Title
Using predicted Spanish preference to target bilingual mailings in a mail survey with telephone follow-up

Journal
Health Services Research, 54(1)

ISSN
0017-9124

Authors
Elliott, Marc N
Klein, David J
Kallaur, Paul
et al.

Publication Date
2019-02-01

DOI
10.1111/1475-6773.13088

Peer reviewed

DOI: 10.1111/1475-6773.13088

RESEARCH ARTICLE

HSR Health Services Research

Using predicted Spanish preference to target bilingual mailings in a mail survey with telephone follow-up

Marc N. Elliott PhD[1] | David J. Klein MS[1] | Paul Kallaur MA[2] | Julie A. Brown BA[1] | Ron D. Hays PhD[1,3] | Nate Orr MA[1] | Alan M. Zaslavsky PhD[4] | Megan K. Beckett PhD[1] | Sarah Gaillot PhD[5] | Carol A. Edwards BA[1] | Amelia M. Haviland PhD[1,6,7]

[1]RAND Corporation, Santa Monica, California

[2]Center for the Study of Services, Washington, District of Columbia

[3]UCLA David Geffen School of Medicine, Los Angeles, California

[4]Harvard Medical School, Boston, Massachusetts

[5]Centers for Medicare & Medicaid Services, Baltimore, Maryland

[6]Carnegie Mellon University, Pittsburgh, Pennsylvania

[7]RAND Corporation, Pittsburgh, Pennsylvania

Correspondence
Marc N. Elliott, PhD, RAND Corporation, Santa Monica, CA 90401.
Email: elliott@rand.org

Funding Information
This study was funded by the Centers for Medicare & Medicaid Services (CMS) contract HHSM-500-2005-00028I to RAND Corporation. Ron D. Hays was also supported in part by a grant from the National Institute on Aging (P30-AG021684).

Objective: Spanish-preferring Medicare beneficiaries are underrepresented in national patient experience surveys. We test a method for improving their representation via higher response rates.

Data Sources/Study Setting: 2009-2010 Medicare CAHPS surveys; Medicare population.

Study Design: We used surname and address to predict Spanish-language preference for a national sample of 177 139 beneficiaries. We randomized half of the 10 000 non-Puerto Rico beneficiaries with the highest predicted probabilities of Spanish preference (>10 percent) to bilingual mailings (intervention) and half to standard English-only mailings (control).

Data Collection: Medicare CAHPS Survey data were collected through mail surveys with telephone follow-up of nonrespondents.

Principal Findings: Mail response rate was higher for intervention (28.7 percent) than control (23.9 percent) ($P < 0.0001$); phone response rates among mail nonrespondents were similar in intervention and control arms (15.8 percent vs 15.7 percent, $P = 0.90$). Targeted bilingual mailings induced 6.5 percent of those who would not have responded to respond by mail and 54.0 percent of those who would have responded in English to respond in Spanish. Beneficiaries with greater Spanish probabilities showed greater increases in response rates, a higher proportion of responses in Spanish, and lower control response rates among.

Conclusions: Targeted bilingual mailing of mixed-mode surveys using commonly available surname and address information can efficiently increase representation of this underrepresented group.

KEYWORDS
indirect estimation, response rates, sample design, sampling, survey research-mail

1 | INTRODUCTION

Health researchers and social scientists use surveys with national probability samples to track population health care experiences and disparities between subgroups. As the United States becomes more ethnically and linguistically diverse,[1] it is becoming more important to accurately estimate characteristics and experiences of those for whom English is not the preferred language. Surveys may underrepresent those unable to respond in English because of the high cost of administering surveys in multiple languages. Linguistically isolated Spanish-language-speaking households had the lowest percentage of response by mail on the American

Community Survey.[2] Research also suggests that bilingual (Spanish-English) patients give fewer "don't know" responses to diagnostic surveys if the survey is completed in Spanish instead of English.[3,4]

Spanish is the second most used language in the United States; the 2010 American Community Survey estimates that it is the primary language spoken at home for 37 million people, almost half of whom have limited English proficiency.[5] Here, we describe and evaluate an effort to efficiently collect information about the health care experiences of Spanish-preferring Medicare beneficiaries with limited English proficiency (SPLEP). These beneficiaries, who are entitled to Medicare coverage due to age (65 or older) and/or disability, often have linguistic and health literacy barriers and report poorer care than other Medicare beneficiaries; for instance, they are immunized at half the rates of English-speaking non-Hispanic Whites.[6]

The Centers for Medicare & Medicaid Services (CMS) implements the Consumer Assessments of Healthcare Providers and Systems (CAHPS®) surveys to measure beneficiaries' perceptions of care and services received through Medicare managed care plans and fee-for-service insurance.[7,8] As the largest nationally representative survey regarding the health care experiences of U.S. seniors and disabled adults, the Medicare CAHPS Survey (MCAHPS) provides a unique opportunity to monitor the health care experiences of seniors and disabled adults.

However, SPLEP are underrepresented in MCAHPS and similar surveys because of low response rates.[9–13] While more than half of those in the MCAHPS sample who are listed administratively as non-Hispanic White complete a survey, less than one-third of those identified as Hispanic or Asian do.[13,14] Compared to beneficiaries of the same race/ethnicity who answer in English, beneficiaries who prefer non-English languages (eg, Spanish, Chinese) have lower scores on CAHPS composites.[14–18] Increasing response rates for those who prefer languages other than English may improve the representativeness of seniors particularly vulnerable to poor care experiences.

"Double-stuffed" mailing of the survey—including both a Spanish and an English copy in the same envelope—increases mailing and printing costs by 35 percent.[19] Due to the large total sample size, it was not feasible to send the Spanish-language 2010 MCAHPS survey to beneficiaries outside of Puerto Rico with the rare exception (<1 percent) of beneficiaries who called a toll-free number on the bilingual prenotification letter but refused the offer to complete the survey on the telephone in Spanish immediately. Nonrespondents to the survey mailings were followed up by telephone and offered the choice to complete the survey on the phone in English or Spanish. Almost all (>99 percent) responses in Spanish from the U.S. mainland occurred during telephone follow-up. Increasing Spanish-language mail responses has the potential to improve the measurement of patient experience both by increasing the representation of underrepresented groups and by making the mix of mail and phone survey modes for Spanish-preferring respondents, and hence their survey mode effects, more similar to those of English-preferring respondents.[20,21] Increasing the representation of Spanish speakers is superior to nonresponse weighting for their underrepresentation, in that it results in greater precision and requires fewer assumptions.

In this study, we report on an experiment to evaluate the response rate effects of mailing a bilingual survey package to targeted Medicare beneficiaries when cost makes it infeasible to implement 100 percent bilingual mailing. Targeting is based on an innovative approach to identify beneficiaries with the highest probabilities of being Spanish-preferring based on information from CMS administrative data, prior MCAHPS survey data, and U.S. Census data.

1.1 | CMS administrative data, race/ethnicity, and Spanish-language preference

Centers for Medicare & Medicaid Services administrative data contain information on beneficiaries' age, gender, and other limited demographic information, including race/ethnicity. The standard CMS administrative race/ethnicity variable is derived from Social Security Administration (SSA) records. For persons assigned a Social Security number prior to 1980, race/ethnicity categories were limited to "Black," "White," or "Other." Starting in 1980, Social Security forms used race/ethnicity categories that conform to current Office of Management and Budget standards.[22] Persons who apply for a replacement Social Security card have an opportunity to update their race/ethnicity reported in the SSA database, and this information is transferred to CMS. Because of how data were collected prior to 1980, CMS administrative records are known to substantially underidentify non-Black minority groups, including Hispanics.[23–27]

In the absence of direct race/ethnicity information, the Institute of Medicine (now the National Academy of Medicine) recommends the use of indirect estimation to monitor health disparities in care as a bridging strategy.[28] Surname analyses and geocoding have been used to infer race/ethnicity when it is otherwise unavailable. Surname analyses use dichotomous dictionaries (eg, Hispanic, not Hispanic) to identify Hispanics[29–32]; although useful, these surname lists do not fully utilize the information that surnames might convey regarding race/ethnicity because dichotomization loses information about relative probabilities of being Hispanic given one's surname. Geocoding links an individual's address to a Census measure of their neighborhood's racial/ethnic composition and uses that measure as a basis for inferring the individual's race/ethnicity. Because Hispanics tend to live in less-segregated neighborhoods than some other racial/ethnic groups (eg, Blacks),[33,34] geocoding alone cannot accurately identify members of these minority groups.

One indirect estimation method recommended by the Institute of Medicine[28] combines probabilities based on residential racial/ethnic information with probabilities based on the latest surname list, both from U.S. Census data, to produce a vector of probabilities that a person is Hispanic, non-Hispanic White, Black, Asian/Pacific Islander, or Multiracial.[26] This Bayesian Improved Surname and Geocoding (BISG) method achieves a weighted average of 93 percent concordance (area under the curve for predicted probabilities vs observed outcome; 95 percent for Hispanic; 94 percent for Asian/Pacific Islander; 93 percent for Black and White) compared to self-report.[26] The BISG approach has been adapted to improve race/ethnicity information from CMS administrative files.[35] In this paper, we adapt that Medicare-specific version of the BISG method (the Medicare BISG or MBISG) to predict Spanish-language preference.

2 | METHODS AND RESULTS

2.1 | Predicting Spanish-language preference

Because Hispanic ethnicity is the primary predictor of Spanish preference but is not well measured in the original administrative CMS race/ethnicity variable,[27] we began by estimating the probability that a beneficiary is Hispanic using the MBISG approach. First, we estimated the population-level distribution of beneficiary race/ethnicity (Hispanic, non-Hispanic White, Black, Asian or Pacific Islander [API], American Indian/Alaska Native [AI/AN], or Multiracial) for each combination of Medicare administrative classification and age-group (18-34, 35-54, 55-64, 65-74, 75-84, 85+), as the average of self-reported race/ethnicity in weighted CAHPS data, collapsing age-groups where needed to avoid imprecise estimation. These population estimates were matched to beneficiaries by Medicare administrative race/ethnicity and age and used as a vector of six initial racial/ethnic probabilities. Second, an independent set of probabilities of each race/ethnicity for each beneficiary were calculated using only surname and address information.[26] Third, these two estimates (from three information sources) were combined using Bayes' theorem to develop a final, "updated" vector of six racial/ethnic probabilities for each beneficiary.

As noted above, virtually all Spanish responses outside of Puerto Rico (>99 percent) were by telephone because obtaining a paper Spanish survey required the beneficiary to call and request it. Very few people initiate a call, and when they do, efforts are made to convert the inbound request to telephone (Spanish) interviews to increase response rates. Hence, we do not have a direct measure of Spanish-language preference for all sampled beneficiaries, but only a proxy based on completing the survey in Spanish by telephone when the offer is made to the few who call in or to the many who receive a call during phone follow-up. We expect that this proxy is positively related to Spanish-language preference, but underestimates it.

We used weighted logistic regression to predict Spanish completion of the 2010 MCAHPS survey among all U.S. mainland completions (n = 397 639) in the prior year's (2009) MCAHPS survey. The independent variables were as follows: the predicted probabilities of being Black, Hispanic, and API, the SSA-based race/ethnicity variable, 2007-2009 county-level proportions of persons who both speak Spanish at home and have limited English proficiency (from the U.S. Census Bureau's ongoing American Community Survey, geocoded to a beneficiary's county of residence), receipt of a low-income subsidy (limited to those with income <150 percent of the federal poverty line), beneficiary age and gender, and coverage type. The concordance of this logistic regression was 97 percent ($P < 0.001$). The most important predictors of Spanish completion (data not shown) were (a) the MBISG predicted probability of being Hispanic (odds ratio [OR] = 1.62 per 10 percent change), county-level percent who speak Spanish at home (OR = 1.38 per 10 percent change), and eligibility for a low-income subsidy (OR = 2.91).

Finally, we used the coefficients from the logistic regression model predicting Spanish response to the 2009 MCAHPS survey to calculate individual (proxy) probabilities of Spanish-language preference among all 2010 MCAHPS beneficiaries for use in the bilingual mailing experiment. For brevity, in what follows we will refer to these predicted probabilities of Spanish telephone completion under 2009 survey protocols as the "predicted probability of Spanish preference."

2.2 | Bilingual mailing experiment sample and survey administration

A randomized bilingual mailing experiment was conducted as part of the 2010 MCAHPS survey. We restricted this experiment to those having fee-for-service with a free-standing prescription drug plan (FFS + PDP), representing approximately 38 percent of Medicare beneficiaries. There were 177 139 FFS + PDP beneficiaries randomly sampled in 2010 for MCAHPS surveys. Among those beneficiaries, the distribution of the predicted probability of Spanish preference is skewed, with a mean of 2.2 percent and median of 0.1 percent. We randomized the 10 000 (5.6 percent) sample members with a predicted probability of Spanish response of at least 10 percent to one of two conditions: a standard mailing arm or a bilingual mailing arm. The two experimental arms had very similar distributions for the probability of Spanish response: The means (SDs) for both the standard mailing arm and bilingual mailing arm were 33.9 percent (13.0 percent) with a median of 33.2 percent.

As noted above, the standard MCAHPS approach (control) has a bilingual (English/Spanish) presurvey notification letter. The prenotification letter provides a telephone number to call in to complete the interview by phone in English or Spanish. Two English-language mailings of the survey were sent to all beneficiaries in the U.S. mainland. Nonrespondents were then followed up by telephone, with options to complete the survey in English or Spanish.

The bilingual mailing (intervention) involved a double-stuffed mailing with English and Spanish surveys, but was otherwise the same as the standard mailing with identical bilingual prenotification and follow-up. Development of the Spanish version of the MCAHPS survey included initial professional translation followed by review by a bilingual committee from the CAHPS Cultural Comparability Team and a professional translator experienced in CAHPS instrument design issues. Studies have found empirical evidence of measurement equivalence for Spanish and English versions of CAHPS surveys.[36]

2.3 | Overall effects of bilingual mailing on response rates

Because the detection of ineligibles might differ by language and mode, and our focus was on increasing total response to the survey, we used a conservative variant of AAPOR RR1 that treated the few ineligibles as nonrespondents. As shown in Table 1, those with a high predicted probability of Spanish preference had markedly lower

TABLE 1 Response mode and language by experiment status, 2010 fee-for-service beneficiaries with prescription drug coverage (n = 177 139)

	≥10% Predicted Spanish-preferring probability Participated in experiment					<10% Predicted Spanish-preferring probability Did not participate in experiment		
	Received bilingual mailing (n = 5000)		Did not receive bilingual mailing (n = 5000)			Did not receive bilingual mailing (N = 167 139)		
	N	%	N	%	*P*-value[b]	N	%	*P*-value[c]
Resolution of cases								
Spanish mail	763	15.3	0	0.0	**<0.0001**	0	0.0	—
English mail	674	13.5	1194	23.9	**<0.0001**	77 895	46.6	**<0.0001**
Spanish phone	436	8.7	455	9.1	0.53	208	0.1	**<0.0001**
English phone	128	2.6	143	2.9	0.39	17 284	10.3	**<0.0001**
Nonresponse/Incomplete[a]	2999	60.0	3208	64.2	**<0.0001**	71 752	42.9	**<0.0001**
Response rates[a]								
Overall		40.0		35.8	**<0.0001**		57.1	**<0.0001**
Total mail		28.7		23.9	**<0.0001**		46.6	**<0.0001**
Total phone		11.3		12.0	0.30		10.5	**<0.0001**
Phone among mail nonrespondents		15.8		15.7	0.90		19.6	**<0.0001**
Percentage of responses in Spanish								
All responses		59.9		25.4	**<0.0001**		0.2	**<0.0001**
Mail responses		53.1		0.0	**<0.0001**		0.0	—
Phone responses		77.3		76.1	0.63		1.2	**<0.0001**

[a]A small number of ineligibles were pooled with nonrespondents to produce conservative response rates, since it is possible that the detection of ineligibles would differ by language and mode.
[b]Fisher's exact test *P*-value comparing the two experimental groups.
[c]Fisher's exact test *P*-value comparing the two groups that did not receive bilingual mailing (participated vs did not participate in experiment).

response rates than other Medicare beneficiaries under the standard mail protocol (35.8 percent vs 57.1 percent, $P < 0.0001$), especially by mail (23.9 percent vs 46.6 percent, $P < 0.0001$). The bilingual mailing improved the mail response rate for the targeted beneficiaries by 20.1 percent or 4.8 percentage points from 23.9 percent (22.7 percent-25.1 percent) to 28.7 percent (27.5 percent-30.0 percent, $P < 0.0001$). Response rates by telephone follow-up for high-Spanish-probability beneficiaries were not significantly different for the two mail protocols (11.3 percent bilingual mailing vs 12.0 percent standard, $P = 0.30$, which translates to 15.8 percent and 15.7 percent of mail nonrespondents, respectively). By calculating the ratio of the 4.2 percent increase (40.0 percent-35.8 percent) in the overall response rate to the proportion who did not respond to the standard mailing (100−35.8 = 64.2 percent), we estimated that 6.5 percent of persons with high (>10 percent) predicted probability of Spanish response who would not have otherwise responded, responded by mail.

2.4 | Overall effects of bilingual mailing on language of response

With respect to telephone responses, 77.3 percent of bilingual mailing and 76.1 percent of other high-probability Spanish phone respondents chose to respond in Spanish compared to only 1.2 percent of all other telephone respondents (see Table 1), providing support for the validity of the predictions of Spanish-language preference. With respect to mail responses, 53.1 percent of responses to the bilingual mailing were in Spanish, compared to none (0.0 percent) in the absence of bilingual mailing (even among those with high predicted probabilities of Spanish preference). By calculating the percentage of beneficiaries who responded in Spanish (40.0 × 59.9 = 24.0 percent in bilingual arm, 9.1 percent in standard arm) and attributing 4.2 percent of the 24.0 percent bilingual arm Spanish response rate to the difference in overall response rates (24.0−4.2 = 19.8 percent), we estimate that 54.0 percent [(19.8−9.1)/19.8] of predicted Spanish respondents who would have responded by English under the standard protocol responded in Spanish instead.

2.5 | Stratifying effects on response rate and language of response by level of predicted probabilities of Spanish preference

The sample of beneficiaries targeted for bilingual mailings is likely to consist of a mixture of beneficiaries with a true Spanish-language preference and false positives-beneficiaries who truly prefer to respond in English or another language. If our proxy predicted probabilities provide useful information about which potential respondents are most likely to be Spanish-preferring, we would expect the proportion of those truly preferring Spanish and the percentage of responses that are in Spanish to be highest for respondents with the highest predicted probabilities and lowest for respondents with the lowest predicted probabilities. We might also expect that the response rates in the absence of bilingual mailing would fall as this probability increased (due to more frequent language mismatch) and perhaps that the increase in mail response rate with bilingual mailing would increase as this probability increased due to greater relevance.

We present more granular data on response rates and propensity to respond in Spanish to examine if patterns differ by level of predicted probability group. Table 2 distinguishes three levels: predicted probabilities of 40 percent or more (2939 cases), 25-39 percent (4637 cases), and 10-24 percent (2424 cases). Table 2 generally shows the same patterns as in Table 1, but graded by predicted probability of Spanish. As expected, standard protocol response rates were lowest in the group with the highest predicted probability of Spanish preference (32.8 percent overall, $P < 0.05$ vs all others 20.1 percent by mail, $P < 0.005$ vs all others). Under bilingual mailing, the percentages of all targeted beneficiaries responding by mail and in Spanish were 18.0 percent, 15.7 percent, and 11.1 percent in the three levels of predicted probability of preferring Spanish. Gains in mail response rates due to bilingual mailing were largest for respondents with >40 percent and 25 percent-40 percent predicted probability (+6.1 percentage points, $P < 0.001$, and +5.2 percentage points, $P < 0.001$, respectively), but nonsignificant for respondents with <25 percent predicted probability (+2.7 percentage points, $P = 0.15$).

As further validation of the predicted probability approach, the percentage of mail responses that were in Spanish increased with the predicted probability of Spanish: 68.8 percent (>40 percent probability of Spanish preference), 53.2 percent (25 percent-40 percent probability), and 36.5 percent (<25 percent probability), respectively. A similar pattern was seen in the proportion of phone responses that were in Spanish: 92.0 percent (>40 percent probability), 79.3 percent (25 percent-40 percent probability), and 56.5 percent (<25 percent probability) for the bilingual mailing arm, and 93.5 percent (>40 percent probability), 74.8 percent (25 percent-40 percent probability), and 53.2 percent (<25 percent probability) for the standard mailing arm ($P > 0.05$ within all three pairs). There is little evidence that bilingual mailings alter total phone response rates: 12.0 percent (>40 percent probability), 10.4 percent (25 percent-40 percent probability), and 12.1 percent (<25 percent probability) for bilingual mailing, vs 12.7 percent (>40 percent probability), 12.3 percent (25 percent-40 percent probability), and 10.4 percent (<25 percent probability) for standard mailing ($P < 0.05$ only for the 25 percent-40 percent comparison in favor of standard mailing).

3 | CONCLUSIONS

As is the case for many health surveys, the Spanish-preferring subgroup of MCAHPS surveys has been greatly underrepresented in mail responses and overall. In the absence of a bilingual mailing intervention, the response rate for a group with a high predicted probability of preferring Spanish was 35.8 percent as compared to 57.1 percent for all other 2010 respondents; the response rate for just those truly preferring Spanish may be lower. Thus, this study shows the value of survey procedures to increase the representation of Spanish-preferring beneficiaries in health surveys. In some cases, budgets may require targeted Spanish mailings when bilingual mailings for all are not affordable.

We showed that our predicted probabilities based on respondent's demographic characteristics, address, and surname drawing on a model predicting Spanish response in the prior year's survey were effective in identifying members of this formerly unidentifiable group. The targeted sample is indeed primarily Spanish-preferring with a majority responding in Spanish by mail or by telephone when given the option.

The randomized intervention of mailing both Spanish- and English-language surveys vs the standard English-only mailing protocol increased mail response rates for this traditionally difficult-to-reach subgroup. Its results imply that 6.5 percent of persons with high (>10 percent) predicted probability of Spanish response, who would not have otherwise responded, responded by mail. Among those with the highest predicted probability (>40 percent), 8.0 percent responded by mail who would not otherwise have done so. These increases represent a 20 percent relative improvement in mail response rates for targeted beneficiaries compared to response rates using the standard English-only mailing protocol. In addition, study results suggest that 54.0 percent of predicted Spanish respondents who would have responded in English under the standard protocol responded in Spanish instead (50.2 percent of the most likely [>40 percent] Spanish-preferring respondents). This shift from English to preferred language may improve data quality if respondents are better able to understand the question and response choices in their preferred language. Additionally, the greater representation of Spanish-preferring beneficiaries who may have had worse experiences improves the overall and plan-specific measurement of patient experiences as well as our ability to measure and address disparities involving these beneficiaries. Although not measurable in this experiment, it should be noted that bilingual mailings to those with low predicted probabilities of Spanish preference (<10 percent) may be more inclusive, but less cost-effective as there is a trade-off between greater cost and greater total effects at lower cutoffs of predicted probabilities and greater efficiency at higher cutoffs.

The higher proportion of Spanish responses by telephone (>90 percent) than by mail (<70 percent) for the group with the highest predicted probabilities of Spanish preference may reflect limited literacy in both English and Spanish for some of the targeted beneficiaries,[6] which may itself place some limits on the potential gains from

TABLE 2 Response mode and language by experimental condition and Spanish-preferring predicted probabilities, 2010 fee-for-service beneficiaries with prescription drug coverage (n = 177 139)

	Received bilingual mailing						Did not receive bilingual mailing						*P*-value (Received vs Did Not Receive)[b]		
	Predicted probability of Spanish preference						Predicted probability of Spanish preference								
	>40% (n = 1470)		25-40% (n = 2318)		10-25% (n = 1212)		>40% (n = 1469)		25-40% (n = 2319)		10-25% (n = 1212)				
	N	%	N	%	N	%	N	%	N	%	N	%	>40%	25-40%	10-25%
Resolution of cases															
Spanish mail	265	18.0	363	15.7	135	11.1	0	0.0	0	0.0	0	0.0	<0.0001	**<0.0001**	**<0.0001**
English mail	120	8.2	319	13.8	235	19.4	296	20.1	561	24.2	337	27.8	**<0.0001**	**<0.0001**	**<0.0001**
Spanish phone	162	11.0	191	8.2	83	6.8	174	11.8	214	9.2	67	5.5	0.49	0.25	0.21
English phone	14	1.0	50	2.2	64	5.3	12	0.8	72	3.1	59	4.9	0.84	0.05	0.71
Nonresponse/Incomplete[a]	909	61.8	1395	60.2	695	57.3	987	67.2	1472	63.5	749	61.8	**0.003**	**0.02**	**0.03**
Response rates[a]															
Overall		38.2		39.8		42.7		32.8		36.5		38.2	**0.003**	**0.02**	**0.03**
Total mail		26.2		29.4		30.5		20.1		24.2		27.8	**0.0001**	**<0.0001**	0.15
Total phone		12.0		10.4		12.1		12.7		12.3		10.4	0.58	**0.04**	0.20
Phone among mail nonrespondents		16.2		14.7		17.5		15.9		16.3		14.4	0.82	0.22	0.09
Percentage of responses in Spanish															
All responses		76.1		60.0		42.2		36.1		25.3		14.5	**<0.0001**	**<0.0001**	**<0.0001**
Mail responses		68.8		53.2		36.5		0.0		0.0		0.0	**<0.0001**	**<0.0001**	**<0.0001**
Phone responses		92.0		79.3		56.5		93.5		74.8		53.2	0.68	0.25	0.63

Note: P-values < 0.05 are bolded.

[a]A small number of ineligibles were pooled with nonrespondents to produce conservative response rates, since it is possible that the detection of ineligibles would differ by language and mode.

[b]Fisher's exact test *P*-value comparing the two experimental groups within each predicted probability stratum.

bilingual mailing and underscore the need for telephone follow-up of mail nonresponse to ensure broad representation.[12,20,37]

A strength of our approach is that MBISG estimates of Spanish preference for all Medicare beneficiaries are freely available from CMS via a Data Use Agreement. Moreover, with additional effort, our approach could be potentially modified to target beneficiaries who prefer other non-English languages, such as Chinese. Doing so would require a modification of the MBISG algorithm to identify national origin groups with the Asian Pacific Islander group, which is technically feasible given the distinctiveness of Chinese surnames but it would require additional information from the Census Bureau. A limitation of our analysis is that it was restricted to beneficiaries with fee-for-service coverage who were members of prescription drug plans, a group that is less than half of all Medicare beneficiaries, and differs from other Medicare beneficiaries to some extent;[38] the pattern of associations between response rate and propensity to respond in Spanish may not hold for other populations or survey versions.

Our approach to predicting the probability of race/ethnicity and language can be used in realms beyond survey design such as targeting public health, educational, or other interventions with language-specific materials. Similarly, it may be useful in other aspects of survey protocols such as allocating bilingual telephone interviewers to cases for follow-up. Our approach can also be adapted to making predictions of language-minority populations even if all the information that we used in the analyses presented here was not available. Future research should test the effects of bilingual mailings on other populations including those younger than the Medicare population examined here.

In summary, we demonstrate that administrative data can be used to accurately predict Spanish-language preference in an older population and that bilingual mailing to this population based on those predictions produces some increases in response rates. In addition, the bilingual mailing substantially shifts responses to the language preferred by the respondent. These results are notable and have important implications for a vulnerable subgroup that is significantly underrepresented in national health surveys. Furthermore, the usefulness of these methods need not be limited to linguistic subgroups. A similar approach might be taken to improving representation of other subpopulations that can be identified with probability models using administrative and/or geographically-linked data, and for which some targeted modification of survey methods might improve response.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This study was funded by the Centers for Medicare & Medicaid Services (CMS) contract HHSM-500-2005-00028I to RAND Corporation. Ron D. Hays was also supported in part by a grant from the National Institute on Aging (P30-AG021684). All authors are responsible for the reported research and have participated in all phases of it. Each author had a significant role in conceptualizing and writing the article and each approved the final manuscript. The authors have no financial or proprietary interest in the subject matter or materials discussed in the manuscript. However, please note that Sarah Gaillot is an employee of the sponsoring agency. The authors would like to thank Biayna Darabidian, BA for administrative assistance in preparation of the manuscript.

CONFLICT OF INTEREST

All authors declare no conflict of interest. However, please note that Sarah Gaillot is an employee of the sponsoring agency.

ORCID

Marc N. Elliott http://orcid.org/0000-0002-7147-5535

REFERENCES

1. Population Projections. U.S. Census Bureau; 2017. https://www.census.gov/programs-surveys/popproj.html.
2. McGovern PD. *A Quality Assessment of Data Collected in the American Community Survey (ACS) From Households With Low English Proficiency*. Washington, DC: U.S. Census Bureau; 2004.
3. Marcos L, Trujillo M. The psychiatric examination of Spanish-Americans. In: Duran RP, ed. *Culture, Language and Communicative Behavior*. Newark, NJ: Alex Publishing Co; 1981:187-194.
4. Karno M, Burnam A, Escobar JI, Hough RL, Eaton WW. Development of the Spanish-language version of the National Institute of Mental Health Diagnostic Interview Schedule. *Arch Gen Psychiatry*. 1983;40(11):1183-1188.
5. U.S. Census Bureau. *Language Use and English-Speaking Ability: 2000*. Washington, DC: US Census Bureau; 2003.
6. Haviland AM, Elliott MN, Hambarsoomian K, Lurie N. Immunization disparities by Hispanic ethnicity and language preference. *Arch Intern Med*. 2011;171(2):158-165.
7. Goldstein E, Cleary PD, Langwell KM, Zaslavsky AM, Heller A. Medicare Managed Care CAHPS(®): a tool for performance improvement. *Health Care Financ Rev*. 2001;22(3):101-107.
8. Schnaier JA, Sweeny SF, Williams VS, et al. Special issues addressed in the CAHPS survey of Medicare managed care beneficiaries. Consumer Assessment of Health Plans Study. *Med Care*. 1999;37(3 Suppl):Ms69-Ms78.
9. Kahn KL, Liu H, Adams JL, et al. Methodological challenges associated with patient responses to follow-up longitudinal surveys regarding quality of care. *Health Serv Res*. 2003;38(6 Pt 1):1579-1598.
10. Neuman P, Strollo MK, Guterman S, et al. Medicare prescription drug benefit progress report: findings from a 2006 national survey of seniors. *Health Aff*. 2007;26(5):w630-w643.
11. Sharkey JR, Haines PS. The feasibility of using a telephone-administered survey for determining nutritional service needs of noninstitutionalized older adults in rural areas: time and costs. *Gerontologist*. 2001;41(4):546-552.
12. Zaslavsky AM, Zaborski LB, Cleary PD. Factors affecting response rates to the Consumer Assessment of Health Plans Study survey. *Med Care*. 2002;40(6):485-499.
13. Klein DJ, Elliott MN, Haviland AM, et al. Understanding nonresponse to the 2007 Medicare CAHPS survey. *Gerontologist*. 2011;51(6):843-855.
14. Carrasquillo O, Orav EJ, Brennan TA, Burstin HR. Impact of language barriers on patient satisfaction in an emergency department. *J Gen Intern Med*. 1999;14(2):82-87.

15. Weech-Maldonado R, Morales LS, Spritzer K, Elliott M, Hays RD. Racial and ethnic differences in parents' assessments of pediatric care in Medicaid managed care. *Health Serv Res*. 2001;36(3):575-594.
16. Weech-Maldonado R, Morales LS, Elliott M, Spritzer K, Marshall G, Hays RD. Race/ethnicity, language, and patients' assessments of care in Medicaid managed care. *Health Serv Res*. 2003;38(3):789-808.
17. Morales LS, Cunningham WE, Brown JA, Liu H, Hays RD. Are Latinos less satisfied with communication by health care providers? *J Gen Intern Med*. 1999;14(7):409-417.
18. Elliott MN, Edwards WS, Klein DJ, Heller AMY. Differences by survey language and mode among Chinese respondents to a CAHPS health plan survey. *Public Opin Quart*. 2012;76(2):238-264.
19. Edgman-Levitan S, Brown J, Fowler FJ Jr, Gallagher PM, Shaller D. *Feedback Loop: Testing a Patient Experience Survey in the Safety Net*. Oakland, CA: California Healthcare Foundation; 2011.
20. Fowler FJ, Gallagher PM, Stringfellow VL, Zaslavsky AM, Thompson JW, Cleary PD. Using telephone interviews to reduce nonresponse bias to mail surveys of health plan members. *Med Care*. 2002;40(3):190-200.
21. Elliott MN, Zaslavsky AM, Goldstein E, et al. Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. *Health Serv Res*. 2009;44(2 Pt 1):501-518.
22. U.S. Office of Management of Budget. *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. Washington, DC: U.S. Office of Management of Budget; 1997.
23. Lauderdale DS, Goldberg J. The expanded racial and ethnic codes in the Medicare data files: their completeness of coverage and accuracy. *Am J Public Health*. 1996;86(5):712-716.
24. Arday SL, Arday DR, Monroe S, Zhang J. HCFA's racial and ethnic data: current accuracy and recent improvements. *Health Care Financ Rev*. 2000;21(4):107-116.
25. Waldo DR. Accuracy and bias of race/ethnicity codes in the Medicare enrollment database. *Health Care Financ Rev*. 2004;26(2):61-72.
26. Elliott MN, Morrison PA, Fremont A, McCaffrey DF, Pantoja P, Lurie N. Using the Census Bureau's surname list to improve estimates of race/ethnicity and associated disparities. *Health Serv Outcomes Res Methodol*. 2009;9(2):69-83.
27. Zaslavsky AM, Ayanian JZ, Zaborski LB. The validity of race and ethnicity in enrollment data for Medicare beneficiaries. *Health Serv Res*. 2012;47(3 Pt 2):1300-1321.
28. Institute of Medicine. *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. Washington, DC: Institute of Medicine; 2009.
29. Abrahamse AF, Morrison PA, Bolton NM. Surname analysis for estimating local concentration of Hispanics and Asians. *Popul Res Policy Rev*. 1994;13(4):383-398.
30. Falkenstein MR. The Asian and Pacific Islander Surname List: As Developed from Census 2000. Bureau of the Census. Joint Statistical Meetings; 2002.
31. Kestenbaum BM, Ferguson R, Elo IT, Turra CM. Hispanic Identification. Southern Demographic Association Meetings; 2000.
32. Perkins RC. *Evaluating the Passel-Word Spanish Surname List: 1990 Decennial Census Post Enumeration Survey Results*. Washington, DC: US Census Bureau; 1993.
33. Lewis Mumford Center. *Ethnic Diversity Grows, Neighborhood Integration Lags Behind*. Albany, NY: University of Albany; 2001.
34. Massey DS, Denton NA. Hypersegregation in U.S. Metropolitan Areas: black and Hispanic segregation along five dimensions. *Demography*. 1989;26(3):373-391.
35. Martino SC, Weinick RM, Kanouse DE, et al. Reporting CAHPS and HEDIS data by race/ethnicity for Medicare beneficiaries. *Health Serv Res*. 2013;48(2 Pt 1):417-434.
36. Hurtado MP, Angeles J, Blahut SA, Hays RD. Assessment of the equivalence of the Spanish and English versions of the CAHPS Hospital Survey on the quality of inpatient care. *Health Serv Res*. 2005;40(6 Pt 2):2140-2161.
37. Peytchev A, Baxter RK, Carley-Baxter LR. Not all survey effort is equal reduction of nonresponse bias and nonresponse error. *Public Opin Q*. 2009;73(4):785-806.
38. Elliott MN, Haviland AM, Orr N, Hambarsoomian K, Cleary PD. How do the experiences of Medicare beneficiary subgroups differ between managed care and original Medicare? *Health Serv Res*. 2011;46(4):1039-1058.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Elliott MN, Klein DJ, Kallaur P, et al. Using predicted Spanish preference to target bilingual mailings in a mail survey with telephone follow-up. *Health Serv Res*. 2019;54:5–12. https://doi.org/10.1111/1475-6773.13088