Age and Race Data on Death Certificates for Older Asian Americans

Diane S. Lauderdale and Bert Kestenbaum

Short Abstract

In this project we assess the quality of age and race reported on death certificates for older Asian Americans. We previously determined death probabilities for the elderly of six subgroups (Chinese, Filipino, Indian, Japanese, Korean, and Vietnamese) using Social Security (SSA) files and identification algorithms based on race, place of birth and names. We link the SSA death records from that study (n=136,350) to death certificates, using the National Death Index and the California Death Statistical Masterfile. Comparing year of birth on death certificates to SSA records, we find agreement is generally good, but varies by ethnicity, nativity and age. To assess ethnic identification on death certificates, we focus on persons for whom our ethnic assignment in the SSA file is most confident (foreign born with distinctive names). Ethnic identification on death certificates varies by ethnicity and markedly by state, with more complete identification in states with larger Asian populations.

Diane S. Lauderdale, Department of Health Studies, University of Chicago, 5841 S. Maryland Ave. MC 2007, Chicago, IL 60337; E-mail: lauderdale@health.bsd.uchicago.edu. Bert Kestenbaum, Social Security Administration. The work of the first author was supported by NIA R01 AG17889-01. The opinions expressed in this paper are those of the authors, and no official endorsement by the Social Security Administration should be inferred.

Long Abstract

Background

Age-specific death rates are typically constructed by combining vital statistics (for the numerator) with census data (for the denominator). Rates constructed in this way by the National Center for Health Statistics (NCHS) for the race category "Asian or Pacific Islander (API)" are extraordinarily low, less than two-thirds those for whites in all age groups (Arias & Smith 2003). Confidence in these rates depends largely on the quality of the age and race information on the death certificates used to construct them. Both age and race are in the section of a death certificate that is filled out by a funeral director, ideally through consultation with the next-of-kin. The quality of age data on death certificates does not appear to have been previously examined for Asian Americans, or for the foreign-born. There is evidence that quality of age data for the elderly varies by race, specifically Black versus White (Preston *et al*, 1996; Hill *et al*, 1999; Kestenbaum, 1992), and that the quality of age data could meaningfully affect comparisons of old age mortality within and between countries (Manton and Vaupel, 1995; Kestenbaum, 1997; Hill *et al ,* 1999).

While the quality of age information on death certificates for Asian Americans is previously unexplored, race information on death certificates for Asian Americans is a well-recognized problem. However, key features of the problem, such as whether under-reporting varies by ethnicity, age, sex or part of the country, have not been examined. NCHS routinely attaches this disclaimer to tables of death rates: "The number of deaths and death rates for Hispanic origin and specified races other than white or black should be interpreted with caution because of inconsistencies between reporting Hispanic origin and race on death certificates and censuses and surveys…" (Arias & Smith 2003). There have been several studies aimed at determining the extent of Asian under-reporting on death certificates. To our knowledge, studies of adult death records have focused on the aggregate Asian (or Asian and Pacific Islander -- API) race category only. From a small sample study, NCHS estimated that under-reporting causes death rates for the total API to be understated by about 11 percent (Rosenberg *et al*, 1999). In the National Longitudinal Mortality Study, Sorlie et al. (1992) compared the race item on 12 Current Population Surveys with death certificates. Of the 272 identifying themselves as API in the Current Population Reports, 224 were identified as any API on death certificates (82.4%). While the 82.4% figure from this study is widely cited, and has been used to adjust death rates (Hahn and Eberhardt, 1995), it is based on a very small sample. Several infant mortality studies have used files of linked birth and death certificates (Frost and Shy, 1980; Yu, 1984; Hahn *et al*, 1992) to examine concordance of race for infants who died. Because of the much larger sample size available for these national linked files, Asian subgroups have been separately assessed. Hahn *et al* found a consistent pattern of *greatly* under-reported Asian ancestries on infant death certificates compared to birth certificates: only 40.4% of Japanese; 52.4% of Chinese, 44.9% of Filipinos, and 54.6% of Other Asian and Pacific Islanders so identified on birth certificates were similarly identified on death certificates. Most of the others were recorded as White. Clearly, under-reporting of Asian ancestry on death certificates relative to the source of denominator data would produce spuriously low death rates, explaining in part the very large observed discrepancy between the API and white groups for both infants and adults.

We have previously determined sex-and-age-specific death probabilities for the elderly of six Asian American subgroups: Chinese, Filipino, Indian, Japanese, Korean, and Vietnamese (Lauderdale and Kestenbaum, 2002). We did this by creating ethnic identification algorithms based on race, place of birth, surname, and given name so that a single file that lacked ethnic detail (the Social Security Administration file of persons enrolled in Medicare) could be used to identify both a population at risk and deaths. The derivation of the surname lists and the ethnic identification algorithms have been previously described (Lauderdale and Kestenbaum 2000, 2002). From that project, we have Social Security death information for 136,353 persons who died from 1990 to 1999

at age 65 or older and whom we have identified as belonging to one of the six Asian American subgroups. In this project, we link the SSA death records to death certificates, using both a complete file of California death certificates for 1990 to 1999 and the National Death Index (NDI). We compare year of birth reported on death certificates with year of birth from the Social Security Administration (SSA), which generally requires written documentation. We investigate whether concordance in birth year varies by ethnicity, sex, age or foreign birth.

Because our ethnic identification in the SSA files involves some uncertainty, discrepancies between our ethnic identification and death certificate identification could partly reflect failures of our algorithm. Therefore, to assess the completeness of Asian subgroup identification on death certificates, we take a conservative approach and only assess death certificate race data for persons with the most certain ethnic assignment in the SSA file (they have both informative place of birth and a distinctive surname/maiden name). Because the NDI only provides information on race category agreement for three of the Asian subgroups (Chinese, Japanese and Filipino), the evaluation of the race data is confined to these groups We investigate whether under-reporting varies by ethnicity, state and sex.

Methods

Data Sources

This project uses two administrative files at SSA to identify Asian American deaths and two files of death certificates, the 1990 to1999 Death Statistical Master files for the state of California and the National Death Index (NDI). The two Social Security files and their application in this project have been previously described in detail (Lauderdale and Kestenbaum 2002). Below is a summary of the SSA files and the ethnic identification algorithms.

SSA files

The primary SSA file is the Master Beneficiary Record (MBR), which houses a record for each person ever entitled to Social Security or Medicare benefits. The accuracy of SSA administrative data items is related to their importance for program administration. Thus, since age is the basis of entitlement for some programs, proof is required for date of birth. The expectation of accurate date of birth information, given the evidentiary requirements, must be tempered by the reality that birth certificates or similar documents often are unavailable, especially for the foreign-born. SSA has detailed policies about the value of various types of records with year of birth or age information, and specific descriptions of the types of documents available from some Asian countries of origin (Social Security Administration, 1995). The second SSA file we use is the NUMIDENT, the file of applications for a social security number (SSN), for a replacement card, or for a revised card. From this file we select data items useful for Asian subgroup identification: race, country of birth, surname, father's surname (for women), and given name. The SSN links the records from the two files (MBR and NUMIDENT).

Asian Subpopulation Identification in the SSA files

There is no explicit *subgroup or ethnic* identification for Asian Americans in the SSA files; we infer subgroup by combining three NUMIDENT data elements: race, country of birth and name.

1. Race in the NUMIDENT is one of six categories: white, black, other (i.e., neither white nor black), Hispanic, Asian or American Indian, or it is blank. The six categories represent the conflation of two coding schemes. Before November 1980, the choices for race on an SSN application were white, black and other. Since November 1980, the choices are white non-Hispanic, black non-Hispanic, Hispanic, Asian, and American Indian. Most elderly persons last completed an application before 1980. From previous work (Lauderdale and

Goldberg, 1996) we know that, except for Asian Indians, Asian Americans generally check the "Asian" box on the current form, and the "other" box on the pre-1980 form.

2. Country of birth is very informative, because the elderly in four of our six subpopulations are overwhelmingly foreign-born. Unfortunately, country of birth is missing for some persons, generally those who filed claims for retirement or disability benefits through the mid-1970s, when the NUMIDENT was computerized. While the NUMIDENT was a paper file, the original application was physically removed when a person applied for benefits and sent to the local office. The replacement form did not include some of the information on the original, including place of birth.

3. For the native-born as well as the foreign-born, **name** may be predictive of Asian subpopulation membership. Because name change with intermarriage could result in misclassification, we replace a woman's own surname with her father's surname. An earlier paper describes in detail our derivation and evaluation of **surname** lists for each of the six Asian populations of interest (Lauderdale and Kestenbaum, 2000). One set of lists is *conditional* on Asian race information, and the other set is *unconditional*. We derived **given name** lists similarly to the surname lists, except that there are separate lists for men and women.

Thus for the purpose of assigning subpopulation membership, we have a race code and 3 items of information: place of birth (if known), whether the surname/maiden name is on the surname lists, 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. 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). 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 ethnic identification when the race code is contra-indicative.

We expect that these methods will identify a majority of persons who would self-identify as belonging to one of the six Asian populations. Because of variation across populations in the inclusiveness of the name lists, and also the percent foreign-born, these methods will not identify the same proportion of each population. This approach assumes that mortality is similar for persons who have names on the lists (that is names which are specific to a single ethnic group and not rare) and persons of the same ethnicity who do not have names on the lists.

These methods identified persons who contributed 4.5 million person-years of observation from 1990 to1999, among whom136, 350 deaths occurred. Social Security death records include date of birth, date of death, sex, given name, surname, father's surname, state of last residence, SSN, state or country of birth (if available), and race code. Our next step was to match the 136, 350 SSA death records to death certificates.

Death Certificate Matching

Because 40 percent of the SSA death records listed California as the last state of residence, we stratified the records into California versus other states. California makes available to researchers their Death Statistical Masterfile, a complete file of death certificates (with name and SSN). We matched SSA death records for persons last residing in California to the California death certificates by SSN. Ninety-five percent of 60,449 California SSA deaths matched a California death record by SSN (and sex). For the California linked death records, all of the information on the death certificate was available. For this project, useful elements from

California death records were year of birth, place of birth (however, some of the Asian countries are grouped in the residual "Rest of World" category), and detailed race code, including all six specific separate Asian ethnic groups.

The non-California deaths, as well as the California deaths that did not link by SSN, were submitted to the NDI The NDI is an index to the complete death record files submitted by all states to the National Center for Health Statistics. The primary purpose is to identify the location and certificate number of death certificates potentially matching submitted information. When a file of individual records is submitted, a set of potential matches is returned for each submitted record. For every one of the potential matches, there is an indicator of whether each item of submitted information (e.g. SSN, surname, race, place of birth) is exactly the same on the potentially matching death certificate. The matching process is described in detail in the National Death Index User's Manual (Bilgrad 2000). Potential matches are returned with probability scores and a numeric classification reflecting the number and type of matching items. If an item does not match (e.g. race), only that fact is returned, not the actual content in the NDI record. The exception to this is year of birth, where the difference in birth years between the submitted information and the NDI record is returned (e.g. –2 or +3).

To select the best match among the set of returned potential matches, we did not just pick the highest probability score. There were two reasons. The first reason was that a matching date of birth would contribute to the probability score, and we wished to assess concordance for year of birth. Thus selecting the record with highest probability score would bias upward our estimate of year-of-birth concordance. (Note that since year of birth was one of the many variables used by NDI to select the set of potential matches, our year of birth concordance may still be biased upward, although not greatly.) Second, we knew the date of death with relative certainty for each record; since most NDI users do not know fact and date of death when submitting records, this information does not factor into the probability score. However, we were able to use date of death as an essential matching variable, since we can assume it to be highly accurate in both the death certificate and the SSA record. Our basic matching algorithm required that at least seven of the nine SSN digits match and that sex, year and month of death, state of death and at least one of the names (allowing for phonetic matches as determined by NCHS) also match This yielded 70,137 unique, likely true matches for the 78,630 submitted records. Below is a table of the proportion matching by ethnicity. A key determinant of matching variation was the proportion of death certificates with an accurate SSN.

Table 1. SSA death records which were successfully matched to a death certificate (DC), either through the California file of death records or through the NDI.

	SSA records	DC matched	% matched
TOTAL	136,350	127,892	93.8
ETHNICITY			
Chinese	42,312	39,978	94.5
Indian	7,460	|6,674	89.5
Japanese	35,327	34,909	98.8
Korean	11,050	10,497	95.0
Filipino	31,423	27,508	87.5
Vietnamese	8,778	|8,326	94.9

Year of Birth Agreement

We compared year of birth for all matching SSA-DC records, and determined the proportion with the *same* year of birth and the proportion with a *close* year of birth, which we defined as within one year (+1, 0 or –1

difference). We separately determine these proportions by sex, ethnicity, foreign birth (information from the death certificates) and age (65-84 and 85+).

Year of birth agreement was very good. Overall 95.1 percent had the same year of birth on the death certificate as was given in the SSA files. Agreement did not vary by sex, but did vary by age, nativity and ethnicity. Predictably, agreement was less for older persons and for the foreign-born. With respect to ethnicity, Chinese and Filipinos had less agreement and Japanese and Vietnamese the most. For persons who did not have the same year on the two sources, the death certificate year of birth was on average 0.65 years older for the US-born and 0.19 years younger for the foreign-born.

Table 2. Year of birth agreement between Social Security records and death certificates for persons who died at age 65 or older (according to Social Security records)

	Year of Birth	
	Same (%)	Within 1 Year (%)
ALL	95.1	97.8
SEX		
Male	95.1	97.7
Female	95.0	97.9
ETHNICITY		
Chinese	93.2	96.8
Indian	96.3	97.8
Japanese	97.4	99.3
Korean	96.0	98.1
Filipino	94.5	96.9
Vietnamese	97.5	98.3
BIRTHPLACE		
U.S.	97.3	99.1
Foreign	94.2	97.3
AGE		
65- 84	96.0	98.1
85+	94.4	97.6

Completeness of Racial Subgroup Identification on the DC

To evaluate the death certificate race information on a national basis, we are limited to the three subgroups (Chinese, Japanese and Filipino) for which the NDI will return specific matching information. The other Asian groups are all in the residual "other Asian" category. To avoid including any SSA records where our ethnic identification was less certain, we only select records for persons with the appropriate foreign country of birth in the SSA files (China, Hong Kong and Taiwan for Chinese, Japan for Japanese, and the Philippines for Filipino), and further require that the person have a surname/maiden name on our conditional name list for the corresponding subgroup. It seems highly likely that these persons are of the appropriate ethnic group. We determine the proportions by ethnicity with matching race category on the death certificate. This evaluation of the death certificate race item confined to the foreign-born with informative names is likely to overestimate correct identification, since the birthplace information is generally also available to the funeral director, and thus presumably he or she may be more likely to record the specific subgroup in the race box. The name may also provide intuitive information. We compare these proportions by ethnicity, sex, and state. We hypothesized that states with larger Asian populations would have more complete Asian race reporting because of familiarity.

We group states into three categories: California (which has by far the largest Asian population), six states with the next largest Asian populations (Hawaii, Illinois, New Jersey, New York, Texas, Washington) and all other states.

We found evidence of ethnic and sex variation in completeness of race identification, and a strong pattern of state variation. Identification is nearly complete in California for all three groups.

Overall, 92.2 percent of persons born in China, Taiwan or Hong Kong and with a distinctively Chinese surname (or maiden name for women) were identified as Chinese on the death certificate. That proportion did not vary by sex. However, it did vary greatly by state. In California, 98.4 percent were identified as Chinese on the DC, the percent so identified was less in the six states with the next largest Asian population (91.1 percent), and the identification was much less complete (75 percent) in all the other states. For Japanese and Filipino the overall identification on the death certificate was lower than for Chinese (88.7 percent for each). For Japanese, there is a pattern of sex variation, with a much higher percentage of men than women identified as Japanese on the death certificate (95.6 versus 87.3 percent). For Filipino, identification does not vary by sex. For Japanese, identification is nearly complete in California, moderate in the other large Asian population states (92.6 percent) and much lower in the rest of the states (67.6 percent). The pattern of less complete identification in states with smaller Asian population is most marked for Filipinos, with only 82.3 percent identified in the six relatively large Asian population states and 61.6 percent identified in the rest of the states.

Table 3. Race identification for persons with high likelihood of being Chinese, Filipino and Japanese (corresponding foreign place of birth in the Social Security files and a distinctive surname/maiden name)

	Chinese		Japanese		Filipino	
	N	% identified as Chinese on DC	N	% identified as Japanese on DC	N	% identified as Filipino on DC
ALL	21474	92.2	3646	88.7	18236	88.7
Male	11685	92.0	612	95.6	10222	89.3
Female	9789	92.5	3034	87.3	8014	87.9
BY STATE						
California	10759	98.4	1496	97.6	11622	95.7
Other states with large Asian populations *	7294	91.1	1284	92.6	4719	82.3
All Other States	3421	75.0	866	67.6	1895	61.6

* Hawaii, Illinois, New Jersey, New York, Texas, Washington

Discussion

The discussion will compare these data with other studies of death certificate age and race, explore reasons for the patterns observed, and describe the implications for calculating death rates.

Cited Literature

Arias E, Smith BL. 2003. Deaths; Preliminary Data for 2001. *National Vital Statistics Reports*; vol 51, no. 5. Hyattsville, Maryland: National Center for Health Statistics.

Bilgrad, R. 2000. National Death Index User's Manual. National Center for Health Statistics: Hyattsville, Maryland.

Frost F, Shy KS. 1980. Racial differences between linked birth and infant death records in Washington state. *American Journal of Public Health* 70:974-976.

Hahn RA, Eberhardt S. 1990. Life expectancy in four U.S. Racial/ethnic populations: *Epidemiology* 1995;6:350-355.

Hahn RA, Mulinare J, Teutsch SM. 1992. Inconsistencies in coding of race and ethnicity between birth and death in US infants. *Journal of the American Medical Association* 267:259-263.

Hill ME, Preston SH, Rosenwaike I. 2000 Age reporting among white Americans aged 85+: results of a record linkage study. *Demography* May;37(2):175-86.

Kestenbaum B. 1992. "A Description of the Extreme Aged Population Based on Improved Medicare Enrollment Data." *Demography* 29:565-580.

Kestenbaum B. 1997. Recent mortality of the oldest old, from Medicare data. Paper presented at the Annual Meeting of the Population Association of American.

Lauderdale DS, Goldberg J. 1996. "The Expanded Racial and Ethnic Codes in the Medicare Data Files: Their Completeness of Coverage and Accuracy." *American Journal of Public Health* 86:712-716.

Lauderdale DS, Kestenbaum B. 2000. "Asian American Ethnic Identification by Surname*." Population Research and Policy Review* 19:283-300.

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

Manton KG, Vaupel JW. 1995. Survival after the age of 80 in the United States, Sweden, France, England, and Japan. New England Journal of Medicine 333:1232-5.

Rosenberg HM, Maurer JD, Sorlie PD, Johnson NJ, MacDorman MF, Hoyert DL, Spitler JF, Scott C. 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).

Preston SH, Elo IT, Rosenwaike I, Hill M. 1996. African-American mortality at older ages: results of a matching study. *Demography* 33:93-209.

Sorlie PD, Rogot E, Johnson NJ. 1992. Validity of demographic characteristics on the death certificate. *Epidemiology*;3:181-184

[United States.] Social Security Administration. Office of Policy. 1995. *POMS, Program Operations Manual System.* [Baltimore, Md.] : U.S. Dept. of Health and Human Services, Social Security Administration , Office of Policy.

Yu E. 1984. The low mortality rates of Chinese infants: some plausible explanatory factors. *Social Science and Medicine* 16:253-262.