often wrong. The mistaken finding that Zidovudine was not as effective in the treatment of human immunodeficiency virus in blacks compared with whites actually caused many blacks to refuse beneficial treatment.8

**Mistaken Belief 2**

**The Defined Populations Are Homogenous**

The current racial/ethnic categories are not homogenous. These broad categories appear to be categories of convenience. The AA and NHOPI categories appear to be especially heterogeneous. What does an AA of Pakistani origin have in common with an AA of Korean origin? They do not have similar habits, culture, or geography in common. They do not even look alike. The same can be said for the NHOPI category.

From 1970 to just before the 2000 census, the federal government used the broader category API. For a time, Native Hawaiians were not in the API category but were included in the Native American category.

Truth be told, one can look at data for cancer incidence or mortality rates for any of these races and find significant variability. American whites had an aggregate age-adjusted SEER cancer incidence rate of 463 per 100,000 population in 2012. Among whites in metropolitan Detroit, it was 509 per 100,000 and in New Mexico, it was 393 per 100,000.9 Native Americans residing in the southwestern United States have cancer demographics very different from those living in the plains or major cities.

**Mistaken Belief 3**

**Racial Categories Tell the Whole Story About Health Care Differences**

There are a number of ways to categorize populations and to study cancer control. They include sex, socioeconomic status (SES) (education, household income, or insurance status), area of geographic origin, or even area of residence. All of these have been used to show a disparity in disease state or treatment outcome.10

Analysis of state cancer registry data show that the annualized age-adjusted cancer death rate during the period from 2008 to 2012 for US residents was 171 per 100,000 population, whereas it was 204 per 100,000 for residents of Kentucky and 127 per 100,000 for residents of Utah.9 These data are perhaps more important than racial data, because they give us a defined geographic area in which to intensify cancer control efforts and demonstrate an achievable goal.

SES is an incredible force upon cancer risk and outcome. Despite this, there are limited data bases that track health outcomes by household income, education, or insurance status. If we are truly going to document poor outcomes and assess the cause and the corrective interventions needed, then we need better data on SES and outcomes. Some European governments do this very well using an index of “social deprivation.”11

In some innovative analyses of black-white disparities, American Cancer Society epidemiologists have shown that education (an element of SES) is a more important influence on health outcome than race. Similarly, they have shown that insurance status correlates with better outcome. For example, insured patients diagnosed with stage II colorectal cancer have a superior 5-year survival rate compared with uninsured patients diagnosed with stage I disease.

I should point out these observations about SES and outcomes were made using data from the National Cancer Database, a joint program of the Commission on Cancer of the American College of Surgeons and the American Cancer Society. These type of data are not available from federal or state registries.

The racial/ethnic data as collected and published do have some usefulness if viewed in a sociopolitical lens. However, the aggregation of data into broad racial/ethnic categories obscures.
This is especially true among Americans of Asian and Pacific Islander decent. As Torre and colleagues note and I want to reemphasize: “reporting data, in aggregate, masks important differences between very heterogeneous populations.” Figure 2 of the article by Torre et al illustrates this by showing the documented incidence rates for all the major cancers in 10 of the populations that make up AAs and NHOPIs (AANHPIs). There are very different cancer risk profiles and cancer rates among AANHPIs, but these differences are obscured by the aggregation of NHOPIs with AAs, who generally have much lower rates.

There are other challenges that make collection and analysis of data very difficult. They include:

- Many of these populations are small and are scattered throughout the United States.
- The small size of the population can prevent adequate assessment. For example, during the 5-year period studied, there were fewer than 50 stomach cancer cases documented among Cambodian, Lao, or Samoan Americans.
- The populations are changing in size very quickly. The denominator population necessary when calculating rates per 100,000 can change dramatically in just a few years. A small population can double in just a few years between census counts.
- Demographic data, such as racial/ethnic identity, are often incorrect or missing for patients. Misclassification often occurs when health care providers guess a person’s race.

In the case of the AA and NHOPI categories, we note that currently available data are imperfect and old. Although data under the older category API are available up to 2014, incidence data for the 2 categories are only available from 2006 to 2008. Disaggregated mortality data are not publically available at all. We need disaggregated incidence and mortality data, and we need more recent data.

In response to an Institute of Medicine study on population categorization, the US Department of Health and Human Services has developed new standards for collecting data on race and ethnicity that will allow for disaggregation of data for some AANHPI subgroups in the future. This is still less than optimal but needs to be implemented as soon as possible.

In the US, health data categorized by race and ethnicity serve an important purpose in highlighting disparities, in spite of information lost through inaccurate classifications or broad groupings. However, while we do need to focus on and collect data by race and ethnicity, we need to realize the extreme limitations of such data and collect population data based on socioeconomics and other factors that can define disparities and assess interventions. There are cultural factors correlating with race and ethnicity that affect cancer risk. There are also differences in food preparation and consumption, in exercise, and even in how people smoke that affect cancer risk. There are also sociopolitical issues, such as discrimination and social isolation, which affect access to and quality of cancer care (to include preventive care).

We in public health must define the populations in our data bases as clearly as possible. Then, we must also define the problems and the interventions to overcome the problems as clearly as possible. Along the way, we must realize the limitations of our data. No population is homogenous. Race/ethnicity has little to nothing to do with genetics and inherent biology and much to do with extrinsic (sociopolitical) influences on the cause and cure of cancer. Other ways of categorizing populations to include area of geographic origin, area of residence, education level, and income level and focused studies on specific populations have to be considered to get a full picture of disease and the human condition.