Classifying Asian Americans by race versus ethnicity: differences in cancer death rates

Diane S. Lauderdale, PhD
Department of Health Studies
University of Chicago

Dezheng Huo, MD PhD
Department of Health Studies
University of Chicago

Corresponding author:
Diane S. Lauderdale, PhD
Department of Health Studies
University of Chicago
5841 S. Maryland Ave. MC2007
Chicago IL 60637

Phone: 773-834-0913
Fax: 773-702-1979
Email: Lauderdale@uchicago.edu

Acknowledgements: This project was supported by NIA R01 AG17889-01. We acknowledge the substantial contribution of Bert Kestenbaum (Office of the Actuary, Social Security Administration) whose prior work identifying Asian American subgroups in the Social Security files, done in collaboration with the first author, is the basis for this project.

ABSTRACT

Most US health statistics group Asian Americans into one race category, although many researchers question the value of the aggregation. However, small sample sizes in primary data collections and classification problems in administrative and vital statistics files generally preclude subgroup detail. In this large data linkage project, we determine site-specific cancer death rates for six Asian American subgroups and test for heterogeneity to better understand whether rates by race are actually misleading. We previously determined death rates for persons aged 65+ in the six largest Asian American subgroups using Social Security files, avoiding vital statistics classification deficiencies; unexpectedly, ethnic death rates were quite similar. In this project we link the Social Security death records to death certificates through the California Death Masterfile and the NDI. We find significant heterogeneity among Asian subgroups for about half of cancer sites. Asian Indians are most often the subgroup with a divergent rate.

INTRODUCTION

The racial categorization of Asian Americans is problematic for health statistics, including cancer rates. The question of whether to aggregate or disaggregate Asian American ethnic groups is shared across most areas of health-related data collection and research concerning Asian Americans, who have never been a “comfortable fit” in US racial categories. [1,2] The subgroups that are aggregated in the race category “Asian/Pacific Islander” or “Asian American” (such as Vietnamese, Asian Indians and Chinese) may be too dissimilar to form a category that can be used meaningfully to track changes in disease occurrence over time or inform etiologic research and health policy. However, there are logistical obstacles to determining cancer rates for separate Asian ethnic groups. The first problem is the accuracy and compatibility of the Asian ethnicity information in the data sources used to construct the numerators of morbidity and mortality rates, typically disease registries, administrative data or death certificates for numerators and the Census for denominators. Focusing on cancer rates, the SEER report *Racial/Ethnic Patterns of Cancer in the United States 1988-1992* notes that “Inconsistencies between the racial/ethnic designations from these different sources, however, may lead to either overstating or understating the true cancer rate for a particular group.” [3] A study by the National Center for Health Statistics estimated that under-reporting of any Asian race on death certificates resulted in death rates being understated by 11% for Asian Americans. [4]

An additional obstacle to producing accurate ethnicity-specific rates is that the Census Bureau does not estimate the population for Asian subgroups by year between decennial censuses, only for the aggregate race category. Thus rates are best determined just in the decennial years (e.g. 1990, 2000); sometimes researchers aggregate three or five years around the decennial year (e.g. 1999-2001 or 1998-2002) to increase sample size, but that assumes constant population change. [5] However, for more specific disease rates even when aggregating over five years, the number of ethnicity-specific, sex-specific occurrences may still be too small to produce stable rate estimates for group comparisons. For this reason, and because of concern about misclassification for subpopulations, the National Program of Cancer Registries reports cancer rates for the aggregate race category only in their most recent report, even though the state cancer registries have codes for each of the subpopulations. [6] Cancer appears to be a relatively more important cause of death for Asians compared to other race groups: it is the leading cause of death for Asian American women.

With the 2000 Census, the new option to check multiple races adds an additional challenge to the calculation of cancer rates. For both Asian race and for each individual subpopulation there are two different ways to calculate the population denominator, one that excludes those checking more than one race and one that includes them. The impact on Asian American population counts is large. Most dramatically, the Japanese population is more than 40% greater with the inclusion of multiple race persons. [7]

This project builds on a prior demographic study that determined all-cause death rates for persons age 65 and older in the six largest Asian American subgroups (Chinese, Indian, Japanese, Korean, Filipino and Vietnamese) for the years 1990-1999. The project uses a single data source for numerators and denominators, avoiding the problem of inconsistent ethnic identification, and aggregates a sufficient number of years to produce stable rates. [8] In the present study, through record linkage to death certificate information, we determine cause of death and calculate cause-specific cancer death rates. We use these rates to test whether there is significant heterogeneity among subgroups for each cancer site. Thus we both provide cancer-specific death rates for older Asian Americans using a methodology that overcomes well-recognized pitfalls in vital statistics data, and we also help answer the question of how misleading it is to present disease rates only for the aggregate Asian American race category.

METHODS

Data

This project uses four data sources: two administrative files at the Social Security Administration (SSA), the 1990-1999 Death Statistical Master file for the state of California and the National Death Index (NDI). The method of identifying Asian Americans in the SSA files and of calculating age- and sex-specific death rates for persons aged 65 and older have been previously described in detail. [8] The general strategy was to link the SSA enrollment file, which includes sex, date of birth, and the fact and date of death to a second SSA file containing information that could be useful for inferring Asian ethnic group: race, country of birth (available for most persons), surname, father's surname (for women), and given name. Below we describe the development of the name lists and the algorithm for inferring ethnicity in the linked SSA file.

Name lists were developed that were specific to each of six subgroups (Chinese, Asian Indian, Japanese, Korean, Filipino, Vietnamese). An earlier paper describes in detail the derivation and evaluation of *conditional* and *unconditional* surname lists for each of the six Asian populations: *conditional* lists when a person is known to be Asian (e.g., a record has an "Asian or Pacific Islander" race code) and *unconditional* lists when race is not known. [9] Most names on a conditional list are also on the corresponding unconditional list, but some are not. For example, if a person named "Bang" is known to be Asian, then that person is very likely Korean; however, the surname "Bang" is generally not Asian. Similarly, persons with Spanish surnames are rarely Asian; however, if someone with a Spanish surname has been identified as Asian, then he is almost certainly Filipino.

The lists were derived from race and country of birth information in an SSA administrative file. A name was added to the conditional list of a particular subpopulation if the name was not rare and if most persons born in Asia with that name were born in the associated country. The frequency of the name by race was used to determine whether the name could also be included on the unconditional list. The lists were evaluated with the help of the Census Bureau, using a file which the Census Bureau had created in conjunction with the 1990 post-enumeration survey: ethnic identification by surname was compared with self-identified ethnicity. Because surnames in different Asian countries vary in how distinctive and how common they are, the lists vary in their inclusiveness (sensitivity). The Vietnamese and Japanese lists are the most inclusive, and the Indian and Filipino the least inclusive. Given name lists were derived similarly to the surname lists, except that there are separate lists for men and women. The lists are not conditional on Asian identification.

Thus to assign subpopulation, there is a race code and 3 additional items: place of birth (if known), whether the surname/maiden name is on the surname lists (conditional or also unconditional), and whether the given name is on a list. Since none of these items of information is perfectly specific, we developed an algorithm for inferring ethnicity. [8] The algorithm has three branches, according to whether the race code is indicative of Asian ("Asian" or "other"), uninformative (blank) or contra-indicative (all other race codes). For Filipino identification, "other" was considered uninformative rather than indicative, because there are many Hispanic persons in the "other" category. For Asian Indians, both "white" and "American Indian" were considered uninformative rather than contra-indicative.

The basic algorithm infers ethnicity from (1) surname – using the conditional list, given name, or place of birth (in that order of precedence) when the race code is indicative of Asian identification, (2) surname – using the unconditional list, given name, or place of birth (in that order of precedence) when the race code is uninformative, or (3) two pieces of information pointing to a particular ethnicity when the race code is contra-indicative. For the third rule, no ethnic assignment was made when surname/maiden name pointed to a different ethnicity than both place of birth and given name.

Only persons enrolled in Medicare Part B at some time during the 1990s were included because death information is highly accurate for Medicare B enrollees. A 10-year period (1990-1999) was chosen to obtain reliable death rates for relatively small populations. Numerators and denominators for the death rates during the decade of the 1990s were based on monthly observations of the Medicare B experience. Each person could potentially contribute as much as 10 person-years of observation. There were over one million person-years of observation each for Chinese, Japanese and Filipino, and about 300,000 each for Indian, Korean and Vietnamese (see Table 1). [8] There were 135,895 Asian American death records. The assignment of individual records to Asian subgroups and the calculation of death rates were done by a colleague at the SSA because of the confidentiality of SSA records for living persons. SSA death records, however, could be released for linkage to external files.

To determine cause-specific death rates, we linked the SSA death records to primary cause of death information recorded on death certificates. Because 40 percent of the SSA death records listed California as the last state of residence and because California makes available to researchers complete files of death certificates with names and social security numbers, we matched SSA death records to the California death master file using social security numbers and sex. There were 57,801 matches. Of persons who last lived in California according to SSA records, 57,367 were matched (89.7%). Of those who last lived outside of California, 434 were matched (0.6%). For matched cases, 96.4% had an exact last name match on the SSA and California files; 76.9% had exact first name match; 99.9% had exact year of death match; and 98% had the same age at death. We examined cases with differences in first name and found that almost all were variant spellings. Therefore, we accepted all matches. We used the primary cause-of-death code from the California death certificates.

To determine cause of death for the rest of the SSA death records, we used the National Death Index (NDI). The NDI is a national database of death records maintained by the National Center for Health Statistics and available to researchers on a fee basis per record searched. We drew a sample of 20,000 unmatched SSA records, stratified by ethnic group and nativity, with higher sampling probabilities for the smaller ethnic groups so that cause-specific rate estimates would be similarly precise for smaller and larger ethnic groups. In total, 19,954 matched records were returned from NDI. Multiple matches were generated for some subjects. For each subject, the first record with highest probability matching score was retained. As a result, 19,072 subjects were matched. However, 7 matched cases had a missing value in the cause of death field and were excluded. In total, 76,866 (56.6%) of the 135,895 SSA death records were matched; 57,801 from the California death master file (57,367 California residents and 434 last resident in other states) and 19,065 from NDI (448 California residents and 18,617 last resident in other states). Essentially, 9 sampling strata were formed, one for California deaths and eight for non-California deaths (see table 2). Sampling weights were defined as the inverse of the probability of selection and matching. We calculated sampling weights for each specific ethnic-, sex-, and age-stratum for each cause of death examined. Age was categorized into 5 groups: 65-69, 70-74, 75-79, 80-84, and ≥85 years old. Because of the sampling scheme, sampling weights varied across ethnic groups but were similar across age and sex groups within ethnic group.

Statistical Analysis

Each ethnic- and sex-specific death rate is the ratio of the number of deaths from specific cancer to the person years of population observed, expressed per 100,000 person-years. The numerator is the number of deaths in the sample and the denominator is the ethnic- and sex-specific population from the SSA, divided by the corresponding sampling weight. The classification code for deaths changed in 1999 from ICD-9 to ICD-10, and we accounted for this change using a method developed by the National Center for Health Statistics. [10, 11]. By double-coding 1996 death records, the National Center for Health Statistics calculated comparability ratios for 113 selected causes of death that were the number of deaths from a specific cause classified by ICD-10 divided by the number of deaths classified by ICD-9. These ratios allow us to estimate the number of deaths that would have been coded to any of the 113 causes under ICD-10 rules for the years 1990-1998. We use the level of cancer site detail available when using the 113 causes, which combines some related sites, such as “lip, oral cavity & pharynx.” We do not include the residual category of “cancers not elsewhere classified,” and we do not present deaths due to breast cancer for men because of extremely small numbers.

For comparison, national death rates for Whites were calculated for 1990-1999, based on publicly-available data. The numerators for the rates were aggregated from single-year cause-specific death counts, [12,13] using the comparability ratios for the ICD-9 to ICD-10 transition. The denominators were the aggregation of the intercensal national population estimates produced by the Bureau of the Census. These are the “bridged” estimates for single race categories that take into account the change in race categories from the 1990 to the 2000 census. [14]

To compare the Asian death rates to the White rates and adjust for age, Poisson regression models were built for each cause of death and stratified by sex. [15] Models took into account the sampling design. Rate ratios and 95% confidence intervals were calculated from the Poisson models with White as the reference group. Two log likelihood ratio tests were performed: (1) to test whether the death rates of Asian Americans as a whole differ from the White rate and (2) to test whether there was significant heterogeneity in death rates among the six Asian ethnic groups. Because of the large sample sizes and number of tests, we only comment on statistical significance when

$p < 0.001$. This study was approved by the Institutional Review Board of the University of Chicago.

Results

The results are presented in Table 3. For all cancers combined and for each of 21 cancer sites (or groups of related sites), the table gives the rate per 100,000 for Whites aged 65 and older and the estimated rates for each of the six ethnic groups by sex. Age-adjusted rate ratios from Poisson regression models compare the specific ethnic rates to the White rates, and the confidence intervals indicate whether each rate differs significantly from the White rate. For each cancer site and sex, there are two p-values, one testing whether the cause-specific death rate for Asian Americans as a whole differs from the White rate and the other testing whether there is significant heterogeneity among the six Asian ethnic groups.

For the comprehensive category of all malignant neoplasms (listed first in Table 3), the rates are significantly lower for each Asian ethnic group relative to Whites. Older Asian Americans considered together have a risk of death due to cancer that is highly significantly lower than Whites. However, there is also significant heterogeneity among the Asian ethnic groups. Indian men and women have the lowest rates among the Asian groups, about half the White rate. Korean men and Chinese women have the highest rates among Asians.

For over half of the specific cancer sites, every Asian group has a lower rate than Whites: larynx; lung, bronchus & trachea; skin; prostate; female breast; corpus & uterus; ovary; kidney & renal pelvis; bladder; brain & CNS; Non-Hodgkins lymphoma; and leukemia. Among these low-risk cancers for Asian Americans, there is nonetheless highly significant heterogeneity ($p < 0.001$) among the Asian groups for three sites: lung, bronchus & trachea, female breast and prostate. For lung cancer, the Indian rate for both men and women is less than half the rate for the other Asian groups (which are generally similar to each other) and about one-fourth the White rate. For prostate cancer, there is a range of rate ratios (relative to Whites) from 0.23 for Vietnamese up to 0.61 for Filipinos. For female breast, Koreans and Vietnamese have extremely low rates, less than one-fifth the White rate, while Japanese and Indian women have the highest rates among Asians, still just half the risk of White women.

There are no cancer sites for which every Asian group has a higher risk than Whites, but there are three where the majority of Asian groups have higher rates: stomach, liver and cervix. For all of these sites there is highly significant heterogeneity among the Asian groups. For stomach cancer, Chinese, Japanese, Korean and Vietnamese all have rates two to four times higher than the White rate. However, Indians and Filipinos do not have high rates compared to Whites; the Filipino male rate is actually significantly lower than the White rate. For liver cancer, all but Indians have higher rates than Whites: Chinese, Korean and Vietnamese have very high rates and the Japanese rate is moderately high. The Indian rate is similar to the White rate. For cervical cancer, the Vietnamese and Indian rates are about twice the White rate, while the rates for the other groups are generally similar to the White rate except for the Japanese who have a significantly lower rate than White women.

For other cancer sites, there is an inconsistent pattern across Asian groups, with some subgroups having rates lower than Whites and some having rates similar to Whites. Heterogeneity is most striking for colon and rectum & anus. For these sites, Vietnamese, Korean, Filipino and Indians have about half the risk of Whites, while Chinese have risk 20% to 30% lower than Whites and Japanese have similar risk to Whites.

Conclusions

We have calculated site-specific cancer death rates for persons aged 65 and older for six Asian American ethnic groups and compared each of these rates to White rates. We have also compared rates for Asian Americans as a whole to Whites and tested whether there is heterogeneity among the Asian American ethnic groups. For less than half of cancer sites, Asian Americans have uniformly lower rates compared to Whites. These cancer sites are larynx; skin; corpus & uterus; ovary; kidney & renal pelvis; bladder; brain & CNS; Non-Hodgkins lymphoma; and leukemia. For these sites, calculating and presenting rates for the race category Asian would reasonably describe the risk for all six ethnic groups. For the other half of cancer sites, though, there is significant heterogeneity among the Asian groups

For the most common cancers – colon, lung, breast and prostate – there is heterogeneity among the Asian groups. This heterogeneity is most noteworthy for colon cancer because one of the Asian groups (Japanese) is not at reduced risk relative to Whites while all of the others are. An Asian aggregate rate would mask the divergent Japanese rate.

For the three cancers where Asians are generally thought to be at high risk – cervical, liver and stomach – there is highly significant heterogeneity among Asian groups, and only some have elevated risk compared to Whites. Of note, the cervical cancer death rate is not significantly different between Asians and Whites when the six ethnic subgroups are aggregated.

While there is no single pattern repeated across cancer sites of the same ethnic group or groups always having cancer death rates that diverge from the other Asian groups, there is one group that clearly is more often the subgroup with the highest or lowest rate – Asian Indians. Indians are the highest risk Asian group for breast, cervix, uterus, ovary, esophagus, and myeloma, and the lowest risk group for colon, lip, stomach and lung. This pattern suggests that aggregating Asian American subgroups for surveillance or etiologic research would be less problematic were Asian Indians separated from East Asians. The racial categorization of Asian Indians in the US has been historically inconsistent, and their grouping with Asians is actually quite recent in federal data collection. [16] The divergent Asian Indian rates are likely a combination of environmental and genetic factors. Some of these cancers have very strong environmental causes, such as lung and cervical cancer, and the Asian Indian rate probably reflects differences in environmental exposure. For those sites where no environmental factors seem to explain much variation in risk, genetic differences may play a role. One study that used statistical analysis of genotypes from more than a thousand individuals to define broad clusters of people based on genetic similarity placed South Asians in the Eurasian cluster, separate from the East Asian cluster. [17]

A few recent studies have examined cancer rates for Asian subgroups, reflecting recent enhancements to ethnic identification in some data sources. Chu and Chu used the best approach that is possible with publicly available data to study cancer mortality. [5] They calculated death rates for 1999-2001 for seven states (CA, HI, IL, NJ, NY, TX and WA) that have added Asian subgroup detail to their death certificate files. These are the states with the largest Asian populations. Chu and Chu present an "upper boundary," counting only those checking a single subgroup code in the 2000 Census in the denominator, and a "lower boundary," using the larger number who report additional race codes as well. They conclude that the aggregate race pattern did mask differences between subpopulations. Their study included all ages, but results are generally similar to our study of older persons. However they do not present White rates for comparison or test whether the subgroup rates differ from each other. There is one fairly consistent difference between the rates they present and the ones we find: Chu and Chu's rates for Indian women are relatively lower compared to the other Asian subgroups than we found. For example, for "all malignant neoplasms," Chu and Chu report Indian women have a much lower overall cancer death rate than any other Asian subgroup, a rate that is 52% - 56% of their highest Asian female rate. However, we found that Indian women had similar overall cancer death rate to Filipino women (but lower than the other subgroups), and the Indian rate was 75% of the highest Asian subgroup rate. This difference could be the result of less complete race identification for Indian women (for example, their being categorized as White) on death certificates, which would systematically understate death rates when combining data sources.

California is home to about 35% of Asian Americans. The California Cancer Registry and the Northern California Cancer Center have both produced recent studies of Asian subgroups in their coverage areas. The California Cancer Registry reported incidence and mortality rates for five subpopulations (Chinese, Filipino, Japanese, Korean and Vietnamese) and for eight cancer sites for the period 1997-2001. [18] They used linear interpolation and extrapolation to estimate the California subgroup populations, averaging the 2000 population derived by including and excluding the multiple race persons. They acknowledge concerns about subgroup misclassification and conclude that aggregate Asian cancer rates would not accurately reflect the cancer burden of specific subgroups. The subgroups they report as having highest and lowest risk for cancer sites are generally the same as we found in our study, given that they do not include Indians. The Northern California Cancer Center used similar methods to calculate cancer incidence rates for six Asian subpopulations (including Asian Indians). [19] They divide their observation period into four-year intervals. The rates – even for relatively common cancers – can be very unstable (e.g. the 95% confidence interval for Vietnamese prostate cancer incidence from 1990-1993 is 36.0-110.7 per 100,000), making subgroup comparisons and trend assessments between time periods underpowered for most cancers. While California is the state with the largest Asian population, there may be differences between Asians in California and the rest of the country that affect cancer risk.

These prior studies, using the best available demographic methods and most complete subgroup data, are nonetheless subject to the two recognized problems with Asian subgroup disease rates: potential mismatch in ethnic identification between numerators and denominators and small sample sizes. While our study avoids these problems, we have other limitations. Our methodology only yields death rates, not incidence rates, and we only examine death rates for the six largest Asian groups, and not others such as Thai and Cambodians. Our study only examines cancer mortality for persons aged 65 and older, and there may be divergent subgroup trends for some cancers at younger ages. Nonetheless, 70% of cancer deaths do occur at age 65 or older. [20] The chief limitation though, is the potential for selection bias, if the Asian Americans whose death experience we examine are not fully representative of the general Asian American population. There are two points at which selection bias could occur. The first is our restriction to the Medicare Part B experience: while the very great majority of elderly are enrolled in Medicare B, enrollment may be less for some Asian populations. The second potential source of selection bias is our method of ethnic identification, which may perform better for the foreign-born (for whom country of birth and given name are informative) than for the US-born (for whom only surname/maiden name may be informative). Therefore the foreign-born are likely better represented in our study. However, because of the recency of sizable Asian immigration, the great majority of Asian American elderly are foreign born. Only two ethnic groups have substantial numbers of US-born elderly: Chinese and Japanese. For these two groups we specifically sampled death certificates by nativity to ensure representation of the US-born. Similarly, persons with distinctive names specific to a single Asian subpopulation are better represented than persons with less distinctive names (e.g. Lee). We are aware of two studies that specifically examined this issue of whether persons with distinctive ethnic names were representative of the whole ethnic group. Shin and Yu argued that Koreans surnamed "Kim" were generally representative of all Koreans, and Rosenwaike found that Medicare enrollees with the most common Chinese, Japanese, Korean, and Vietnamese surnames had demographic and geographic characteristics similar to all elderly persons who identified themselves with these race categories in the census. [21, 22] Our final limitation is that our unique linked data set cannot be readily updated, so the years of observation are limited to the 1990s. Also, statistical tests have more power to detect heterogeneity for the more common cancers, due to greater numerator sample size. However, differences we find in relative risk for those cancer sites are large, greater than 2-fold, for colon, breast, lung and prostate cancer.

We hope that this opportunity to determine cancer death rates for Asian American subpopulations with sufficient numbers to test heterogeneity across subpopulations and with consistent ethnic identification can inform ongoing efforts to improve public health data for Asian subgroups. We have found that cancer death rates vary significantly among Asian subpopulations for about half of cancer sites, including the most prevalent cancers and those recognized as high among Asian Americans (stomach, liver and cervix). We also found that Asian Indians were more often at the high or low end of risk for the Asian subpopulations than any other subgroup, suggesting that disaggregating Asian Indians from East Asians would improve the interpretability of aggregate Asian rates.

Literature Cited

1.King RC. Racialization, recognition, and rights: Lumping and splitting multiracial Asian Americans in the 2000 Census. Journal of Asian American Studies 2000;3:191-217

2. The Asian Population: 2000. Census 2000 Brief. By. Jessica S. Barnes and. Claudette E. Bennett. U. S. Census Bureau 2002.

3. Miller BA, Kolonel LN, Bernstein L, et al. (eds). Racial/Ethnic Patterns of Cancer in the United States 1988-1992, National Cancer Institute. NIH Pub. No. 96-4104. Bethesda, MD, 1996.

4. Rosenberg HM, Maurer JD, Sorlie PD, et al. 1999. "Quality of Death Rates by Race and Hispanic Origin: a Summary of Current Research, 1999." National Center for Health Statistics. Vital Health Stat 2(128).

5. Chu KC, Chu KT. 1999-2001 Cancer mortality rates for Asian and Pacific Islander ethnic groups with comparisons to their 1988-1992 rates. Cancer. 2005 Dec 15;104(12 Suppl):2989-98.

6. U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999-2002 Incidence and Mortality Web–based Report Version. Atlanta: Department of Health and

Human Services, Centers for Disease Control and Prevention, and National Cancer Institute; 2005. Available at: www.cdc.gov/cancer/npcr/uscs.

7. Terrance J. Reeves and Claudette E. Bennett, We the People: Asians in the United States, Census 2000 Special Reports, CENSR-17. U.S. Census Bureau, Washington, DC (2004), p. 1. Archived at: http://www.census.gov/prod/2004pubs/censr-17.pdf

8. Lauderdale DS, Kestenbaum B. Mortality of Elderly Asian American populations based on Medicare and Social Security Data. Demography 2002;39:529-540.

9. Lauderdale DS, Kestenbaum B. Asian American ethnic identification by surname. Population Research and Policy Review. 2000;19:283-300.

10. Anderson RN, Minino AM, Hoyert DL, Rosenberg HM. Comparability of cause of death between ICD-9 and ICD-10: Preliminary estimates. National Vital Statistics Reports; Vol 49 No. 2. Hyattsville, Maryland: National Center for Health Statistics. 2001

11. National Center for Health Statistics. Documentation for the public use multiple cause of death file on comparability between ICD-9 and ICD-10: A double-coded file based on the 1996 data year multiple cause of death file. Hyattsville, Maryland: National Center for Health Statistics.

12. National Center for Health Statistics. HIST002. Deaths for 113 Selected Causes by 5-Year Age Groups, Race, and Sex: United States, 1979-98
http://www.cdc.gov/nchs/data/statab/hist002_2.pdf

13. National Center for Health Statistics. GMWKH210R Death Rates from 113 Selected Causes, United States, Specified Hispanic Origin, Race for Non-Hispanic Population, 1999, 2001, 2002.
http://www.cdc.gov/nchs/datawh/statab/unpubd/mortabs/gmwkh210_10.htm

14. National Center for Health Statistics. Bridged-race intercensal estimates of the July 1, 1990-July 1, 1999, resident population of the United States by year, single-year of age (0,1,2,…, 85 years and over), bridged-race (White, Black or African American, American Indian or Alaska Native, Asian or Pacific Islander), Hispanic origin (not Hispanic or Latino, Hispanic or Latino), and sex. (Released 7/26/2004)
ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datasets/nvss/bridgepop/icen_natA1.txt

15. Breslow NE, Day NE. Statistical Methods in Cancer Research. Volume II-The Design and Analysis of Cohort Studies. IARC Scientific Publications No. 82. International Agency for Research on Cancer. Lyon: 1987

16. S.M. Lee. 1993. "Racial Classifications in the U.S. Census, 1890 to 1990." Ethnic and Racial Studies, vol. 16, no. 1: 75-94.

17. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. Genetic structure of human populations. Science. 2002;298:2381-5.

17. Kwong SL, Chen MS Jr, Snipes KP, Bal DG, Wright WE. Asian subgroups and cancer incidence and mortality rates in California. Cancer. 2005;104(12 Suppl):2975-81.

18. Gomez SL, Le GM, Miller T, et al. Cancer incidence among Asians in the Greater Bay Area, 1990-2002. Fremont, CA: Northern California Cancer Center, 2005.

19. Ries LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Clegg LX, et al (eds). SEER Cancer Statistics Review,1973-1998. Bethesda, MD: National Institute of Health, 2000; NIH publication no. 00-2789.

20. Shin EH, Yu EY. 1984. “Use of Surnames in Ethnic Research: the Case of Kims in the Korean American Population.” Demography 21:347-359.

21. Rosenwaike I. 1994. “Surname Analysis as a Means of Estimating Minority Elderly: an Application Using Asian Surname.” Research in Aging 16:212-227.

Table 1. Person-years of observation, by sex and age, for persons age 65 and older for six Asian American population, based on Medicare B enrollment data: 1990-1999

	Age group					
	65-69	70-74	75-79	80-84	85+	Total
Men						
Chinese	184,674	170,833	123,775	68,944	41,072	589,298
Indians	45,566	33,834	20,798	10,576	5,212	115,986
Japanese	134,556	142,691	101,290	49,500	31,284	459,321
Koreans	49,860	41,853	28,477	14,902	8,266	143,358
Filipinos	114,583	126,890	91,610	60,342	41,873	435,298
Vietnamese	52,825	39,145	24,936	12,852	7,079	136,837
Women						
Chinese	226,074	202,388	146,792	91,878	70,570	737,702
Indians	50,999	41,683	25,376	12,151	6,482	136,691
Japanese	262,742	204,985	126,575	63,080	59,482	716,864
Koreans	80,935	67,713	49,387	29,725	18,142	245,902
Filipinos	175,204	168,535	119,649	66,812	35,025	565,225
Vietnamese	58,429	47,658	35,050	21,702	14,563	177,402

Table 2. Sampling weight by 9 strata

Stratum	No. of deceased	No. of matched	% of sampling	Sampling weight
Lived in California	63,978	57,815	90.4	1.11
Lived out of California				
Foreign-born Chinese	15,067	2,481	16.5	6.07
Native-born Chinese	6,891	2,385	34.6	2.89
Indians	5,186	2,365	45.6	2.19
Foreign-born Japanese	3,049	2,261	74.2	1.35
Native-born Japanese	19,724	2,498	12.7	7.90
Koreans	5,848	2,398	41.0	2.44
Filipinos	12,652	2,352	18.6	5.38
Vietnamese	3,500	2,311	66.0	1.51
Total	135,895	76,866	56.6	

Table 3. Cancer death rates (per 100,000) for persons aged 65+ from six Asian American groups and age-adjusted rate ratios relative to the White rate: 1990-1999.

Cancer site	Men		Women	
	Rate	RR (95% CI)	Rate	RR (95% CI)
All malignant neoplasms				
White	1432.0	1.00	907.8	1.00
Chinese	1023.2	0.71 (0.69-0.74)	625.3	0.72 (0.69-0.75)
Indians	594.0	0.45 (0.40-0.49)	468.0	0.58 (0.52-0.64)
Japanese	1063.4	0.74 (0.71-0.77)	587.1	0.70 (0.68-0.73)
Koreans	1136.1	0.82 (0.77-0.87)	569.7	0.67 (0.63-0.72)
Filipinos	901.2	0.60 (0.57-0.62)	465.2	0.55 (0.52-0.58)
Vietnamese	883.7	0.66 (0.62-0.70)	534.7	0.63 (0.59-0.67)
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*<0.0001*		*<0.0001*
Lip, oral cavity & pharynx				
White	19.8	1.00	9.4	1.00
Chinese	15.3	0.74 (0.57-0.95)	10.1	1.03 (0.76-1.38)
Indians	10.0	0.48 (0.23-1.01)	5.9	0.68 (0.26-1.81)
Japanese	24.9	1.25 (0.89-1.75)	4.3	0.42 (0.26-0.67)
Koreans	13.7	0.74 (0.45-1.23)	2.2	0.22 (0.07-0.68)
Filipinos	15.9	0.79 (0.56-1.12)	7.5	0.70 (0.47-1.05)
Vietnamese	19.2	0.98 (0.65-1.47)	11.1	1.17 (0.72-1.91)
p for testing difference from White		*0.01*		*0.0002*
*p for testing heterogeneity**		*0.11*		*0.0008*
Esophagus				
White	35.2	1.00	9.6	1.00
Chinese	23.7	0.65 (0.52-0.81)	8.4	0.72 (0.49-1.06)
Indians	37.0	1.09 (0.72-1.64)	10.4	1.16 (0.61-2.24)
Japanese	22.7	0.60 (0.47-0.78)	6.0	0.55 (0.37-0.82)
Koreans	37.8	1.07 (0.76-1.49)	1.0	0.08 (0.01-0.55)
Filipinos	16.2	0.47 (0.35-0.63)	5.7	0.49 (0.29-0.83)
Vietnamese	20.9	0.62 (0.41-0.94)	5.6	0.62 (0.31-1.25)
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*0.003*		*0.016*
Stomach				
White	37.1	1.00	19.8	1.00
Chinese	73.3	1.95 (1.71-2.22)	38.7	2.00 (1.71-2.33)
Indians	24.9	0.73 (0.46-1.17)	17.5	1.14 (0.66-1.97)
Japanese	114.8	3.09 (2.74-3.49)	58.1	3.51 (3.06-4.02)
Koreans	176.9	4.92 (4.21-5.74)	70.7	4.01 (3.32-4.85)
Filipinos	28.7	0.69 (0.55-0.86)	20.0	1.11 (0.87-1.42)
Vietnamese	88.0	2.54 (2.09-3.09)	51.3	2.92 (2.32-3.67)
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*<0.0001*		*<0.0001*

*Log likelihood ratio tests for difference within six Asian American ethnicities

Table 3. (Continued)

Cancer site	Men		Women	
	Rate	RR (95% CI)	Rate	RR (95% CI)
Colon and rectum				
White	150.7	1.00	115.1	1.00
Chinese	112.1	0.73 (0.66-0.81)	74.7	0.70 (0.63-0.77)
Indians	48.2	0.33 (0.24-0.46)	41.7	0.46 (0.33-0.65)
Japanese	160.5	1.05 (0.95-1.17)	87.2	0.91 (0.82-1.02)
Koreans	89.5	0.61 (0.50-0.76)	57.5	0.57 (0.47-0.70)
Filipinos	86.4	0.55 (0.48-0.63)	46.6	0.46 (0.39-0.54)
Vietnamese	62.6	0.44 (0.35-0.56)	52.8	0.52 (0.42-0.65)
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*<0.0001*		*<0.0001*
Liver & intrahepatic bile ducts				
White	28.2	1.00	14.6	1.00
Chinese	97.1	3.41 (3.07-3.79)	42.3	3.02 (2.61-3.50)
Indians	30.0	1.08 (0.69-1.70)	12.6	0.89 (0.49-1.61)
Japanese	53.1	1.87 (1.53-2.27)	32.9	2.52 (2.15-2.97)
Koreans	125.0	4.54 (3.82-5.40)	64.0	4.72 (3.91-5.70)
Filipinos	49.6	1.57 (1.32-1.87)	16.5	1.23 (0.94-1.60)
Vietnamese	141.8	5.24 (4.50-6.09)	61.9	4.55 (3.73-5.56)
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*<0.0001*		*<0.0001*
Pancreas				
White	65.1	1.00	54.8	1.00
Chinese	55.3	0.85 (0.74-0.99)	45.5	0.87 (0.75-1.01)
Indians	50.6	0.83 (0.60-1.17)	29.3	0.54 (0.36-0.82)
Japanese	73.0	0.97 (0.82-1.15)	54.8	1.05 (0.91-1.21)
Koreans	56.9	0.89 (0.68-1.17)	47.6	0.93 (0.74-1.15)
Filipinos	53.0	0.78 (0.65-0.94)	37.1	0.72 (0.60-0.87)
Vietnamese	37.8	0.62 (0.46-0.82)	32.0	0.64 (0.48-0.85)
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*0.12*		*0.0008*
Larynx				
White	13.3	1.00	2.7	1.00
Chinese	7.2	0.45 (0.29-0.68)	0.0	0.00
Indians	9.6	0.74 (0.33-1.65)	0.0	0.00
Japanese	1.2	0.11 (0.05-0.27)	0.3	0.15 (0.04-0.60)
Koreans	7.9	0.55 (0.28-1.11)	0.0	0.00
Filipinos	6.3	0.37 (0.21-0.66)	0.6	0.29 (0.09-0.91)
Vietnamese	6.1	0.49 (0.23-1.03)	0.6	0.26 (0.04-1.81)
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*0.009*		*0.27*

*Log likelihood ratio tests for difference within six Asian American ethnicities

Table 3. (Continued)

Cancer site	Men		Women	
	Rate	RR (95% CI)	Rate	RR (95% CI)
Lung, bronchus, & trachea				
White	454.3	1.00	206.8	1.00
Chinese	300.2	0.66 (0.62-0.70)	147.0	0.71 (0.66-0.77)
Indians	101.0	0.23 (0.18-0.29)	57.0	0.28 (0.21-0.38)
Japanese	257.0	0.53 (0.49-0.58)	99.0	0.48 (0.44-0.53)
Koreans	335.5	0.75 (0.67-0.83)	112.5	0.55 (0.48-0.64)
Filipinos	262.5	0.56 (0.52-0.61)	84.6	0.39 (0.35-0.44)
Vietnamese	285.5	0.64 (0.58-0.72)	114.2	0.56 (0.48-0.65)
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*<0.0001*		*<0.0001*
Skin				
White	16.9	1.00	7.7	1.00
Chinese	2.9	0.15 (0.08-0.26)	2.3	0.27 (0.15-0.51)
Indians	2.8	0.17 (0.04-0.70)	0.8	0.16 (0.02-1.15)
Japanese	1.5	0.10 (0.05-0.20)	1.5	0.14 (0.05-0.36)
Koreans	0.0	0.00	1.7	0.25 (0.08-0.77)
Filipinos	8.9	0.41 (0.24-0.70)	3.1	0.33 (0.14-0.79)
Vietnamese	1.9	0.11 (0.03-0.44)	0.0	0.00
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*0.006*		*0.21*
Breast				
White			129.3	1.00
Chinese			47.8	0.38 (0.33-0.44)
Indians			62.9	0.54 (0.41-0.71)
Japanese			56.2	0.45 (0.40-0.52)
Koreans			21.2	0.17 (0.12-0.24)
Filipinos			41.2	0.33 (0.28-0.39)
Vietnamese			23.5	0.19 (0.14-0.27)
p for testing difference from White				*<0.0001*
*p for testing heterogeneity**				*<0.0001*
Prostate				
White	213.5	1.00		
Chinese	88.6	0.42 (0.37-0.47)		
Indians	85.9	0.48 (0.36-0.62)		
Japanese	120.5	0.56 (0.49-0.63)		
Koreans	70.5	0.35 (0.28-0.45)		
Filipinos	147.0	0.61 (0.55-0.68)		
Vietnamese	42.7	0.23 (0.17-0.30)		
p for testing difference from White		*<0.0001*		
*p for testing heterogeneity**		*<0.0001*		

*Log likelihood ratio tests for difference within six Asian American ethnicities

Table 3. (Continued)

Cancer site	Men		Women	
	Rate	RR (95% CI)	Rate	RR (95% CI)
Cervix uteri				
White			7.7	1.00
Chinese			12.6	1.37 (1.05-1.79)
Indians			16.4	2.18 (1.27-3.76)
Japanese			4.9	0.58 (0.39-0.87)
Koreans			11.8	1.50 (0.98-2.31)
Filipinos			8.1	1.00 (0.69-1.46)
Vietnamese			13.7	1.74 (1.12-2.69)
p for testing difference from White				*0.10*
*p for testing heterogeneity**				*0.0002*
Corpus & uterus, NOS				
White			23.2	1.00
Chinese			10.0	0.40 (0.29-0.55)
Indians			19.9	0.92 (0.56-1.53)
Japanese			13.0	0.55 (0.42-0.72)
Koreans			8.2	0.38 (0.23-0.63)
Filipinos			10.3	0.43 (0.30-0.60)
Vietnamese			7.5	0.35 (0.19-0.63)
p for testing difference from White				*<0.0001*
*p for testing heterogeneity**				*0.083*
Ovary				
White			45.6	1.00
Chinese			24.8	0.54 (0.44-0.67)
Indians			29.4	0.70 (0.46-1.05)
Japanese			21.3	0.48 (0.40-0.59)
Koreans			16.0	0.35 (0.23-0.52)
Filipinos			22.5	0.48 (0.38-0.61)
Vietnamese			22.3	0.50 (0.35-0.72)
p for testing difference from White				*<0.0001*
*p for testing heterogeneity**				*0.26*
Kidney & renal pelvis				
White	30.7	1.00	15.5	1.00
Chinese	15.9	0.45 (0.34-0.60)	10.7	0.68 (0.49-0.96)
Indians	18.3	0.57 (0.32-1.00)	7.4	0.51 (0.23-1.13)
Japanese	21.4	0.55 (0.39-0.78)	7.4	0.35 (0.23-0.52)
Koreans	17.7	0.53 (0.33-0.87)	11.2	0.79 (0.49-1.25)
Filipinos	12.5	0.34 (0.25-0.48)	6.5	0.40 (0.26-0.60)
Vietnamese	11.4	0.40 (0.23-0.70)	3.2	0.22 (0.09-0.53)
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*0.38*		*0.01*

*Log likelihood ratio tests for difference within six Asian American ethnicities

Table 3. (Continued)

Cancer site	Men		Women	
	Rate	RR (95% CI)	Rate	RR (95% CI)
Bladder				
White	48.1	1.00	16.1	1.00
Chinese	25.3	0.46 (0.36-0.57)	9.3	0.59 (0.44-0.80)
Indians	28.0	0.67 (0.42-1.07)	4.9	0.42 (0.16-1.12)
Japanese	15.1	0.29 (0.21-0.39)	10.9	0.75 (0.54-1.06)
Koreans	18.7	0.39 (0.25-0.60)	6.4	0.45 (0.23-0.86)
Filipinos	13.1	0.23 (0.16-0.34)	4.5	0.25 (0.14-0.42)
Vietnamese	13.4	0.31 (0.19-0.51)	7.4	0.53 (0.30-0.97)
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*0.002*		*0.012*
Brain & CNS				
White	22.8	1.00	15.8	1.00
Chinese	7.8	0.29 (0.20-0.42)	6.5	0.43 (0.31-0.60)
Indians	8.5	0.37 (0.15-0.89)	11.3	0.64 (0.34-1.19)
Japanese	5.9	0.23 (0.13-0.40)	3.2	0.17 (0.10-0.28)
Koreans	12.3	0.49 (0.29-0.85)	4.4	0.27 (0.13-0.54)
Filipinos	12.3	0.47 (0.32-0.68)	6.0	0.34 (0.22-0.51)
Vietnamese	9.4	0.42 (0.23-0.76)	4.7	0.31 (0.16-0.63)
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*0.187*		*0.014*
Hodgkins disease				
White	2.3	1.00	1.7	1.00
Chinese	2.7	0.56 (0.21-1.48)	1.2	0.59 (0.22-1.58)
Indians	3.9	1.57 (0.39-6.26)	0.0	0.00
Japanese	1.6	0.76 (0.36-1.60)	1.6	0.61 (0.25-1.46)
Koreans	0.0	0.00	0.0	0.00
Filipinos	0.3	0.16 (0.02-1.12)	0.4	0.34 (0.08-1.35)
Vietnamese	3.1	1.54 (0.58-4.12)	0.6	0.43 (0.06-3.02)
p for testing difference from White		*0.032*		*0.001*
*p for testing heterogeneity**		*0.079*		*0.52*
Non-Hodgkins lymphoma				
White	54.2	1.00	41.8	1.00
Chinese	37.6	0.66 (0.55-0.78)	23.8	0.60 (0.49-0.61)
Indians	19.9	0.38 (0.22-0.64)	19.5	0.53 (0.33-0.70)
Japanese	45.0	0.79 (0.65-0.97)	20.6	0.50 (0.40-0.56)
Koreans	33.2	0.59 (0.41-0.84)	22.0	0.57 (0.41-0.67)
Filipinos	51.1	0.88 (0.74-1.04)	26.2	0.65 (0.52-0.66)
Vietnamese	25.7	0.51 (0.35-0.72)	18.0	0.46 (0.32-0.48)
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*0.002*		*0.47*

*Log likelihood ratio tests for difference within six Asian American ethnicities

Table 3. (Continued)

Cancer site	Men		Women	
	Rate	RR (95% CI)	Rate	RR (95% CI)
Leukemia				
White	55.8	1.00	33.4	1.00
Chinese	28.4	0.50 (0.41-0.61)	19.4	0.62 (0.49-0.78)
Indians	23.5	0.44 (0.27-0.70)	17.8	0.60 (0.36-1.00)
Japanese	28.5	0.44 (0.34-0.56)	14.2	0.40 (0.30-0.52)
Koreans	22.5	0.44 (0.29-0.67)	10.2	0.31 (0.19-0.50)
Filipinos	32.0	0.53 (0.43-0.66)	19.8	0.65 (0.51-0.84)
Vietnamese	15.9	0.30 (0.19-0.48)	14.1	0.46 (0.30-0.70)
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*0.26*		*0.01*
Multiple myeloma & immunoproliferative neoplasms				
White	27.1	1.00	19.1	1.00
Chinese	13.3	0.44 (0.32-0.61)	7.7	0.33 (0.24-0.47)
Indians	26.9	1.07 (0.66-1.71)	15.2	0.93 (0.55-1.57)
Japanese	14.2	0.47 (0.34-0.66)	9.8	0.42 (0.30-0.60)
Koreans	14.9	0.57 (0.33-0.96)	10.6	0.52 (0.32-0.87)
Filipinos	18.6	0.64 (0.49-0.83)	16.6	0.93 (0.70-1.24)
Vietnamese	10.2	0.40 (0.22-0.72)	7.8	0.43 (0.24-0.77)
p for testing difference from White		*<0.0001*		*<0.0001*
*p for testing heterogeneity**		*0.05*		*0.0001*

*Log likelihood ratio tests for difference within six Asian American ethnicities