

Identifying Latino and Asian Populations through the use of Surnames

Betsy A. Kohler MPH, CTR

Executive Director, NAACCR

2009 Collaborative NCCCP & NPCR
Conference

April 15-17 2009



Background- The Problem

- Cancer rates vary in different population sub groups
- For calculation of cancer rates, need to have compatible numerators (cases) and denominators (population)
- Denominator = census data
 - Self-reported race
 - Up to 5 races, 2 ethnicities per person

Race Determination for Cancer Cases

- Race and ethnicity are determined in a variety of ways
 - Medical record
 - Observation
 - Self-identification
 - On death certificate
- Up to 5 races/ 2 ethnicities – rarely more than one specified
- Rules-Pick minority race/ethnicity

NAACCR Race Codes

- 01 White
- 02 Black
- 03 American Indian, Aleutian, or Eskimo
- 04 Chinese
- 05 Japanese
- 06 Filipino
- 07 Hawaiian
- 08 Korean
- 10 Vietnamese
- 11 Laotian
- 12 Hmong
- 13 Kampuchean (Cambodian)
- 14 Thai
- 15 Asian Indian or Pakistani, NOS

NAACCR Race Codes, Continued

- 16 Asian Indian
- 17 Pakistani
- 20 Micronesian, NOS
- 21 Chamorran
- 22 Guamanian, NOS
- 25 Polynesian, NOS
- 26 Tahitian
- 27 Samoan
- 28 Tongan
- 30 Melanesian, NOS
- 31 Fiji Islander
- 32 New Guinean
- 96 Other Asian, including Asian, NOS and Oriental, NOS
- 97 Pacific Islander, NOS
- 98 Other
- 99 Unknown

NAACCR Ethnicity Codes

- 0 Non-Spanish; non-Hispanic
- 1 Mexican (includes Chicano)
- 2 Puerto Rican
- 3 Cuban
- 4 South or Central American (except Brazil)
- 5 Other specified Spanish/Hispanic origin (includes European; excludes Dominican Republic)
- 6 Spanish, Hispanic, Latino, NOS
- 7 Spanish surname only.
- 8 Dominican Republic
- 9 Unknown whether Spanish or not



Background



- Convened expert panel in 2001 to see if we could develop a standardized method for identifying H/L in central cancer registries
- Panel members from states with large H/L populations
- Knew there was undercounting of Latinos in general, and specific H/L groups even more undercounting

Issues for Consideration

- No gold standard for comparison of H/L rates
- Cancer rates vary among H/L sub-groups by country of origin
- H/L sub-groups tend to cluster geographically
- Age structure of H/L differ from US general population, and from group to group
- Length of time in US varies geographically and by subgroup-important etiologically
- Surname algorithms have difficulty with Portuguese, Italian and Filipino names
- Problems with 2000 census
- Issues with self-identification

Available Data

- Last Name
- Maiden Name
- Place of Birth
- Reported Ethnicity
- Percent population in county of residence that is H/L

Basic Steps- One

- Retain ethnicity reported by facility if classified as Mexican, Cuban, Puerto Rican, Dominican, South or Central American or Other Spanish,
- Use indirect method for cases who are Non-Hispanic/Latino, unknown, and Hispanic/Latino NOS

Step Two

- Filter cases based on birthplace
 - Some areas have high prevalence of Spanish surnames, but are not H/L (eg. Phillipines, Carribbean, Brazil)-Exclude
 - Other areas have high prevalence of H/L (eg. Mexico, most of South America, Latin America, etc.)--Include

Steps Three and Four

- Exclude certain races (Hawaiian, American Indian/ Alaskan Native, Filipino)
- Option to exclude anyone living in county with less than 5% of population Hispanic/Latino

Step Five-- Surname Matching

- Cases run through an algorithm based on:
 - Last Name or
 - Maiden Name for Females
- If names are considered “Heavily Hispanic” they will be considered Hispanic/Latino by algorithm
 - If women have no Maiden Name available, will be classified as H/L based on spouse’s name

Identification of Asian Pacific Islanders



- Similar problems
 - Variation of cancer rates by API race groups
 - Race assigned by observation
 - Known to be undercounting
 - No gold standards for calculating rates
- Task: Assign Specific Race to cases classified as “Asian/Pacific Islander, NOS”

Available Data

- First Name
- Last Name
- Maiden Name
- Sex
- Place of Birth

Steps for Assigning Specific API Race Codes

- 1. Identify all cases coded to 96=Other Asian, Asian NOS, Oriental NOS
- 2. Evaluate other Race codes-if other API race code is listed, use this specific code
- 3. For Asian NOS (Code 96) evaluate place of birth (eg. Birthplace= China, Hong Kong, Taiwan or Macao then Race=Chinese)
- 4. Exclude ambiguous birthplaces such as Nepal, Hawaii, Indonesia and Hispanic/Latino birthplaces

Step 5 Use API Name Lists

- Lauderdale and Kestenbaum published lists of surnames and first names strongly predictive of Chinese, Japanese, Korean, Filipino, Asian Indian, or Vietnamese race, based on an examination of 1.8 million Social Security applications for persons born in Asia before 1941—these races account for about 91% of Asian population

Other Lists

- NAACCR List- 80,000 cancer cases among Asians from 1997-2001 from seven NAACCR registries (Hawaii, Los Angeles, Louisiana, Illinois, Nevada, New York, Texas)
- List of Hmong Names from Cancer Registry of Central California

Both Surnames and Given names are checked



Identifying American Indians/Alaskan Natives



- Agreement was reached with the Indian Health Service
- Link State Cancer Registry Data with files of those receiving healthcare from Indian Health Services
- If link made, Race is considered American Indian/ Alaskan Native
- Annual match for states with CHSDA counties

Conclusions

- By using available data, careful considerations, and empirical validation, we can improve the identification of minority race and ethnicity coding in cancer registries
- Important for Cancer Control Groups to use the same classification systems
- Match with your central registry to obtain Race/Ethnicity

Next Steps

- Making data available for Race/Ethnic groups
 - AI/AN data in CINA 2009 by region, not state
 - API data available for Hawaii in CINA in 2009
- Improved numerators, need good denominators
- Focus attention on other groups

A Cautionary Tale



Merry Christmas
From

The McDonalds