



Published in final edited form as:

J Health Care Poor Underserved. 2010 November ; 21(4): 1354–1381. doi:10.1353/hpu.2010.0939.

Methodological Issues in the Collection, Analysis, and Reporting of Granular Data in Asian American Populations: Historical Challenges and Potential Solutions

Nadia Shilpi Islam, PhD[Deputy Director],

New York University (NYU) Center for the Study of Asian American Health and the Research Director of the NYU Health Promotion and Prevention Research Center

Suhaila Khan, MD, PhD[Chief Executive Officer],

SHK Global Health, LLC

Simona Kwon, DrPH,

Director of the NYU B Free Center of Excellence in the Elimination of Hepatitis B Disparities

Deeana Jang, JD[Policy Director],

Asian and Pacific Islander American Health Forum

Marguerite Ro, DrPH[Deputy Director], and

Asian and Pacific Islander American Health Forum

Chau Trinh-Shevrin, DrPH[Director]

NYU Center for the Study of Asian American Health

Abstract

There are close to 15 million Asian Americans living in the United States, and they represent the fastest growing populations in the country. By the year 2050, there will be an estimated 33.4 million Asian Americans living in the country. However, their health needs remain poorly understood and there is a critical lack of data disaggregated by Asian American ethnic subgroups, primary language, and geography. This paper examines methodological issues, challenges, and potential solutions to addressing the collection, analysis, and reporting of disaggregated (or, granular) data on Asian Americans. The article explores emerging efforts to increase granular data through the use of innovative study design and analysis techniques. Concerted efforts to implement these techniques will be critical to the future development of sound research, health programs, and policy efforts targeting this and other minority populations.

Keywords

Asian American; methodology; health disparities; minority health

Asian Americans are among the fastest growing populations in the country, constituting 5% of the total United States (U.S.) population.¹ By the year 2050, there will be 33.4 million residents who identify as Asian only, representing a 213% population increase compared with a 49% increase in the total U.S. population.¹ Despite these trends, Asian Americans remain poorly understood and are among the most understudied racial/ethnic minority groups in the U.S.^{2,3} One stereotype that leads to the invisibility of Asian Americans is the

model-minority myth, a generalization that all Asian Americans are self-sufficient, well-educated, and upwardly mobile. Such generalizations do not account for the tremendous diversity among Asian Americans, or for differences in socioeconomic status, access to resources, migration patterns, and immigration histories that characterize various Asian American ethnic groups.^{4,5}

The Institute of Medicine and National Research Council of the National Academies have issued two reports calling for data collection on patient race, ethnicity, and language as a strategy for improving quality of care and reducing health disparities in racial/ethnic minority groups.^{6,7} Inadequate data collection and a gross representation of the Asian American population continues to be a concern for researchers, policymakers, and advocates. Ghosh's landmark study reported that a Medline and PubMed search of published research between 1986 and 2000 on six major health disparity areas revealed that only 0.01% of articles included any Asian Americans in the study sample.³ The inadequate representation of Asian Americans in national and regional surveys and research studies renders and reinforces inaccurate views of their health status.³

Although there have been significant developments in federally-funded research initiatives and an increase in the peer-reviewed literature since 2000, the existing peer-reviewed literature is not representative of the Asian American population as a whole and fails to reflect its most pressing health issues.⁸ There remains a critical lack of data disaggregated by Asian ethnic groups, primary language, and geographic locations.

Significance of paper

This paper critically reviews methodological issues and identifies challenges and potential solutions to addressing health data collection, analysis, and reporting of Asian Americans. The paper focuses primarily on large national data collection efforts, particularly those with publicly available datasets, including but not limited to the U.S. Census Bureau (Decennial Census, American Community Survey, Current Population Survey),⁹ National Health Interview Survey (NHIS),¹⁰ National Health and Nutrition Examination Survey (NHANES),¹¹ Behavioral Risk Factor Surveillance System (BRFSS),¹² and (at the state level) the California Health Interview Survey (CHIS).¹³ Such large-scale data collection efforts, including federal and private, provide valuable population-based data that drive policy, program, and funding decisions. The capacity to capture data on Asian Americans in these national and state surveys, including sample design, sample size, and mode of data collection is particularly important.

Asian Americans in the U.S

There are approximately 14.9 million Asian Americans in the U.S. Asian Americans represent more than 50 different ethnicities and 100 different languages. Asian Americans are diverse in characteristics such as country of origin, religion, generational status (e.g., U.S.-born versus foreign-born), duration of residence in the U.S., and socioeconomic indicators (see Table 1). Census data indicate that there is a bimodal distribution of Asian Americans by income and other socio-demographic characteristics. For example, the annual median income ranges from \$31,934 (for Hmong) to \$61,322 (for Asian Indians).

Asian Americans are geographically dispersed across the U.S. but primarily concentrated within urban areas. California, New York, and Texas are home to the three largest Asian American populations in the U.S. Smaller cities throughout the U.S., including cities in North Carolina, Georgia, Nevada, Arkansas, Utah, Tennessee, Nebraska, Colorado, Arizona, and Kentucky—states that are not historically significant immigrant settlement areas—have seen some of the largest population increases of Asian Americans in the last decade. Much

of our current state of knowledge on Asian American health is gleaned from data on Asian Americans living primarily in the West Coast in California, Hawaii, and Washington. There is limited literature available on the health of Asian Americans in New York City, home to the largest Asian American population in the U.S., or other locations in the region with large Asian American populations, such as New Jersey, Maryland, or Philadelphia. There are even fewer studies that represent Asian Americans living in the rest of the U.S., in such regions as the Midwest, the Southwest, and the Southeast. This is particularly troublesome given recent demographic trends indicating large subgroups of Asian Americans concentrated in these geographic regions. For example, there are large pockets of Southeast Asians, including the Hmong and Cambodian communities, living in Minnesota; significant numbers of South Asians in urban areas of Texas; and a growing Filipino population in Florida.¹⁴ These demographic shifts and trends should be reflected in current and future data collection efforts in order to capture accurately the social and health needs of this growing and diverse population.

Limitation of large data collection efforts

The major factors hindering data collection and analysis efforts are described below.

Lack of or limited subgroup categorizations—Most surveys lack or have limited subgroup categorizations for Asian Americans. The Census Bureau makes the most comprehensive data collection effort and has made numerous provisions to ensure representation of Asian Americans through outreach efforts, in-language interviewing, subgroup categorization of Asian Americans, and oversampling in some areas. At the state level, CHIS has made similar efforts in California.

Several national surveys only collect limited Asian subgroup information, including NHIS, NHANES, Medical Expenditure Panel Survey (MEPS) (which is linked to NHIS), and the Early Childhood Longitudinal Survey (ECLS). For example, NHIS collects ethnicity data only for six specific Asian American subgroups (Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese). Unfortunately, the sample sizes by ethnic groups are often insufficiently powered to allow for meaningful analysis. Further, the datasets do not provide data for smaller, but rapidly growing Asian American subgroups, such as South Asian (e.g., Bangladeshi, Pakistani, or Sri Lankan) or Southeast Asian (e.g., Cambodian, Thai, Indonesian) populations. Sub-grouping options for *Asian* are not available in BRFSS, National Household Education Survey (NHES), Survey of Income and Program Participation (SIPP), National Survey of Family Growth (NSFG), National Immunization Survey (NIS), or the Medicare Current Beneficiary Survey.¹⁵ Box 1 presents a summary of several major federal datasets and registries and their Asian subgroup categorization.

Box 1

SUMMARY OF DATA COLLECTION METHODS AND ASIAN SUBGROUP CLASSIFICATION IN SELECTED FEDERAL DATA SETS AND REGISTRIES

| Survey | Brief Description | Mode/ Language of Interviewing | Asian Race/ Ethnic Subgroup Classification | Oversample Asians? |
|-------------|--|---|--|--------------------|
| U.S. Census | The decennial census is the oldest data collection effort in the United States. In addition to providing a | In-person interviews. The decennial census makes special efforts to hire indigenous | Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, | no |

| Survey | Brief Description | Mode/ Language of Interviewing | Asian Race/ Ethnic Subgroup Classification | Oversample Asians? |
|--|---|--|--|-----------------------|
| | <p>“snapshot” of the United States, the decennial census provides information at all levels of geography, from the large to the small, ranging from political entities such as states, counties, cities, and local governments, to small areas such as blocks and tracts.</p> | <p>interviewers, especially so in areas containing large numbers of non-English speaking respondents. Where a bilingual interviewer is not immediately available and another family member is unable to bridge the language gap, a callback visit is scheduled and the required language skill is located and made available. Partnerships also are established with the local community and with public interest groups in order to ensure the availability of the needed language skills, and to obtain assistance in seeking public cooperation in responding to the census. Census questionnaires are available in five languages other than English—that is, in Chinese, Korean, Spanish, Tagalog, and Vietnamese upon request. In addition, questionnaire assistance booklets are available in over 30 languages</p> | <p>Other Asian (with write-in option)</p> | |
| <p>American Community Survey (ACS)</p> | <p>The American Community Survey (ACS) is planned as a continuing sample survey designed to replace the Census long form in 2010. It is intended to provide reliable annual estimates of the detailed social, economic, and housing characteristics for all states, and for cities, counties, metropolitan areas, and population groups of 65,000</p> | <p>The data collection operation for housing units (HUs) consists of three modes: mail, telephone, and personal visit. The language assistance program for the American Community Survey (ACS) includes a set of methods and procedures designed to assist sample households with</p> | <p>Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Other Asian (with write-in option)</p> | <p>no</p> |

| Survey | Brief Description | Mode/ Language of Interviewing | Asian Race/ Ethnic Subgroup Classification | Oversample Asians? |
|---------------------------|---|--|--|-----------------------|
| | persons or more. For smaller areas, multi-year average data covering the most recent 2-to-5 years will be used to generate the estimates | limited English proficiency in completing the ACS interview. The program includes assistance in a wide variety of languages during the telephone and personal visit nonresponse follow-up stages. Efforts to expand language assistance in the mail mode were postponed; the current focus in the mail mode is limited to supporting Spanish-language speakers. In 2005, interviewer language capabilities included English, Spanish, Portuguese, Chinese, Russian, French, Polish, Korean, Vietnamese, German, Japanese, Arabic, Haitian Creole, Italian, Navajo, Tagalog, Greek, and Urdu. | | |
| Current Population Survey | The Current Population Survey (CPS) is a monthly survey conducted by the Bureau of the Census to produce the official government statistics on the Nation's employment and unemployment. In March of each year, the survey includes the questions found in the Census decennial long-form | In-person interviews. In areas containing large numbers of non-English speaking respondents, the Bureau generally attempts to locate, hire, and train members of the group who are bilingual, and they are assigned as needed. Where a bilingual interviewer is not available, the interviewer attempts to locate another member of the family who is bilingual to assist in the interview, or arranges to call back when a translator can be obtained. | In connection with the extended March supplement, persons identified as Asian or Pacific Islanders are further classified into subgroups (e.g., Chinese, Japanese, Filipino, Asian Indian, Vietnamese, Guamanian, Hawaiian | no |

| Survey | Brief Description | Mode/ Language of Interviewing | Asian Race/ Ethnic Subgroup Classification | Oversample Asians? |
|---|--|--|---|---|
| Survey of Income and Program Participation (SIPP) | The Survey of Income and Program Participation (SIPP), is a major continuing household survey, providing information on the detailed sources of income, on participation in a wide range of government programs, and on program eligibility. | In-person interviews. The survey is administered through a computer assisted personal interview (CAPI). In areas containing large numbers of non-English speaking respondents, the Census Bureau generally attempts to locate, hire, and train members of the group who are bilingual, and they are assigned as needed. Where a bilingual speaker is not available, the interviewer attempts to locate another member of the household who is bilingual to assist in the interview, or arranges for a callback with a bilingual interviewer (or translator). | Asian/Pacific Islander Race asked only; None of the specific subgroups comprising Asians or Pacific Islanders is collected | no |
| National Health Interview Survey (NHIS) | The National Health Interview Survey is based on a sample of the civilian non-institutionalized population of the United States on American health indicators, health care access and use, and health-related behaviors. | In-person interviews. The NHIS policy regarding the use of bilingual interviewers parallels that of the Bureau of the Census, given that the survey is conducted for NCHS by the Bureau. Bilingual interviewers are recruited routinely for those areas known to be predominantly non-English speaking, with Spanish as the most important second language. Where feasible, other members of the household who are bilingual are asked to assist. Other language skills are provided, as the situation requires successful, when the respondent requests a | In a question regarding race, respondents are asked to classify themselves as: Chinese, Filipino, Japanese, Korean, Asian Indian, and Vietnamese; or Other Asian Pacific Islander | Over-samples were conducted as part of the 2006–2009 NHIS surveys |

| Survey | Brief Description | Mode/ Language of Interviewing | Asian Race/ Ethnic Subgroup Classification | Oversample Asians? |
|---|--|--|---|---|
| | | telephone interview, when part of the interview needs to be completed and it is not possible to schedule another personal visit, or when road conditions or travel distances would make it difficult to schedule a personal visit before closeout. | | |
| National Vital Statistics System — Natality/Mortality | A national database of all birth and death records collected in conjunction with local registration offices. | Chart Review (no interviewing required) | The race of the mother and father (but not child) is obtained for natality records. The following categories for race are separately identified: Chinese, Japanese, Hawaiian, Filipino, and other Asian/Pacific Islander. a total of nine states, which contain about two-thirds of the U.S. population of these additional API groups, code births as Vietnamese, Asian Indian, Korean, Samoan, Guamanian, and other API groups. The same racial categories are used for mortality records | N/A |
| National Survey of Family Growth (NSFG) | The National Survey of Family Growth (NSFG) is focused on factors affecting pregnancy—including sexual activity, contraceptive use, and infertility—the use of family planning and other | In-person interviews. Some Spanish-English bilingual interviewers are hired and made available as needed. Respondents who cannot be interviewed in English or in | Only Asian or Pacific Islander Race is recorded | no. The emphasis in this study has been on Black and Hispanic reproductive health |

| Survey | Brief Description | Mode/ Language of Interviewing | Asian Race/ Ethnic Subgroup Classification | Oversample Asians? |
|------------------------------------|--|--|---|-----------------------|
| | medical services such as prenatal care, and the health of women and infants | Spanish are classified as eligible, but non-respondents. Because of the sensitive content of the interview, family members or other third party translators are not allowed to be present during the interview. Thus, if an eligible person speaks only other than English or Spanish, that person cannot be interviewed in the NSFG. | | |
| National Immunization Survey (NIS) | The National Immunization Survey (NIS) collects specific vaccination data for children 19 through 35 months of age | List-assisted random-digit-dialing telephone survey followed by a mailed survey to children's immunization providers. English- and Spanish-speaking interviewers are used to collect the information. In addition, selected other language skills are available. A Spanish-language version of the questionnaire is available, when required. Where the required language skills are not available, an effort is made to obtain the assistance of an English-speaking family member or the AT&T language line translators are used | Only Asian, Native Hawaiian, or Other Pacific Islander race is collected | no |
| NHANES | The National Health and Nutrition Examination Survey (NHANES) obtains information about the health and nutritional status of a representative national sample of the civilian, non-institutionalized | Health interviews are conducted in respondents' homes. Examinations are performed in specially-designed and equipped mobile examination centers, which travel to survey locations | In a question regarding race, respondents are asked to classify themselves as: Chinese, Filipino, Japanese, Korean, Asian Indian, and Vietnamese; | no |

| Survey | Brief Description | Mode/ Language of Interviewing | Asian Race/ Ethnic Subgroup Classification | Oversample Asians? |
|--|--|---|--|-----------------------|
| | population of all ages through direct interviews, physical examinations which obtain a wide variety of standardized medical information, and selected laboratory analyses | throughout the country. The survey team consists of a physician, medical and health technicians, dietary and health interviewers. A large staff of trained bilingual interviewers conducts the household interviews. (Spanish only) | or Other Asian Pacific Islander | |
| Medical Expenditure Panel Survey (MEPS) | The Medical Expenditure Panel Survey (MEPS) is a nationally representative survey of health care use, medical expenditures, sources of payment, and insurance coverage for both the U.S. civilian non-institutionalized population and nursing homes and their residents. Both individual and family level information on health care utilization and expenses are collected | In-person survey. Bilingual interviewers, especially Spanish-speaking, are used regularly. Other language skills are located as required. The CAPI system contains a Spanish-language version of the interview form. | Since MEPS draws its sample from persons interviewed previously in the National Health Interview Survey, the race/ethnicity data collected during the NHIS interview are available | no |
| Medicare Current Beneficiary Survey (MCBS) | The objective of the study is to determine expenditures and sources of payment for all services used by Medicare beneficiaries, including co-payments, deductibles, and noncovered services; to ascertain all types of health insurance coverage and relate coverage to sources of payments; and to trace processes over time, such as changes in health status, spending down to Medicaid eligibility, and the impacts of program changes | In-person survey. The interviewing staff includes resident, bilingual interviewers, especially in highly concentrated Spanish-speaking populated areas, such as California, Florida, Texas and Puerto Rico. When the necessary language skills are not immediately available, translators are obtained. | Asian or Native Hawaiian or other Pacific Islander assessed only | no |
| National Household | The National Household Survey | In-person survey. The CAPI system | In a question regarding | no |

| Survey | Brief Description | Mode/ Language of Interviewing | Asian Race/ Ethnic Subgroup Classification | Oversample Asians? |
|--|---|---|--|-----------------------|
| Survey on Drug Abuse (NHSDA) | on Drug Abuse (NHSDA) provides statistical information on the use of illegal drugs, collected through interviews with a national household sample of persons 12 years old and older | displays both English and Spanish language versions of the interviewer form, and the staff contains interviewers with English/Spanish language capability. Other household members are asked to assist when bilingual capability is not available, or arrangements are made to callback in order to obtain the interview. | race/ethnicity, respondents are asked to classify themselves as: Chinese, Filipino, Japanese, Korean, Asian Indian, and Vietnamese; or Other | |
| National Household Education Survey (NHES) | The National Household Education Survey (NHES) is designed to provide information on selected educational issues that are best addressed by contacting households directly, rather than schools or other educational institutions | Telephone survey. NHES is conducted in English and in Spanish, as required. The questionnaires are available on the CATI system in a Spanish language version, with bilingual interviewers trained to complete the interview in either English or Spanish. Telephone surveys may be answered by someone who does not speak English; if the interviewer is not bilingual in the language of the respondent, such cases are noted by the interviewer as "language problem" and, if the language is recognized, it is recorded. If the initial interviewer is functional in the respondent's language (usually Spanish), the interview is immediately carried out. In cases involving "language problem," efforts are made to identify and locate an English (or Spanish) speaking | Asian or Pacific Islander race assessed noly | no |

| Survey | Brief Description | Mode/ Language of Interviewing | Asian Race/ Ethnic Subgroup Classification | Oversample Asians? |
|---|---|--|--|-----------------------|
| | | household member to assist with the interview; failing that approach, translators or persons with the unique language skill are used to complete the interview. | | |
| Early Childhood Longitudinal Study—Birth Cohort | The Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), is designed to track the development of 15,000 children born in the year 2000 through their first grade year of school. the primary objectives of ECLS-B are: 1) To understand the growth and development of children in critical domains; 2) To understand how children transition to out-of-home programs and to school; and 3) To understand children's school readiness. | In-person survey. If the parent or provider feels more comfortable in a language other than English or Spanish, then translators are utilized when available. Additionally, at 9-months and 2-years, the child assessments were administered in the child's primary language either by a bilingual interviewer or with a translator. | Race/ethnicity for the parents will be collected directly from the Birth Certificate. Since the certificate does not record the child's race, the child will be assigned the mother's race. Respondents that select Asian race are asked to further classify themselves as Chinese, Filipino, Japanese, Korean, Asian Indian, Vietnamese, or Other | yes |
| Early Childhood Longitudinal Study—Kindergarten | The Early Childhood Longitudinal Study, Kindergarten Class of 1998–1999 (ECLS-K) is designed to track the performance of some 22,000 children from kindergarten through the fifth grade. Its objective is to study the "whole child," including health, social and emotional development, and educational experiences | Data were collected in a variety of formats, including one-on-one assessments, computer-assisted telephone interviews (CATI), and self-administered paper and pencil questionnaires. Language minority children will be identified through a Home Language Survey; classroom teachers also will be asked questions about the child's home language. Children with home language other than English will complete the Oral Language Development | Information is collected from the child; from parents and guardians; from teachers; and from school administrators. Children that are reported as Asian race are asked to be further classified as Chinese, Filipino, Japanese, Korean, Asian Indian, Vietnamese, or Other | yes |

| Survey | Brief Description | Mode/ Language of Interviewing | Asian Race/ Ethnic Subgroup Classification | Oversample Asians? |
|---|---|--|---|-----------------------|
| | | Scale (OLDS). Above a given score, children will be assessed in English; Hispanic children below the score will be assessed using Spanish language subtests. Bilingual interviewers are available as needed for all other interviews and, to the extent possible, in the required languages. For languages other than Spanish, the interviewer does direct translation for the respondent. | | |
| Behavioral Risk Factor Surveillance System (BRFSS) 2007 | The BRFSS is a cross-sectional telephone survey conducted by state health departments with technical and methodological assistance provided by the CDC. Every year, states conduct monthly telephone surveillance using a standardized questionnaire to determine the distribution of risk behaviors and health practices among noninstitutionalized adults. The states forward the responses to the CDC, where the monthly data are aggregated for each state. The data are returned to the states, then published on the BRFSS Web site. Data are collected monthly in all 50 states, the District of Columbia, Puerto Rico, the U.S. Virgin Islands, and Guam. | Telephone survey. 53 states used computer-assisted telephone interviewing (CATI). The survey is administered in English and Spanish. Some states utilize the ATT interpretation line for other languages when necessary | Only Asian Native Hawaiian or Other Pacific Islander assessed | no |
| SEER | The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI) is an authoritative source | N/A | Based on Asian race and country of birth | N/A |

| Survey | Brief Description | Mode/ Language of Interviewing | Asian Race/ Ethnic Subgroup Classification | Oversample Asians? |
|--------|--|--------------------------------------|---|-----------------------|
| | of information on cancer incidence and survival in the United States. SEER currently collects and publishes cancer incidence and survival data from population-based cancer registries covering approximately 26 percent of the US population. SEER coverage includes 23 percent of African Americans, 40 percent of Hispanics, 42 percent of American Indians and Alaska Natives, 53 percent of Asians, and 70 percent of Hawaiian/Pacific Islanders. | | | |

Inconsistent/non-standardized definitions of Asian American—Aggregate classification of *Asian Americans* raises questions about the comparability of data. For example, a survey may present data on *Asian Americans* but include only Chinese and Korean respondents. Similarly, surveys may include samples with a variety of Asian American subgroups; however, particular subgroups may be overrepresented due to which groups live in the place where sampling occurred.

Additionally, surveys and studies use inconsistent methods of classifying Asian American subgroups, often based on ethnicity or country of birth. Data may be collected by directly asking a respondent for his or her ethnic identity. When ethnic identity information is not captured, country of birth and/or parent's country of birth may be used as a proxy for ethnicity. However, ethnic identification and nativity may not be synonymous when considering the Asian American diaspora, which includes such individual mixtures as Chinese born in Vietnam or Asian Indians born in Malaysia. Lin and colleagues underscore this substantial methodological challenge within the Surveillance, Epidemiology and End Results (SEER) datasets, an important resource for the study of cancer incidence and mortality, where information on birthplace is often used as a proxy for immigrant status and ethnic identity.² Similarly, Waksberg and colleagues reported that there is considerable variation in the way Asians are asked to describe themselves in federal surveys.¹⁵

Limited data collection in Asian languages—Most national surveys collect data only in English and Spanish. With the exception of several surveys administered by the U.S. Census Bureau, there is currently no national federally-sponsored data collection effort where survey administration is consistently conducted in a language other than English or Spanish. The policy for a surprising number of national data collection efforts is to use family members for interpretation if necessary. Additionally, some surveys (such as the NSFG) classify non-English or Spanish individuals as “non-respondents.”¹⁵ Sampling bias

may result from the high rates of limited English proficiency (LEP) and linguistic isolation experienced in these communities. Limited English proficiency is defined as the inability to speak, read, write or understand English “very well.” Similarly, linguistically isolated households are defined by the Census Bureau as households where no person 14 years of age or older speaks English very well. According to the Census 2000, more than one-third of Asian Americans identified themselves as LEP, a rate four times higher than the general population. Limited English proficiency rates are especially high among particular Asian American subgroups. Similarly, about one-quarter of the Asian American population live in linguistically isolated households. These rates vary considerably across subgroups, ranging from rates as low as 10% in the Filipino population to as high as 45% in the Vietnamese population.¹⁴ Additionally, relying on family members for interpretation of health information is problematic.¹⁶

Limited English proficiency and linguistically isolated individuals often have lower socioeconomic status and poorer access to health care, and suffer from a larger burden of health disparities and inequities compared with individuals who speak English fluently or very well or those who do not live in linguistic isolation. Thus, a majority of the national surveys may be underestimating the prevalence of chronic illness and health care barriers for this population. Box 1 portrays the language in which interviews are administered for several major federal studies.

Uneven distribution of geographic representation—Concern has also been expressed for the lack of geographic representation in existing surveys on Asian American health disparities and inequities.¹⁷ For example, the SEER registry/dataset, the only federally-funded national registry of cancer incidence in the U.S., currently has sites located in a limited number of regions, such as California.* The SEER registry has no sites in New York despite New York’s substantial Asian American population. Indeed, the only SEER registry sites on the East Coast are located in Connecticut, which includes a small Asian American population, and New Jersey, which was added as a SEER site in late 2001. Consequently, this supposedly national dataset over-represents cancer data from the West Coast for Asian Americans.¹⁷

Small sample sizes and oversampling of Asian Americans or ethnic subgroups—Studies that include Asian Americans within their sample often suffer from small sample sizes, precluding the inclusion of Asian Americans in the final analyses. In their assessment of major federal data sets on minority populations, Waksberg and colleagues found that analysis by Asian or Pacific Islander subgroup is not possible for a majority of federal data sets.¹⁵ The national vital statistics records, Census and the American Community Survey (ACS) are the only federal data sets where detailed cross-classification is possible for Asian or Pacific Islander subgroups. It is important to note that the Census and the ACS capture demographic data but very little information related to health and or health care.

In studies where there are sufficient samples of Asian Americans overall, it is often not possible to conduct meaningful subgroup analyses due to insufficient sample sizes by Asian ethnic group. For example, the sample size of individuals reporting Asian race in the 2005 NHIS sample was 3,748. The sample sizes for the subgroups available in NHIS, however, ranged from 261 to 704, which are sufficiently powered for bivariate analyses but are inadequate for more sophisticated multivariate models.¹⁸ Even within the same survey,

*Current SEER sites include: Connecticut, Iowa, New Mexico, Utah, New Orleans, Louisiana, New Jersey, Puerto Rico, and Hawaii and the metropolitan areas of Detroit and San Francisco–Oakland, Atlanta, Seattle–Puget Sound, Los Angeles County, and four counties in the San Jose–Monterey area south of San Francisco.

sample sizes can differ dramatically across years, with the trend of sample sizes decreasing over time. For example, NHIS interviewed 27% fewer house-holds in 2007 than in 2005 (45,000 in year 2005; 33,000 in year 2007).¹⁰

There have been notable exceptions in non-federal data collection efforts where adequate sample sizes have been collected. The National Latino and Asian American Study (NLAAS) funded in 2002 is a national probability sample of 2,095 Asian Americans ages 18 and older, residing in the U.S. The survey was administered in Cantonese, English, Mandarin, Spanish, Tagalog, and Vietnamese. The survey oversampled Chinese, Vietnamese, and Filipinos, but also includes other Asian (e.g., Bangladeshi) and Pacific Islanders (e.g., Samoans). The study utilized novel sampling and survey administration techniques to address challenges in data collection for racial and ethnic minority groups, which are described elsewhere.¹⁹ The CHIS is a population-based survey study modeled after the NHIS conducted in California and was designed especially to capture data from hard-to-reach ethnic subgroups. The survey is administered in Cantonese, English, Korean, Mandarin, Spanish, or Vietnamese. For each wave of data collection, respondents living in particular urban locations and particular racial groups (including Asian Americans) are oversampled.

Limitations related to RDD sampling—Most data collection efforts employ telephone-based random digit dialing (RDD) methods primarily. There is growing evidence regarding the potential biases of RDD. Those without a landline phone are more likely to be members of a minority group, less educated, more likely to report worse self-rated health, poor, new immigrants, and younger adults.^{20–27} Potential biases include non-coverage of households without a landline telephone,²⁸ lower survey response rates,²⁹ and non-coverage of some household members.^{30,31} Although growing non-coverage of households without telephones may not be a major concern in the general U.S. population as 86% of U.S. households have at least one landline, non-coverage in telephone surveying among Asian Americans has not been addressed.^{32,33} Box 1 contains information on major federal studies that utilize telephone versus in-person surveying methods.

Lack of implementation of OMB standards—In 1997, the Office of Management and Budget (OMB) revised the standards for collection of race and ethnicity data by the federal government, and required that data on Asian Americans be collected separately from data on Native Hawaiians and Other Pacific Islanders (NHPI). Following the revision, the U.S. Department of Health and Human Services (DHHS) adopted its Policy Statement on Inclusion of Race and Ethnicity in DHHS Data Collection Activities.³⁴ That policy clarified that the OMB standards do not require that race and ethnicity data be collected and reported, but that HHS's policy is that data on race and ethnicity be included in data collection and reporting activities. Because this policy is not a requirement and only applies to HHS's data collection activities, most recipients of HHS funding, including states and the private sector, are not required to collect or report these data. Furthermore, this policy does not require HHS programs to collect data on primary languages spoken by the beneficiaries of HHS services and programs. The Patient Protection and Affordable Care Act of 2010 signed into law on March 23, 2010 includes a provision requiring that within two years after the date of enactment of the law that “any federally conducted or supported health care or public health program, activity or survey (including Current Population Surveys and American Community Surveys conducted by the Bureau of Labor Statistics and the Bureau of the Census) collects and reports, to the extent practicable” data on race, ethnicity, sex, primary language, and disability status for applicants, recipients, or participants.³⁵ The law further requires that sufficient data be collected to generate statistically reliable estimates by race, ethnic, sex, primary language, and disability status subgroups and, if needed, statistical oversamples of these subpopulations. While this would seem to be a huge victory and greatly improve and increase the availability of subpopulation data, the law includes a

provision that the data may not be collected under this section unless funds are directly appropriated to collect the data.

Medicare data have proven to be a rich source of information about racial, ethnic, and socioeconomic disparities in health and health care among beneficiaries. An analysis of 2002 Medicare administrative data, however, shows that only 52% of Asian beneficiaries and 33% of both Hispanic and American Indian/Alaska Native beneficiaries were identified correctly.³⁶ Medicare obtains its race and ethnicity data from the Social Security Administration (SSA) which collects the data when an individual applies for a social security number. Due to various changes in the way that SSA collects these data and when they collect the data, the data are either not available, not collected, or not collected in a manner consistent with the OMB revised standard.

Lack of or inconsistent reporting of race and ethnicity data in disease registries, health plans and hospitals—Reporting of race and ethnicity variables within disease registries, health plans, or hospitals is inconsistent or does not occur.^{37,38} Accurate reporting of race and ethnicity data in these settings is important both for enhancing patient care and for providing valuable estimates of disease prevalence and health outcomes in particular populations. The issue of reporting of Asian American race and Asian American subgroup data within these settings is understudied. However, some work indicates that Asian Americans are often misclassified or misidentified within hospital settings. For example, studies using Department of Veterans Affairs data compared patient self-reports with administrative reporting of race and ethnicity; they observed that Asians experience some of the highest rates of misclassification or classification as “unknown” race.^{39,40} Another concern is that hospital databases, disease registries, or clinical trial registries may use inconsistent definitions of Asian Americans. Similarly, reporting of race and ethnicity in the national vital statistics records is inconsistent. The National Vital Statistics record the identification of Chinese, Japanese, Hawaiian, and Filipinos in all 50 states, but identify the other ethnic groups—including Vietnamese, Asian-Indian, Korean, Samoans, and Guamanians—in only nine states, which contain about two-thirds of the U.S. population in each of these groups.¹⁵

Potential Solutions to Obtaining Disaggregated Data on Asian Americans

Pooling/linking of datasets

Researchers have pooled datasets across several years in order to create sufficiently powered samples that are large enough to allow for meaningful analysis of health issues among Asian American subgroups. Barnes and colleagues pooled three years (2004–2006) of NHIS data from the Family Core and the Sample Adult Core components, yielding a final sample size of 87,029 completed interviews with adults aged 18 years in order to increase reliability of estimates for some of the smaller population subgroups. Findings from the analysis confirmed that aggregating data for all Asian Americans masks significant variation that exists between Asian Americans subgroups with respect to health access, utilization, behaviors, and outcomes. For example, rates of uninsurance ranged from 12% in the Japanese population to 25% in the Korean population.⁴¹ The Kaiser Family Foundation conducted an analysis of health care coverage access among Asian Americans and NHPs that pooled three years (2004–2006) of data from the Current Population Survey (CPS).⁴² The analysis demonstrated that aggregate data on Asian Americans provides an inaccurate picture of true disparities in health care access and coverage faced by this population. For example, rates of uninsurance ranged from 12% in the Asian Indian community to 31% in the Korean community. However, for many federal data collection efforts, pooling data across five years will only meet minimal precision standards for Asian American subgroups.

Implementing the Institute of Medicine's recommendations on the standardization of race and ethnicity data

In August 2009, the Institute of Medicine released a report that identified standardized categories for the variables of race, ethnicity, and language that can be used to facilitate the sharing, compilation, and comparison of high-quality data.⁴³ As a result of the report, a series of recommendations were made to facilitate the collection of improved data on race and ethnicity. Importantly, the report highlighted the collection of “granular ethnicity data—defined as a person’s ethnic origin or descent, roots, heritage, or the place of birth of the person or the person’s parents or ancestors.”⁴³ The report enumerated several strategies to ensure the standardization of this type of granular data. Second, the report cited the need to collect better data on the role of language proficiency on health care access, and suggested that data collection entities, at minimum, should assess how well a respondent speaks English and the preferred language of communication. Third, the report recommended that when directly collected race and ethnicity data are not available, entities should use indirect estimation strategies—for example, geocoding or surname analysis—to assess race or ethnicity. Finally the report concluded with several strategies that federal agencies might use to ensure standardized collection of race and ethnicity data.⁴³

Using innovative sampling strategies (e.g., oversampling, targeted, respondent-driven, cumulations, add-ons)

Innovation in the design of studies can lead to improvements in the effective sample size of Asian Americans in research study. This can be achieved through targeted and efficient increases in sample sizes for these groups. These strategies have been successfully applied in NLAAS and CHIS. Other regions around the country have also begun to establish state or city-based data collection efforts that provide population-based data on Asian Americans. Examples include the New York City Community Health Survey, the New York City NHANES survey, and others. Additionally, Waksberg and colleagues note that many federal data collection efforts could benefit from supplementing their A/PI samples. Although the authors admit that this can often be a costly procedure, they highlight several surveys that can be supplemented at minimal additional cost.¹⁵

A few studies have explored sampling issues (such as precision sample size) related to data collection, analysis, and reporting of small populations. Most have reported that most national health surveys cannot measure the health of racial/ethnic subgroups for Asian Americans.¹⁵ One study on Chinese and Native Americans/Alaskan Indians suggested three promising sample design strategies for NHIS data. These sample design strategies include complete sampling of targets within households, oversampling selected macro-geographic units, and oversampling from an incomplete list frame. The study reported that all three strategies provide effective sample sizes for the Chinese population.⁴⁴

Innovative community-based sampling strategies, such as targeted sampling and respondent-driven sampling^{45,46} are appropriate for capturing data on hard-to-reach populations, particularly for addressing community-level interventions.⁴⁷ Targeted sampling, particularly venue-based sampling⁴⁸ has been successfully used in HIV-related men who have sex with men (MSM) studies and is advocated as an evidenced-based method to engage hard-to-reach populations by the CDC’s National HIV Behavioral Surveillance program.⁴⁹ Respondent-driven sampling is another approach for locating and sampling hard-to-reach populations. Biases can result, however, from oversampling respondents with larger personal networks. Incorporating a randomized sampling frame can reduce biases in different types of community-based sampling.⁵⁰

Few studies have tested the effectiveness of these different sampling strategies in Asian American communities. However, some studies suggest that there are no significant differences between different sampling strategies. Ngo-Metzger and colleagues distributed surveys to Vietnamese and Chinese patients with LEP from community health centers that were randomly assigned to either receive a telephone or mail survey.⁵¹ Both data collection methods were shown to be feasible and to produce statistically similar results. A comparison of random sampling using area-based sampling of Chinese and Korean participants through telephone surveying and convenience sampling through ethnic venues indicated no large, statistically significant differences in participants' demographic characteristics.⁵² The study authors concluded that convenience sampling at well-chosen ethnic venues can yield a representative sample.

Another successful example is the B Free National Center of Excellence in the Elimination of Hepatitis B Disparities (B Free CEED) community health information survey data set. The goal of this survey was to collect information from a target community of Korean and Chinese immigrants in metropolitan New York City. Using a community-based participatory approach, the partner members of B Free CEED collectively decided to utilize a combination of street intercept and venue-based sampling. Potential aggregation points (e.g., high-traffic areas of the target population, community centers, religious institutions, and commerce areas frequented by the target population) were identified. Environmental scans were conducted of the aggregation points by fieldworkers and a final list of three-to-four street intercept and three-to-four venue-based locations for each the Chinese and Korean target groups were selected. To address issues of sampling bias, a randomization plan was implemented. Interviewers were trained to approach and interview every third eligible person who crossed a pre-determined recruitment marker. After the completion of an interview, the next third person to cross the recruitment marker was eligible for participation. Systematic information on sex, age group (i.e., young adult, middle-aged, or senior), sub-ethnic group, time of refusal, and reason for refusal was collected for all individuals who refused participation. This allowed for an estimation of the survey sample and provided descriptive data on the population that refused to participate in the survey. Using this methodology, B Free CEED has collected nearly 600 surveys in the target population. As for reaching the target population, 96% of the Korean respondents and 94% of the Chinese respondents reported that they were born outside the U.S. Moreover, over 30% of those individuals surveyed did not have a landline telephone, indicating that a significant portion of this population would not have been captured via the majority of federal data collection efforts.* These community-based sampling strategies may be particularly effective in achieving increased sample sizes.

Other low-cost innovative sampling options recommended by Santos⁵³ are: a) cumulations (exploit continuous surveys, promote small health modules for non-health surveys, promote government/private sector research partnerships); b) add-ons in specific cities with high rates of Asian American populations; and c) spreading the costs through "add and stretch" methods (superimposing Asian American national samples on continuous surveys in replications over time, e.g., five years). Finally, the coordination of data products and reports within NCHS would facilitate the dissemination of meaningful findings for the Asian American population.

New analytic approaches can offer more accurate findings from existing data. As noted, pooling data from multi-year datasets is one such approach. However, there are limitations to consider. First, pooled data can still yield large standard errors for small sub-groups.⁴¹ Further, sampling frames between various surveys may vary, different questions may be

*Unpublished data provided by author (S. Kwon).

asked across surveys and across years, and variables may be defined in distinct ways across surveys. Finally, analyses using pooled data do not allow for the examination of trends in morbidity and mortality over time, and such analyses are limited by the variables that are consistently collected across years. Trend-analysis may be particularly important for evaluating the effectiveness of policy changes or systems-level interventions to affect health at a national or regional level. Similarly, examining changes across time will continue to be important as migration from Asian countries to the U.S. continues. The HHS Office of Minority Health has engaged the RAND Corporation to conduct groundbreaking research on analytic and other approaches to improve the precision of health estimates for small racial and ethnic groups, such as applying a Kalman Filter to analyses of large datasets like NHIS, which would allow for the examination of trends across time even when pooling data across years. That agency and organizations involved with large-scale national surveys should provide technical guidance and conduct workshops for investigators to learn how to pool data and link existing databases. Combining estimation procedures from various surveys helps to address non-coverage and non-response issues and to estimate prevalence rates of other factors.⁵⁴

Creating community-based clinical registries

Health plans, disease registries, hospitals, and other providers can play a role in increasing knowledge of Asian American health status, increasing our understanding of disparities, and developing interventions to improve health care quality and health outcomes. Given the limitations on race reporting in many hospital and disease registries, another potential method for addressing this problem is to create community-based local or regional clinical data repositories. An example of this type of repository is the Asian American Hepatitis B Program, a program created to provide hepatitis B screening, vaccination, and follow-up care to Asian Americans and other communities in New York City at high risk of hepatitis B infection. Study participants were recruited through community-based screening events held throughout New York City. A standardized implementation and quality assurance protocol was systemically created for each area. The demographic and epidemiologic survey was translated into English, Chinese, Korean, Vietnamese, Hindi, Urdu, and Bengali and was self-administered with the assistance of bilingual volunteers. The centralized clinical repository currently has data on approximately 10,000 individuals screened and approximately 1,200 individuals identified with chronic hepatitis B, which has allowed for many sub-group epidemiological analyses.⁵⁵

Funding

Adequate funding of data collection programs at national, state, and local level and the sustainability of such funding are central to ensuring the collection of relevant data for Asian American populations. Data collection efforts such as the 2000 U.S. Census were highly successful in large part due to the robust outreach components that were built into the research design. Appropriate outreach to Asian American communities in the 2010 U.S. Census is essential in order to ensure their full participation. Other large national survey projects should consider incorporating culturally-specific outreach strategies (such as the use of ethnic media, utilizing lay community members in the data collection, and endorsement of the research process from respected community agencies and leaders). National survey projects should increase language access by hiring bilingual interviewers and translators and translate and administer the surveys in Asian American and NHPI languages. For example, the Census Bureau hires bilingual enumerators and enters into community partnerships to ensure an accurate count for the decennial Census. The CHIS incorporates Korean and Vietnamese household oversamples and interviews in Chinese (Mandarin, Cantonese), Korean, and Vietnamese languages. The Hawaii BRFSS and the Los Angeles County health survey also administer them in several Asian languages.

Federal data collection institutions should continue funding large-scale epidemiological studies in Asian American communities. The NLAAS provides an important model, but disaggregated data are primarily limited to the Chinese, Filipino, and Vietnamese subgroups. Future studies like the NLAAS should employ efforts to collect disaggregated data on other (including smaller) Asian American subgroups. Similarly, multi-site studies offer unique opportunities to coordinate data collection across geographic areas to ensure both geographic as well as ethnic group representation of Asian Americans in research.

Policymakers must also continue to support and fund efforts to conduct oversamples of Asian American and other racial and ethnic minority groups in national surveys such as the NHIS and BRFSS. Oversampling can increase screening and total survey costs, though screening data could be spread across surveys to reduce costs. Targeting high geographic concentrations of Asian Americans and NHPIs could also reduce costs and increase sample size, although it might also introduce bias in the sample.⁵⁶ Ultimately, obtaining an accurate national picture may require coordinated regional, state, and local efforts. The federal government should fund and collaborate with community-based organizations, tribal governments, tribal and native epidemiology centers, historically Black colleges and universities, Hispanic-serving institutions, Alaska Native and Native Hawaiian institutions and Asian American and Pacific Islander serving institutions. These entities should be engaged in data collection, analysis, and dissemination efforts.

The following box (Box 2) presents a summary of the challenges associated with data collection, reporting, and analysis in this population as well as potential solutions and recommendations for improvement.

Box 2

SUMMARY OF CHALLENGES, POTENTIAL SOLUTIONS AND RECOMMENDATIONS FOR DATA COLLECTION, REPORTING, AND ANALYSIS AMONG ASIAN AMERICANS

| Challenges for Data Collection, Analysis, and Reporting among Asian Americans | Potential Solutions & Recommendations |
|---|--|
| Lack or limited subgroup categorization | Implement collection of Asian subgroup ethnicity across datasets |
| | Standardization of data collection and implementation of OMB standards |
| Inconsistent definition/classification of Asian American | Standardization of data collection and implementation of OMB standards |
| | Avoid classification of Asian Americans by country of birth |
| | Implement collection of Asian subgroup ethnicity across datasets |
| Limited data collection in Asian languages | Provide bilingual interpreters for Asian languages |
| | Liason with community based organizations to conduct data collection |
| | Utilize language interpretation lines when using telephone based technologies |
| Uneven geographic distribution of geographic representations | Ensure health registry sites are geographically representative of the Asian population |
| | Create data repositories across geographic regions |

| Challenges for Data Collection, Analysis, and Reporting among Asian Americans | Potential Solutions & Recommendations |
|--|--|
| | Support oversampling Asian populations in traditional settlement areas (e.g. New York City) as well as emerging communities (e.g. Atlanta, GA) |
| Small sample sizes of Asian American and Asian American subgroups | Increase support national studies focused on the Asian American population |
| | Increase funding for oversamples of Asian American populations in national datasets |
| | Pooling and linking datasets |
| Limitations to RDD sampling | Employ innovative methodologies for capturing small populations |
| | Include cell phone only households in telephone based surveys |
| Lack of implementation of OMB standards | Increase standardization of data collection and implementation of OMB standards by holding agencies that are not compliant accountable |
| Lack of/inconsistent reporting of race and ethnicity data in registries and health plans | Support community based health registries |

OMB = Office of Management and Budget
RDD = Random digit dialing

Conclusions

Accurate data that reflect the diverse Asian American community are critical to developing sound research, health programs, and policy efforts. Additionally, accurate data that is truly representative of the population will correct the stereotype of Asian Americans face as a “model minority” population without of health problems. Though there are challenges that remain in the collection, analysis, and reporting of data for this population, researchers and advocates have made important strides it attempting to remedy these problems. Concerted and ongoing effort to ensure that there is a consistent set of standards for collection and reporting of data on race, ethnicity, and primary language across federal programs is essential.

While the Asian American population faces unique challenges for data collection and analysis due to issues of language proficiency and the heterogeneity of the community, it is important to note that such problems are also relevant for other immigrant and minority populations in the U.S. For example, Hispanic Americans also have a variety of ethnic backgrounds, and research has demonstrated that health status and outcomes vary considerably among groups⁵⁷ Issues regarding data collection and analysis have long been voiced as a major concern in NHPI populations.⁵⁸ Additionally, in recent years, the U.S. has experienced large waves of migration from countries in Africa, Eastern Europe, and the Middle East. Issues of LEP and subgroup differences may also be relevant to the collection and analysis of data in these emerging populations. Given the applicability of these issues to other minority populations, it may be beneficial to develop collaborative efforts between Asian Americans and other populations to address data collection issues. Improvements in the collection, analysis, and reporting of disaggregated or granular data will continue to be relevant in our quest to eliminate health disparities and promote health equity across all populations.

Acknowledgments

This publication was made possible by Grant Numbers P60 MD000538 and R24MD001786 from the National Institutes of Health, National Center on Minority Health and Health Disparities (NIH-NCMHD), 1UL1RR029893 from the National Center for Research Resources, National Institutes of Health, and 1U48DP001904-0 and U58 DP001022 from the Centers for Disease Control and Prevention (CDC). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH-NCMHD or the CDC. The authors would like to acknowledge Nandita Sabnis for her assistance with the paper.

Notes

1. U.S. Census Bureau. Profile of general demographic characteristics: 2000. Washington, DC: U.S. Census Bureau; 2000. Available at: http://factfinder.census.gov/servlet/QTTable?_bm=y&-geo_id=%200%201000US&-qr_name=DEC_2000_SF1_U_DP1&-ds_name=DEC_2000_SF1_U&-redoLog=false
2. Lin-Fu J. Asian and Pacific Islander Americans: an overview of demographics characteristics and health care issues. *Asian Am Pac Isl J Health*. 1993 Summer;1(1):20–36. [PubMed: 11567240]
3. Ghosh C. Healthy People 2010 and Asian Americans/Pacific Islanders: defining a baseline of information. *Am J Public Health*. 2003 Dec; 93(12):2093–8. [PubMed: 14652340]
4. Chen MS Jr, Hawks BL. A debunking of the myth of the health Asian Americans and Pacific Islanders. *Am J Health Promot*. 1995 Mar–Apr; 9(4):261–8. [PubMed: 10150729]
5. National Heart, Lung, and Blood Institute. Addressing cardiovascular health in Asian American and Pacific Islanders: a background report. Bethesda, MD: National Institutes of Health; 2000. Available at: <http://www.nhlbi.nih.gov/health/prof/heart/other/aapibkgd/aapibkgd.pdf>
6. Smedley, BD.; Stith, AY.; Nelson, AR., editors. Unequal treatment: confronting racial and ethnic disparities in health. Washington, DC: The National Academies Press; 2003.
7. Ver Ploeg, M.; Perrin, E., editors. Eliminating health disparities: measurement and data needs. Washington, DC: The National Academies Press; 2004. Panel on DHHS Collection of Race and Ethnic Data.
8. Ghosh, C. Asian American health research: baseline data and funding. In: Trinh-Shevrin, C.; Islam, NS.; Rey, MJ., editors. *Asian American communities and health: context, research, policy and action*. San Francisco, CA: Jossey Bass; 2009. p. 50-72.
9. U.S. Census Bureau. Current Population Survey: decennial census. Washington, DC: U.S. Census Bureau; 2009. Available at: <http://www.census.gov/population/www/socdemo/hh-fam.html#decennial>
10. National Center for Health Statistics. National Health Interview Survey public use data release. Hyattsville, MD: National Center for Health Statistics; 2008. Available at: <http://www.cdc.gov/nchs/nhis>
11. National Center for Health Statistics. National Health and Nutrition Examination Survey. Atlanta, GA: Centers for Disease Control and Prevention; 2009. Available at: <http://www.cdc.gov/nchs/nhanes>
12. National Center for Chronic Disease Prevention and Health Promotion. Behavioral Risk Factor Surveillance System (BRFSS) annual survey data. Atlanta, GA: Centers for Disease Control and Prevention; 2009. Available at: <http://www.cdc.gov/BRFSS>
13. UCLA Center for Health Policy Research. California Health Interview Survey website. Los Angeles, CA: UCLA Center for Health Policy Research; 2008. Available at: <http://www.chis.ucla.edu>
14. U.S. Census Bureau. American Community Survey. Washington, DC: U.S. Census Bureau; 2006. 2007
15. Waksberg, J.; Levine, D.; Marker, D. Assessment of major federal data sets for analyses of Hispanic and Asian or Pacific Islander subgroups and Native Americans: extending the utility of federal data bases. Washington, DC: U.S. Department of Health and Human Services; 2000. Available at: <http://aspe.hhs.gov/hsp/minority-db00/task3/section2.htm>

16. Bajaj, R. Recording voices: stories of Asian Pacific American youth as language brokers in New York City. New York, NY: Coalition for Asian American Children and Families; 2008. Available at: <http://www.cacf.org/documents/Recording%20Voices%20Policy%20Brief.pdf>
17. Islam, NS.; Trinh-Shevrin, C.; Rey, MJ. Toward a contextual understanding of Asian American health. In: Trinh-Shevrin, C.; Islam, NS.; Rey, MJ., editors. Asian American communities and health: context, research, policy and action. San Francisco, CA: Jossey Bass; 2009. p. 4-22.
18. Lucas, J. Race and ethnicity data collection in federal surveys: the national health interview survey experience. Presented at: Office of Minority Health National Leadership Summit on Eliminating Racial and Ethnic Disparities in Health: A Blueprint for Change; National Harbor (MD). February 25-27, 2009;
19. Alegria M, Takeuchi D, Canino G, et al. Considering context, place, and culture: the National Latino and Asian American Study. *Int J Methods Psychiatr Res.* 2004; 13(4):208-20. [PubMed: 15719529]
20. Freeman H, Kiecolt KJ, Nicholls WL 2nd, et al. Telephone sampling bias in surveying disability. *Public Opin Q.* 1982 Fall; 46(3):392-407. [PubMed: 10258862]
21. Weaver CN, Holmes SL, Glenn ND. Some characteristics of inaccessible respondents in a telephone survey. *J Applied Psychology.* 1975 Apr; 60(2):260-2.
22. Marcus AC, Crane LA. Telephone surveys in public health research. *Med Care.* 1986 Feb; 24(2): 97-112. [PubMed: 3945137]
23. Aquilino WS, LoSciuto LA. Effects of interview mode on self-reported drug use. *Public Opin Q.* 1990; 54:362-95.
24. Groves, RM.; Kahn, RL. Surveys by telephone: a national comparison with personal interviews. New York, NY: Academic Press, Inc; 1979. p. 79-120.
25. Kempf AM, Remington PL. New challenges for telephone survey research in the twenty-first century. *Annu Rev Public Health.* 2007; 28:113-26. [PubMed: 17094769]
26. Thornberry, OT.; Massey, JT. Trends in United States telephone coverage across time and subgroups. In: Groves, RM.; Biemer, PP.; Lyberg, LE., et al., editors. Telephone survey methodology. New York, NY: John Wiley and Sons, Inc; 1988. p. 25-49.
27. Blumberg SJ, Luke JV, Cynamon ML. Telephone coverage and health survey estimates: evaluating the need for concern about wireless substitution. *Am J Public Health.* 2006 May; 96(5):926-31. [PubMed: 16571707]
28. Massey, JT.; Botman, SL. Weighting adjustments for random digit dialed surveys. In: Groves, R.; Biemer, P.; Lyberg, L., et al., editors. Telephone survey methodology. New York, NY: John Wiley and Sons, Inc; 1988. p. 143-60.
29. Massey, JT.; O'Connor, D.; Krotki, K. 1997 proceedings of the Survey Research Methods Section, American Statistical Association. Arlington, VA: American Statistical Association; 1998. Response rates in random digit dialing (RDD) telephone surveys; p. 707-12.
30. Maklan, D.; Waksberg, J. Within-household coverage in RDD surveys. In: Groves, R.; Biemer, P.; Lyberg, L., et al., editors. Telephone survey methodology. New York, NY: John Wiley and Sons, Inc; 1988. p. 51-69.
31. Massey, JT. Estimating the response rate in a telephone survey with screening. Arlington, VA: American Statistical Association; 1996. p. 673-7.
32. Blumberg, SJ.; Luke, JV. Wireless substitution: early release of estimates from the National Health Interview Survey. Atlanta, GA: National Center for Health Statistics; January-June. 2007 Available at: <http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200712.pdf>
33. Federal Communication Commission. Trends in telephone service. Washington, DC: Federal Communication Commission; 2005.
34. U.S. Department of Health and Human Services. Policy statement on inclusion of race and ethnicity in DHHS data collection activities. Washington, DC: U.S. Department of Health and Human Services; 1997. Available at: <http://aspe.hhs.gov/datacncl/inclusn.htm>
35. Patient Protection and Affordable Care Act: PL-111-148, Section 3101, March 23, 2010.
36. McBean, AM. Improving Medicare's data on race and ethnicity. Washington, DC: National Academy of Social Insurance; 2006.

37. Shanawani H, Dame L, Schwartz DA, et al. Non-reporting and inconsistent reporting of race and ethnicity in articles that claim associations among genotype, outcome, and race or ethnicity. *J Med Ethics*. 2006 Dec; 32(12):724–8. [PubMed: 17145914]
38. Hasnain-Wynia R, Baker DW. Obtaining data on patient race, ethnicity, and primary language in health care organizations: current challenges and proposed solutions. *Health Serv Res*. 2006 Aug; 41(4p1):1501–18. [PubMed: 16899021]
39. Kressin NR, Chang BH, Hendricks A, et al. Agreement between administrative data and patients' self-reports of race/ethnicity. *Am J Public Health*. 2003 Oct; 93(10):1734–9. [PubMed: 14534230]
40. Boehmer U, Kressin NR, Berlowitz DR, et al. Self-reported vs. administrative race/ethnicity data and study results. *Am J Public Health*. 2002 Sep; 92(9):1471–3. [PubMed: 12197976]
41. Barnes, PM.; Adams, PF.; Powell-Griner, E. Health characteristics of the Asian adult population: United States, 2004–2006. Hyattsville, MD: National Center for Health Statistics; 2008.
42. Henry, J. Kaiser Family Foundation. Health coverage and access to care among Asian Americans, Native Hawaiians, and Pacific Islanders. Washington, DC: Henry J. Kaiser Family Foundation; 2008. Available at: <http://www.kff.org/minorityhealth/upload/7745.pdf>
43. Institute of Medicine. Race, ethnicity, and language data: standardization for health care quality improvement. Washington, DC: The National Academies Press; 2009.
44. Elliott MN, Finch BK, Klein D, et al. Sample designs for measuring the health of small racial/ethnic subgroups. *Stat Med*. 2008 Sep 10; 27(20):4016–29. [PubMed: 18351713]
45. Watters JK, Biernacki P. Targeted sampling: options for the study of hidden populations. *Social Problems*. 1989 Oct; 36(4):416–30.
46. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. *Social Problems*. 1997 May; 44(2):174–99.
47. O'Connell AA. Sampling for evaluation: issues and strategies for community-based HIV prevention programs. *Eval Health Prof*. 2000 Jun; 23(2):212–34. [PubMed: 10947526]
48. Muhib F, Lin LS, Stueve A, et al. A venue-based method for sampling hard-to-reach populations. *Public Health Rep*. 2001; 116 (Suppl 1):216–22. [PubMed: 11889287]
49. Kalton, G. Sampling considerations in research on HIV risk and illness. In: Ostrow, DG.; Kessler, RC., editors. *Methodological issues in AIDS behavioral research*. New York, NY: Plenum Press; 1993. p. 53-74.
50. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. *Social Problems*. 1997 May; 44(2):174–99.
51. Ngo-Metzger Q, Kaplan SH, Sorkin DH, et al. Surveying minorities with limited-English proficiency: does data collection method affect data quality among Asian Americans? *Med Care*. 2004 Sep; 42(9):893–900. [PubMed: 15319615]
52. Lee SK, Cheng YY. Reaching Asian Americans: sampling strategies and incentives. *J Immigr Minor Health*. 2006 Jul; 8(3):245–50. [PubMed: 16791534]
53. Santos, R. Oversampling AAPI populations in federal health surveys. *Proceedings of the Health Brain Trust on Data and Research Conference*; Bethesda (MD). May 2007;
54. Elliott, MN.; McCaffrey, D.; Finch, BK., et al. Using the Kalman Filter to improve disparity estimates for rare racial/ethnic groups. Washington, DC: U.S. Department of Health and Human Services; 2007.
55. Centers for Disease Control and Prevention (CDC). Screening for chronic hepatitis B among Asian/Pacific Islander populations—New York City, 2005. *MMWR Morb Mortal Wkly Rep*. 2006 May 12; 55(18):505–9. [PubMed: 16691180]
56. Ku, L. Improving national health survey data about Asians and Pacific Islanders. *Proceedings of the Health Brain Trust on Data and Research Conference*; Bethesda (MD). May 2007;
57. Khan LK, Sobal J, Martorell R. Acculturation, socioeconomic status, and obesity in Mexican Americans, Cuban Americans, and Puerto Ricans. *Int J Obes Relat Metab Disord*. 1997 Feb; 21(2):91–6. [PubMed: 9043961]
58. Panapasa, SV. Social, demographic, and cultural characteristics of U.S. Pacific Islanders. In: Trinh-Shevrin, C.; Islam, NS.; Rey, MJ., editors. *Asian American communities and health: context, research, policy and action*. San Francisco, CA: Jossey Bass; 2009.

Table 1
ASIAN AMERICAN SOCIODEMOGRAPHIC CHARACTERISTICS BY ETHNIC SUBGROUP

| Asian American Subgroups | Total Population | Total Asian (%) | Median Household Income (\$) | Average Household Size | Below the Federal Poverty Line | |
|-------------------------------|------------------|-----------------|------------------------------|------------------------|--------------------------------|------------------------|
| | | | | | (%) | Per Capita Income (\$) |
| Total U.S. | 299,398,485 | 41,994 | 2.6 | 12 | 21,587 | |
| Total Asian | 14,656,608 | 100.0 | 3.1 | 10 | 26,415 | |
| Asian Indian | 2,662,112 | 18.2 | 3.7 | 23 | 13,532 | |
| Bangladeshi | 69,687 | 0.5 | 4.4 | 29 | 10,215 | |
| Cambodian | 238,943 | 1.6 | 2.9 | 13 | 22,385 | |
| Chinese (including Taiwanese) | 3,565,458 | 24.3 | 3.4 | 7 | 19,259 | |
| Filipino | 2,915,745 | 19.9 | 6.1 | 38 | 6,613 | |
| Hmong | 209,866 | 1.4 | 2.7 | 19 | 18,819 | |
| Indonesian | 86,246 | 0.6 | 2.4 | 8 | 24,819 | |
| Japanese | 1,221,773 | 8.3 | 2.3 | 9 | 25,576 | |
| Korean | 1,520,703 | 10.4 | 2.8 | 14 | 18,027 | |
| Laotian | 216,789 | 1.5 | 4.2 | 19 | 11,454 | |
| Pakistani | 211,356 | 1.4 | 3.7 | 18 | 17,685 | |
| Thai | 199,559 | 1.4 | 2.8 | 15 | 25,139 | |
| Vietnamese | 1,559,394 | 10.9 | 2.6 | 14 | 17,232 | |
| Other Asian | 66,751 | 0.5 | 3.7 | 16 | 15,385 | |

Source: U.S. Census Bureau. 2006 American Community Survey. Washington, DC: U.S. Census Bureau, 2007.