Rena J. Pasick, DrPH ■ Susan L. Stewart, PhD

Joyce A. Bird, PhD ■ Carol N. D'Onofrio, DrPH

Quality of Data in Multiethnic Health Surveys

All authors are with the Northern California Cancer Center, Union City. Dr. Pasick is Director, Prevention Sciences. Dr. Stewart is a Biostatistician. Dr. Bird is a Research Scientist. Dr. D'Onofrio is an Adjunct Investigator.

SYNOPSIS

Objective. There has been insufficient research on the influence of ethno-cultural and language differences in public health surveys. Using data from three independent studies, the authors examine methods to assess data quality and to identify causes of problematic survey questions.

Methods. Qualitative and quantitative methods were used in this exploratory study, including secondary analyses of data from three baseline surveys (conducted in English, Spanish, Cantonese, Mandarin, and Vietnamese). Collection of additional data included interviews with investigators and interviewers; observations of item development; focus groups; think-aloud interviews; a test-retest assessment survey; and a pilot test of alternatively worded questions.

Results. The authors identify underlying causes for the 12 most problematic variables in three multiethnic surveys and describe them in terms of ethnic differences in reliability, validity, and cognitive processes (interpretation, memory retrieval, judgment formation, and response editing), and differences with regard to cultural appropriateness and translation problems.

Conclusions. Multiple complex elements affect measurement in a multiethnic survey, many of which are neither readily observed nor understood through standard tests of data quality. Multiethnic survey questions are best evaluated using a variety of quantitative and qualitative methods that reveal different types and causes of problems.

Address correspondence to:
Dr. Pasick, Northern California Cancer Ctr, PO Box 5033, Union City CA 94587-3106; tel. 510-429-2500; fax 510-429-2550; e-mail <Rpasick@nccc.org>.

Throughout the population sciences, extensive research has focused on disparities in health status associated with racial and ethnic diversity and socioeconomic position.[1-7] It is generally agreed that data that accurately reflect the needs and barriers of culturally and linguistically diverse communities are a critical component of programs and policies designed to resolve these inequities. Historically, however, health survey practices in the United States have been based on research with mainstream Anglo Americans. The applicability of such research methods and data to other cultures and languages in the US has yet to be adequately explored and remains poorly understood. Fundamental questions remain unresolved, such as how best to assess the cultural appropriateness of surveys, or how to identify gains and losses from tailoring survey questions by culture rather than developing and applying universal items across ethnic groups. Without a better understanding of these overarching issues, the quality of data produced using mainstream methods among diverse populations should be regarded as questionable. The vigorous exploration of these issues through methodologic research merits priority status in public health.

An extensive research literature covers all aspects of health survey methods, including standards for evaluating the quality of survey questions and resulting data. Problems common to many surveys, such as non-response bias, sample representativeness, and interview bias, have been well documented.[8-13] There exists a growing body of work on the study of the thought processes of respondents in answering survey questions.[14-23] Yet the recent proliferation of multiethnic health surveys, a response to an increasingly diverse population, has not been matched, with some exceptions,[24-34] by commensurate interest or research on the implications of ethnocultural and language differences in measurement. Indeed, all the problems that might be found in any standard English language survey designed for an Anglo sample will also be found in multiethnic and multilingual survey data. However, diversity in cultural meanings, questionable cultural appropriateness of survey items, non-equivalent connotations of translated items, and demographic characteristics of respondents add complications to questionnaire construction that can severely limit the quality of data generated.

Quality of survey questions is usually described in terms of reliability (the reproducibility or stability of survey data), and validity (how well survey items measure what they are intended to measure). In multiethnic surveys, the comparability of data (the extent to which survey items are equally valid and reliable across ethnic and language groups), is just as important. Statistical tests used for these purposes, however, can indicate only the presence or absence of comparability across samples; they cannot shed light on the underlying reasons for the lack of comparability.

After decades of research on attitude surveys, Sudman et al., concluded that the most critical element of response effects is the wording of the question, because the respondent must process the wording first in order to understand what information is being sought.[9] Psychologists pursuing this line of inquiry have identified the cognitive processes associated with problem questions: respondent interpretation of a question (what the item means to the respondent); memory retrieval (recall strategies used by the respondent to arrive at information needed to reply); judgment formation (cognitive processes used by the respondent to arrive at an opinion); and response editing (decision process by respondent on what and how much information to reveal).[35,36] Warnecke, et al., examined the effect of race/ethnicity on these processes in English-language cognitive interviews with African Americans, Puerto Rican Americans, Mexican Americans, and non-Hispanic white Americans. They found that all of these processes, except memory retrieval, were affected by race/ethnicity differences.[35] Cognitive interviews such as this are also known as "think-aloud" interviews in which individual respondents are asked a series of unstructured probes to ascertain how they understood questions and arrived at answers.[9,35,36] Those authors were able to attribute some race/ethnicity effects to broad cultural factors (social desirability, for example), although translation related problems could not be addressed in their study.

Research opportunity. In 1991, the National Cancer Institute (NCI) simultaneously funded three community-based intervention studies to evaluate strategies to increase use of breast and cervical cancer screening using surveys in different ethnic communities in the San Francisco Bay area. At that time, the National Center for Health Statistics (NCHS) issued a request for proposals to improve the quality and quantity of "minority health data." Recognizing a rich context in which to examine the development and quality of survey data across cultures and languages, investigators from the NCI-funded intervention study teams submitted a proposal to NCHS that resulted in the current methodologic study, *Improving Health Surveys for Multiethnic Populations*, a secondary

The applicability to other cultures and languages of health survey practices based on research with mainstream Anglo Americans has yet to be adequately explored and remains poorly understood.

analysis based on the concurrent NCI studies. The NCHS team was composed of investigators from each of the three NCI studies.

Our objectives were to identify survey questions in the three NCI studies that did and did not work well across multiple language and cultural groups, and to explore underlying causes for the problematic questions. An additional aim that emerged during the study was to assess the usefulness of various quantitative and qualitative methods to identify and explain problems that may be unique to multiethnic surveys. This study was necessarily exploratory, as the data available for analyses had been collected within the objectives, structure, and time lines of the three independent studies.

The project has generated numerous findings, with a substantial proportion reflecting problems that might be expected in any survey of any population. Because we are interested in the issues that are unique to multiethnic surveys, we focus here on common validity and reliability problems, but specifically those that are distributed differently across ethnic groups. Where data permit, we explain variations in terms of cognitive processes, and the cultural or translation origins of these problems. Each type of finding represents a progressively deeper level of understanding of the cultural origins of problems.

We report on questions that worked well across ethnic groups and those that did not, and provide examples of how various methods served to identify not only the problems but their underlying themes and causes. Based on these findings, we draw preliminary conclusions about the types of problems that can be anticipated in multiethnic surveys, and we suggest methodologic strategies to assess the quality of data that researchers and evaluators use. We also make recommendations for continued research on this topic.

METHODS

Our *Improving Surveys* study was conducted in two phases. We used standard statistical procedures for assessing reliability and validity, standard qualitative procedures for survey question development and testing, such as focus groups, and emerging methods for improving survey questions (cognitive interviews and interviewer debriefing). In addition, systematic tracking of the development of the survey items in one study, *Pathways to Early Cancer Detection for Four Ethnic Groups (Pathways)*, served as the basis for interviews with the researchers who worked with these items, along with the rating of questions by these investigators.

In Phase I, a combination of qualitative and quantitative techniques identified questions in each survey that presented the greatest difficulty within or across ethnic groups. In Phase II, the most problematic questions were subjected to more intensive and multi-method scrutiny to identify underlying causes.

Three baseline surveys. The survey data sets used in this study originated in the following projects: (1) The *Breast and Cervical Cancer Intervention Study (BACCIS)*, a 5-year controlled intervention trial to improve screening rates through inreach (clinic-based) and outreach (community-based) strategies. Data were provided by baseline household interviews with 1599 African American, Latina, Chinese and white women, ages 40–75, residing in low-income areas of San Francisco and Contra Costa counties; (2) The *Vietnamese Prevention Study (VPS)*, a 5-year controlled intervention trial to improve preventive health practices, including breast and cervical cancer screening, through a media campaign targeting Vietnamese women 18 years or older in Alameda and Santa Clara Counties (Orange and Los Angeles Counties served as the control area). Data were collected through telephone interviews with 933 Vietnamese women; and (3) *Pathways*, a 4-year program project that included three studies of women ages 18–74. Two of the studies were controlled intervention trials to increase breast and cervical cancer screening, one targeting Latina and the other Vietnamese women. A third study was a survey of cancer screening practices among

In Phase I, a combination of qualitative and quantitative techniques identified questions in each survey that presented the greatest difficulty within or across ethnic groups. In Phase II, the most problematic questions were subjected to more intensive and multi-method scrutiny to identify underlying causes.

775 Chinese American women. Baseline surveys were conducted with 1601 Latina and 645 Vietnamese women and, through supplemental funding, parallel surveys were conducted with 602 African American and 605 white women. A common set of core questions was developed by the *Pathways* team and administered by telephone to samples of Latina, Chinese, African American and white women, and in household interviews to samples of Vietnamese women.

Of the 4228 interviews completed in the *Pathways* surveys, 3583 were conducted by telephone and 645 were in person. All interviews with African American and white women were conducted in English, and all interviews with Vietnamese women were conducted in the Vietnamese language. Chinese women were interviewed in their choice of Mandarin, Cantonese or English. and Latina respondents could answer in Spanish or English. The methods used in these studies, including translation problems and procedures and findings from the baseline surveys, have been reported elsewhere.[37-43]

PHASE I

Identification of problematic survey questions. Either a specific question or a group of several related questions can be considered one variable.

We used a combination of quantitative and qualitative techniques to assess problems with survey questions. We applied several approaches to allow for crosschecking and corroboration of findings. For example, we used qualitative methods to re-examine questions with high rates of missing data, and quantitative analyses to test variables identified qualitatively as difficult to measure. In addition, some methods identified problems that were not picked up by others. The quantitative methods included descriptive and analytic statistics and numerical ratings of question difficulty by researchers and interviewers. We used interviews with researchers and interviewers and

participant observations of item development to identify problems qualitatively.

Statistical analyses. For each of the three baseline surveys, we calculated frequencies and means, and examined bivariate and multivariate relationships using cross-tabulations, factor analysis, or logistic regression as appropriate. We referred variables with higher than average rates of missing data and those that did not behave as expected in analysis to the *Improving Surveys* team as potentially problematic. For the methodological study, we further analyzed data from the *BACCIS* and *Pathways* surveys by ethnicity and language of interview to identify variations in missing data rates, variable correlations, and factors extracted from scalable items.

Rating of questions. Key researchers and staff (those who had been involved in developing, translating and back-translating items) from each of the three survey projects were asked to independently rate the measurement difficulty of all questionnaire items in their survey. We requested ratings on three factors using a 5-point Likert-type scale: ease in wording, accuracy of variable measurement, and cultural appropriateness. Interviewers were also asked to use a 5-point scale to rate their projects' survey questions on aspects of survey administration: ease or difficulty in asking the question, perceived ease or difficulty with which respondents answered, and cultural appropriateness. To identify the most problematic variables, we calculated frequencies and mean scores for each item on all six factors.

Researcher, translator, and interviewer interviews. Upon completion of the rating process, we initiated the first phase of qualitative assessments by conducting two-hour individual interviews with each investigator-rater to discuss items that had been rated as difficult. Also, debriefing sessions[44] of two or more hours duration were held

separately for each team of survey workers. In these group meetings, accounts of workers' experiences and observations in the course of interviewing elucidated their ratings of questions that had been consistently assessed as difficult. Six tape-recorded sessions were held, involving a total of 55 interviewers.

Participant observation of item development. Because the baseline survey for *Pathways* was being developed during the first phase of this methodological study, the *Pathways* development process was studied through participant observation. Reasons for adding, dropping, modifying, or re-ordering particular questions were carefully recorded as the multiethnic, multi-lingual *Pathways* team met over a six-month period to draft the survey questionnaire, report difficulties encountered in translation and pretesting, and to make needed revisions. Questions that evoked the most discussion and compromise were regarded as potentially troublesome.

Findings from Phase I were documented question by question for each survey. Question summaries included missing data rates by ethnicity, notes about other issues encountered in quantitative analyses, investigator and interviewer ratings, and comments made during interviews with survey raters or at *Pathways* meetings. Variations in the length of question summaries proved to be good indicators of relative level of difficulty, as later confirmed from other data sources and methods. The questions presenting the most difficulty in each survey were singled out and systematically compared to data on questions measuring the same or similar variables in the two other surveys. Often, it was evident that the problem with a given question was one of unclear wording since similar but differently worded items in another survey did not present problems. Other items, however, appeared consistently to present different issues for different ethnic groups or to involve multiple sources of difficulty. These were flagged for more intensive study in Phase II. At the end of Phase I, a short list of the 12 most problematic questions had been identified.

PHASE II

Identification of causes of problems. The second phase of study concentrated on the short list of questions assessed as most difficult by both the quantitative and first-phase qualitative methods. This phase included a test-retest assessment survey of *Pathways* respondents, focus group discussions, think-aloud interviews, and a pilot test of alternative forms of questions.

Test-retest/assessment survey. An overlap in timing with the *Pathways* survey made it possible to call back either a 10% random sample or a minimum of 75 respondents from each of three ethnic groups (African American, Chinese, and Latina) and the comparison sample of white women within five days after completing the baseline *Pathways* survey. Vietnamese respondents completed baseline interviews in person and were called back within two weeks. We took advantage of this timing to further explore difficult items and to corroborate findings from interviewer de-briefings.

Focus groups and think-aloud interviews. Eight bilingual, bicultural graduate student research assistants (RAs), two from each ethnic group targeted in this study (African American, Chinese, Latino, and Vietnamese), were recruited and intensively trained to collect qualitative and survey data. They were graduate students in health sciences, recruited by the research team through campus fliers. Training for all methods was conducted jointly by the research team and representatives of a private research and consulting firm that specialized in survey research and focus group training and implementation. Training for focus groups consisted of eight 4-hour classroom sessions designed specifically for this study. The training included one practice focus group conducted by each team of two research assistants in a focus group facility, observed by the entire study team. Additional trainings included two 4-hour sessions on basic interviewing techniques and telephone interviewing, and two 4-hour sessions on think-aloud interviewing. Debriefing and procedural meetings also were conducted throughout this process on an as-needed basis.

Each team of two RAs conducted three ethnic-specific focus groups with low-income women in sites around the San Francisco Bay area. Each RA also conducted 15 think-aloud interviews with low-income, monolingual women from her respective ethnic group, for a total of 120 think-aloud interviews. Since our principal aim was to explore problem questions among populations other than English-speaking white women, the focus groups, think-aloud interviews, and alternative forms tests were performed only among the target groups. Funding constraints in this exploratory study precluded extending these methods to white women for comparison purposes.

Pilot test of alternative question forms. Based on data from the methodological analyses, we developed alternative wording for the most problematic questions. These were pilot-tested by the RAs in brief telephone surveys. Trans-

lated, back-translated, reconciled versions of these questions were asked in English, Cantonese, Spanish, and Vietnamese by telephone in a convenience sample of 120 women (30 from each ethnic group). At the end of each survey, interviewers were asked to assess the questions by indicating if respondents hesitated in answering any question, asked for clarification, or seemed confused with any question. Interviewers were also asked to note whether any questions felt awkward to ask or if the respondent seemed uncomfortable or dishonest in answering any questions.

In all instances of data collection, including baseline surveys, focus group, test-retest, think-aloud, and alternative forms interviews, participants were informed by the interviewers of their rights as research participants. All study protocols were approved for the protection of human subjects by the Northern California Cancer Center, or the University of California (Berkeley and San Francisco) Institutional Review Boards.

RESULTS

We report the criteria used to distinguish between problematic and non-problematic questions, followed by a list of three concepts and the associated 12 questions that were identified as problematic. Frequencies or means and standard deviations are shown for all troublesome questions. To demonstrate the interplay and relative contribution of qualitative and quantitative data, we provide detailed findings on questions of household size, routine health checkups, and fatalism. Where applicable, we also present results from tests of alternative questions developed through a combination of analyses.

Questions that work well. The types of survey questions that appear to work well across cultures and languages— that were not rated as difficult by researchers and interviewers, did not result in significant missing data, and did not yield unexpected frequencies or patterns of association—were clearly defined concepts that asked for factual information in clear and simple language.

For example, interviewers indicated that some non-English speaking women needed the definition of mammogram that was provided in the question, "Have you ever had a mammogram?" However, once defined, it was quite easy for all women to determine whether or not they ever had this test.

Problematic questions. Conversely, survey questions emerged as troublesome for the following reasons:

- the concept or the original wording was unclear
- the translation was difficult
- the question did not allow for cultural differences in meaning
- the concept or wording was culturally inappropriate
- the memory recall or judgment required of the respondent was too difficult
- the request for sensitive information led to untruthful responses

Many of these problems contributed to poor reliability or validity. Interviewers rated as very troublesome a question about the number of respondent's visits to the doctor in the past 12 months, a question that involved both recall and interpretation of what constitutes a doctor and a visit. To illustrate the complementary nature of the qualitative and quantitative methods in identifying troublesome questions, Table 1 shows the level of agreement between the *Pathways* baseline results and the test-retest survey for two questions that required respondent recall. These results support interviewer assessments for the questions about ever having had a mammogram and the number of doctor visits during the past 12 months.

The survey questions identified as most problematic, warranting further exploration through multiple methods, were operational measures of the following concepts: (a) sociodemographics (education, self-identification of "race" and ethnicity, household size, income, and health insurance status); (b) preventive behaviors (receipt of routine checkups including number of doctor visits in the past 12 months, respondent had heard of or had a Pap test, heard of a mammogram, performed breast self-examination); and (c) attitudes and beliefs, especially fatalism (measured by several questions, such as the belief that one can have cancer without knowing it and that there is a chance of surviving breast cancer if detected early).

Tables 2, 3, and 4 show the distribution of all problematic questions by survey and by ethnicity, and provide the context for all the findings that follow. Demographic differences, evident from survey to survey, are due primarily to variations in sampling. For example, part of the *Pathways* sample of Chinese women was drawn from a health plan list that included women with higher levels of insurance coverage, income, and education than does the sample of Chinese women for *BACCIS*, which was a randomized household survey with mainly low-income, uninsured, non-English speaking respondents. For this reason, demographic comparisons can be made only within, rather than across, studies. The preventive behaviors shown in Table 3 are consistent with these sociodemo-

Table 1. Test-retest reliability: agreement of responses between the *Pathways* baseline and callback surveys, by race/ethnicity, for two questions (N = 467)

Self-reported race/ethnicity	Number called back	Have you ever had a mammogram?			In the past 12 months, how many times have you been to see a doctor?		
		n	Kappa	95% CI	n	ICC	95% CI
African American.............	75	73	0.93	0.72, 0.99	72	0.80	0.70, 0.87
Chinese.....................	84	84	0.90	0.74, 0.97	82	0.65	0.51, 0.76
Latina......................	160	160	0.90	0.79, 0.95	152	0.80	0.73, 0.85
Vietnamese.................	74	74	0.74	0.51, 0.88	69	0.56	0.37, 0.70
White......................	74	74	1.00	0.86, 1.00	73	0.90	0.84, 0.93

CI = confidence interval
ICC = intra-class correlation coefficient

graphic differences, with higher income samples showing higher levels of prevention.

Nevertheless, ethnic differences are conspicuous. The extent to which there is consistency within populations across studies can be an indicator of consistency in both the behavior and its measurement. Differences, however, are more difficult to interpret. There was considerable variation by question and by ethnicity in the proportion of respondents who agreed with the attitudinal statements shown in Table 4. For example, agreement with the statement "There is little one can do to prevent getting cancer" ranged from a low of 28% among white women to 54% among Latina women in the *BACCIS* survey. In addition, respondents generally were more likely to refuse to answer these questions than demographic and screening questions. All within-study differences by ethnicity shown in Tables 2 through 4 are statistically significant ($P = 0.001$).

Sources of difficulty. The most problematic questions as listed above represent those explored in greatest depth. All were discussed in focus groups and pilot-tested in alternative forms. Three were explored also in think-aloud interviews: receipt of routine check-up, chance of surviving cancer, and education. Interviews with investigators and interviewers, and the statistical analyses, generated data on many other items as well. We use examples of results from all sources to illustrate the range of items affected by problems that are specific to multiethnic surveys.

Household size. The following questions were used to measure household size in the three studies. This variable

is used in combination with data on income to determine a respondent's socioeconomic or poverty level status.

Questions from *BACCIS*:
Q52A. **Do you live alone or with others? (If alone, SKIP to QUESTION 54).**
Q52B. **Please indicate how each of the people you live with are related to you (husband, daughter, son, boarder, nephew, friend), their sex, and their age (at their last birthday). You don't need to tell me their names.**
Question from *VPS*:
Q107. **How many people in your household share meals together? (Note: If some people in the household prepare meals separately from the meals of the respondent, they are in another household unit.)**
Question from *Pathways*:
Q99. **In total, how many people live in your household?**

Although missing data were low for the above questions, interviewer feedback indicated that these items posed difficulties. The process of enumerating all members with whom the respondent lived was time-consuming, laborious and, to some respondents, overly personal. While the *VPS* question apparently worked well for the Vietnamese women, *Pathways* investigators considered it inappropriate for other ethnic groups. In many African American and white households, for example, members of a household may have been too busy to share meals. As a result, the *Pathways* question is subject to different

Table 2. Distribution of selected sociodemographic characteristics in three studies, by ethnicity (N = 6760)

	Study	African American n = 602 (P) n = 475 (B) n	Stat	Chinese n = 775 (P) n = 279 (B) n	Stat	Latina n = 1601 (P) n = 230 (B) n	Stat	Vietnamese n = 645 (P) n = 933 (V) n	Stat	White n = 605 (P) n = 500 (B) n	Stat	Other race/ ethnicity n = 115 (B) n	Stat
Household income													
>$20K (percent)....	P	503	50	632	60	1253	37	602	11	501	78		
≥$20K (percent)....	B	407	31	220	39	162	45			450	68	80	45
Above FPL (percent)....	P	541	81	701	81	1399	49	615	31	534	93		
Above FPL (percent)....	V							832	66				
Health insurance													
Insured (percent)....	P	591	91	755	85	1598	66	643	74	597	89		
Insured (percent)....	B	475	87	279	66	230	74			500	89	114	77
Insured (percent)....	V							929	85				
Education level													
Years, mean (± SD)......	P	596	13.4 (± 2.8)	773	11.3 (± 5.0)	1601	9.4 (± 4.3)	644	7.6 (± 4.8)	604	15.3 (± 2.7)		
Range of years.......			1–22		0–28		0–20		0–23		0–23		
Years, mean... (± SD)......	B	473	12.1 (± 2.9)	275	8.1 (± 5.2)	229	9.6 (± 4.9)			498	14.1 (± 2.9)	112	12.8 (± 4.0)
Range of years.......			0–21		0–20		0–22				0–22		2–21
High school graduate or higher (percent)....	V							923	45				

interpretations about the meaning of *household*. Interviewer feedback and focus group discussions revealed considerable confusion about who should be counted... sons and daughters who don't live at home, but contribute money? Older children who live at home between jobs? People who stay with the family for a while and other temporary residents? Unrelated persons or just family members? Also, respondents often forgot to count themselves as household members.

Interviewer debriefings and focus groups identified the nature of problems around household size as twofold. First, there was considerable distrust of the survey

process and concern for the safety of household members that differed by ethnic group; and second, there was confusion over whom to include in the count. Given that few data were missing, and based on respondent problems, interviewers suggested that some respondents probably had provided untruthful or inaccurate responses.

In the Latina focus groups it was reported that the question about household size would be answered honestly in one's homeland but not in the US, due to fear of landlords and immigration officials. According to interviewers in all studies, the household size question

Table 2. (continued)

	Study	African American n = 602 (P) n = 475 (B) n	Stat	Chinese n = 775 (P) n = 279 (B) n	Stat	Latina n = 1601 (P) n = 230 (B) n	Stat	Vietnamese n = 645 (P) n = 933 (V) n	Stat	White n = 605 (P) n = 500 (B) n	Stat	Other race/ ethnicity n = 115 (B) n	Stat
Household size													
Mean	P	593	2.4	767	3.8	1599	4.0	644	4.4	602	2.4		
(± SD)......			(± 1.4)		(± 1.8)		(± 2.0)		(± 1.9)		(± 1.2)		
Range........			1–10		1–12		1–15		1–12		1–8		
Mean	B	470	2.5	279	3.1	228	3.2			500	2.3	115	3.5
(± SD)......			(± 1.6)		(± 1.7)		(± 1.6)				(± 1.3)		(± 2.1)
Range........			1–11		1–11		1–8				1–9		1–11
Mean	V							933	4.4				
(± SD)......									(± 2.0)				
Range........									1–14				
Visits to doctor													
Mean	P	587	6.8	760	4.0	1574	4.8	627	5.2	597	5.3		
(± SD)......			(± 10.0)		(± 4.6)		(± 7.0)		(± 5.1)		(± 8.3)		
Range of visits			0–104		0–30		0–96		0–60		0–100		

NOTE: p = 0.001 for within-study ethnic differences in frequencies (by chi-square tests) and means (by ANOVA) for tabulated variables

P = Pathways, conducted in 1993–1994
B = Breast and Cervical Cancer Intervention Study (BACCIS), conducted in 1992–1993
V = The Vietnamese Prevention Study (VPS), conducted in 1992
FPL = federal poverty level
SD = standard deviation
Stat = summary statistic(s): percent, mean ± standard deviation, or range
Range = minimum-maximum

touched on concerns regarding eligibility for public assistance, limitations on the number of occupants allowed in a residence, and other legal issues.

Of Latina and Vietnamese women, 14–15% thought that most people would not truthfully answer questions about household size (Table 5). Almost one-fourth (23%) of the Vietnamese women who were called back refused to answer this question. Upon further inspection of these data, we found that women interviewed in languages other than English were more likely to report that most people would not answer questions about household size truthfully. These findings suggest that non-English speaking women may be concerned about survey questions on household size because of their or a family member's immigration status. While immigration status *per se* is not a "cultural" factor, it is highly associated with immigrants of specific cultural origins and may easily be overlooked in questionnaire preparation as a potential source of problems.

Analyses of the callback data showed no association between reported household size and attitudes toward the question. The proportion with household size of five or more was greatest among Vietnamese women (46%), followed by Latinas interviewed in Spanish (39%), Chinese interviewed in Chinese (32%), Latinas (27%), Chinese (18%), African Americans (9%) and white women (6%) interviewed in English. Based on the qualitative data, it is possible that the actual household sizes of immigrant families are larger than reported.

Cognitive problems of interpretation (respondent's understanding of what is meant by "members of the household" and "sharing of meals," for example) were a major source of difficulty for questions on household size,

Table 3. Distribution of selected reported preventive behaviors in three studies, by ethnicity (N = 6760)

Survey questions	Study	African American n = 602 (P) n = 475 (B)		Chinese n = 775 (P) n = 279 (B)		Latina n = 1601 (P) n = 230 (B)		Vietnamese n = 645 (P) n = 933 (V)		White n = 605 (P) n = 500 (B)		Other race/ ethnicity n = 115 (B)	
		n	Percent	n	Percent	n	Percent	n	Percent	n	Percent	n	Percent
Checkups													
Goes when well	P	602	74	775	65	1601	53	645	41	605	74		
Past 12 months	B	475	85	279	52	230	72			500	76	115	73
Past 12 months	V							933	65				
Heard of Pap													
Yes...........	P	602	99	775	62	1601	81	644	28	605	100		
Yes...........	B	475	99	279	42	230	86			500	98	115	92
Yes...........	V							933	60				
Had Pap													
Ever	P	602	98	775	67	1601	76	641	42	605	99		
Ever	B	475	97	279	54	230	97			500	97	115	87
Ever	V							933	63				
Heard of mammogram													
Yes...........	P	602	97	775	82	1601	89	644	59	605	98		
Yes...........	B	473	98	279	64	227	90			500	100	115	89
Yes...........	V							932	66				
Did BSE													
Ever	P	602	93	775	70	1601	82	644	34	605	93		
Ever	B	474	94	278	63	229	85			497	92	115	83
(Not asked)	V												

NOTE: p = 0.001 for within-study ethnic differences in distribution of responses for tabulated variables, by chi-square tests

P = Pathways, conducted in 1993–1994
B = Breast and Cervical Cancer Intervention Study (BACCIS), conducted in 1992–1993
V = The Vietnamese Prevention Study (VPS), conducted in 1992
BSE = breast self-examination

as were judgment formation (respondents' assessment of which persons fit into the interpreted category) and response editing (respondents' decisions on how many people they decide to report).

Routine health checkups. Receipt of routine checkups is a key predictor and correlate of cancer screening. Each survey asked about use of health checkups, but different questions were used. Extensive revisions (second in number only to income questions) in the evolution of

the final *Pathways* question, as well as reports by researchers and interviewers, flagged this as a difficult item.

Question from *BACCIS*:

Q11. During the last 12 months, how many times have you visited a doctor or other health professional just for a checkup (physical examination)—even when you were feeling well?

Questions from *VPS*:

Q32. Have you ever heard of a routine checkup?

Q33. A checkup is when you go to see a doctor while you are feeling well. Have you ever had a routine checkup?

Q34. (IF YES TO Q33) When did you have the last checkup?

Questions from *Pathways*:

Q13. In the past 12 months, how many times have you been to see a doctor?

Q19. Do you go to the doctor for checkups even when you're well?

To evaluate the validity of these items in the *BACCIS* and *Pathways* surveys, we assessed the strength of the association between reported receipt of check-ups and a clinical breast examination (CBE) in the past year by computing odds ratios for respondents ages 40 years and older. Because a CBE typically occurs as part of a routine check-up, we expected that receiving a CBE would be quite rare among women who did not receive check-ups, and that the odds of receiving a CBE would be much higher for those who had a check-up than for those who did not. We were surprised to find that in both surveys, a substantial proportion of women who did not get check-ups did get a CBE. Nevertheless, in all ethnic groups there was a significant association between receipt of check-ups and a CBE. As shown in Table 6, this association was much stronger in *BACCIS*, where ethnic-specific odds ratios ranged from 6.3 to 15.1, than in *Pathways*, where odds ratios ranged from 2.1 to 3.3. The stronger association between checkups and CBE in the *BACCIS* survey suggests that the *BACCIS* question about health checkups had better validity. This is important to note, perhaps reflecting that the *BACCIS* questions about health checkups were worded more specifically than the *Pathways* more general questions.

Interviewer feedback, focus groups, and think-aloud interviews indicated that the concept of health checkup was unfamiliar to some Chinese, Latina, and Vietnamese immigrants. Some Chinese women wondered why anyone would see a doctor if not to treat an illness. Researchers and translators reported that the Vietnamese language contains no parallel concept, and in fact both *routine* and *checkup* had to be explained. These findings reflect the lack of emphasis on prevention in the health systems of home countries, according to the researchers interviewed and the interviewer debriefings.

In the think-aloud interviews with Latina women, many respondents indicated they did not go to the doctor

for checkups. Interestingly, these women said they thought that most people would not be truthful in answering this question because they know they are "supposed to get checkups." When asked to re-phrase the question in their own words, many Latina respondents did not understand the question and could not rephrase it. Other Latina women offered good suggestions for rephrasing, such as "Do you go to the doctor when you are not sick?" Several Latina women also suggested defining "checkup" as "going to the doctor to see if everything is okay."

Most Chinese women said they knew what a checkup is but did not get them because they either lacked insurance or they believed they were healthy and did not need them. Most Chinese women indicated that other people would answer this question honestly. The few who thought otherwise noted that some people might not be truthful if they thought having checkups indicated the presence of an illness that is associated with shame: "Some people may not want to tell the truth, because if they let people know that they get regular checkups, people may think that they get sick often and despise them."

African American women participating in think-aloud interviews clearly understood what a checkup is. However, so many were dealing with hypertension that they had difficulty distinguishing the monitoring of that condition from a general health checkup. In calculating the number of checkups, African American women included those done by nurse practitioners and herbalists. Many African American women thought that people would not be truthful in answering a question about checkups since everyone knows they should go, yet many do not. An interesting interpretation of going to the doctor when well also emerged in the African American focus groups. To some respondents, people who claim to have regular checkups may be regarded as dishonest or hypochondriacs, thus evoking suspicion.

Focus groups with Chinese women revealed that visiting a doctor may be taboo, particularly among older women, as "They are afraid that the doctor will not be able to cure them, but will put them to death instead." There was confusion about what a checkup is, and the word *regular* was preferred over *routine*. The majority would include visits to an herbalist, but there were varying opinions about counting visits to a chiropractor or acupuncturist. Vietnamese focus group participants also did not understand the word *routine*. Vietnamese and Latina women considered checkups only in the context of visits to a Western physician, not to alternative providers. In most focus groups, the issue of routine checkup was inex-

Table 4. Agreement with fatalistic attitudes and beliefs in three studies, by ethnicity (N = 6760)

Survey questions	Study	Response	African American n = 602 (P) n = 475 (B)	Chinese n = 775 (P) n = 279 (B)	Latina n = 1601 (P) n = 230 (B)	Vietnamese n = 645 (P) n = 933 (V)	White n = 605 (P) n = 500 (B)	Other race/ ethnicity n = 115 (B)
Attitudes					Percent			
Do you think that life and death are entirely beyond our control?............	P	Yes	68	64	44	39	31	
		No	29	33	55	28	66	
		No Answer	4	3	1	33	3	
I have no control over the things that affect my health.	B	Agree	35	55	43		18	35
		Disagree	65	44	57		82	63
		No Answer	0	1	0		0	2
If I am going to get sick, I will get sick no matter what I do.	B	Agree	61	59	58		40	56
		Disagree	38	39	41		59	43
		No Answer	1	2	1		1	1
There is little one can do to prevent getting cancer.	B	Agree	44	52	54		28	39
		Disagree	52	27	43		69	55
		No Answer	4	21	3		2	6
Do you think that illness is a matter of chance or fate?. . .	P	Yes	48	30	50	38	34	
		No	34	66	49	42	56	
		No Answer	18	4	2	20	10	
Getting cancer is due to bad luck..	B	Agree	10	57	28		20	23
		Disagree	88	35	71		79	74
		No Answer	2	8	0		1	3
Do you think that cancer is God's punishment?........	P	Yes	5	10	3	18	0	
		No	90	78	96	57	98	
		No Answer	5	12	0	25	1	
Getting cancer is God's punishment.............	B	Agree	7	21	7		2	10
		Disagree	91	62	92		97	86
		No Answer	2	17	1		1	4
Cancer is due completely to God's will?.............	V	Yes				27		
		No				53		
		No Answer				19		

tricably linked to insurance and considered a luxury if one is uninsured. Among Latina women, having a checkup was considered in the context of the concept of *decidiosas*, taking care of others before taking care of oneself.

Based on the above findings, we developed a new measure composed of two questions designed to assist the respondent in separating doctor visits due to a problem from those for a checkup:

Alternative Question:
In the last year, have you been to see a doctor about a health problem?

Table 4. (continued)

Survey questions	Study	Response	African American n = 602 (P) n = 475 (B)	Chinese n = 775 (P) n = 279 (B)	Latina n = 1601 (P) n = 230 (B)	Vietnamese n = 645 (P) n = 933 (V)	White n = 605 (P) n = 500 (B)	Other race/ ethnicity n = 115 (B)
Beliefs					*Percent*			
Do you think you could have cancer without knowing it?.............	P	Yes	78	68	63	51	83	
		No	19	25	36	24	15	
		No Answer	3	6	1	25	2	
A woman can have breast cancer without knowing it..............	B	Agree	94	78	91		97	92
		Disagree	4	13	9		2	4
		No Answer	2	9	0		1	3
You can be sick but not have symptoms?.............	V	Yes				37		
		No				33		
		No Answer				31		
In your opinion, if breast cancer is detected early, what is a person's chance of surviving?.............	P	Good	76	71	69	44	86	
		Fair/Poor	20	24	30	34	12	
		No Answer	4	4	1	22	2	
Breast cancer can be cured if found early enough.......	B	Agree	95	88	90		97	96
		Disagree	4	4	10		2	3
		No Answer	2	9	1		1	1
Some cancers can be cured if detected early?...........	V	Yes				69		
		No				12		
		No Answer				20		

NOTES: Percentages may not add to 100% due to rounding; *p* = 0.001 for within-study ethnic differences in distribution of responses for tabulated variables, by chi-square tests

P = *Pathways*, conducted in 1993–1994
B = *Breast and Cervical Cancer Intervention Study (BACCIS)*, conducted in 1992–1993
V = *The Vietnamese Prevention Study (VPS)*, conducted in 1992

IF YES: In addition to taking care of your health problem, has a doctor given you a complete checkup in the last year?

IF NO: In the past year, did you go to the doctor for a general health checkup?

Respondents reported no difficulty in answering these items. In addition, it appears that respondents made use of the distinction between problem and checkup visits.

Taken together, these findings provide evidence of a wide range of underlying causes of difficulty with the original items. These include (*a*) lack of validity, which varied considerably by ethnicity; (*b*) interpretation, or respondent's understanding of the concept "checkup"; (*c*) judgment formation, as in respondent's assessment of which medical provider visits were to be included; (*d*) response editing, in this case the result of cultural inappropriateness due to cultural proscriptions; and (*e*) translation, or the absence of parallel concept or words in a non-English language.

Table 5. Test-retest survey. Distribution by ethnicity of responses to *Pathways* callback question: "For the people you know, how truthful do you think they would be in answering the questions about size of household?" (N = 467)

Race/ethnic group	n	Very/somewhat truthful		A little/not at all truthful		No answer	
		Percent	95% CI	Percent	95% CI	Percent	95% CI
African American..............	75	85	75, 92	9	4, 19	5	2, 14
Chinese....................	84	90	82, 96	5	2, 12	5	2, 12
Latina.....................	160	86	79, 90	14	10, 21	0	0, 3
Vietnamese.................	74	62	50, 73	15	8, 25	23	14, 34
White.....................	74	95	86, 98	1	0, 8	4	1, 12

NOTES: Percentages may not add to 100% due to rounding; $p = 0.001$ for ethnic differences in distribution of responses, by chi-square test.

CI = confidence interval

Fatalism. The most problematic questions for interviewers were those that attempted to measure the cultural construct of fatalism, the belief that all events are determined by fate and are thus inevitable. Widely discussed and studied, this concept is thought be a critical factor in the adoption of positive health practices for many cultures, but successful measurement remains elusive. Interviewers reported an inordinate amount of time spent by respondents trying to comprehend these questions.

Questions to measure fatalism, from *BACCIS:*

Q10B. **If I am going to get sick, I will get sick no matter what I do.**

Q10C. **I have no control over the things that affect my health.**

Q27G. **There is little one can do to prevent getting cancer.**

Q27L. **Getting cancer is God's punishment.**

Q27M. **Getting cancer is due to bad luck.**

Question from *VPS:*

Q18. **Do you believe that cancer is due completely to God's will?**

Questions from *Pathways:*

Q18. **Do you think that life and death are entirely beyond our control?**

Q21. **Do you think that illness is a matter of chance or fate?**

Mean interviewer ratings identified these items as difficult to ask and answer in surveys of African American, Chinese, and Latina women, and of questionable cultural appropriateness for the latter two groups. Interviewers for

Pathways regarded Question 21 as a little easier for respondents to answer than Question 18; however, both questions had to be repeated to respondents and were a source of confusion. In responding to Question 21, Chinese and Vietnamese women often tried to choose either fate or chance. Strictly speaking, these are two different concepts; thus, the question may be considered "double-barreled." However, in mainstream surveys they appear to have been easily accepted as a single concept referring to forces outside individual control. Because Vietnamese and Chinese women perceived them as separate concepts, a "Yes" or "No" answer would not be appropriate.

When the meaning and understanding of *Pathways* Question 18 was probed in focus groups, most people agreed that one can indeed do things to protect one's health. However, the wording of this question was deemed highly offensive to Chinese women, who felt that the implication was personally threatening.

Chinese women were far more comfortable with alternative versions tried out in the focus groups:

Alternative Qa:

Do you think that going to a doctor helps your health?

Alternative Qb:

Does fate or destiny decide how long we live?

While less emotionally charged for the Vietnamese women, *Pathways* Question 18 was considered too complex for a simple "Yes" or "No" answer. Women found it annoying and confusing. They preferred the alternative Question *a* over both *Pathways* Question 18 and alterna-

tive Question *b*. Latinas also found *Pathways* Question 18 difficult and confusing. For African Americans, *Pathways* Question 18 was difficult to answer and to rephrase. They were ambivalent due to beliefs that while there are things one can do to control health, there also is much beyond one's control.

If the five *BACCIS* variables reliably measured fatalism cross-culturally, we would expect that (*a*) they would be highly correlated with one another, and (*b*) the pattern of association would be consistent across ethnic groups. To evaluate this, we obtained ethnic-specific factor loadings, shown in Table 7, by performing principal components analysis with a varimax rotation. In all ethnic groups, two factors explained 54% to 62% of the variance, indicating a fair amount of correlation among the variables. There were both similarities and differences in the patterns of factor loadings, indicating ethnic differences in patterns of association among the five variables. Items 10B and 10C loaded on the same factor in all groups, and items 27L and 27M loaded on the other factor, suggesting that these two pairs of items may represent two different aspects of fatalism. However, item 27G loaded with 10B and 10C for African American, Latina, and white women, whereas it loaded with 27L and 27M for Chinese and other women; this inconsistency may be due either to real cultural differences, or to differences in measurement.

To evaluate the fatalism questions in BACCIS, we used logistic regression to model agreement with each of five fatalistic statements as a function of demographic factors known to be associated with fatalistic beliefs (Table 8). The results indicated that agreement was more common among women with less than 12 years of educa-

tion or who were foreign born. For the three statements that directly expressed a lack of control over health, respondents were more likely to agree if they were African American, older, or reported that their health was poor. The associations found were consistent with our expectations and tended to support the validity of the survey questions as indicators of fatalism.

Alternative forms survey. In the alternative forms survey, we offered three options for fatalism questions based on input from interviewers and focus groups:

Alternative Q1:

We're interested in learning whether or not people believe that they can do things to stay healthy.

    (*a*). **Do you think there are things YOU can do to prevent sickness like cancer or heart disease?**   **Yes/No**

    (*b*). **Do you think there are things YOU can do to help you live longer?**   **Yes/No**

    (*c*). **Is length of life mostly determined by fate or mostly by things we do?**

                Fate/Things we do/ Don't know.

This question set, along with an item on income, were the only items in this pilot test of alternate wording that continued to generate problems for the respondents, according to interviewer responses to the assessment questions at the end of the survey. Problems with the alternative fatalism questions were reported among 13% of African American, 33% of Vietnamese, 20% of Chinese, and 3% of Latina respondents.

Table 6. Association of receipt of clinical breast examination (CBE) during past year with receipt of checkups, among women ages 40 and older, by ethnicity, BACCIS and Pathways surveys (N = 3669)

Race/ethnic group	BACCIS			Pathways		
	n	*Odds ratio*	*95% CI*	*n*	*Odds ratio*	*95% CI*
African American.................	474	6.9	4.0, 11.9	405	2.1	1.3, 3.4
Chinese.........................	278	6.3	3.5, 11.4	535	2.6	1.8, 3.8
Latina.........................	230	13.7	6.6, 28.1	837	3.3	2.4, 4.4
White	499	15.1	9.3, 24.7	411	3.0	1.8, 5.0

NOTE: $p = 0.054$ for *BACCIS*, and $p = 0.44$ for *Pathways*, within-study ethnic differences, by Breslow-Day tests for homogeneity of the odds ratios

CI = confidence interval

Table 7. Factor loadings for five fatalism variables using principal components analysis with varimax rotation, by ethnicity, BACCIS (N = 1599)

Survey statements	African American		Chinese		Latina		White		Other race/ethnicity	
If I am going to get sick, I will get sick no matter what I do....	**0.70**	-0.04	0.04	**0.80**	**0.75**	0.03	**0.74**	0.09	-0.09	**0.88**
I have no control over the things that affect my health.....	**0.75**	0.11	0.11	**0.78**	**0.69**	0.09	**0.75**	0.04	0.34	**0.72**
There is little one can do to prevent getting cancer.........	**0.72**	0.10	**0.64**	-0.04	**0.65**	0.05	**0.70**	0.06	**0.66**	0.08
Getting cancer is God's punishment................	0.13	**0.81**	**0.69**	0.08	0.03	**0.80**	0.29	**0.64**	**0.82**	0.01
Getting cancer is due to bad luck....................	-0.00	**0.83**	**0.79**	0.19	0.10	**0.78**	-0.12	**0.86**	**0.75**	0.15
Variance explained		0.59		0.56		0.54		0.57		0.62

NOTE: Each item was coded on a scale from 0 ("strongly disagree") to 3 ("strongly agree"), with "don't know" responses coded as 1.5.

It is not surprising that there are differences in personal reactions to these questions. Fatalism is a cultural construct, appearing to have different meanings, relevance and implications in different cultural groups. Thus, it is difficult to find wording that clearly indicates the concept of fatalism. However, in addition to the general cultural inappropriateness of the items, lack of validity also was an issue, as were problems of interpretation (for example, respondent's understanding that she was being asked to choose between two different concepts when the intention of the question was to assess her opinion of a single concept), and judgment formation (respondent's ambivalence regarding her own opinion.)

Other selected findings. A few additional results highlight other issues of interest, such as the intractable problem in health surveys regarding measurement of income. Responses to the item, "How truthful do you think people would be in answering questions about their income?" (Table 9) indicate even greater problems with truthfulness regarding income than for questions about household size.

For African Americans, the problems reported by interviewers and in focus groups were respondents' questions about why a health survey would ask about income, wanting to know how the information would be used. Chinese women in the focus groups indicated they would avoid providing the information, saying they "definitely don't like to answer" and would intentionally "underestimate." Among Latina women, there was much discussion about the difficulty in estimating family income because they do not have all the information, "...difficult...have to sit and think." Distrust was strong among Vietnamese women, with interviewers noting that "They'll give you a number, but it will be wrong."

Finally, questions to measure knowledge and receipt of a Pap test, main dependent variables in all three studies, were the source of multiple problems. We learned from researchers and translators that there is no equivalent term for *Pap smear* in Chinese or Vietnamese. In addition, there is evidence in all groups, including African American and white women, that women often confuse the Pap test with a pelvic examination, or assume that if they have had a pelvic examination, they have also had a Pap test. While many less acculturated women have never heard of a Pap test, interviewers reported that white and African American women were frequently insulted by being asked if they had ever "heard of" a Pap test. Our procedure was to follow that question with a definition of the test so women who are not familiar with the term could still ascertain whether or not they had the test. The proportion of women by ethnicity who recognized the names of the tests for mammogram and Pap smear among those reported as having had each test are stratified by whether or not the interview was in English (Table 10). Almost everyone who had had a mammogram had heard of the test, whether interviewed in English or not. For Pap, however, women responding in English were almost certain to have heard of the test, while many women interviewed in other languages had not. Among

Table 8. Characteristics independently associated with agreement with five fatalistic statements in logistic regression models, BACCIS (N = 1557)

Characteristics	If I am going to get sick, I will get sick no matter what I do.		I have no control over the things that affect my health.		There is little one can do to prevent getting cancer.		Getting cancer is God's punishment.		Getting cancer is due to bad luck.	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Latina............	1.6	0.9, 2.6	1.8	0.9, 3.3	1.6	0.9, 2.9	0.5	0.1, 3.3	0.8	0.3, 1.8
African American...	2.0	1.4, 3.0	2.3	1.4, 3.7	1.8	1.2, 2.8	1.6	0.5, 4.9	0.6	0.3, 1.2
Chinese..........	1.0	0.5, 2.0	0.7	0.3, 1.6	1.1	0.5, 2.2	1.7	0.2, 13	1.4	0.5, 3.5
Other race/ ethnicity.........	1.2	0.7, 2.3	0.8	0.3, 1.7	1.1	0.6, 2.1	2.0	0.4, 9.7	1.3	0.5, 3.0
Age, per decade....	1.2	1.0, 1.3	1.2	1.1, 1.4	1.2	1.0, 1.3	1.2	0.9, 1.5	1.1	0.9, 1.2
<12 years of education........	1.7	1.3, 2.2	2.1	1.6, 2.9	1.3	1.0, 1.7	2.7	1.4, 5.1	1.6	1.1, 2.4
Excellent/good health	0.6	0.5, 0.8	0.5	0.4, 0.7	0.8	0.6, 1.0	1.2	0.7, 2.0	1.1	0.8, 1.5
Foreign born	1.8	1.1, 3.1	2.1	1.1, 3.9	1.8	1.0, 3.2	2.1	0.5, 8.6	2.3	1.1, 5.0
Resided ≤10 years in California:										
US born.........	0.9	0.5, 1.7	1.2	0.6, 2.5	1.3	0.7, 2.4	3.7	1.2, 12	0.9	0.3, 2.5
Foreign born	0.9	0.6, 1.4	0.8	0.5, 1.3	1.0	0.6, 1.5	0.6	0.3, 1.3	1.1	0.7, 1.7
English interview ...	1.1	0.6, 1.8	0.6	0.3, 1.2	0.8	0.4, 1.4	0.3	0.1, 1.3	0.4	0.2, 0.7

NOTES: Controlling for area of residence and all variables tabulated. Referent categories are: White race/ethnicity, 12 or more years education, poor or fair self-rated health status, US born, more than 10 years in California, and non-English interview.

OR = odds ratio
CI = confidence intervals

women who had a Pap smear, only 51% of Vietnamese women interviewed in Vietnamese and 68% of Chinese women interviewed in their native language had heard of a Pap smear. This compares with 98% for Chinese women interviewed in English. Knowing the label of the tests indicates some basic understanding of what the test is, and enables women to ask their doctors about it.

DISCUSSION

Many of the complex elements that affect measurement in multiethnic surveys are neither readily observed nor understood through standard tests of data quality. Moreover, ethnic differences in some aspects of data quality do not in themselves imply the cause of difficulties. Identification of ethnic differences by quantitative methods (descriptive statistics, measures of reliability, and validity) is a useful first step in exposing the cultural nature of potential problems in methodology. However, explication of the cognitive processes employed by respondents and

the association of these processes with ethnic differences is needed to gain insights that can lead to a broad cultural explanation.

Our methodologic study progressed to a deeper understanding of the cultural appropriateness and comparability of wording through two exploratory approaches: by combining quantitative and cognitive analyses with insights from interviews of ethnically diverse investigators and interviewers, and by experimenting with alternate forms of troublesome survey questions. In some cases, we were able to identify ethnic differences only in cognitive processes, but in many others we could associate these differences with cultural factors that suggested why items were culturally inappropriate.

Implications for public health. Subjecting survey questions to extensive qualitative and quantitative formative research, particularly within each culture to be studied, is a critical step to improve the quality of multiethnic survey instruments. It still will be necessary to assess the

Table 9. Test-retest survey. Distribution of responses by ethnicity to *Pathways* callback question: "For the people you know, how truthful do you think they would be in answering the questions about income?" (N = 467)

Race/ethnic group	n	Very/somewhat truthful		A little/not at all truthful		No answer	
		Percent	95% CI	Percent	95% CI	Percent	95% CI
African American.............	75	68	56, 78	21	13, 33	11	5, 20
Chinese.....................	84	75	64, 84	23	14, 33	2	0, 9
Latina......................	160	66	58, 73	34	27, 42	0	0, 3
Vietnamese..................	74	38	27, 50	30	20, 42	32	22, 44
White......................	74	81	70, 89	14	7, 24	5	2, 14

NOTE: p = 0.001 for ethnic differences in distribution of responses, by chi-square test

CI = confidence interval

quality of data generated from the completed survey, however, again using multiple methods. It must be expected that more time and resources will be required to develop and evaluate a multiethnic survey, and even more so for one conducted in multiple languages compared with a survey targeted to one ethnic group. Researchers should be aware of the resource intensity of this work during a study's earliest planning stages. In the absence of sufficient resources to conduct the testing recommended here, investigators are advised to seek out existing survey questions for which adequate development and testing already has been done. If appropriately tested items cannot be found, it is preferable to focus on objective rather than attitudinal questions, thus reducing the likelihood of difficulties. Also, debriefing of interviewers appears to be quite informative yet inexpensive, making it a useful and efficient procedure to weed out potentially problematic survey items. If adequately designed questions cannot be identified through these means, it would be preferable to reduce the scope of a study rather than to produce data of questionable quality.

Each of our methods added to our understanding of measurement problems, with one exception. The participant observation of the *Pathways* surveys development was quite labor intensive, but served only to indicate or corroborate problems that emerged in other ways, particularly through investigator interviews. Among the other procedures, we found the most valuable to also be cost-efficient. In particular, the debriefing of bilingual and bicultural interviewers provided some of our best information. Even with the highest quality pretesting, additional and important insights are likely to be generated during the actual administration of a survey. However, it is important to corroborate findings obtained during administration in order to avoid biasing the results with the views of individual interviewers. While the actual ratings of questions by researchers and interviewers were of limited value due to their subjective nature and small numbers, the ratings served as good indicators of items requiring more inquiry.

Although some researchers[9] advocate the use of focus groups and think-aloud interviews together, arguing that the latter are appropriate for people with little formal education, our team of research assistants who conducted both types of interviews found the think-aloud interview far too challenging for respondents with minimal education. According to the RAs, many think-aloud participants had difficulty understanding the questions since they were unaccustomed to such inquiries. This caused them considerable embarrassment, particularly because of the individualized nature of the interview. The focus group discussions, on the other hand, permitted women who did not understand the questions to listen and to contribute when they could do so comfortably. Furthermore, the discussion by their peers helped to advance their understanding.

Limitations. As an exploratory study attempting to employ and assess multiple methods, there are many limitations to these findings. First, time and resources did not permit application of all methods to all troublesome items. This limited both the degree of understanding of the nature of the problems and our ability to fully assess the value of the methods. Second, although a large body

of data was produced, the small-scale use of each method precludes generalization of our findings. Third, and very importantly, the methods we employed should be part of an iterative process in which the alternative questions derived from all the previous tests are themselves subjected to the full complement of tests. Finally, a single exploratory study is insufficient for constructing a comprehensive model of the patterns of culturally-based problems and their causes that is needed to establish principles for multiethnic survey development. Considerably more methodologic research is needed before problem identification can lead to solutions.

CONCLUSIONS

Despite its limitations, this study generated many illuminating and useful observations. Based on the findings, we have drawn the following conclusions regarding both the nature of problem questions and the value of methodologic strategies in studying the quality of multiethnic data:

- Different categories of items present different problems and vary in the degree to which they are likely to be difficult. The more objective an item is, the less it is subject to the array of problems described here, although these items may still be affected by translation, cultural history, and recall problems. Attitudinal questions are likely to pose the most problems.

- No ethnocultural group is immune to cultural and wording measurement problems.
- For populations with a recent history of immigration, we believe that cultural factors are best explored among those who are the least acculturated, as indicated by being monolingual in their native language. Collecting qualitative data from non-English speaking people clearly will illuminate translation problems, and is likely to be highly effective in identifying cultural differences.
- Data from cross-cultural surveys should be used with caution, regardless of the rigor with which they were developed and evaluated.
- No single test can assess all aspects of item functioning. The optimal evaluation of questions should include multiple quantitative and qualitative methods that complement one another and reveal different types and causes of question problems.
- Improved questions resulting from this process should also be rigorously tested.

Ultimately, it may not be possible to attain both cultural appropriateness and comparability of wording. Nevertheless, the procedures described here can be expected to yield much improved and thus valuable data for use in research and programs to promote the health of ethnically diverse communities.

We have discussed previously the value of multiethnic research.[45] Yet, the complex and varied nature of the

Table 10. Test name recognition: proportion of respondents who had heard of mammogram and Pap tests among those who reported ever having had the test, stratified by whether interview was conducted in English, Pathways

| | Mammogram (N = 2330) | | | | | | Pap smear (N = 3187) | | | | | |
| | All (p = 0.001) | | English interview Yes (p = 0.47) | | No (p = 0.074) | | All (p = 0.001) | | English interview Yes (p = 0.043) | | No (p = 0.001) | |
Race/ethnic group	n	Percent	n	Percent	n	Percent	n	Percent	n	Percent	n	Percent
African American....	438	99	438	99			592	100	592	100		
Chinese......	422	94	69	100	353	93	517	74	107	98	410	68
Latina........	865	96	260	100	605	95	1210	88	460	99	750	82
Vietnamese ...	177	97			177	97	271	51			271	51
White	428	100	428	100			597	100	597	100		

NOTE: p-values for ethnic differences by chi-square test

The complex and varied nature of the problems identified in this study raise the question of the extent to which it is possible to meet rigorous standards of validity, reliability, and comparability in multiethnic research.

problems identified in this study raise the question of the extent to which it is possible to meet rigorous standards of validity, reliability, and comparability in such endeavors. The literature on the study of culture contains a much discussed dilemma regarding the tension between the in-depth study within one culture to understand what is unique about it, and the comparative study of more than one culture in search of similarities between them.[46] In the former case, measures are derived entirely from within the culture and have meaning and validity in that, but not necessarily in other, cultures. In the latter, in order to conduct comparisons measures must be applied from outside the culture, with some degree of uncertainty regarding the cross-cultural equivalence in meanings. The multi-lingual surveys reported here represent varying approaches to this problem. The method used by each had its strengths and limitations. Research is needed to systematically compare these methods to establish what is gained and lost by each.

References

1. Adler NE, Marmot M, McEwen BS, Stewart J, editors. Socioeconomic status and health in industrial nations: social, psychological and biological pathways. New York: Academy of Sciences; 1999.
2. Haynes MA, Smedley BD. The unequal burden of cancer: an assessment of NIH research and programs for ethnic minorities and the medically underserved. Washington, DC: National Academy Press; 1999.
3. Hogue CJR, Hargraves MA, Collins KS, editors. Minority health in America. Baltimore (MD): Johns Hopkins University Press; 2000.
4. Kramer EJ, Ivery SL, Ying Y-W, editors. Immigrant women's health: problems and solutions. San Francisco: Jossey-Bass; 1998.
5. Marmot M, Wilkinson RG, editors. Social determinants of health. New York: Oxford University Press; 1999.
6. Braithwaite RL, Taylor SE, editors. Health issues in the black community. San Francisco: Jossey-Bass; 1992.
7. Zane NWS, Takeuchi DT, Young KNJ, editors. Confronting critical health issues of Asian and Pacific Islander Americans. Thousand Oaks (CA): Sage Publications; 1994.
8. Aday LA, editor. Designing and conducting health surveys. San Francisco: Jossey-Bass; 1989.
9. Sudman S, Bradburn NM, Schwarz N, editors. Thinking about answers: the application of cognitive processes to survey methodology. San Francisco: Jossey-Bass; 1996.
10. Tanur JM, editor. Questions about questions: inquiries into the cognitive bases of surveys. New York: Russell Sage Foundation; 1994.
11. Warnecke RB, editor. Proceedings of Health Survey Research Methods 1995 June 24-26; Breckenridge, Colorado. Hyattsville (MD); 1996. DHHS Pub. No. PHS 96-1013.
12. Biemer PP, Groves RM, Lyberg LE, Mathiowetz NA, Sudman S, editors. Measurement errors in surveys. New York: Wiley; 1991.
13. Fowler FJ Jr. How unclear terms affect survey data. Public Opinion Q 1992;56:218-31.
14. Dippo CS, Norwood JL. A review of research at the Bureau of Labor Statistics. In: Tanur JM, editor. Questions about questions: inquiries into the cognitive bases of surveys. New York: Sage; 1991. p. 271-90.
15. Campanelli PC, Martin E, Creighton K. Respondents' understanding of labor force concepts: insights from debriefing studies. Proceedings of the 5th Annual Research Conference; 1989 Washington DC. Department of Commerce (US), Bureau of the Census; 1989. p. 361-74. Available from: URL: http://www.bls.gov/pdf/orebib.pdf
16. Rothwell N. New ways of learning how to improve self-enumerative questionnaires: a demonstration project [paper]. Washington (DC): Bureau of the Census (US); 1983.

17. Rothwell N. Laboratory and field response research studies for the 1980 Census of Population in the United States. J Official Stat 1985;1:137-57.

18. McKenney NR, Bennett CE. Issues regarding data on race and ethnicity: the Census Bureau experience. Public Health Rep 1994;109:16-25.

19. Sirken M, Jabine T, Willis G, Martin E, Tucker C. A new agenda for interdisciplinary survey research methods. Proceedings of the CASM II Seminar; 1999; Hyattsville, Maryland. Department of Health and Human Services (US), National Center for Health Statistics; 1999. Available from: URL: http://www.bls.gov/pdf/orebib.pdf

20. Gaertner G, Cantor D, Gay N. Tests of alternative questions for measuring industry and occupation in the CPS. Proceedings of the Section on Survey Research Methods, American Statistical Association; 1989. Available from: URL: http://www.bls.gov/pdf/orebib.pdf

21. Martin EA, Polivka AE. Diagnostics for redesigning survey questionnaires: measuring work in the current population survey. Public Opinion Q 1995;59:547-67.

22. Palmisano M. Respondents understanding of key labor force concepts used in the CPS. Proceedings of the Section on Survey Research Methods, American Statistical Association; 1989. Available from: URL: http://www.bls.gov/pdf/orebib.pdf

23. Department of Labor (US). Current population survey, design and methodology. Bureau of Labor Statistics report. Washington (DC); 2000. Technical Paper 63.

24. McKay RB. Cultural factors affecting within household coverage and proxy reporting in Hispanic (Salvadoran) households. Proceedings of the Section on Survey Research Methods, American Statistical Association; 1992. p. 614-8. Available from: URL: http://www.bls.gov/pdf/orebib.pdf

25. McKay RB. Cognitive research in reducing nonsampling errors in the current population survey supplement on race and ethnicity. Proceedings of Statistics Canada Symposium 96; 1996 November, Ottowa, Canada.

26. McKay RB, de la Puente M. Cognitive testing of racial and ethnic origin questions for the CPS Supplement. Monthly Labor Rev 1996;119(9):8-12.

27. Tucker NC, McKay RB, Kojetin BA, Harrison R, de la Puente M, et al. Testing methods of collecting racial and ethnic information: results of the current population survey supplement on race and ethnicity. Washington (DC): Bureau of Labor Statistics (US); 1996. Statistical Note Series; Vol. 40.

28. Devins GM, Beiser M, Dion R, Pelletier LG, Edwards RG. Cross-cultural measurements of psychological well-being: the psychometric equivalence of Cantonese, Vietnamese, and Laotian translations of the affect balance scale. Am J Public Health 1997;87:794-9.

29. Keller SD, Ware JE Jr, Bentler PM, Aaronson NK, et al. Use of structural equation modeling to test the construct validity of the SF-36 Health Survey in ten countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998;51:1179-88.

30. Ware JE Jr, Kosinski M, Gandek B, Aaronson NK, et al. The factor structure of the SF-36 Health Survey in ten countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998;51:1159-65.

31. Karvonen S, Abel T, Calmonte R, Rimpela A. Patterns of health-related behavior and their cross-cultural validity—a comparative study on two populations of young people. Soz Praventimed 2000;45:35-45.

32. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993;46:1417-32.

33. Guillemin F, Briancon S, Pourel J. Validity and discriminant ability of a French version of the Health Assessment Questionnaire in early RA. Disabil Rehabil 1992;14:71-7.

34. Chang RH, Dodder RA. The modified purpose in life scale: a cross national validity study. Int J Aging Hum Devel 1983;18:207-17.

35. Warnecke RB, Johnson TP, Chávez N, Sudman S, O'Rourke DP, Lacey L, et al. Improving question wording in surveys of culturally diverse populations. Ann Epidemiol 1997;7:334-42.

36. Warnecke RB, Sudman S, Johnson TP, O'Rourke D, Davis AM, Jobe JB. Cognitive aspects of recalling and reporting health-related events: Papanicolaou smears clinical breast examinations, and mammograms. Am J Epidemiol 1997;146:982-92.

37. Hiatt RA, Pasick RJ, Stewart S, Bloom J, Davis P, Gardiner P, et al. Community-based cancer screening for underserved women: design and baseline findings from the Breast and Cervical Cancer Intervention Study (BACCIS). Prev Med. In press 2001.

38. Hiatt RA, Pasick RJ, Pérez-Stable EJ, McPhee SJ, Engelstad L, Lee M, et al. Pathways to early cancer detection in the multiethnic population of the San Francisco Bay Area. Health Educ Q, 1996;23: S10-S27.

39. Pérez-Stable, EJ, Otero-Sabogal R, Sabogal F, Nápoles-Springer A. Pathways to early cancer detection for Latinas: En Acción Contra el Cancer. Health Educ Q, 1996;23:S41-S59.

40. McPhee SJ, Bird JA, Ha NT, Jenkins CNH, Fordham D, Le B. Pathways to early cancer detection for Vietnamese Women: Suc Khoe La Vang! (Health is Gold!) Health Educ Q, 1996;23:S60-S75.

41. Lee M, Lee F, Stewart S. Pathways to early breast and cervical cancer for Chinese American women. Health Educ Q 1996;23:S76-S88.

42. Jenkins CN, McPhee SJ, Bird JA, Pham GQ, Nguyen BH, Nguyen T, et al. Effect of a media-led education campaign on breast and cervical cancer screening among Vietnamese-American women. Prev Med 1999; 28:395-406.

43. Pasick RJ, Sabogal F, Bird JA, D'Onofrio C, Jenkins CHN, Lee M, et al. Problems and progress in development and translation of health survey questions across ethnic groups: the Pathways experience. Health Educ Q 1996;23:S28-S40.

44. Campanelli PC, Martin EA, Rothgeb JM. The use of interviewer debriefing studies as a way to study response error in survey data. Statistician 1991;40:253-64.

45. Pasick RJ, D'Onofrio CN, Otero-Sabogal R. Similarities and differences across cultures: questions to inform a third generation for health promotion research. Health Educ Q 1996;23:S142-S161.

46. Berry JW. Imposed etics, emics and derived etics: their conceptual and operational status in cross-cultural psychology. In: Headland TN, Pike KL, Harris M, editors. Emic and etics: the insider/outsider debate. Newbury Park (CA): Sage; 1990. p. 84-99. ■