ELSEVIER

Problems With the Collection and Interpretation of Asian-American Health Data: Omission, Aggregation, and Extrapolation

ARIEL T. HOLLAND, BA, AND LATHA P. PALANIAPPAN, MD, MS

Asian-American citizens are the fastest growing racial/ethnic group in the United States. Nevertheless, data on Asian American health are scarce, and many health disparities for this population remain unknown. Much of our knowledge of Asian American health has been determined by studies in which investigators have either grouped Asian-American subjects together or examined one subgroup alone (e.g., Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese). National health surveys that collect information on Asian-American race/ethnicity frequently omit this population in research reports. When national health data are reported for Asian-American subjects, it is often reported for the aggregated group. This aggregation may mask differences between Asian-American subgroups. When health data are reported by Asian American subgroup, it is generally reported for one subgroup alone. In the Ni-Hon-San study, investigators examined cardiovascular disease in Japanese men living in Japan (Nippon; Ni), Honolulu, Hawaii (Hon), and San Francisco, CA (San). The findings from this study are often incorrectly extrapolated to other Asian-American subgroups. Recommendations to correct the errors associated with omission, aggregation, and extrapolation include: oversampling of Asian Americans, collection and reporting of race/ethnicity data by Asian-American subgroup, and acknowledgement of significant heterogeneity among Asian American subgroups when interpreting data.
Ann Epidemiol 2012;22:397–405.

KEY WORDS: Asian American, Data Collection, Research Design, Research Methodology.

INTRODUCTION

Asian Americans are the fastest growing racial/ethnic group in the United States, with a population of more than 14 million as of 2010, and projected to grow to nearly 38 million in 2050 (1). The six largest Asian-American subgroups (Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese) comprise approximately 97% of the Asian American population (single race) (2). Asian Americans are also a heterogeneous group, with unique socioeconomic profiles and language abilities (Table 1) (3). There is a wide spectrum of education, household income, and language ability with Asian Indians in the higher and Koreans and Vietnamese in the lower range of these standard sociodemographic indicators. However, Asian-American subgroups are frequently combined into a single Asian category, masking heterogeneity among the subgroups. Data on Asian American health, particularly for the Asian subgroups, are scarce, and many health disparities for this population remain unknown.

Other researchers and advocacy groups have highlighted the importance of collecting and reporting data by Asian-American subgroups (4–7). The Federal Government has recently taken steps to improve the collection of Asian American health data. In 2009, President Obama signed Executive Order 13515, reinstating President Clinton's Executive Order 13125 that established the President's Advisory Commission, the Federal Interagency Working Group, and the Office of the White House Initiative to improve the health, education, and economic status of the Asian-American and Pacific Islander community (8). President Obama highlighted the need to disaggregate data by Asian American subgroup (8).

As of 2010, Section 4302 of the Affordable Care Act requires that all health surveys sponsored by the Department of Health and Human Services (HHS), such as the National Health Interview Survey (NHIS), the National Medical Expenditure Panel Survey, and the National Immunization Survey, include standardized information on race, ethnicity, sex, primary language, and disability status (Table 2) (9). Disaggregation is only the first step in providing meaningful health data for this group. We must also seek to adequately sample Asian subgroups to provide statistically stable estimates across groups and to offer surveys in appropriate languages and through translators to ensure adequate representation of limited English proficiency and low health literacy populations.

From Health Policy Research, Palo Alto Medical Foundation Research Institute, Palo Alto, CA.

Address correspondence to: Latha P. Palaniappan, 795 El Camino Real, Ames Building, Palo Alto, CA 94301. Tel.: (650) 853-4752; Fax: (650) 853-4835. E-mail: lathap@pamfri.org.

Received December 16, 2011. Accepted April 6, 2012.

1047-2797/$ - see front matter
doi:10.1016/j.annepidem.2012.04.001

AEP Vol. 22, No. 6
June 2012: 397–405

Selected Abbreviations and Acronyms

CHIS = California Health Interview Survey
EHR = Electronic Health Records
FDA = Food and Drug Administration
HHS = Department of Health and Human Services
LEP = limited English proficient
NHANES = National Health and Nutrition Examination Survey
NHIS = National Health Interview Survey
NHW = non-Hispanic white

Previous research and recent federal policy changes address the important issue of disaggregation of Asian-American health data by subgroup. However, few papers have examined the errors in interpretation of Asian-American health data, which are ongoing in the literature. This article will address the implications of recent federal policy changes data collection and reporting, as well as identify methods to improve the collection and interpretation of Asian-American health data, focusing on omission, aggregation, and extrapolation.

HISTORY OF DATA COLLECTION FOR ASIAN AMERICANS

A brief history of data collection for Asian Americans will provide context for our discussion of the problems with the collection and interpretation of Asian American health data. The U.S. Census population data are extremely important for providing denominator data for disease incidence and prevalence statistics by race/ethnicity. The U.S. Census Bureau has collected data on race since the first U.S. decennial census in 1790 (10). Race data for Asian Americans were first collected in 1860 for Chinese, in 1870 for Japanese, and other Asian "races" (Filipino, Hindu, and Korean) were added starting in 1910 (10). The U.S. Census has never used race as a purely biological or genetic classification but rather to reflect common social usage. Although demographic information has been collected on Asian-American subgroups for quite some time, the U.S. Census often reports on population characteristics for Asian Americans as a group. In addition, it was not until 2000 that the Census separated Asians and Pacific Islanders in data reports per the Office of Management and Budget directive issued in 1997 (11). Although the separation of Asian from Pacific Islanders is an improvement, differences among diverse Asian-American subgroups are often masked when data are reported for these subgroups as an aggregated group.

National disease and death registries are an important source of information for high-quality monitoring of population health and health disparities. Few studies have examined leading causes of mortality among Asian-American subgroups (12, 13) because few states collect Asian subgroup information on death records (14). Most studies of mortality in Asian-American subgroups have been conducted using California mortality records because of the high concentration of Asian Americans in the geographic region, and availability of Asian subgroup information on this state's death records. Before 2003, only seven states required the reporting of specific Asian racial/ethnic subgroups (California, Hawaii, Illinois, New Jersey, New York, Texas, and Washington) (15). In 2003, the Secretary of the HHS approved the separation of Asian race category from the Pacific Islander race category and added the following Asian subcategories on U.S. death and birth certificates and reports: Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, other Asian (specify) (15). Although these forms now allow for the collection of Asian American subgroup data, it is important that medical examiners and the coroner understand the importance of collecting this data. Coroner misclassification of race/ethnicity on death certificates is known to be greater in Asian Americans, with 13% of deaths misclassified, compared with 7% of deaths for Hispanic, and <1% of deaths for black and Non-Hispanic white (NHW) individuals (16).

The recent passage of the Affordable Care Act requires the Secretary of HHS to establish data collection standards for race, ethnicity, sex, primary language, and disability status. The HHS Standards Workgroup reviewed the Office of Management and Budget data collection standards and the Institute of Medicine report "Race, Ethnicity, and Language Collection: Standardization for Health Care Quality and Improvement" to inform their standards. The categories for HHS data standards for race and ethnicity are determined by the disaggregated Office of Management

TABLE 1. Demographic characteristics of Asian subgroups in the United States (3)

Subgroup	Total population (margin of error)	Foreign-born population (%)	Bachelor's degree or greater* (%)	Median household income (margin of error)	Speak English less than "very well" (64)
Asian Indian	2,495,998 (±40,554)	1,828,381 (73.3%)	1,174,852 (70.1%)	$90,528 (±1,539)	419,161 (22.2%)
Chinese	3,077,783 (±42,257)	2,133,367 (69.3%)	1,119,349 (52.1%)	$68,202 (±1,607)	1,156,236 (46.0%)
Filipino	2,425,697 (±36,981)	1,608,949 (66.3%)	837,167 (48.0%)	$79,840 (±1,157)	352,437 (22.2%)
Japanese	710,063 (±18,966)	407,034 (57.3%)	265,778 (46.3%)	$61,743 (±1,448)	84,658 (24.8%)
Korean	1,344,267 (±26,878)	977,262 (72.7%)	466,816 (49.8%)	$53,887 (±2,431)	461,241 (46.1%)
Vietnamese	1,431,980 (±32,667)	984,626 (68.8%)	266,734 (27.8%)	$55,667 (±1,302)	670,120 (54.8%)

*Based on the population >25 years of age.