UCLA

Title

Validity of Race, Ethnicity, and National Origin in Population-based Cancer Registries and Rapid Case Ascertainment Enhanced With a Spanish Surname List

Authors

Clarke, Lisa C
Rull, Rudolph P
Ayanian, John Z
et al.

Peer reviewed

Validity of Race, Ethnicity, and National Origin in Population-Based Cancer Registries and Rapid Case Ascertainment Enhanced with a Spanish Surname List

Lisa C. Clarke, PhD,
Epidemiology Program, County of Marin, San Rafael, CA

Rudolph P. Rull, PhD,
School of Community Health Sciences, University of Nevada, Reno, Reno, NV

Dept of Health Research and Policy, Stanford School of Medicine, Stanford, CA

John Z. Ayanian, MD,
Brigham and Women's Hospital, Boston, MA

Dept of Health Policy, Harvard Medical School, Boston, MA

Robert Boer, MS,
Dept of Public Health, Erasmus Medical Center, Rotterdamn, The Netherlands

Dennis Deapen, DrPH,
Department of Preventive Medicine, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA

Dee W. West, PhD, and
Cancer Prevention Institute of California, Fremont, CA

Katherine L. Kahn, MD
The RAND Corporation, Santa Monica, CA

UCLA School of Medicine, Los Angeles, CA

Lisa C. Clarke: drlisacclarke@gmail.com; Rudolph P. Rull: rrull@unr.edu; John Z. Ayanian: ayanian@hcp.med.harvard.edu; Robert Boer: fietsrob@gmail.com, dutchrob@yahoo.com; Dennis Deapen: ddeapen@usc.edu; Dee W. West: Dee.West@cpic.org; Katherine L. Kahn: kahn@rand.org, kkahn@mednet.ucla.edu

Abstract

Background—Accurate information regarding race, ethnicity, and national origins is critical for identifying disparities in the cancer burden.

The corresponding author's address, phone number, fax number, and email address are as follows: Katherine L. Kahn, M.D., Senior Natural Scientist, RAND Corporation, 1776 Main Street, Santa Monica, CA 90407, Phone: (310) 794 2287, Fax: (310) 794-0732, kahn@rand.org.

Objectives—To examine the use of a Spanish surname list to improve the quality of race-related information obtained from rapid case ascertainment (RCA) and to estimate the accuracy of race-related information obtained from cancer registry records collected by routine reporting.

Subjects—. Self-reported survey responses of 3,954 participants from California enrolled in the Cancer Care Outcomes Research and Surveillance Consortium (CanCORS).

Measures—Sensitivity, specificity, positive predictive value (PPV), and percent agreement. We employed logistic regression to identify predictors of under-reporting and over-reporting of a race/ethnicity.

Results—Use of the Spanish surname list increased the sensitivity of RCA for Latino ethnicity from 37% to 83%. Sensitivity for cancer registry records collected by routine reporting was 95% for Whites, Blacks, and Asians, and specificity was high for all groups (86–100%). However, patterns of misclassification by race/ethnicity were found that could lead to biased cancer statistics for specific race/ethnicities. Discordance between self- and registry-reported race/ethnicity was more likely for women, Latinos, and Asians.

Conclusion—Methods to improve race and ethnicity data, such as using Spanish surnames in RCA and instituting data collection guidelines for hospitals, are needed to ensure minorities are accurately represented in clinical and epidemiological research.

Keywords

cancer; race/ethnicity; Spanish surname list; rapid case ascertainment

INTRODUCTION

Accurate and reliable information on patients' race, ethnicity, and national origin is critical to the identification of disparities in the burden of cancer, particularly for studies that require oversampling of certain groups. Although all states in the United States now have cancer registries, rapid case ascertainment (RCA) is a commonly employed method used by researchers to identify incident cancer cases for studies involving active patient contact soon after diagnosis.[1] Under RCA, central registry technicians regularly review pathology reports at hospitals, other medical facilities and free-standing pathology laboratories where microscopic verification of cancer is recorded. These technicians then collect a copy of each pathology report describing a verified neoplasm and also review the patient's admission record to obtain demographic information, including race/ethnicity. Pathology reports and demographic information made available to approved researchers provides patient contact information more quickly than would occur through routine registry reporting.[1] These data, however, are usually not quality controlled since the medical record for these cancers are just being established. Admission records reviewed using RCA may also contain race/ethnicity information reflecting observations of health care workers, rather than patients' self-reported race-related information.[2,3]

Characteristics such as American Indian/Alaska Native race, Latino (or Hispanic) ethnicity, and Latino and Asian national origins are frequently missing or incorrect in medical record, which may hamper the identification and recruitment of cancer patients from these

populations.[2,4–14] A simple and economically feasible method to potentially improve the validity of race/ethnicity from RCA is the use of a Spanish surname list.[15]

Data collected by cancer registries using routine reporting are typically subjected to additional data inputs that become apparent with time, as well as additional quality checks than those collected using RCA. With routine reporting, as compared with RCA, hospitals ascertain, abstract and report all cancer cases diagnosed or treated at their facility. Hospitals collect data from many sources and these data are subjected to edits and then the reports are submitted electronically to the local central cancer registry where quality control is performed through a combination of automated and manual methods. Data gathering and quality assurance steps are likely to include edit checks across the multiple sources from which hospitals receive data inputs, and electronic submission to the local central cancer registries where quality control is performed using both automated and manual methods. In addition, routine reporting utilizes an algorithm that incorporates maiden and surname, race, sex, and place of birth to distinguish between Latino and non-Latino patients and improve the validity of race/ethnicity information obtained from medical records.[2,15–17] With these methods in place, studies have found a sensitivity of 69% and 79% for Latino patients diagnosed in Northern California from 1973–1990 and in eleven SEER regions across the U.S. from 1973–2001, respectively.[5,6]

Although accurate information on race, ethnicity and national origin facilitates the identification and inclusion of minorities in clinical research, a comprehensive evaluation of the accuracy of RCA-derived race/ethnicity used in conjunction with a Spanish surname list has not been conducted. However, several studies have concluded that race/ethnicity collected through medical records can be improved by using surname lists. One study of patients in Northern California diagnosed with cancer in 1990 found that the sensitivity of Latino ethnicity increased from 59% to 68% when ethnicity obtained from medical records was used in conjunction with surname.[2]

In addition, the validity of registry-reported race, ethnicity, and origin is largely unknown for several minority populations and is necessary for accurate cancer statistics.[2,4–6,18] The only known study of Asian origin in cancer registries observed that accuracy varied greatly from 70% to 90% among Asians patients of Chinese, Japanese, Filipino, Vietnamese, and other origin diagnosed between 1973 and 1999 in Northern California.[5]

The accuracy of data on Latino origin, including Mexicans, Cubans, Puerto Ricans, Dominicans, Central or South Americans, or other national or regional origins has not been comprehensively validated against self-reported origin. Consequently, reports of cancer statistics aggregated for multiple nations of origin within a specific race/ethnicity will mask important origin-specific differences, while origin-specific statistics may be inaccurate.[19–21] Both scenarios potentially impede efforts to identify and reduce health disparities.[22]

In addition, one study of Latino cancer patients in Florida examined characteristics of Latinos who were misclassified as non-Latino. Pinheiro et al. found that Latinos of Black race and women were more likely to be misclassified as non-Hispanic than other Latinos.[23] Acculturation is the adoption of language, culture, beliefs, and behaviors of the dominant

culture.[24,25] Latinos born in the United States, one aspect of acculturation, were also more likely to be misclassified.[23] These results are consistent with numerous studies that have identified disparities in cancer rates, screening, and outcomes associated with one's level of acculturation and demonstrate the importance of examining acculturation when possible.[24,26–31]

To assess the extent of misclassification of race, ethnicity and origin for a diverse multiethnic population, we compared self-reported data collected from telephone interviews of 3,954 California participants diagnosed between 2003–2005 in the Cancer Care Outcomes Research and Surveillance (CanCORS) Consortium[32,33] with 1) information collected via RCA using a Spanish surname list [15] and with 2) race, ethnicity and national origin as reported in cancer registries. In both comparisons, self-reported information was used as the gold standard. In the comparison to registry records, we also identified socio-demographic characteristics associated with discordance between these sources of race/ethnicity data.

METHODS

Study Population

The CanCORS Consortium is a population-based study of 10,547 patients with incident lung or colorectal cancer from multiple regions of the U.S.[32] This study's data collection methods have been previously described in detail.[34] Briefly, study participants at least 21 years of age with a histologically-confirmed diagnosis in 2003–2005 were identified and interviewed.[34] A baseline telephone interview was conducted in English, Spanish, Mandarin or Cantonese four to seven months after the date of diagnosis with either the cancer patient or a surrogate (relative or household member) if the patient was deceased or too ill to complete the questionnaire.[34] During the interview, trained interviewers collected self-reported data on patient demographics, treatment, health history, and other characteristics.[34]

Race and ethnicity are considered distinct by the U.S. Office of Management and Budget (OMB).[35] Race includes five categories: White, Black or African American, American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander.[35] Ethnicity includes two categories: Latino or non-Latino.[36] A person is Latino if they or their family originated Mexico, Puerto Rico, Cuba, Spanish speaking Central and South American countries, and other Spanish cultures[36]; Latinos can be of any racial background.[36]

Race/ethnicity was initially obtained using RCA and later assessed during the telephone interview. Following these guidelines, participants were asked theses questions regarding race, ethnicity, and national origin: 1) "Are you of Latino or Hispanic origin?"; 2) If yes, "which group best describes your Latino or Hispanic origin?" (Mexican, Puerto-Rican, Cuban or Cuban-American, or Other Latino); 3) "Would you describe yourself as Native Hawaiian, Other Pacific Islander, American Indian, Alaska Native, Asian, Black, African American, or White? Or more than one of these?"; 4) If Asian, "[w]hat specific ethnic group are you? "(Chinese, Japanese, Filipino, Vietnamese, Cambodian, Korean, Indian, Pakistani, Other). The study protocol was approved by Institutional Review Boards at all participating research sites.[37]

Among the 10,547 CanCORS participants, 4,168 resided in California counties covered by the Los Angeles County Cancer Surveillance Program and selected counties in the Greater Bay Area Cancer Registry (including Alameda, Contra Costa, San Francisco, San Mateo, Santa Clara and Monterey counties). Of these 4,168 participants, 3,954 (95%) were successfully linked with the central registries' race-related data that were coded according to Surveillance, Epidemiology, and End Results (SEER) guidelines.[38] For 2,071 California patients for whom race-related information was not collected by the hospital, surname was used to assist in the identification of race/ethnicity.

Statistical Analysis

To optimize consistency with SEER guidelines for racial and ethnic categorization of cancer registry data, we grouped each participant into one of six mutually exclusive categories of self-reported race/ethnicity: White, Black, Asian/Pacific Islander, American Indian/Alaska Native, Latino (regardless of race), or "other".[38] The "other" race category included non-Latino participants who: 1) identified more than one race, 2) did not answer the question on race, or 3) did not know their race, ethnicity, or origin.

We then calculated the sensitivity, specificity, positive predictive value (PPV), and percent agreement for race/ethnicity collected using RCA with and without the use of a Spanish surname list[15], using self-reported information from CanCORS as the reference or gold standard. We also calculated the sensitivity, specificity, positive predictive value (PPV), and percent agreement for race, ethnicity, and national origin collected by cancer registries from routine reporting, again using self-reported information from CanCORS interviews as the reference or gold standard. In addition, we identified factors associated with discordance using unconditional logistic regression to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for each of the following predictor variables, controlling for the other variables in the list: interview type (self or surrogate), age, sex, education, race/ethnicity, English language use at home, cancer site, cancer registry region, and marital status. In this study, participants who spoke only English at home were assumed to be more acculturated than those who spoke a different language at home. Within categories of race/ethnicity assigned using routine registry reporting, patients who should have been included in a racial/ethnic category (based on their self-report) but were excluded from that category in the registry are defined as under-reported. For example, patients who stated in the interview that they were Latino but were not identified via routine reporting as Latino would be classified as having their ethnicity under-reported. In contrast, within categories of self-reported race/ethnicity, patients who were included in a racial/ethnic category in the registry, but should have been excluded (based on their self-report), are defined as over-reported. For example, patients who identified themselves as non-Latino, but were categorized as Latinos in the registry, would be classified as over-reported among registry-reported Latinos.

We employed logistic regression to identify predictors of: 1) under-reporting in the registry of self-reported race/ethnicity, and 2) over-reporting of a race/ethnicity in the registry different from a patient's self-reported race/ethnicity. Statistical analyses were conducted using SAS version 9.0 statistical software.[14]

RESULTS

The distributions of self-reported sociodemographic characteristics and registry-recorded type of cancer of the 3,954 interviewed participants are listed in Table 1. Approximately half of the study participants were 70 years of age or over, male, diagnosed with lung cancer, had more than a high school education, and were married. Almost 65% exclusively spoke English at home and over 70% answered the questionnaire themselves. Most participants reported White race/ethnicity (58%), followed by Latino, Asian/PI, Black and American Indian/Alaska Native, respectively, while 5% did not report race, ethnicity, or origin or were coded as "other" race. The largest proportion of Asian/Pacific Islander participants were of Chinese origin (38%) and most self-reported Latinos were of Mexican origin (62%).

Using self-reported race/ethnicity as the reference or gold standard, RCA supplemented with a Spanish surname list increased sensitivity of recorded Latino ethnicity from 37% to 83% and decreased specificity from 98% to 95% (Table 2). Agreement increased from 89% to 94% for Latinos and from 78% to 82% for Whites. For "other" race, specificity increased from 86% to 88%. In contrast, using the Spanish surname list did not appear to change the sensitivity and specificity of the identification of Blacks or Asian/Pacific Islander. American Indian/Alaska Native race was not included in the analysis due to small sample size.

Measures of agreement for race, ethnicity, and national origin data collected using routine reporting compared to self-reported information indicate that specificity in each racial/ethnic group ranged from 86% to 100%, while sensitivity varied from 95% for Whites, Blacks, and Asians to 86% for Latinos and 7% for American Indian/Alaska Native (Table 3). Among subgroups of Asian/Pacific Islander national origin, sensitivity ranged from 84% for Chinese and Filipinos to 33% for "other" Asians. Among subgroups of Latin national origin, sensitivity ranged from 76% for Cubans to 18% for Central and South Americans.

Self-reported and routinely-recorded race/ethnicity was discordant in 11% of the 3,954 participants (Table 4). Self-reported Latinos, Asian/Pacific Islanders, and women were more likely to have their race/ethnicity misclassified in cancer registry records. Overall, participants who spoke a language other than English were less likely to be misclassified. We adjusted for geographic location in the models, but there were no meaningful differences. A separate model (not shown) indicated significant interaction between race/ethnicity and language, therefore the analysis was stratified by race/ethnicity.

Odds ratios for the under- and over-reporting of race/ethnicity in the registry are listed in Table 5. Of the four racial/ethnic groups, under-reporting was most common among self-identified Latinos (14%). Under-reporting among self-identified non-Hispanic Whites was more likely among females, adults with less than a high-school education, and speakers of a language other than English. In contrast, among self-identified Latinos, under-reporting was less likely if a surrogate completed the interview. Females and English-speakers were more likely to be under-reported among self-identified Asian/PIs. Among self-identified non-Latinos, 16% were misclassified or over-reported as Latino in the registry; this proportion was 10% for self-reported non-Whites, non-Blacks, and non-Asian/PIs. Misclassification as White among non-Whites was more likely among younger adults, females, adults with less

than a high-school education, and non-English speakers while self-reported non-Blacks who were non-English speakers were more likely to be misclassified as Black. Among self-identified non-Latinos, misclassification was more likely to occur among older adults, females, and English speakers. Again we adjusted for geographic location in the models, but there were no meaningful differences.

DISCUSSION

This study found that using a Spanish surname list with RCA enhanced the identification of eligible Latino patients for health studies. Furthermore, agreement of race/ethnicity between self-report and cancer registry records was generally good, but was quite low for some Asian/Pacific Islander and Latino origins and, especially, for American Indian/Alaska Native race. In addition, women were more likely to have their self-reported race/ethnicity misclassified in registry data and the impact of English language use on misclassification differed by group.

An increased sensitivity of RCA when using a Spanish surname list is consistent with similar studies.[2,5,6] However, our observed sensitivity of 86% for Latino ethnicity was higher than previously reported.[5,6] This is likely because this study uses data from a more recent time period during which the health community has placed more emphasis on culturally and linguistically appropriate care.

Regarding cancer registries, misclassification varied by race, ethnicity, and origin. The results for Asian origins from our study differed from previous research by Gomez and Glaser[5], likely reflecting small sample sizes for some groups, changes in the collection of race/ethnicity information over time, and regional variation in study populations.

Furthermore, acculturated self-reported Asian/Pacific Islanders were more likely to be classified as a different race/ethnicity in the cancer registry. As a result, cancers common among less acculturated Asian/Pacific Islanders may appear higher than they actually are, while rates of cancers common among acculturated Asian/Pacific Islanders may appear lower.

Misclassification of ethnicity among Latinos has been previously highlighted and may reflect uncertainty regarding the difference between race and ethnicity.[2,39] In this case, self-reported Latinos were more likely reported as a different race/ethnicity if they were older or had more education. In contrast, Pinheiro found that younger Latinos were more likely to be misclassified as non-Latino.[23] These differences perhaps exist because the predominant Latino population in Florida and patterns of migration differs drastically from those found in California. Similar to our results, others have found that Latinos of higher socio-economic status, which includes education, are more likely to be misclassified as non-Latino.[40] Both age and education may be capturing nuances missed when English language alone as used as the proxy for acculturation. It is possible that older and more educated Latinos are more acculturated and, therefore, their Latino ethnicity is less obvious to health providers. Interestingly, we found that non-Latinos with higher levels of education were not more likely misclassified as Latino, indicating that participants with higher levels of education

were more likely to be excluded from Latino cancer statistics. As a result, cancers associated with higher SES, such as breast cancer, may appear artificially low for Latinos.

We have identified two possible explanations for the misclassification of self-reported Asian/Pacific Islanders in the registry, most of who were classified as White. Participants of mixed White and Asian/Pacific Islander racial backgrounds may be more likely to self-identify as Asian/Pacific Islander than as White. It may also be possible that when an interpreter is not required, hospital staff may be inclined to identify non-Black English-speakers as White.

This study is the first to comprehensively validate Mexican, Puerto Rican, Cuban, Central America, South American, and "other" Latino origins in cancer registry data. This study examined only two types of cancer and only included residents from two regional registries in California and thus may not be generalizable to other cancer types or cancer patients in other parts of California or other states. However, lung and colorectal cancers make up over 20% of all cancer cases in California and the regions in this study include approximately 45% of all cancer patients in the state.[41] Although we could not link 5% of study participants identified through RCA with routinely recorded registry data, we expect a minimal impact on the results of the study.[5]

In considering enhanced race/ethnicity data collection methods, costs need to be considered in relation to the magnitude of gains in observed accuracy. The low cost of using the Surname List in relation to its observed gains in accuracy is sizeable. Alternative strategies for improving the accuracy of race/ethnicity rely upon implementing new systems to facilitate self- report of race/ethnicity at the point of care. Gains from this latter approach are likely to be widespread, and notably are expected to benefit all racial and ethnic groups. However, costs associated with this transition are substantial, especially in today's system when most health care settings do not use electronic health records that allow patient to self report. With time, as health care systems become more technologically advanced and interoperable, it is likely that the cost-benefit ratio for supporting self-report of race/ethnicity will become favorable. However, for now, the use of Surname Lists remains a low cost strategy for enhancing gains in accuracy.

These findings provide valuable information for investigators using RCA and registry data to study cancer-related disparities. It also underscores the Institute of Medicine's case regarding the importance of standardizing the collection of race and ethnicity data for improving the quality of care and reducing health disparities for a variety of heath conditions.[22] Our findings demonstrate the need for more consistent policies and approaches for hospitals and other health care providers to collect self-reported data on race, ethnicity, and origin, so these data can be used more effectively to identify and address racial and ethnic disparities in the quality of cancer care and cancer outcomes.

References

1. Pearson ML, Ganz PA, McGuigan K, et al. The case identification challenge in measuring quality of cancer care. J Clin Oncol. Nov 1; 2002 20(21):4353–4360. [PubMed: 12409335]

2. Stewart SL, Swallen KC, Glaser SL, et al. Comparison of methods for classifying Hispanic ethnicity in a population-based cancer registry. Am J Epidemiol. Jun 1; 1999 149(11):1063–1071. [PubMed: 10355383]

3. Gomez SL, Le GM, West DW, et al. Hospital policy and practice regarding the collection of data on race, ethnicity, and birthplace. Am J Public Health. Oct; 2003 93(10):1685–1688. [PubMed: 14534222]

4. Espey DK, Wiggins CL, Jim MA, et al. Methods for improving cancer surveillance data in American Indian and Alaska Native populations. Cancer. Sep 1; 2008 113(5 Suppl):1120–1130. [PubMed: 18720372]

5. Gomez SL, Glaser SL. Misclassification of race/ethnicity in a population-based cancer registry (United States). Cancer Causes Control. Aug; 2006 17(6):771–781. [PubMed: 16783605]

6. Clegg LX, Reichman ME, Hankey BF, et al. Quality of race, Hispanic ethnicity, and immigrant status in population-based cancer registry data: implications for health disparity studies. Cancer Causes Control. 2007; 18(2):177. [PubMed: 17219013]

7. West CN, Geiger AM, Greene SM, et al. Race and ethnicity: comparing medical records to self-reports. J Natl Cancer Inst Monogr. 2005(35):72–74. [PubMed: 16287889]

8. Sweeney C, Edwards SL, Baumgartner KB, et al. Recruiting Hispanic women for a population-based study: validity of surname search and characteristics of nonparticipants. Am J Epidemiol. 2007; 166(10):1210. [PubMed: 17827445]

9. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. Annu Rev Public Health. 2006; 27:1–28. [PubMed: 16533107]

10. Ashing-Giwa KT, Padilla GV, Tejero JS, et al. Breast cancer survivorship in a multiethnic sample: challenges in recruitment and measurement. Cancer. Aug 1; 2004 101(3):450–465. [PubMed: 15274058]

11. Des Jarlais G, Kaplan CP, Haas JS, et al. Factors affecting participation in a breast cancer risk reduction telephone survey among women from four racial/ethnic groups. Prev Med. Sep-Oct; 2005 41(3–4):720–727. [PubMed: 15936066]

12. Rowland ML, Forthofer RN. Investigation of nonresponse bias: Hispanic Health and Nutrition Examination Survey. Vital Health Stat 2. Dec.1993 (119):1–75. [PubMed: 8135023]

13. Vernon SW, Roberts RE, Lee ES. Ethnic status and participation in longitudinal health surveys. Am J Epidemiol. Jan; 1984 119(1):99–113. [PubMed: 6691340]

14. Hoopes M, Petersen P, Vinson E, et al. Regional differences and tribal use of American Indian/Alaska Native cancer data in the Pacific Northwest. J Cancer Educ. Apr; 2012 27(Suppl 1):S73–S79. [PubMed: 22281722]

15. Word, D.; Perkins, R, Jr. Building a Spanish Surname List for the 1990's – A New Approach to an Old Problem. Population Division, U. S. Bureau of the Census; Washington D.C: 1996.

16. NAACCR Latino Research Work Group. , editor. NAACCR Guideline for Enhancing Hispanic-Latino Identification: Revised NAACCR Hispanic/Latino Identification Algorithm [NHIA v2]. Springfield (IL): North American Association of Central Cancer Registries; Sep. 2005

17. Stewart SL, Swallen KC, Glaser SL, et al. Adjustment of cancer incidence rates for ethnic misclassification. Biometrics. 1998; 54(2):774–781. [PubMed: 9629656]

18. Swallen KC, West DW, Stewart SL, et al. Predictors of misclassification of Hispanic ethnicity in a population-based cancer registry. Ann Epidemiol. 1997; 7(3):200. [PubMed: 9141643]

19. Lauderdale DS, Huo D. Cancer death rates for older Asian-Americans: classification by race versus ethnicity. Cancer Causes Control. Mar; 2008 19(2):135–146. [PubMed: 17992577]

20. McCracken M, Olsen M, Chen MS Jr, et al. Cancer incidence, mortality, and associated risk factors among Asian Americans of Chinese, Filipino, Vietnamese, Korean, and Japanese ethnicities. CA Cancer J Clin. Jul-Aug;2007 57(4):190–205. [PubMed: 17626117]

21. Kwong SL, Chen MS Jr, Snipes KP, et al. Asian subgroups and cancer incidence and mortality rates in California. Cancer. Dec 15; 2005 104(12 Suppl):2975–2981. [PubMed: 16247792]

22. Ulmer, C.; McFadden, B.; Nerenz, DR. Race Ethnicity and Language Data: Standardization for Health Care Quality Improvement. Washington, DC: National Academies Press; 2009.

23. Pinheiro PS, Sherman R, Fleming LE, et al. Validation of ethnicity in cancer data: which Hispanics are we misclassifying? J Registry Manag. Summer;2009 36(2):42–46. [PubMed: 19694116]

24. Berry, J. Acculturation and mental health. In: Dasen, PR.; Berry, J.; Sartorius, N., editors. Health and Cross-Cultural Psychology. London: Sage Publications; 1988.

25. Cabassa LJ. Am J Public Health. 2003 Jul.93(7):1034. author reply 1034–5.Integrating cross-cultural psychiatry into the study of mental health disparities. [PubMed: 12835167]

26. Gorin SS, Heck JE. Cancer screening among Latino subgroups in the United States. Prev Med. May; 2005 40(5):515–526. [PubMed: 15749133]

27. Kudadjie-Gyamfi E, Consedine NS, Magai C. On the importance of being ethnic: coping with the threat of prostate cancer in relation to prostate cancer screening. Cultur Divers Ethnic Minor Psychol. Jul; 2006 12(3):509–526. [PubMed: 16881753]

28. Goodman MJ, Ogdie A, Kanamori MJ, et al. Barriers and facilitators of colorectal cancer screening among Mid-Atlantic Latinos: focus group findings. Ethn Dis. Winter;2006 16(1):255–261. [PubMed: 16599380]

29. Lilienfeld, AM.; Levin, ML.; Kessler, II. Cancer in the United States. Cambridge, Mass: Harvard University Press; 1972.

30. Polednak AP. Identifying newly diagnosed Hispanic cancer patients who use a physician with a Spanish-language practice, for studies of quality of cancer treatment. Cancer Detect Prev. 2007; 31(3):185–190. [PubMed: 17706369]

31. Swan J, Breen N, Coates RJ, et al. Progress in cancer screening practices in the United States: results from the 2000 National Health Interview Survey. Cancer. Mar 15; 2003 97(6):1528–1540. [PubMed: 12627518]

32. Ayanian JZ, Chrischilles EA, Fletcher RH, et al. Understanding cancer treatment and outcomes: the Cancer Care Outcomes Research and Surveillance Consortium. J Clin Oncol. Aug 1; 2004 22(15):2992–2996. [PubMed: 15284250]

33. Malin JL, Ko C, Ayanian JZ, et al. Understanding cancer patients' experience and outcomes: development and pilot study of the Cancer Care Outcomes Research and Surveillance patient survey. Support Care Cancer. 2006; 14(8):837–848. [PubMed: 16482448]

34. Ayanian JZ, Z A, Arora NK, Kahn KL, Malin JL, Ganz PA, Van Ryn M, Hornbrook MC, Kiefe CI, He Y, Urmie J, Weeks JC, Harrington DP. Patients' experiences with care for lung cancer and colorectal cancer: findings from the cancer care outcomes research and surveillance consortium. J Clin Oncol. Sep 20; 2010 28(27):4154–4161. [PubMed: 20713876]

35. U.S. Census Bureau. What is Race?. 2012. http://www.census.gov/population/race/ Accessed May 7 2013

36. U.S. Department of Commerce. United States Census Bureau People and Households Hispanic Origin. http://www.census.gov/population/hispanic/ Accessed May 1 2013

37. Nielsen S, He Y, Ayanian J, et al. Quality of cancer care among foreign-born and US-born patients with lung or colorectal cancer. Cancer. 2010; 116:5497–5506. [PubMed: 20672356]

38. U.S. Department of Health and Human Services. Appendix D – Race and Nationality Descriptions from the 2000 Census and Bureau of Vital Statistics. SEER Coding And Staging Manual. 2004

39. Grieco, E.; Cassidy, R. Overview of Race and Hispanic Origin Census 2000 Brief. U.S. Census Bureau; 2001.

40. Hazuda HP, Comeaux PJ, Stern MP, et al. A comparison of three indicators for identifying Mexican Americans in epidemiologic research. Methodological findings from the San Antonio Heart Study. Am J Epidemiol. Jan; 1986 123(1):96–112. [PubMed: 3940446]

41. Kwong, SL.; Allen, M.; Wright, WE. Cancer in California. Sacramento, CA: California Department of Health Services, Cancer Surveillance Section; p. 1988-2002.2005/12/01/2008

Table 1

Self-Reported Baseline Characteristics of the CanCORS Participants in California, 2003–2005 (N=3954).

	No.	%
Age		
21 – 54	708	17.9
55 – 69	1,359	34.4
70	1,887	47.7
Sex		
Male	2,054	52.0
Female	1,900	48.0
Cancer Site		
Lung	1,995	50.5
Colorectal	1,959	49.5
Education		
< High School	717	18.1
High School degree/GED	1,030	26.0
> High School	2,072	52.4
Unknown	135	3.4
Marital status		
Married or living with a partner	2,118	53.6
Divorced, widowed, separated	1,193	30.2
Never married	216	5.5
Unknown	427	10.7
Language [a]		
English	2,534	64.1
Other	1,036	26.2
Unknown	384	9.7
Surrogate		
No	2,843	71.9
Yes	1,111	28.1
Race/ethnicity		
White	2,272	57.5
Black	381	9.6
Latino	611	15.5
Asian/PI	469	11.9
AI/AN	27	0.7
Multiple Race/"Other"/Unknown	194	4.9
Latino National Origin (n=611)		
Mexican	376	61.5
Puerto Rican	15	2.5
Cuban	17	2.8

	No.	%
Central or South American	73	11.9
"Other"	130	21.3
Asian National Origin (n=469)		
Chinese	180	38.4
Japanese	54	11.5
Filipino	100	21.0
Vietnamese	12	0.3
"Other"	123	26.2

Abbreviations: Asian/PI, Asian/Pacific Islander; AI/AN, American Indian/Alaskan Native; CSA, Central and South American; GED, general educational development.

[a]Participants who only spoke English at home were classified as speaking English. Participants who spoke any other language at home or answered their questionnaire in Spanish or Cantonese were classified as speaking a language other than English.

Table 2

Accuracy of Race/ethnicity Ascertained From RCA, Using Medical Records Alone or Supplemented by a Spanish Surname list [15].

Race/Ethnicity	Classification Method	Percent Agreement	95% CI	PPV	95% CI	Sensitivity	95% CI	Specificity	95% CI
White	RCA	78	77, 79	81	79, 83	81	79, 83	74	72, 76
	RCA + surname	82	81, 83	89	88, 90	79	77, 81	87	86, 88
Black	RCA	97	97, 98	89	86, 92	83	82, 84	99	98, 100
	RCA + surname	97	97, 98	90	87, 93	83	82, 84	99	98, 100
Latino	RCA	89	88, 90	79	74, 84	37	35, 39	98	97, 99
	RCA + surname	94	93, 95	77	74, 80	83	82, 84	95	93, 97
Asian/PI	RCA	96	95, 97	80	77, 86	84	82, 84	97	96, 99
	RCA + surname	96	95, 97	84	81, 88	78	83, 85	98	97, 99
"Other"	RCA	82	81, 83	4	2, 6	12	11, 13	86	81, 90
	RCA + surname	84	83, 85	4	2, 6	10	9, 11	88	83, 93

Abbreviations: RCA, rapid case ascertainment; PPV, positive predictive value; Asian/PI, Asian/Pacific Islander.

Note: Self-reported race, ethnicity, origin as reference. Data on AI/AN not collected in RCA and, therefore, could not be presented.

Table 3

Validation of Race, Ethnicity, and Origin Data in Cancer Registries for CanCORS Participants in California, 2003–2005.

Cancer Registry Data	Self-report														
	Race/ethnicity					Latino Origins					Asian/PI Origins				
	White	Black	Latino	AI/AN	Asian/PI	Mexican	Puerto Rican	Cuban	CSA	Other	Chinese	Japanese	Filipino	Vietnamese	Other
Total	2272	381	611	27	469	376	15	17	73	93	180	54	100	12	111
Race/ethnicity															
White	2170	14	61	16	19	31	3	1	3	18	2	5	5	0	7
Black	10	366	8	3	2	0	2	0	1	1	1	1	0	0	0
Latino	70	0	528	6	2	343	10	16	68	65	0	0	2	0	0
AI/AN	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
Asian/PI	9	1	12	0	446	2	0	0	0	8	177	48	93	12	104
Latino origins															
Mexican	3	0	169	0	0	153	1	0	2	9	0	0	0	0	0
Puerto Rican	2	0	8	0	0	2	6	0	0	0	0	0	0	0	0
Cuban	0	0	14	4	0	0	0	13	0	0	0	0	0	0	0
CSA	0	0	62	2	0	1	0	0	40	16	0	0	0	0	0
Other	5	0	275	0	2	187	3	3	26	40	0	0	5	0	0
Asian/PI origins															
Chinese	1	0	0	0	188	0	0	0	0	0	151	1	0	3	30
Japanese	0	0	0	0	50	0	0	0	0	0	0	42	0	0	8
Filipino	6	1	9	0	115	2	0	0	0	3	0	0	86	0	29
Vietnamese	1	0	0	0	20	0	0	0	0	0	6	0	0	7	0
Other	1	0	3	0	73	0	0	0	0	2	20	5	7	2	37
Validity measures															
Sensitivity[a]	96	96	86	7	95	41	40	76	18	42	84	77	85	58	33
Specificity[a]	86	99	97	100	98	99	100	100	99	92	99	100	99	100	99

	Self-report														
	Race/ethnicity					Latino Origins					Asian/PI Origins				
Cancer Registry Data	White	Black	Latino	AI/AN	Asian/PI	Mexican	Puerto Rican	Cuban	CSA	Other	Chinese	Japanese	Filipino	Vietnamese	Other
PPV	90	90	84	100	89	87	67	93	60	11	78	81	60	32	45

Abbreviations: CSA, Central and South American; AI/AN, American Indian/Alaskan Native; Asian/PI, Asian/Pacific Islander; % agreement, percent agreement; PPV, positive predictive value.

[a] Self-reported race, ethnicity, origin as reference.

Note: Confidence intervals and "Other" race not included due to space limitations.

Table 4

Odds Ratios and 95% Confidence Intervals of Discordance Between Race/ethnicity Recorded in Cancer Registries and Self-report from CanCORS Participants[a] in California, 2003–2005

	Discordance[b]	
	OR[c]	95% CI
Interview type		
Self	Ref	
Surrogate	0.90	0.62, 1.28
Age		
21 – 54	Ref	
55 – 69	1.11	0.76, 1.63
70	0.70	0.46, 1.05
Sex		
Male	Ref	
Female	2.05	1.51, 2.78
Education		
< High school degree	Ref	
High school degree/GED	1.18	0.75, 1.86
> High school degree	1.39	0.91, 2.14
Language[d]		
English	Ref	
Other	0.45	0.29, 0.70
Self-reported Race/ethnicity [e]		
White	Ref	
Black	0.83	0.45, 1.52
Latino	6.33	4.16, 9.63
Asian/PI	2.12	1.21, 3.74

Abbreviations: OR, odds ratio; CI, confidence interval; GED, general educational development; Asian/PI, Asian/Pacific Islander; AI/AN, American Indian/Alaskan Native.

[a]All race/ethnicities combined

[b]Discordance between race/ethnicity in routinely recorded cancer registry data and self-report.

[c]Multivariable analysis with mutual adjustment of all listed variables. Registry region was controlled for in the analysis and results not shown.

[d]Participants who only spoke English at home were classified as speaking English. Participants who spoke any other language at home or answered their questionnaire in Spanish or Cantonese were classified as speaking a language other than English.

[e]Odd rations for AI/AN and multiple race/"Other"/unknown not presented due to unstable estimates.

Table 5

Odds Ratios and 95% Confidence Intervals of Under-reporting and Over-reporting Race/ethnicity in Cancer Registries for CanCORS Participants in California, 2003–2005.

	UNDER-REPORTING RACE/ETHNICITY		OVER-REPORTING RACE/ETHNICITY	
	I. WHITE			
	Self-report White, Routine Reporting Misclassified as Non-White		Self-report Non-White, Routine Reporting Misclassified as White	
	(93 of 1,930 participants (5%))		(210 of 2,140 participants (10%))	
	OR	**95% CI**	**OR**	**95% CI**
Interview type				
Self	Ref		Ref	
Surrogate	0.90	0.55, 1.49	0.82	0.58, 1.17
Age				
21– 54	Ref		Ref	
55 – 69	0.92	0.52, 1.63	0.54	0.36, 0.81
70	0.35	0.19, 0.64	0.44	0.30, 0.64
Sex				
Male	Ref		Ref	
Female	2.03	1.30, 3.19	1.31	0.97, 1.78
Education				
> High school	Ref		Ref	
High school	1.33	0.80, 2.19	1.06	0.75, 1.51
degree/GED	2.63	1.46, 4.74	2.07	1.36, 3.15
< High school				
Language[a]				
English	Ref		Ref	
Other	4.03	2.30, 7.06	3.23	2.20, 4.75
	II. BLACK			
	Self-report Black, Routine Reporting Misclassified as Non-Black		Self-report Non-Black, Routine Reporting Misclassified as Black	
	(13 of 329 participants (4%))		(36 of 352 participants (10%))	
	OR	**95% CI**	**OR**	**95% CI**
Interview type				
Self	Ref		Ref	
Surrogate	2.36	0.71, 7.86	0.76	0.32, 1.82
Age				

	UNDER-REPORTING RACE/ETHNICITY		OVER-REPORTING RACE/ETHNICITY	
	I. WHITE			
	Self-report White, Routine Reporting Misclassified as Non-White		**Self-report Non-White, Routine Reporting Misclassified as White**	
	(93 of 1,930 participants (5%))		**(210 of 2,140 participants (10%))**	
	OR	**95% CI**	**OR**	**95% CI**
21– 54	Ref		Ref	
55 – 69	0.76	0.18, 3.29	1.00	0.39, 2.58
70	0.69	0.14, 3.50	1.09	0.40, 2.97
Sex				
Male	Ref		Ref	
Female	0.61	0.19, 2.02	1.30	0.62, 2.73
Education				
> High school	Ref		Ref	
High school	1.20	0.35, 4.14	1.26	0.55, 2.87
degree/GED	0.24	0.03, 2.18	0.65	0.24, 1.77
< High school				
Language[a]				
English	Ref		Ref	
Other	2.64	0.27, 26.32	4.98	1.53, 16.22
	UNDER-REPORTING RACE/ETHNICITY		OVER-REPORTING RACE/ETHNICITY	
	III. LATINO			
	Self-report Latino, Routine Reporting Misclassified as Non-Latino		**Self-report Non-Latino, Routine Reporting Misclassified as Latino**	
	(75 of 529 participants (14%))		**(87 of 541 participants (16%))**	
	OR	**95% CI**	**OR**	**95% CI**
Interview type				
Self	Ref		Ref	
Surrogate	0.34	0.16, 0.76	0.63	0.32, 1.22
Age				
21– 54	Ref		Ref	
55 – 69	1.87	0.91, 3.88	2.06	1.05, 4.06
70	2.80	1.30, 6.04	1.86	0.88, 3.95
Sex				
Male	Ref		Ref	
Female	3.37	1.83, 6.19	1.86	1.07, 3.24

	UNDER-REPORTING RACE/ETHNICITY		OVER-REPORTING RACE/ETHNICITY	
	I. WHITE			
	Self-report White, Routine Reporting Misclassified as Non-White		**Self-report Non-White, Routine Reporting Misclassified as White**	
	(93 of 1,930 participants (5%))		**(210 of 2,140 participants (10%))**	
	OR	**95% CI**	**OR**	**95% CI**
Education				
> High school	Ref		Ref	
High school	0.36	0.18, 0.73	0.92	0.48, 1.75
degree/GED	0.24	0.12, 0.49	0.49	0.24, 1.01
< High school				
Language[a]				
English	Ref		Ref	
Other	0.22	0.80, 2.47	0.11	0.06, 0.19
	IV. ASIAN/PI			
	Self-report Asian/PI, Routine Reporting Misclassified as Non- Asian/PI		**Self-report Non- Asian/PI, Routine Reporting Misclassified as Asian/PI**	
	(22 of 379 participants (6%))		**(44 of 423 participants (10%))**	
	OR	**95% CI**	**OR**	**95% CI**
Interview type				
Self	Ref		Ref	
Surrogate	0.95	0.31, 2.89	0.91	0.45, 1.86
Age				
21– 54	Ref		Ref	
55 – 69	0.93	0.31, 2.80	2.49	0.99, 6.27
70	0.56	0.16, 1.98	1.98	0.75, 5.24
Sex				
Male	Ref		Ref	
Female	3.29	1.26, 8.61	0.94	0.49, 1.82
Education				
> High school	Ref		Ref	
High school	0.67	0.18, 2.50	0.64	0.27, 1.54
degree/GED	1.17	0.29, 4.84	0.67	0.25, 1.82
< High school				
Language[a]				
English	Ref		Ref	
Other	0.15	0.06, 0.38	0.72	0.32, 1.61