## **Behavior Coding Real-time Survey Interpreters**

Michael W. Link<sup>1</sup>, Larry Osborn<sup>2</sup>, Marta Induni,<sup>3</sup> Michael P. Battaglia<sup>2</sup>, and Martin R. Frankel<sup>4</sup>

<sup>1</sup>Centers for Disease Control and Prevention

<sup>2</sup> Abt Associates Inc.

<sup>3</sup> Survey Research Group, Policy Research Institute <sup>4</sup> Abt Associates and Baruch College, CUNY

Note: The findings and conclusions of this report are those of the authors and do not necessarily represent the views of Abt Associates, Baruch College, or the Centers for Disease Control.

Paper prepared for presentation at the 2006 annual conference of the American Association for Public Opinion Research, Montreal, Canada, May 18-21.

#### **Behavior Coding Real-time Survey Interpreters**

Despite the dramatic growth in the non-English/non-Spanish-speaking population in the United States over the past decade, most surveys are still conducted only in English or Spanish. One approach to expanding the number of languages offered in telephone surveys is to use a thirdparty interpreter, offering real-time translations on the telephone as the survey is conducted. This approach allows the interview to be conducted in a broader range of languages and typically makes more effective use of language specialists than does hiring native speakers as interviewers. While research has shown this to be an effective means of improving survey response among non-English/Spanish speakers, there are still questions about the quality of the translation process given that for most languages the interview is not pre-translated which raises concerns about how real-time translation might affect survey responses (Murray, Battaglia, and Cardoni 2004).

We provide the first detailed assessment of the quality of the real-time interpreter approach, using behavior coding of interviews conducted as part of the 2005 California Behavioral Risk Factor Surveillance System (BRFSS). The interviews were conducted with a set of respondents that otherwise would have been given a final code of "language barrier" and counted as survey nonrespondents. With the permission of the respondents, interviews were recorded and later behavior coded, quantifying for each question administered (1) the accuracy of the question translation, (2) the accuracy of the translated response, (3) the degree of difficulty administering the question, (4) the number of times the question needed to be repeated, and (5) the number of times the interpreter engaged in dialogue with the respondent which was not relayed to the interviewer.

#### Methods

#### Data Collection Procedures

The assessment was conducted in California in three phases: (1) identification of eligible telephone numbers, (2) contacting and interviewing sample members using a real-time interpreter, and (3) behavior coding of the recorded interviews. Telephone numbers were subsampled from the regular, monthly BRFSS sample records if the case was finalized as a "language barrier problem," meaning that no one in the household could be reached who spoke English or Spanish (the two languages in which BRFSS is conducted). In order to increase the potential sample size for the pilot, cases initially sampled from January 2005 onward were considered eligible for inclusion in the pilot even though interviewing did not begin until September 2005. Telephone numbers sampled from January through August were considered "retrospective" cases for the purposes of the analysis presented here, because contacts via the interpreter were initiated a month or more after the last BRFSS call attempt. Those numbers sampled September through December were considered "concurrent" cases as contact attempts via the interpreter with these cases were generally initiated within a month of the last BRFSS call attempt. Although the pilot study is on-going through June 2006, our analysis focuses only on those cases which were initially sampled for BRFSS during 2005.

Contacting and interviewing the language follow-up cases began September 1, 2005. Interviewers from the same survey research group that conducts the California BRFSS initially contacted all retrospective telephone numbers to determine as best as possible the language spoken in the household. This was accomplished by contacting the households and attempting to obtain an answer to one of the following questions: 1) What language do you speak? or 2) What country are you from? It was anticipated that many non-English speakers would still be able to answer one of these questions, even when posed in English. To initiate follow-up contact,

interviewers first contacted the interpreter service to obtain the assistance of an interpreter fluent in the language thought to be spoken in the household. Next a three-way call was established between the interviewer, the interpreter, and the sampled telephone number. In situations where the language could not be determined during the initial contact, or was different than initially indicated, a language specialist at the interpreter service came onto the line to assist the interviewer in identifying the correct language and accessing an appropriate interpreter. For concurrent cases, the BRFSS interviewers asked the same questions and entered the likely language into case notes for use by the language follow-up survey interviewers. Contacting and interviewing, for numbers resulting in an eligible household, then proceeded with the interviewer administering the survey and the language specialist providing interpretation of the question for the sample member and the response for the interviewer. With permission from the sample member and interpreter, the interviews were recorded for later processing.

Because the fielding period extended over two calendar years, with the bulk of the concurrent interviews occurring in 2006, the September 1, 2006 BRFSS core questionnaire was used for all interviews (the final 2006 BRFSS core questionnaire differed somewhat from the version used here in terms of question content and placement). This was logistically easier and less expensive than using different questionnaires for different years. The BRFSS core questionnaire consists of approximately 75 items, which are asked in all 50 states, focusing on health conditions and risk behaviors associated with morbidity and death. Approximately two-thirds of the content of the core questionnaire remains identical from year to year, with the remaining one-third of questions alternating even and odd years. Further details on BRFSS survey design, methodology, and questionnaire are available elsewhere (Mokdad, Stroup, and Giles 2003) and at http://www.cdc.gov/brfss.

Next, the taped interviews were behavior coded by trained language specialists from an outside language service vendor (not the vendor providing the interview interpreters). The coders were fluent in the language in which the interview was conducted and were trained on the procedures for behavior coding. The coders assessed administration of each question on the following five attributes:

- Was the question interpreted accurately? (1 = least accurate, 4 = most accurate);
- Was the response relayed accurately? (1 = least accurate, 4 = most accurate);
- Were there concepts in the question that appeared to be difficult to translate accurately? (yes/no);
- How many times did the question need to be repeated to the sample member? (number of times);
- Were there side conversations between the interpreter and the respondent which were not translated for the interviewer?(yes/no).

#### Statistical analysis

The final dataset contained the following information: (1) call history information from the original BRFSS survey; (2) call history and case disposition information from the language follow-up; (3) questionnaire responses for the completed questionnaires; and (4) the behavior coding assessments for each question. Because data collection is on-going, the analysis presented here should be considered preliminary. Only cases originally sampled for the 2005 California BRFSS are included.

The analysis focused on four key areas (final case dispositions, demographic characteristics, question-level and interview-level assessments of the quality of the

interpretations, and survey estimates) with comparisons made, when appropriate, to the 2005 California BRFSS survey results.

First, we compared response rates and final case outcomes between the language followup and the 2005 BRFSS. Response rates were calculated using response rate formula #4 recommended by the American Association for Public Opinion Research (AAPOR 2004).

Second, demographic characteristics of the respondents were compared between the 2005 California BRFSS and the language follow-up. To adjust for survey design, both sets of cases were weighted inversely by the number of landline telephones in the household and the number of adults in the household.

Third, we assessed the quality of interpretations at both the question level and the interview level using the five behavior coding attributes. Each attribute was recoded to form five dichotomous quality indicators. Questions which were graded as 1 or 2 for the interpretation accuracy and response accuracy measures were respectively coded 1, indicating there was a problem with interpretation of that question. Questions where the coder indicated that there were concepts which were difficult to translate accurately and where there were side conversations between the interpreter and respondent were each coded as 1. Finally, if the question was repeated one or more times, the question was coded as 1 in terms of the need for the question to be repeated. At the question level, we calculated for each of the five dichotomous indicators the proportion of responses to each question where an error or problem with interpretation was noted. In effect, these measures showed the number of times we found an error or problem with interpretation administration for every 100 times the question was administered. To determine if the errors or problems identified were related to the types of questions being asked, we compared these measures in terms of position of the question in the questionnaire (first third,

middle third, or final third), if the question was a primary question (or "gate" question) asked of everyone or a follow-up question asked only of a subset of those responding to the primary question, and the type of response options provided in the question (yes/no, categorical, numeric, or Likert scale). Significance was assessed by comparing proportions, using an F-test of means.

To assess quality at the interview level, we calculated the proportion of questions in each interview where an error or problem was indicated. This was calculated for each of the five measures individually, thereby giving us interview-level scores for each measure. We then compared how these measures varied across different types of respondents (in terms of sex, age, education, family income, and language spoken). Significance was assessed by comparing proportions, using an F-test of means.

Fourth, we used self-reports of survey participants to assess the prevalence of eight key health and risk behavior questions. Health care coverage was determined by asking: "Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or governmental plans such as Medicare?" Asthma and diabetes were assessed by asking participants, "Have you ever been told by a doctor, nurse, or other health professional that you have [condition]?" Obesity was assessed on the basis of respondents' body mass index (BMI), which was calculated from their self-reported height and weight, with respondents classified as obese if their BMI was  $\geq 30 \text{ kg/m}^2$ . Respondents were classified as "current smokers" if they reported currently smoking every day or some days; they were considered "binge drinkers" if they reported having consumed five or more drinks at least once during the preceding 30 days; they were considered to have been tested for HIV if they responded "yes" to the question, "Have you ever been tested for HIV?"; and, they were considered to have engaged in behaviors linked to the transmission of HIV if they indicated that they had, within the previous year, used

intravenous drugs, been treated for a sexually transmitted or venereal disease, given or received money or drugs in exchange for sex, or had anal sex without a condom. The prevalence estimates from the language follow-up were compared with those from the 2005 California BRFSS, using Chi-squared test of significance. Like the demographic comparisons, these data were weighted to adjust for the survey design.

All analyses were conducted using SPSS version 13, with Complex Samples (SPSS, Inc., 2004).

#### Results

The preliminary dataset included 736 cases finalized in the 2005 California BRFSS as having a language barrier and treated as nonrespondents. Completed interviews were obtained with 171 of these respondents, of which 126 had been behavior coded at the time of analysis. Assessment of participation rates are based on the entire set of cases, while subsequent analyses of demographics, interpretation quality, and survey estimates are limited to the cases for which both a complete interview and complete behavior coding were available. A final report will be produced once all data collection and behavior coding has been completed in June 2006.

## Response Rates

Approximately one-fourth (23.2%) of the 736 telephone number identified for the followup resulted in a completed interview (Table 1). This percentage was considerably higher among the concurrent cases (28.0%) compared to the retrospective cases (19.6%). Conversely, there were nearly three times more cases deemed ineligible among the retrospective cases (18.6%) compared to the concurrent cases (6.8%), with a large majority of these telephone numbers being no longer in service. Overall, the response rate for the language pilot was 30.4%.

The response rate for the 2005 California BRFSS was 27.7%. Using the final case distribution for the language follow-up cases that were called concurrently with the BRFSS survey, we estimated that if the interpreter approach had been used throughout the year the 2005 California BRFSS response rate would have increased by less than 1 percent to 28.5%.

#### Demographic characteristics of respondents

Among the initial 126 language follow-up respondents, interviews were conducted in 18 different languages (Table 2). Asian languages predominated with Vietnamese, Korean, Cantonese, and Mandarin being the most prominent, comprising more than 60% of the interviews. Interviews were conducted in Russian approximately 12% of the time.

Respondents to the language follow-up also differed from the larger set of California BRFSS respondents significantly in several demographic characteristics (Table 3). More than three-fourths (75.6%) of respondents were Asian, compared to 6.1% in the BRFSS. Language follow-up respondents were also more likely to be aged 65 or older (37.3% versus 15.2%), to have a high school or less education (62.8% versus 40.4%), to have a household income below \$25,000 (66.3% versus 30.9%), and to live in households with three or more adults (45.2% versus 33.9%).

#### Quality of Interpretation

We examined the quality of the interpretation and respondent-interpreter-interviewer interaction at both the question level and the interview level. First, focusing on the question-level assessment, we examined the proportion of errors or problems associated with each of the five quality measures for each of the 75 questions.

On average, problems were found with the initial interpretation 4% of the time a question was administered (Figure 1). Problems were most prevalent for the questions "How often do you

get the social and emotional support you need?" (12.5%), "Has a doctor, nurse, or other health professional ever told you that you had angina or coronary heart disease?" (11.3%), and "A Pap test is a test for cancer of the cervix. Have you ever had a Pap test?" (10.9%). No interpretation errors were reported for 6 of the 75 questions.

Errors in relaying the response back to the interviewer were less prevalent, occurring, on average 1.3% of the time (Figure 2). Error rates were highest for questions on binge drinking, "Considering all types of alcoholic beverages, how many times during the past 30 days did you have [five (for men) / four (for women)] or more drinks on an occasion?" (8.2%), and race, "Which one of these groups would you say best represents your race?" [asked of those who indicated in an initial question that they were of multiple races] (6.5%). For 45% of the questions (34 of 75), there were no problems found in relaying the response.

Conceptual problems making translation difficult occurred, on average, 4.3% of the time (Figure 3). This was most problematic for the question on social and emotional support (13.5%) and angina and cardiovascular disease (11.9%), as well as a question asking "About how long has it been since you last visited a doctor for a routine checkup?" (10.3%). Only two questions had no reported conceptual problems.

Questions needed to be repeated, on average, 11.1% of the time, with repeat rates of 10% or higher for 33 of the 75 questions (Figure 4). Among the most often repeated were the questions on social and emotional support (38.4%), binge drinking (28.0%), and last routine checkup (27.4%). Another alcohol consumption question was also repeated quite often: "One drink is equivalent to a 12-once beer, a 5-ounce glass of wine, or a drink with one shot of liquor. During the past 30 days, on the days when you drank, about how many drinks did you drink on

average?" (30.2%). Only one question did not require that it be repeated for any of the respondents.

Side conversations between the interpreter and the respondent which were not translated for the interviewer occurred 4.5% of the time (Figure 5). Side conversations were most likely to occur during administration of the questions on last routine checkup (15.3%) and social and emotional support (15.2%). Side conversations did not occur during administration of 11 of the 75 questions.

We also assessed the degree to which question placement within the questionnaire, the type of question, and the response format might have influenced the proportions on these five quality measures (Table 4). Errors in interpreting questions for the respondent did not appear to be significantly related to question placement, question type, or response format. Errors in relaying responses to the interviewer were higher for questions with categorical responses (such as type of employment or marital status) than they were for other types of response formats. Primary questions, that is, those asked of all respondent or those which serve as "gate questions" to a set of follow-up questions, were more likely to have concepts that were difficult to translate than were follow-up questions (5.7% versus 2.7%). Similarly, questions administered during the first two-thirds of the interview, were more likely to have conceptual problems that made translation difficult compared to questions in the final third of the interview. Questions with a "yes/no" format were less likely than other response formats to require that the question be repeated and less likely to generate side conversations between the interpreter and the respondent. Conversely, primary questions were more likely than follow-up questions to stimulate side conversations.

Next we examined the quality of the interpretation approach looking at the interviewlevel. For each of the five measures individually the proportion of questions within an interview that resulted in errors or problems was calculated for each respondent. These interview-level proportions were then compared across a number of different demographic groups (Table 5). The language spoken appears to have had the greatest impact, with significant variation found on all five measures. Respondents speaking less prominent languages (that is, a language other than the five most prevalent languages) were more likely to have interviews with errors in interpretation of the question and of the response. They also had a significantly higher proportion of questions with conceptual issues making it more difficult to interpret the question. Russian-speaking respondents were the group most likely to have questions repeated and to engage in side conversations with the interpreter. Among the other demographic characteristics examined, women were more likely than men to experience problems with interpretation and to have questions repeated. Respondents aged 70 or older were more likely than younger respondents to engage in side conversations with the interpreter. Neither education nor income was significantly related to any of the five quality measures.

#### **Difference in Survey Estimates**

Finally, we made comparisons between the language follow-up and the 2005 California BRFSS across eight key health and risk behavior variables (Table 6). Language follow-up respondents were more likely to report having had an influenza shot in the past 12 months (47.0% versus 24.6%), and less likely to be obese (10.0% versus 19.6%) or to have ever been tested for HIV (18.5% versus 42.6%).

#### Discussion

Since most health surveys of the general population in the United States are conducted in English, or in some cases Spanish as well, persons who speak a language other than these two predominant languages are often under-represented or not represented at all by these surveys. As a result the health risks and problems that they face may be inadequately described in public health statistics. The potential for such problems has increased over the past several decades as the U.S. population has grown more diversified. As of 2002, the U.S. Census reported that 11.7% of U.S. residents were foreign-born, with 53.3% of those being born in Latin America, 25.0% in Asia, 13.7% in Europe, and 8.0% in some other region of the world (Larsen 2004). The foreign born account for more than a quarter (26%) of the population of California (Malone et al 2003). There has also been a corresponding growth in the percentage of U.S. residents who primarily speak a language other than English. According to the 2000 census, 47.0 million (18%) of the 262.4 million people aged 5 years or older spoke a language other than English at home (Shin and Bruno 2003). In 2000, approximately 4.5% of the U.S. population could be considered "linguistically isolated," meaning all members of the household aged 14 years or older speak a non-English language and also speak English less than "very well" (i.e., have difficulty with English) (US Department of Commerce 2004). Among certain subpopulations, the percentage of people who said they speak English less than "very well" is quite high: 51% of those who speak primarily an Asian or Pacific Island language, 49% of those who speak Spanish, and 34% of those who speak another Indo-European language (Shin and Bruno 2003). Although the percentage of such individuals may still be modest on a national scale, non-English/Spanishspeakers often comprise a significant percentage of the population in local areas.

Real-time interpretation may be an effective technique for including these individuals in survey research efforts. As this pilot has demonstrated, we were able to complete interviews with just over 25% of the cases initially finalized in the California BRFSS as "nonrespondent, language barrier" cases. These individuals would have been classified as nonrespondents due to a survey design limitation (not offering the survey in the respondent's language) rather than to their unwillingness to participate. Moreover, these individuals had a very different demographic profile than those typically interviewed as part of the BRFSS in California. Language follow-up respondents were much more likely to be Asian, older, of lower socioeconomic status, and live in households with three or more adults. Given that nonresponse bias is a product of the level of nonresponse and the degree to which respondents and nonrespondents differ, the real-time interpreter approach appears to have reduced the potential for nonresponse bias by addressing both issues - reducing the level of nonresponse and improving participation among respondents with characteristics different from the original pool of respondents. Further analyses of the larger, final data set will be required, however, to determine if some of the differences noted in survey estimates are age-dependent or the product of real differences between respondents to the regular BRFSS and respondents to the language follow-up.

In terms of quality assessment, the real-time interpretation approach appears to produce favorable results as well. The overall percentage of error in administering the questions appears modest at 4%, while error in interpretation of the responses were much lower at just over 1%. These error rates were even lower for the more prevalent languages (Vietnamese, Korean, Cantonese, Mandarin, and Russian). Error rates varied considerably by question type, placement, content, and response format. Questions which required repeating and/or which stimulated additional side conversation between the sample member and the interpreter (without including

the interviewer) also tended to be those which had apparent conceptual difficulties, making translation more problematic and requiring additional explanation. Cognitive testing and interpreter training could improve the process in both of these areas. Currently the BRFSS questions are cognitively tested only in English. Researchers need, however, to be cognizant of the customs, values, and beliefs of persons in minority communities, particularly because they relate to the sharing of personal information, including health care practices and health conditions (Hilton and Skrutkowski 2002). Focus groups and cognitive interviews of people from various backgrounds can help determine whether respondents will interpret and respond to survey requests and questions as intended (Chang, Chau, and Holroyd 1999; Eyton and Neuwirth 1984). Likewise, additional training of the interpreters stressing the importance of their remaining neutral third-party facilitators in the interview process could help to reduce the number of side conversations in which the interviewer is excluded.

Researchers need to develop survey designs that better address the increasingly complex linguistic mix of the U.S. population. A U.S. Department of Health and Human Services (U.S. DHHS) report recommended that "culturally and linguistically appropriate interviewing techniques need to be employed at all times when conducting surveys on racial and ethnic issues" (U.S. DHHS 1999). The report further recommended that relevant language requirements and cultural factors be incorporated into survey designs when feasible. Real-time interpretation is one approach for expanding the reach of telephone surveys beyond those who speak only English or Spanish, ensuring that the opinions, needs, and behaviors of those who speak other languages are appropriately accounted for in survey statistics.

### References

American Association for Public Opinion Research. Standard definitions: Final dispositions of case codes and outcome rates for surveys (3<sup>rd</sup> ed). Ann Arbor, MI: AAPOR, 2004.

Chang AM, Chau JP, Holroyd E. 1999. "Translation of questionnaires and issues of equivalence." J Adv Nurs; 29(2):316-22.

Eyton J, Neuwirth G. 1984. "Cross-cultural validity: ethnocentrism in health studies with special reference to the Vietnamese." Soc Sci Med;18(5):447-53.

Hilton A, Skrutkowski M. 2002. "Translation instruments into other languages: development and testing processes." Cancer Nursing; 25(1):1-7.

Larsen LJ. 2004. "The foreign-born population in the United States: 2003." Washington, D.C.: U.S. Census Bureau. Report available at: <u>www.census.gov</u>.

Malone N, Baluja KF, Costanzo JM, Davis CJ. 2003. The foreign-born population: 2000 (Census 2000 Brief C2KBR-34). Washington, DC: U.S. Census Bureau. Report available at: <u>www.census.gov</u>.

Mokdad AH, Stroup D, Giles HW. 2003. "Public health surveillance for behavioral risk factors in a changing environment: Recommendations from the behavioral risk factor surveillance team." Morbidity and Mortality Weekly Report 52 (No. RR-9):1-12.

Murray MC, Battaglia M., Cardoni J. 2004. "Enhancing data collection from "other language" households." In Cohen SB, Lepkowski JM, eds., Eighth Conference on Health Survey Research Methods, pp. 109-114. Hyattsville, MD: National Center for Health Statistics.

Shin HB, Bruno R. 2003. "Language use and English-speaking ability: 2000." Washington, D.C.: U.S. Census Bureau. Report available at: <u>www.census.gov</u>.

SPSS Inc. 2004. SPSS Complex Samples 13.0. Chicago, IL: SPSS.

US Department of Commerce. 2004. Census 2000 summary file technical documentation. Washington, DC: US Census Bureau. Data and tables available at: <u>www.census.gov</u>.

U.S. Department of Health and Human Services, National Committee on Vital and Health Statistics. 1999. "Improving the collection and use of racial and ethnic data in HHS." Washington, DC: U.S. DHHS. Report available at: www.aspe.hhs.gov/datacncl/racerpt.index.htm.

			Language Pilot		
			Retrospective	Concurrent	
Measures	BRFSS	All Cases	cases	cases	
Response rate <sup>1</sup>	27.7	30.4	27.3	33.8	
% Completed interview	16.9	23.2	19.6	28.0	
% Eligible, non-interview	15.7	4.8	4.6	5.0	
% Unknown eligibility	46.5	58.6	57.2	60.2	
% Ineligible	20.9	13.5	18.6	6.8	
(n)	(36.225)	(736)	(414)	(322)	

## Table 1. Case dispositions and response rates, language follow-up and2005 California BRFSS

(n) (36,225) (736) (414) (322) <sup>1</sup> Response rate calculated using American Association for Public Opinion Research response rate formula #4 (AAPOR 2004). For all language pilot response rates, the percentage of cases with unknown eligibility estimated to be eligible households (often referred to as the "e" factor) was based on case dispositions for concurrent cases only.

Language	n
Vietnamese	27
Korean	20
Cantonese	15
Mandarin	15
Russian	13
Japanese	9
Farsi	5
Tagalog	5
Punjabi	4
Armenian	3
Amharic	2
French	2
Cambodian	1
Hindi	1
Hirudhi	1
Hmong	1
Thai	1
Turkish	1
(Total)	(126)

Table 2. Languages in which follow-up interviews were conducted

	BRFSS		Language Pilot	
	%	95% CI	%	95% CI
Characteristics	(n)		(n)	
Sex				
Male	40.9	39.3, 42.4	47.7	38.5, 57.0
Female	59.1	57.6, 60.7	52.3	43.0, 61.5
[n]	[6,134]		[126]	
Age				
18 - 34	28.7	27.3, 30.2	7.9	4.0, 14.8
34 - 54	32.7	31.3, 34.2	24.0	16.8, 33.1
55 - 64	23.3	22.1, 24.6	30.8	22.6, 40.5
65+	15.2	14.3, 16.2	37.3	29.3, 46.0
[n]	[6,131]		[126]	
Race/ethnicity				
Hispanic	34.9	33.4, 36.5	1	
White, non-Hispanic	53.1	51.6, 54.6	15.4	9.9, 23.2
Black, non-Hispanic	4.4	3.8, 5.1	<b></b> <sup>1</sup>	
Asian	6.1	5.4, 6.9	75.6	67.0, 82.6
Other race/ethnicity	1.5	1.1, 1.9	9.0	5.2, 15.0
[n]	[5,982]		[126]	
Education				
Less than high school	17.9	16.6, 19.3	30.5	23.0, 39.1
High school diploma / GED	22.5	21.2, 23.9	32.3	24.1, 41.7
Some college or more	59.6	58.0, 61.1	37.3	29.1, 46.3
[n]	[6,096]		[126]	
Income				
< \$25,000	30.9	29.3, 32.4	66.3	54.8, 76.2
\$25,000 - \$49,999	24.4	23.1, 25.8	16.3	9.5, 26.7
\$50,000 - \$74,999	15.2	14.1, 16.3	5.6	2.1, 14.1
\$75,999+	29.6	28.2, 31.0	11.7	5.9, 21.9
[n]	[5,628]		[89]	
Number of adults in household				
One	14.2	13.4, 15.0	8.2	5.4, 12.4
Two	51.9	50.3, 53.4	46.6	37.5, 55.9
Three	33.9	32.3, 35.6	45.2	35.8, 54.9
[n]	[6,134]		[126]	
Number of children in household				
None	53.6	52.0, 55.1	57.4	47.7, 66.5
One or more	46.4	44.9, 48.0	42.6	33.5, 52.3
[n]	[6,134]		[126]	·

Table 3. Comparison of Demographic Characteristics of Respondents,language follow-up and 2005 California BRFSS

<sup>1</sup> Included in "other race/ethnicity" category: Hispanics (n = 1), black, non-Hispanic (n = 2).

	Proportion of responses per question with a problem					
			Concepts	Side		
			were			
			difficult	Question	between	
	Error	Error	to	repeated	interpreter	
	interpreting	relaying	translate	one or	and	
Measures	question	response	accurately	more times	respondent	
Mean across all questions	4.0	1.3	4.3	11.1	4.5	
Question position in						
questionnaire:						
Q1 – Q25	3.4	1.3	4.9	11.4	5.2	
Q26 – Q50	3.8	1.6	5.0	12.2	4.8	
Q51 – Q76	4.7	0.9	3.1	9.6	3.6	
(p-value)	(.310)	(.317)	(.026)	(.467)	(.304)	
Primary or follow-up						
question:						
Primary	3.4	1.3	5.7	11.6	5.5	
Follow-up	4.6	1.2	2.7	10.5	3.4	
(p-value)	(.086)	(.742)	(.001)	(.536)	(.019)	
Type of question:						
Yes/no format	3.9	0.9	4.1	7.1	2.7	
Categorical	5.9	3.0	5.6	16.9	7.2	
Numeric	3.7	1.2	4.4	14.9	5.7	
Likert	3.6	1.4	4.2	12.5	5.6	
(p-value)	(.322)	(.017)	(.630)	(<.001)	(.002)	
(n = 75)						

## Table 4. Question-level interpretation quality assessment

	Proportion of questions per interview with a problem					
		Concepts Side				
		were				conversations
				difficult	Question	between
		Error	Error	to	repeated	interpreter
		interpreting	relaying	translate	one or	and
Measures	n	question	response	accurately	more times	respondent
Mean across all interviews	126	3.4	1.3	5.8	11.5	5.4
Sex						
Male	56	1.9	1.2	6.8	8.6	4.6
Female	68	4.6	1.5	5.0	13.8	6.1
(p-value)		(.044)	(.543)	(.356)	(.061)	(.368)
Age (years)						
18-49	37	4.6	1.8	4.5	10.0	2.7
50 - 69	52	1.7	1.0	5.1	10.6	4.0
70 or older	35	4.4	1.4	8.2	14.3	10.2
(p-value)		(.102)	(.467)	(.310)	(.432)	(<.001)
Education		. ,				
Less than high school	42	2.4	1.1	5.2	9.7	3.9
High school diploma/GED	37	4.7	1.4	5.9	12.9	6.9
Some college or more	45	3.2	1.5	6.3	11.9	5.6
(p-value)		(.377)	(.759)	(.891)	(.636)	(.383)
Annual family income						
Less than \$15,000	39	3.7	1.0	6.6	10.6	6.0
\$15,000-\$24,999	23	1.3	1.0	2.5	9.0	5.2
\$25,000 or more	27	4.4	1.6	8.8	12.4	6.5
(p-value)		(.345)	(.581)	(.196)	(.685)	(.911)
Language						
Vietnamese	27	0.5	0.2	2.3	4.0	0.3
Korean	20	1.8	0.1	5.4	15.5	0.6
Cantonese	15	1.3	0.2	3.0	0.8	0.5
Mandarin	15	1.3	0.3	2.2	8.0	4.3
Russian	13	1.0	0.2	7.1	22.1	23.1
Other	36	8.8	4.1	10.6	16.2	7.7
(p-value)		(<.001)	(<.001)	(.031)	(<.001)	(<.001)

## Table 5. Interview-level interpretation quality assessment

Significance based on F-test of means.

	BRFSS		Language Pilot	
Health condition / risk	$\frac{\%}{(n)}$	95% CI	$\frac{\%}{(n)}$	95% CI
Have a health plan	83.7	82 / 85 0	80.7	72 5 86 8
Trave a nearth plan	(6,128)	02.7, 05.0	(120)	72.5, 00.0
Asthma	13.6	12.6, 14.6	8.2	4.1, 15.7
	(6,132)		(124)	
Diabetes	9.1	8.2, 10.0	15.0	9.6, 22.8
	(6,131)		(122)	
Flu shot past 12 months	24.6	23.3, 25.9	47.0	37.7, 56.4
-	(6,129)		(124)	
Obese (BMI $> 30$ )	19.6	18.4, 20.9	10.0	5.6, 17.1
	(5,829)		(115)	
Current smoker	14.4	13.3, 15.6	15.5	9.4, 24.5
	(6,130)	-	(123)	-
Binge drinking	11.9	10.9, 13.1	5.1	2.0, 12.2
	(6,066)		(123)	
Tested for HIV <sup>1</sup>	42.6	40.8, 44.3	18.5	10.8, 29.8
	(4,490)		(66)	
HIV risk behaviors <sup>1</sup>	5.2	4.5, 6.2	2.9	0.7, 11.5
	(4,505)		(70)	

# Table 6. Comparison of prevalence of key health and risk indicators,language follow-up and 2005 California BRFSS

<sup>1</sup> Questions not asked of respondents age 65 years or older



Figure 1. Percent of responses with error in question interpretation



Figure 2. Percent of Responses with an Error Relaying Response

Questions ( n = 75)



Figure 3. Percent of responses with difficult to translate concepts



Figure 4. Percent of responses where questions repeated 1+ times

Questions (n = 75)



Figure 5. Percent of Responses with Side Conversations