IDENTIFYING AND OVERSAMPLING HISPANICS BY THE PASSEL-WORD SURNAME LIST FOR ENROLLMENT IN A WEB-BASED NUTRITIONAL INTERVENTION

Nikki M. Carroll, MS; Debra P. Ritzwoller, PhD; Melanie A. Stopponi, MPA; Christine Cole Johnson, PhD

Objective: To describe the enrollment rates and characteristics of Hispanics and non-Hispanics from Kaiser Permanente Colorado invited to participate in a web-based intervention promoting increased fruit and vegetable consumption.

Design: Hispanics were identified by the Passel-Word Spanish surname list. Characteristics associated with the likelihood of enrollment overall and by ethnicity were examined by logistic regression.

Results: A total of 174 (6.1%) probable Hispanics and 340 probable non-Hispanics (11.8%) enrolled. Hispanics were 48% less likely to enroll than non-Hispanics, females were almost four times as likely to enroll as males, and those living in a census tract associated with higher income levels were 41% more likely to enroll than other income groups. Among Hispanics, females were 87% more likely to enroll than males and those living in a census tract associated with higher income levels were 62% more likely to enroll than other income groups. Among non-Hispanics, the odds for enrolling increased 14% for each decade increase of age, females were 43% more likely to enroll than males and those living in a census tract associated with higher income levels were 68% more likely to enroll than those in other income groups.

Conclusion: Identifying Hispanics through surname for oversampling can be successful in terms of sampling yield and accuracy. However, our results suggest that Hispanics are less likely to enroll in a web-based nutritional intervention. Additional research is needed to identify methods of attracting more Hispanic subjects to these kinds of interventions. (*Ethn Dis*. 2010;20:15–21)

Key Words: Hispanics, Logistic Regression, Internet, Ethnicity, Passel-Word Surname List, Enrollment, Nutrition

From Kaiser Permanente Colorado, Institute for Health Research (NMC, DPR, MAS), Henry Ford Health Systems, Biostatistics and Research Epidemiology (CCJ).

Address correspondence and reprint requests to Nikki Carroll, MS; Institute for Health Research; Kaiser Permanente, PO Box 378066; Denver, CO 80237-8066; 303-614-1251; 303-614-1225 (fax); Nikki.M.Carroll@kp.org

INTRODUCTION

Recruitment of minorities into clinical research studies can be challenging. In particular, Hispanics are often underrepresented.[1–4] The Hispanic population in the United States has increased dramatically in recent years and, in 2000, persons of Hispanic origin were the largest and fastest growing minority group, comprising nearly 13% of the US population.[5–7] With a growth of almost 40%, from 9% to 13%, sufficient sample size of Hispanics in research studies is crucial due to the unique culture, beliefs, and behaviors of the population and the impact these may have on answers to research questions.[8–9] Not only are Hispanics disproportionately affected by overweight and obesity, but Hispanic women may be at highest risk having close to 50% obesity rate.[10] Therefore, including adequate numbers of Hispanics in studies enhances our abilities to apply research findings to a more diverse population.

Not only are Hispanics disproportionately affected by overweight and obesity, but Hispanic women may be at highest risk having close to 50% obesity rate.[10]

Literature suggests a wide range of enrollment rates from ethnic minorities into clinical research studies. Across all National Cancer Institute cancer treatment trials, Hispanics represented, on average, 5.6% of the total sample[3] while participation in a breast cancer case-control study targeting female Hispanics obtained a 34% participation rate.[11]

Literature also suggests minority recruitment is less successful when the study design does not specifically plan to augment such recruitment[12–14] and recruitment rates can be affected by recruitment methods. The Internet is a growing medium for registry into clinical trials databases.[15] Despite Hispanic internet use reaching 50 percent,[16] these users frequently do not represent research populations that are targeted. These users tend to be a select sample of well-educated, literate, and articulate people who are skilled users of computers and have access to the Internet.[17] In addition, researchers have reported that Hispanic populations are unaware of the possible benefits of clinical research participation[2] which makes effective recruitment strategies more important.

Colorado presents an excellent opportunity for studying ethnic disparities in that the Hispanic population accounts for over 17% of the total population.[5] Member surveys coupled with census data suggest that approximately 16% of the total membership at Kaiser Permanente Colorado (KPCO) are Hispanic. In single or multi-site studies recruiting from this population, it is often advantageous to oversample in such a way so as to achieve a larger Hispanic representation. Oversampling minorities using surname algorithms can increase the absolute numbers enrolled[18] and enhance the ability to discern the characteristics of individuals who respond to a trial.

Beginning in 1950, the US Census Bureau produced and released a decadal Spanish surname list. The basis for

including a specific surname on that list was the similarity of that name's geographic distribution to the Hispanic origin population within the United States.[19] The 1970 census was the first that allowed the opportunity for people to self-identify as a person of Spanish origin, which was later changed to Hispanic. The Passel-Word Spanish surname list is a previously validated surname program for identifying Hispanic subpopulations in the southwest.[11,20–22] Published sensitivities associated with the Passel-Word Spanish surname list range from 82–95% for males, and 67–82% for females.[11,20–22]

This article describes the enrollment rates and individual characteristics of Hispanics and non-Hispanics from KPCO who were invited to participate in a web-based nutritional intervention that promoted increased intake of fruits and vegetables.

METHODS

Setting

Data for this analysis are based on data collected as part of Making Effective Nutritional Choices (MENU). The MENU study was one of the core projects associated with the National Cancer Institute funded HMO Cancer Research Network (CRN) (http://crn.cancer.gov/). The CRN consists of the research programs, enrollee populations, and databases of 12 health care organizations in the HMO Research Network at the time of the study.[23] The MENU study was conducted in five CRN sites including: Henry Ford Health System/Health Alliance Plan in Michigan; HealthPartners Research Foundation in Minnesota; Kaiser Permanente Colorado; Kaiser Permanente Georgia; and Group Health Cooperative in Washington state. The University of Michigan's Center for Health Communications Research worked with MENU investigators to develop and maintain the online intervention programs.

Although MENU was implemented across five sites, and several sites oversampled for minorities, only KPCO oversampled for Hispanics. Therefore, only members from KPCO were included in the analysis for this paper. KPCO is a group model health plan that provides health care for over 400,000 members in the Denver-Boulder-Longmont, Colorado metropolitan area. This study was reviewed and approved by KPCO's Institutional Review Board.

The purpose of the MENU randomized clinical trial was to develop and evaluate an individually tailored, web-based program to promote daily fruit and vegetable consumption. In brief, intervention content was based on principles from Social Cognitive Theory, the Transtheoretical Model, and the Health Belief Model.[24–26] Constructs were incorporated into tailoring the web site including: motivation to change, specific motives for changing (eg, health improvement, weight loss, role modeling), barriers to changing (eg, expense of produce, dislike taste, lack of preparation skills), and cues to action (eg, listing pros and cons of consuming more fruits and vegetables, keeping produce visible, providing recipes). The intervention phase was designed to last four months from the point of enrollment and participants continued to have website access through the end of the study.

Further details associated with the MENU recruitment, intervention, and outcomes are well described elsewhere.[27,28]

Study Sample

Using administrative databases, we identified a pool of potential participants comprising individuals aged 21–65 years at the beginning of the study enrollment period (September, 2005) who were current members with at least one year of enrollment with no gaps greater than 90 days at one time. Diagnostic codes were used to exclude from the sample anyone with a medical condition that could be affected negatively by increasing consumption of fruits and vegetables. These conditions included current cancer treatment, gastroparesis, neurological conditions, mental health conditions, and use of anticoagulant medications.

Race/ethnicity/language preference data in KPCO administrative databases were not available at the time of sampling; therefore, we identified persons likely to be Hispanic on the basis of surname, using the Passel-Word Hispanic surname file.[20,29–30] Surname was captured through administrative databases for all members. We coded the surname to prepare for comparison by: removing embedded titles (eg, Jr., Sr., III); removing embedded blanks so that De La Cruz became DELACRUZ; changing all names to uppercase; and, separating multiple last names or hyphenated names into separate last name fields. Each last name field was compared with the Passel-Word list and a probable Hispanic indicator flag was created.

Sampling Scheme

Once member eligibility and probable Hispanic ethnicity were determined, KPCO generated a random sample of approximately 6,000 individuals (invitees). Fifty percent of the sample was weighted as probable Hispanic and stratified equally by sex, with an overall 10% response goal for study enrollment. Specifically, 1,500 eligible members were selected randomly from four distinct populations from which MENU study participants were recruited: 1) male probable non-Hispanics; 2) female probable non-Hispanics; 3) male probable Hispanics; and 4) female probable Hispanics. The sample was weighted such that 25% of the introductory letters went to each subgroup. Further details associated with the sampling strategy are described elsewhere.[27,28]

Study Enrollment

Invitees were mailed an introductory letter that contained an invitation to participate in the study, a $2 enrollment incentive, a description of the study, and MENU's URL and toll-free phone number. The introductory letter also explained that over the course of one year, three online follow-up surveys would be requested of participants, with a promised $20 incentive at the completion of each.

Responders to the invitation letter logged onto the MENU study website and completed an eligibility survey, which included 9–12 questions (depending on personal tailoring) on health plan membership status, age, accessibility to the Internet for personal use, frequency of use of personal email address, and history and treatment of certain conditions. The eligibility survey took approximately 5 minutes to complete and responders had 40 days to start the eligibility survey. Eligible responders were presented with an online informed consent form.

After responders completed the consent process, the website prompted the users to complete the baseline survey. The baseline questionnaire took approximately 25 minutes to complete and responders were given 28 days to begin the baseline survey and 28 days to complete the survey once started.

The baseline survey collected demographic and fruit and vegetable intake information. Demographic factors captured included self-reported age, race/ethnicity, household income and education. Two measures of fruit and vegetable intake were used. The primary measure was the NCI 19-item fruit and vegetable food frequency questionnaire which queries frequency and portion size over the past month.[31] A second shorter assessment, which appeared before the NCI measure, was a two-item food frequency questionnaire measure which included one question each for total servings of fruit and vegetables consumed on a typical day[32] Validity of these scales has been previously reported.[31–32] Further details associated with baseline measures of fruit and vegetable intake are described elsewhere.[27] Participants were considered enrolled after completion of the online baseline survey.

Measures

Residential address was obtained from administrative databases for all invitees. Geocoding techniques were employed to create individual area-based proxies for income and education as markers of socioeconomic status. Geocoding was performed using MapMarker® Plus and 2000 Census data. Each invitee's address was mapped to census tract and the corresponding median household income and proportion of the census tract attaining various education levels were calculated. Indicator variables were created for each invitee with cut points at census area median household income and post high school vs less education attainment levels.

Statistical Analysis

We computed descriptive statistics to examine variation in enrollment by ethnicity, age, sex and census-based measures. Statistical significance of differences was tested using the Chi-square test or Wilcoxon rank sum test. The alpha level for all statistical tests was set at 0.05.

For the full sample, we used logistic regression modeling to examine factors associated with the likelihood of enrollment including ethnicity, age, sex and census-based socioeconomic status proxies. We also estimated separate models for probable Hispanic and probable non-Hispanic invitees. Customary residual and influential statistics were examined to assess model fit and evaluate outliers.

In order to examine potential identification bias associated with the use of Hispanic surname, the probable Hispanic indicator flag was compared to self-reported ethnicity derived from baseline survey data collected at the time of enrollment for those who participated. Self-reported ethnicity was considered the gold standard so that sensitivity and specificity could be computed to assess the accuracy of the probable Hispanic indicator flag. Sensitivity is defined as the percentage of self-identified Hispanics who were classified as Hispanic by the indicator flag; specificity as the percentage of self-identified non-Hispanics who were classified as such. All analyses were performed using SAS Software version 9.1 (SAS Institute, Cary, NC).

RESULTS

Table 1 presents the characteristics and enrollment proportions of all KPCO invitees. Nearly 6,000 (N=5,751) KPCO members were invited to join MENU, with 2,871 (49.9%) members identified as probable Hispanic and 2,880 (50.1%) members identified as probable non-Hispanic. Those identified as probable Hispanic were slightly younger with a mean age of 42 years, compared to probable non-Hispanics with a mean age of 44 years ($P<.0001$). Probable non-Hispanics were more likely to be from households with higher education (76.1% vs 23.9%, $P<.0001$) and income (58.8% vs 41.2%, $P<.001$). As expected due to the study design, there was no statistical difference in sex proportions by Hispanic ethnicity among the invitees.

Fewer of the probable Hispanics, (6.1%, n=174) enrolled compared to 11.8% (n=340) of the probable non-Hispanics ($P<.0001$) for total enrollment of 514 and an overall enrollment rate of 8.9%. Slightly more than one-third (174/514, 33.9%) of the enrolled participants were probable Hispanics. Probable Hispanics who enrolled were younger with a mean age of 42 years, compared to a mean age of 45 years for probable non-Hispanics (P=.0028). Of

Table 1. Characteristics of invitees and enrollees in the MENU study (KPCO)

Characteristic	Total	Probable non-Hispanic	Probable Hispanic
Invitees	5751	2880 (49.9%)	2871 (50.1%)
Mean age (SD)	43 (12)*‡	44 (12)	42 (12)
Female	2876§	1437 (49.9%)	1439 (50.1%)
Median income >$41,994	2611*§	1535 (58.8%)	1076 (41.2%)
College & above education	602*§	458 (76.1%)	144 (23.9%)
Responses			
Enrolled	514*‡	340 (66.1%)	174 (33.9%)
Not enrolled	5237	2540 (48.5%)	2697 (51.2%)
Enrollees	514	340 (66.1%)	174 (33.9%)
Mean age (SD)	44 (11) †‡	45 (11)	42 (12)
Female	308§	196 (63.6%)	112 (36.4%)
Median income >$41,994	287 †§	203 (70.7%)	84 (29.3%)
≥College education	73†§	61 (83.6%)	12 (16.4%)
Self-reported Hispanic	146*§	13 (8.9%)	133 (91.1%)

* $P <$. 0001.
† $P <$0.01.
‡ Wilcoxon Rank Sum Test.
§ Chi-square test.

the probable Hispanics who enrolled, 64.4% were female (n=112) while 57.7% of the enrolled probable non-Hispanics were female (n=196, P=.1412). The percent of enrolled probable Hispanics mapping to census tracts with a median income greater than $41,994 was lower than enrolled probable non-Hispanics (48.3% versus 59.7%, respectively, P=.001). Similarly, fewer enrolled probable Hispanics than enrolled probable non-Hispanics mapped to census tracts where education levels were college or greater (6.9% versus 17.9%, respectively, P=.0031). Because education was correlated with income and with probable Hispanic ethnicity (Spearman rank correlation coefficient [ρ] range for income and probable Hispanic ethnicity, [−0.38 to −0.68]), education was not considered further in the analysis.

Variables associated with the likelihood of enrollment are shown in Table 2. All variables were significant in the univariate models. In the adjusted model and including an age-sex interaction term, three factors were significantly associated with enrollment. Probable Hispanics were less likely to enroll (adjusted odds ratio [OR], 0.52; 95% Confidence Interval [CI]: 0.42, 0.63) while persons from higher income levels were more likely to enroll (adjusted OR, 1.41; 95% CI: 1.17, 1.70). There was a significant interaction between age and sex in that older age was associated with increased likelihood of enrollment for females (adjusted OR, 3.88; 95% CI: 2.66, 5.68), but the opposite was true for males (adjusted OR, 0.77; 95% CI: 0.65, 0.90).

Separating the data by ethnicity (Table 3) gave us the chance to consider those enrolling vs those not enrolling within Hispanic vs non-Hispanic ethnicity. For Hispanics, two factors were associated with enrollment: higher income levels (adjusted OR, 1.62; 95% CI: 1.19, 2.20) and female gender (adjusted OR, 1.88; 95% CI: 1.36, 2.58). For non-Hispanics, three factors were associated with enrollment: age in increments of 10 years (adjusted OR, 1.14; 95% CI: 1.03, 1.26), female sex (adjusted OR, 1.43; 95% CI: 1.14, 1.80), and higher income levels (adjusted OR, 1.32; 95% CI: 1.04, 1.66).

The sensitivity and specificity of the probable Hispanic flag in comparison to self-reported ethnicity was excellent (sensitivity = 91% [133 of 146], specificity = 89% [326 of 367]). When broken down by sex, sensitivity and specificity remained high, although the sensitivity and specificity dropped for females (sensitivity = 85% [78 out of 88], specificity = 89% [186 out of 220]) while sensitivity and specificity for males increased (sensitivity = 95% [55 out of 58], specificity = 95% [140 out of 147]).

DISCUSSION

The oversampling approach we used employing the Passel-Word Spanish surname list was successful in increasing the total number of Hispanics consenting and enrolling in an online behavioral intervention. To our knowledge,

Table 2. Characteristics associated with enrollment

Characteristic	Univariate		Adjusted*	
	Odds Ratio	95% CI	Odds Ratio	95% CI
Ethnicity (Non-Hispanic referent)	0.48	0.40–0.58	0.52	0.42–0.63
Higher census area income (< median referent)	0.76	0.62–0.94	1.41	1.17–1.70
Age effect for males (10 yr increments)	1.03	1.02–1.04	0.77	0.65–0.90
Age effect for females (10 yr increments)	1.00	0.99–1.01	3.88	2.66–5.68

* Adjusted for ethnicity, higher census area income, and age

Table 3. Enrollment characteristics by ethnicity

Characteristic	Hispanics Adjusted		Non-Hispanic Adjusted*	
	Odds Ratio	95% CI	Odds Ratio	95% CI
Age (10 yr increments)	1.02	0.89–1.16	1.14	1.03–1.26
Sex (Male referent)	1.88	1.36–2.58	1.43	1.14–1.80
Higher census area income (< Median referent)	1.62	1.19–2.20	1.32	1.04–1.66

* Adjusted for age, sex and higher census area income.

this is the first study to demonstrate a mailed online recruitment method targeting Hispanics for oversampling.

Sweeney et al, found older age positively associated with success in contacting Hispanic controls but negatively associated with cooperation in the study interview.[11] Our study did not find any association between age and enrollment for Hispanics. Our results also contrast with Sweeney who found that US Census data, including income, education, and urban/rural residence, did not significantly influence participation in their study.[11] The sensitivity of identifying Hispanics through surname analysis vs self-reported ethnicity validation was greater than other published sensitivities: 89% [326/367] vs ranges of 1.0%–88.99%.[11,19,33–34] The specificity was slightly lower than other published specificities.[11] When broken down by sex, sensitivity for females was on the higher end of the published range of 62%–95.2%[19,33–34] and within the published range for specificity of 88%–89.9%.[19,35–36] Sensitivity and specificity are expected to be lower in females since married women may have taken a non-Hispanic surname. Similarly for males, our study was on the higher end of the published range for both sensitivity (82%–97.7%)[19,35–36] and specificity (92%–92.7%).[19,36] This success is partially due to sampling in a population that already has a rich base of Hispanics and partially due to the health plan being located in a large metropolitan area.[18–19]

A potential limitation to this study was that ethnicity was determined based on self-report by the research participant. We cannot guarantee that the person we intended to enroll in the study was the actual enrollee (eg, a spouse or other household member may have enrolled in place of the intended invitee). Also, we do not have detailed information about our survey nonrespondents.

In addition, the Passel-Word listing was used instead of the more encompassing GUESS listing (RW Buechley, Generally Useful Ethnic Search System with the University of New Mexico Cancer Research and Treatment Center, 1976.) Passel-Word lists the most traditional Spanish surnames while GUESS adds regional variations to that basic list, however, the difference is minimal. Calculating a flag for whether or not the individual was considered probable Hispanic according to the GUESS surname listing added 5 individuals (out of 514, 1.0%) to the probable Hispanic column, where they were not identified as such with the Passel-Word listing.

Methods from this study may not be generalizable to areas outside of our sampling frame, particularly those areas with extremely low concentrations of Hispanics or non-metropolitan areas. Perkins (1993) suggests that a Spanish surname match is less efficient in areas with low concentrations of Hispanics and in non-metropolitan areas.[30]

Also, this intervention was not available in Spanish, even though we were oversampling for Hispanic enrollment. This potentially limited the number of Hispanics who enrolled in the study. We oversampled the potential KPCO Hispanic population, thus the proportion enrolled is not representative of the underlying population. However, we were potentially able to target the more at-risk population of Hispanic women than if we hadn't oversampled.[10] This finding has potential implications for cancer prevention interventions that target Hispanics given the link between increased BMI and cancer risk.[37]

In conclusion, this study demonstrates that identifying Hispanics through surname for oversampling can be successful in terms of sampling yield and accuracy. However, results from this study suggest that Hispanics are significantly less likely than non-Hispanics to enroll in a web-based nutritional intervention. Additional research is needed to identify methods of attracting more Hispanic subjects to these kinds of interventions.

ACKNOWLEDGMENTS

Grants or Financial Support: This work was supported by the National Cancer Institute [U19CA079689] and was performed in affiliation with the Cancer Research Network (CRN). We acknowledge and thank the MENU Study Team, Elizabeth Staton and Barbara McCray for their contributions. The authors have no relationship, financial or otherwise (eg, stock ownership, advisory board, speaker's bureau, honorarium), that might be viewed as creating a conflict of interest.

…results from this study suggest that Hispanics are significantly less likely than non-Hispanics to enroll in a web-based nutritional intervention.

REFERENCES

1. Im EO, Chee W. Methodological issues in the recruitment of ethnic minority subjects to research via the Internet: a discussion paper. *Int J Nurs Stud*. 2005;42(8):923–929.
2. Gilliss CL, Lee KA, Gutierrez Y, et al. Recruitment and retention of healthy minority women into community-based longitudinal research. *J Womens Health Gend Based Med*. 2001;10(1):77–85.
3. Tejeda HA, Green SB, Trimble EL, et al. Representation of African-Americans, Hispanics, and Whites in National Cancer Institute cancer treatment trials. *J Natl Cancer Inst*. 1996;88(12):812–816.
4. Larkey LK, Ogden SL, Tenorio S, Ewell T. Latino recruitment to cancer prevention/screening trials in the Southwest: setting a research agenda. *Appl Nurs Res*. 2008;21(1):30–39.
5. U.S. Census Bureau. Detailed Tables, DP-1. United States Profile of General Demographic Characteristics: 2000. Available at: http://factfinder.census.gov. Last accessed October 2009.
6. Haskins AR, Kirk-Sanchez N. Recruitment and retention of students from minority groups. *Phys Ther*. 2006;86(1):19–29.
7. Sheppard VB, Cox LS, Kanamori MJ, et al. Brief report: if you build it, they will come: methods for recruiting Latinos into cancer research. *J Gen Intern Med*. 2005;20(5):444–447.
8. Epidemiological Research in Hispanic Populations. Opportunities, Barriers and Solutions. Available at: http://www.nhlbi.nih.gov/meetings/workshops/hispanic.htm. Last accessed: October 19, 2009.
9. Menella JA, Ziegler P, Briefel R, Novak T. Feeding Infants and Toddlers Study: the types of food fed to Hispanic infants and toddlers. *J Am Diet Assoc*. 2006;106:S96–S106.
10. Calzada PJ, Anderson-Worts. The obesity epidemic: are minority individuals equally affected? *Prim Care*. 2009;Jun;36(2):307–317.
11. 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):1210–1219.
12. Becker DM, Hill DR, Jackson JS, Levine DM, Stillman FA, Weiss SM. *Health Behavior Research in Minority Populations: Access Design and Implementation*. 1992;U.S.Department of Health and Human Services, NIH Publication No. 92-2965.
13. McDougall GJ, Holston EC, Wilke P. Recruiting African Americans into research on cognitive aging. *Ethn Dis*. 2001;11(1):124–133.
14. Harris KJ, Ahluwalia JS, Catley D, Okuyemi KS, Mayo MS, Resnicow K. Successful recruitment of minorities into clinical trials: The Kick It at Swope project. *Nicotine Tob Res*. 2003;5(4):575–584.
15. Wilson JJ, Mick R, Wei J, et al. Clinical trial resources on the internet must be designed to reach underrepresented minorities. *Cancer J*. 2006;12:475–481.
16. HispanicBusiness.com. Study: Hispanic Internet use reaches 50 percent. Available at: http://www.hispanicbusiness.com/news/newsbyid.asp?id=6135. Last accessed October 2009.
17. Pew Internet. Reports: online activities and pursuits. Available at: http://www.pewinternet.org/PPF/r/113/report_display.asp. Last accessed October 2009.
18. Judkins D, Massey JT, Smith V. Using Surnames to Oversample Hispanics from a List Frame. In: *Proceedings of the Survey Research Methods Section of the American Statistical Associations: 1992*. Washington DC: American Statistical Association, 1992;847–852.
19. U.S. Census Bureau. Evaluating the Passel-Word Spanish surname list: 1990 Decennial Census Post Enumeration Survey Results (Population Division Technical Working Paper 4). Available at: http://www.census.gov/population/www/documentation/twps0004.html. Last accessed: October 19, 2009.
20. Howard CA, Samet JM, Buechley RW, Schrag SD, Key CR. Survey research in New Mexico Hispanics: some methodological issues. *Am J Epidemiol*. 2007;117[1]:27–34.
21. Perez-Stable EJ, Hiatt RA, Sabogal F, Otero-Sabogal R. Use of Spanish surnames to identify Latinos: comparison to self-identification. *J Natl Cancer Inst Monogr*. 1995;18:11–15.
22. Morgan RO, Wei II, Virnig BA. Improving identification of Hispanic males in Medicare: use of surname matching. *Medical Care*. 2004;42(8):810–816.
23. Wagner EH, Greene SM, Hart G, et al. Building a research consortium of large health systems: the Cancer Research Network. *J Natl Cancer Inst Monogr*. 2005;35:3–11.
24. Rothert K, Strecher VJ, Doyle LA, et al. Web-based weight management programs in an integrated health care setting: a randomized, controlled trial. *Obesity*. 2006;14:266–272.
25. Janz NK, Becker MH. The Health Belief Model: a decade later. *Health Educ Q*. 1984;11:1–47.
26. Resnicow K, DiIorio C, Soet JE, Ernst D, Borrelli B, Hecht J. Motivational interviewing in health promotion: it sounds like something is changing. *Health Psychol*. 2002;21:444–451.
27. Alexander GL, Divine GW, Couper MP, et al. Effect of incentives and mailing features on online health program enrollment. *Am J Prev Med*. 2008;34:382–388.
28. Stopponi M, Ritzwoller D, Alexander G, et al. Recruitment to a randomized web-based nutritional intervention trial: characteristics of participants compared to non-participants. *J Med Internet Res*. 2009;11(3):e38.
29. Passel JS, Word DL. *Constructing the list of Spanish surnames for the 1980 Census: an application of Bayes' Theorem* Paper presented at: Annual Meeting of the Population Association of America: Denver, CO, 1980.
30. Word DL, Perkins RC. *Building a Spanish surname list for the 1990's – A New Approach to An Old problem (Population Division Technical Working Paper 13)*. Available at: http://www.census.gov/population/documentation/twpno13.pdf. Last accessed October 2009.
31. Thompson FE, Subar AF, Smith AF, et al. Fruit and vegetable assessment: performance of 2 new short instruments and a food frequency questionnaire. *J Am Diet Assoc*. 2002;102(12):1764–1772.
32. Resnicow K, Odom E, Wang T, Dudley WN, Mitchell D, Vaughan R, Jackson A, Baranowski T. Validation of three food frequency questionnaires and 24-hour recalls with serum carotenoid levels in a sample of African-American adults. *Am J Epidemiol*. 2000;Dec 1;152(11):1072–1080.
33. HMO Research Network. Utilization of a managed Medicaid population to validate a software program designed to determine ethnicity. Available at: http://www.hmoresearchnetwork.org/archives/2002/2002abst.htm. Last accessed October 2009.
34. Stewart SL, Swallen KC, Glaser SL, Horn-Ross PL, West DW. Comparison of methods for classifying Hispanic ethnicity in a population-based cancer registry. *Am J Epidemiol*. 1999;149(11):1063–1071.
35. Coronado GD, Koepsell TD, Thompson B, Schwartz SM, Wharton RS, Grossman JE. Assessing cervical cancer risk in Hispanics. *Cancer EpidemiolBiomarkers Prev*. 2002;11(10), 1:979–984.
36. Wei II, Virnig BA, John DA, Morgan RO. Using a Spanish surname match to improve identification of Hispanic women in Medicare administrative data. *Health Serv Res*. 2006;41(4), Pt 1:1469–1481.
37. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet*. 2008;371:569–578.

AUTHOR CONTRIBUTIONS

Design concept of study: Carroll, Ritzwoller, Stopponi, Cole Johnson

Acquisition of data: Carroll, Ritzwoller, Stopponi, Cole Johnson

Data analysis and interpretation: Carroll, Ritzwoller, Stopponi, Cole Johnson

Manuscript draft: Carroll, Ritzwoller, Stopponi, Cole Johnson

Statistical expertise: Carroll, Ritzwoller, Cole Johnson

Acquisition of funding: Cole Johnson

Supervision: Ritzwoller