A Comparison of Forced-Choice and Mark-All-That-Apply Formats for Gathering Information on Health Insurance in the 2006 American Community Survey Content Test[1]

Leah Ericson
Computing Services, Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213, lericson@andrew.cmu.edu

Chuck Nelson
Housing and Household Economic Statistics Division, U.S. Census Bureau
Washington, D.C. 20233, Charles.T.Nelson@census.gov

Keywords: forced-choice and mark-all-that-apply formats, systematic response error, reinterview data

1. Introduction

The 2006 American Community Survey (ACS) Content Test included two different versions of a new health insurance question for testing. The health insurance item intends to estimate the overall rate of health insurance coverage as well as compare rates of private and public health care coverage. Version 1 of the question used a forced-choice format. The forced-choice format required respondents to indicate "yes" or "no" for each of seven health care types. Version 2 of the question used a mark-all-that-apply format. In this format, respondents read or heard a list of seven types of health care coverage, selecting as many as apply to them. To determine which version would be most suitable for the ACS, responses to these two versions were compared in terms of item non-response, coverage rates, and systematic response error. Through these comparisons, we realized that the results almost always favored one question format over the other. Given the amount of interest in comparing forced-choice formatted questions to their mark-all-that-apply alternatives, we thought it might be beneficial to discuss our empirical findings in terms of the methodological differences between the formats.

2. Background

2.1 The 2006 ACS Content Test

In January through March of 2006, the American Community Survey (ACS) conducted the first test of new and modified content since the ACS reached full implementation levels of data collection. The results of that testing helped determine the content for the 2008 ACS[2]. The year 2008 marks the first year of a three-year aggregated data product that includes data from the same year as the 2010 decennial census (2008 - 2010). Similarly, 2008 is the midpoint year for the first five-year data product that includes data from 2010 (2006-2010). Given the significance of the year 2008, the ACS committed to a research program during 2006 that will result in final content determination in time for the 2008 ACS. This research is the 2006 ACS Content Test.

Through the Office of Management and Budget (OMB) Interagency Committee on the ACS, the Census Bureau included subject matter experts and key data users from other federal agencies in identifying questions for inclusion in the Content Test. In general, the Content Test evaluated alternatives for questions which showed some indication of a problem, for example, high missing data rates, estimates which differed systematically from other sources of the same information, or high simple response variance as measured in the Census 2000 Content Reinterview survey. In addition, the Content Test also included testing of three new topics proposed by other federal agencies for inclusion in the ACS; one of these new topics was health insurance coverage.

---

[1] This report is released to inform interested parties of research and to encourage discussion. Any views expressed on statistical or methodological issues are those of the authors and not necessarily those of the U.S. Census Bureau.

[2] NOTE: The Census Bureau submitted the proposed 2008 ACS questionnaire and the results of the Content Test to the Office of Management and Budget in Spring 2007. The Office of Management and Budget used these findings, along with input from Federal agencies and other sources, to approve the final set of questions that will be on the 2008 ACS.

To meet the primary objective of the 2006 ACS Content Test, analysts evaluated changes to question wording, response categories, instructions, or examples relative to the current version of the questions. One unique feature of the Content Test was the use of a Content Follow-Up (CFU) reinterview. The data from the CFU provided a measure of error by evaluating differences between original responses and reinterview responses.

2.2 Previous Testing or Analysis for Health Insurance

To our knowledge, until 2005, no major national federal survey had ever attempted to ask about health insurance coverage in a mail survey. Therefore, before implementation of a major field test, two sets of cognitive interviews were conducted on this topic. One set was conducted in 2004 and 2005 with funding support from the Department of Health and Human Services. This set of interviews was conducted in order to design a health insurance question that would be appropriate for an ACS-style questionnaire (see Pascale, 2005). The other set was conducted with funding from the Department of Veterans' Affairs and examined how those currently or formerly in the military answered the questions on veteran status and health insurance coverage (see Westat, 2005). The questions tested in these two sets of cognitive interviews were similar to the mark-all-that-apply format that was used in the Content Test (see Appendix A for the questions tested in the 2006 ACS Contest Test).

From the cognitive interviews, three major changes in methodology and wording were suggested. The most significant of these suggestions was the elimination of an answer category on supplemental health plans. This category would have been used to exclude people from the "covered" universe who reported coverage, but only reported supplemental plan coverage. The second suggestion involved the expansion of the mark-all-that-apply question about "any coverage." The question was changed from "Is this person CURRENTLY covered by any type of health insurance?" to "Is this person CURRENTLY covered by any type of health insurance or health coverage plan?" The reason for this change was that many types of public coverage are not technically health insurance plans. There was hope that the expanded wording would better capture public health plan coverage. The last suggestion was the simplification of the military and VA health plan answer categories by excluding CHAMPUS from the military care category and CHAMPVA from the VA health plan category.

Based on the results of the cognitive interviews, two versions of the health insurance question were developed (see Appendix A). Version 1 of the question used a forced-choice format. The forced-choice format requires respondents to indicate "yes" or "no" for each of seven health care types. Version 2 of the question used a mark-all-that-apply format. In this format, respondents were first asked a screener question about whether they have health insurance or a health coverage plan. After this screener question, respondents who have health coverage were asked to "mark all that apply" for seven health care types.

2.3 Previous Research on Forced-Choice vs. Mark-All-That-Apply Questions

The mark-all-that-apply question format is commonly used as a way to reduce respondent burden. With this format, respondents are presented a list of items and are asked to indicate which of these items applies to them. A logical alternative to this format is the forced-choice format. With the forced-choice format, respondents are presented with the same list of items, but are asked to indicate whether or not each item applies to them. Several studies have aimed at determining which question format provides better responses. Many of these studies have found that, although seemingly more burdensome, the forced-choice format usually provides more responses (see, Rasinski et al., 1994; Smyth et al., 2005; Feindt et al., 1997).

An example of this finding is presented by Smyth et al. (2005). In this study, a web survey of students at Washington State University was used to evaluate theoretical explanations for differences between responses to mark-all-that-apply and forced-choice question formats. In support of work by Rasinski et al. (1994), the Smyth et al. study found that weak satisficing (the act of not optimally answering questions) was a cause of fewer affirmative responses to the mark-all-that-apply format when compared to the forced-choice format of the question. In fact, Smyth et al. found that increased response to the forced-choice format is most likely due to deeper thought processing than with the mark-all-that-apply format. Meaning, respondents tend to take more time to carefully consider each item and the appropriate response in the forced-choice format than in the alternative format.

Another similarity between the Rasinski et al. (1994) study and the Smyth et al. (2005) study is the increased occurrence of responses in the forced-choice format. In the Rasinski study, responses to a field test of the National Education Longitudinal Study of 1988 were analyzed. Here, it was found that when yes/no instructions are used, more responses are obtained. Similarly, Smyth et al. (2005) found that more options were marked in the forced-choice format than in the mark-all-that-apply format. A limitation of these studies is that although more responses are captured with the forced-choice format, more

is not necessarily better, and there is no mechanism with which to determine the accuracy of responses in these studies; therefore we are still left with the question, which format produces more accurate responses?

Feindt et al. (1997) helps to answer this question by using the Schools and Staffing Survey (SASS). The SASS used reinterview data to estimate systematic error in tested questions. Feindt et al. (1997) found that the forced-choice format of a question asking teachers to indicate the degrees they have earned elicited responses with less response variance than the alternative mark-all-that-apply version of the question. In addition, four of six response options to the question "What other school positions, if any, did you hold before you became a principal?" were found to elicit more accurate data when respondents were asked to indicate "yes" or "no" for the six items than when they were asked to indicate all items that applied to them.

As previously mentioned, forced-choice format questions usually result in more responses, but it is unclear whether more responses are necessarily better in terms of accuracy. Similar to the Feindt et al. (1997) study, this evaluation of two different health insurance question formats adds to the research on forced-choice vs. mark all that apply by using reinterview data to determine which format provides more accurate responses.

3. Methodology

3.1 Data Collection Methods

The 2006 ACS Content Test Data Collection

The 2006 ACS Content Test consisted of a national sample of approximately 62,900 residential addresses in the contiguous United States (the sample universe did not include Puerto Rico, Alaska and Hawaii). To meet the primary test objective of evaluating question wording changes, approximately half of the sample addresses were assigned to a test group (31,450) and the other half to a control group (31,450). For the topics already covered in the ACS, the test group included the proposed alternative versions of the questions, and the control group included the current version of the questions as asked on the ACS. To keep context and questionnaire length consistent between the two versions both the test and control questionnaires included the health insurance topic (and the other two new topics not currently on the ACS). The control group contained the forced-choice format of the health insurance item and the test group contained the mark-all-that-apply format.

The Content Test mailed English-only questionnaires to sampled households around December 28, 2005, coinciding with the mailing for the ACS January 2006 sample panel. Beginning in February 2006, Census Bureau field representatives visited a sample of households that did not respond by mail in an attempt to collect the data through Computer Assisted Personal-visit Interviewing (CAPI). The CAPI operations, which were conducted in both English and Spanish, ended March 2, 2006. Between January 17 and March 17, 2006 a content reinterview called Content Follow-Up (CFU) was conducted via Computer Assisted Telephone Interviewing (CATI) as a method of measuring response error. As with CAPI operations, the CFU was conducted in both English and Spanish. The CFU is discussed in greater detail below. More information on the data collection methods of the 2006 ACS Content Test and how these methods differ from production ACS data collection methods can be found in Nelson and Ericson (2007).

Content Follow-Up data collection

The CFU reinterview, conducted by the Census Bureau's three telephone centers, provided a method for measuring response error. About two weeks after receiving the returned questionnaire or completed CAPI interview, the responding unit entered the CFU operation. Telephone staff completed the CFU interviews between January 17 and March 17, 2006. At the first contact with a household, interviewers asked to speak with the original respondent. If that person was not available, interviewers scheduled a callback at a time when the household member was expected to be home. If at the second contact we could not reach the original respondent, interviewers completed the interview with another adult household member.

The CFU reinterview did not replicate the full ACS interview. Rather, the CFU used the roster and basic demographic information from the original interview and only asked questions specific to the analytical needs of the Content Test. Reinterview questions were of two general formats: (1) the same question as asked in the original interview, or (2) a different set of questions providing more detail than those asked in the original interview. For topics in which the CFU asked the same question as the original interview, the CFU asked the test or control version of the question based on the original treatment. For topics using a different question or set of questions than the original interview, we asked the same detailed series of

questions regardless of the original treatment condition. The CFU health insurance questions, which were based on prior testing results, reflect this more detailed format.

Generally, the goal of asking the same question twice—called replicated reinterviewing by Biemer and Lyberg (2003)—is to measure simple response variance. Examples of this type of reinterviewing are shown in surveys like the Current Population Survey (see Biemer and Forsman, 1992), the Schools and Staffing Survey (see Bushery et al., 1992), and the National Household Education Survey (see Brick and West, 1992). In these surveys, respondents are reinterviewed with a subset of questions from the original interview in an effort to obtain a second independent measure of the original question. This second measure is used gauge how consistent responses to the question will be over repeated trials (Groves, 2004). Contrastingly, the goal of asking more detailed questions—called gold standard reinterviewing—is to obtain a more accurate measure of the underlying construct (Biemer and Lyberg 2003). Here, the second measure is used to determine the accuracy of the response to the original question. It is important to note this difference in evaluative methodology, as our aim with the health insurance reinterview data was to determine which version of the question elicited more accurate responses.

3.2 2006 ACS Content Test Sample Design

For the original mail survey a stratified systematic sample of households was selected. Households were stratified into high and low response strata within sampled Primary Sampling Units (PSUs) based on tract-level mail response rates to the Census 2000 long form. For CAPI data collection, a two-stage sampling technique was used to help contain field costs. This sampling technique was based on percentage of foreign-born population within the PSU since the majority of that target population responds via CAPI. At least one item undergoing testing in the Content Test required an adequate sample of this population. In the majority of PSUs, we assigned cases to both the control and test groups. For more information on the 2006 ACS Content Test sample design, see Asiala (2006).

There was no sampling for CFU. A CFU interview was attempted for all responding households to the Content Test for which we had a phone number.

3.3 Health Insurance Evaluation Methodology

As noted above, the mark-all-that-apply version of the question used a screener question that asked whether a respondent is covered by any type of health insurance/coverage plan. Respondents answering "yes" to this question were then asked to specify the type(s) of insurance/coverage plan(s). The forced-choice version asked respondents to respond "yes" or "no" to a series of questions on specific types of insurance/coverage plans, without an initial screening question. The evaluation of the health insurance question in the Content Test focused on determining which of the two question formats resulted in superior health insurance coverage estimates. Superiority of a question format was determined in two main ways: 1) the format that produced less systematic error as defined by accuracy of response and 2) the format that produced greater reporting of public health care coverage.

Similar to "gold standard reinterviewing," the method used to determine accuracy of response to the health insurance question relied on obtaining a more accurate measure of the underlying constructs (health insurance coverage status and type). Therefore, the CFU questions for health insurance were more detailed and comprehensive than the questions that were used in either version of the Content Test. By comparing the more comprehensive CFU measure to the original Content Test response we were able to determine the accuracy of the original response as measured by the net difference rate (NDR).

The NDR is used when we assume the CFU interview, which asks more questions and collects more detailed data about a topic, provides a better measure than original version of a question. The NDR reflects the net change between the original response and the response given for the more detailed CFU questions. In other words, since we assume the CFU provides better data, the NDR indicates to what extent the test or control version of a question over- or under-estimates the topic (or category) of interest. Relative to the CFU estimate, an NDR with a negative value indicates an underestimate and a positive value indicates an overestimate by the original question. An NDR that does not statistically differ from "0" indicates that the question asked in the original test or control interview produces results similar to the more detailed question set asked in CFU. In other words, the question should not result in a systematic over- or under-estimate of the topic (or category) of interest. For the purpose of this evaluation, the NDRs calculated for the mark-all-that-apply version of the question were compared to those calculated for the forced-choice version in order to assess which version of the question resulted in more systematic error, regardless of whether the error reflected an over- or under-estimate.

4. Limitations

4.1 General Content Test and Content Follow Up Limitations

The Content Test maintained the same general mail data collection methodology as the ACS, but differed in the mail nonresponse follow-up operations. Some aspects of the Content Test implementation should be considered in evaluating the data.

- The Content Test did not include CATI data collection in order to meet field data collection constraints. While the design of the Content Test allowed all sampled housing units an opportunity to participate even without CATI, questions administered differently over the phone did not get the benefit of a full CATI operation (though some of the CAPI interviews actually do occur by phone). However, since only ten percent of ACS data is collected by CATI and CATI interviewers are trained to help respondents understand question intent and response categories, overall ACS data quality should not suffer when questions are implemented using CATI.

- Though the test design required that field interviewers work only control or only test cases, interviewers in both conditions worked regular ACS production interviews at the same time they completed the Content Test cases. By design the control instrument very closely replicated the ACS production instrument, only differing in the addition of the three newly proposed topics. As a result, interviewers in the test condition had to learn and use two very different instruments, while control interviewers used basically the same instrument between their Content Test cases and ACS production. Thus, test interviewers experienced more challenges in completing their overall caseload. Interviewer debriefing suggested that test interviewers had some difficulty dealing with the two very different instruments simultaneously which may have some impact on the administration of the test version. The test version of the instrument contained the mark-all-that-apply version of the health insurance question.

- The CFU universe did not include non-telephone households and vacant housing units. This only affects those question topics included in the CFU study that are related to the non-telephone household or vacant universes. This is relevant to analysis of the health insurance question in that some respondents on public health care coverage, such as Medicaid, may not have a telephone.

4.2 Limitations Related to the Analysis of Health Insurance Questions on the Content Test

To our knowledge, health insurance questions have never been asked on a major federal mail survey. As a result, while most other subject areas were able to build their test questions from existing questions that had already been used in a mail survey, the health insurance area did not have a standard set of questions to use as a starting point. Thus, the questions presented in this report are likely to have less history under a production environment than other questions that were recommended as a result of the Content Test. Therefore comparisons of these results to similar production style questions may not be applicable.

Also, the analysis in this report is based on two key assumptions which have one main limitation. For one, since the CFU used a more detailed and comprehensive set of health insurance questions, we consider a lower absolute net difference rate between the original interview and the CFU to indicate superiority of a question format. In addition, since the health insurance questions in the Content Test are less detailed than the questions used by most national surveys, underreporting of health insurance coverage is a great concern. Therefore, the set of questions generating higher estimates of coverage is assumed to be superior. The limitation of these assumptions is that, short of obtaining a perfect set of administrative data, we are unable to determine the true health insurance coverage status of the survey respondents.

5. Results

5.1 Response to the Content Test and Content Follow-Up

Control and test treatment groups obtained equivalent response rates overall, and for each mode of collection. Below, Table 1 gives the weighted response rates for each data collection operation and a test of differences between the control and test groups. There were no significant differences between response rates for the control and test groups. Note that the denominator for each calculation included only eligible cases for each mode.

Table 1. Content Test Response Rates, Control vs. Test

Response Rate	Control (%)	Test (%)	Difference (%)	Margin of Error (%)	Significant[3]
Overall response rate	95.8	95.5	-0.3	± 0.9	No
Mail response rate	51.5	51.2	-0.3	± 2.2	No
CAPI response rate	92.6	92.1	-0.4	± 1.7	No
CFU response rate	75.9	76.4	0.5	± 1.6	No

5.2 Forced-Choice vs. Mark-All-That-Apply

5.2.1 Dual Coverage

Respondents with dual health insurance coverage are those who indicated having two or more health coverage types. In the Content Test evaluation of the health insurance question, there was not a direct examination of instances where respondents indicated having more than one coverage type. However, through the evaluation of this question, we noticed that estimates of health coverage status (whether respondents had health insurance coverage) did not differ significantly between formats, while estimates of health coverage type were almost always significantly higher in the forced-choice version than in the mark-all-that-apply version. To understand the importance of this finding, it is essential to know the difference between our definition of health coverage status and health coverage type.

Estimates of health coverage status (whether or not a respondent has health insurance) are defined differently for each version. In the forced-choice version, a respondent is said to have health insurance coverage if they mark "yes" to any of the seven response options, or if they provided a codeable response to the "other" option. In the mark-all-that-apply version, a respondent qualifies as having health insurance coverage if they indicate that they have health insurance coverage in the screener question, if they mark one of the seven health coverage types, or if they provide a codeable response to the "other" option.

In contrast, estimates of health coverage type define which health insurance coverage a respondent has. In the forced-choice version, a respondent is said to have a specific type of health insurance if they indicated "yes" for that type of insurance, or if their response to the "other" category was codeable to one of the original seven categories. Similarly, in the mark-all-that-apply version, a respondent is identified as having a specific health insurance if they indicate the type in the mark-all-that-apply portion of the question, or if their response to the "other" category was codeable to one of the original seven categories.

In other words, for both versions of this question, indicating health coverage type necessarily means that the respondent has health coverage and includes them in the estimate of health coverage status. Therefore it is not possible to have a health coverage type and not have health coverage; however, it is possible for a respondent to indicate that they have health coverage, but not indicate a type of health coverage. This means that the 86% of respondents who indicated having health coverage in both versions (see Table 2), are the same respondents whose health coverage type is analyzed in Table 3.

Table 2 compares estimates of health coverage status for each version of the question, overall, and by demographic characteristics. There was not a significant difference between the two question formats in the overall estimates of health coverage status. Similarly, results across demographic characteristics were mostly non-significant. Respondents who were 65 years old or older produced the only significant difference in estimates of health coverage status. Here, 99% of the forced-choice version respondents in this age category indicated that they have health insurance, while 98% of the mark-all-that-apply version respondents indicated they have health insurance. This is an interesting finding considering there are more opportunities to indicate having health insurance coverage in the mark-all-that-apply version than in the forced-choice version.

[3] A 10% significance level was used for all analyses in this paper.

Table 2. Reported Health Coverage Status by Selected Content Test Characteristics[4]

	Forced-Choice (%)	Mark-all that-Apply (%)	Difference (%)	Margin of Error (%)	Significant
All persons	86.1	85.5	-0.6	± 1.2	No
Age					
Under 18	89.5	88.6	-1.0	± 2.2	No
18-24	68.7	70.0	1.3	± 4.6	No
25-44	78.8	78.3	-0.5	± 2.3	No
45-64	89.6	88.6	-1.0	± 1.4	No
65+	99.3	98.1	-1.2	± 0.8	Yes
Employment Status					
Employed	85.5	85.1	-0.5	± 1.4	No
Unemployed	55.2	55.6	0.4	± 7.7	No
Not in Labor Force	86.2	86.4	0.2	± 1.9	No

Since the overall estimates of health coverage status did not differ between versions, the finding that estimates of health coverage type were significantly higher in the forced-choice version than in the mark-all-that-apply version indicates that more respondents to the forced-choice version marked more than one type of coverage. Table 3 shows estimates of health coverage type for three broad categories as well as their respective sub-categories. For each of the broad categories the forced-choice version produced significantly higher estimates of health coverage type. Only two of the sub-categories had non-significant differences between versions. The direction of these non-significant differences favored the forced-choice version.

Table 3. Reported Health Coverage Types, Forced-Choice vs. Mark-all-that-Apply

	Forced-Choice (%)	Mark-all-that-Apply (%)	Difference (%)	Margin of Error (%)	Significant
Private Coverage	70.7	67.2	-3.5	± 1.7	Yes
Employer Coverage	59.8	57.6	-2.2	± 1.8	Yes
Purchased Coverage	17.3	10.2	-7.1	± 1.2	Yes
Government Coverage	26.2	23.2	-3.1	± 1.4	Yes
Medicare	13.9	11.3	-2.5	± 0.8	Yes
Medicaid	11.3	10.7	-0.6	± 1.2	No
TRICARE	2.8	1.8	-1.0	± 0.6	Yes
VA	2.2	1.1	-1.1	± 0.4	Yes
Indian Health Service	0.4	0.3	-0.1	± 0.2	No
Other Coverage	0.2	0.1	-0.2	± 0.1	Yes

[4] While the results of the Content Test are not strictly comparable to those of national surveys (due to sample size and the test nature of the survey in general), it is important to note that the estimates of health coverage status shown in Table 2 (around 86% for each version) are reasonable when compared to an outside source. For example, in the 2005 National Health Interview Survey (NHIS), 86% of persons reported some kind of coverage at the time of the survey (see NCHS, 2006).

These results indicate that the forced-choice version resulted in more instances of dual health insurance coverage (respondents indicating more than one health coverage type). This finding supports the research of Rasinski et al. (1994) and Smyth et al. (2005) in that the forced-choice format produced more responses than the mark-all-that-apply alternative. As previously mentioned, more responses to the forced-choice version does not necessarily indicate better quality of responses; therefore the systematic error from these two question formats will be discussed in the next section.

5.2.2 Systematic Error

As defined by the net difference rate (NDR), the systematic error of both versions of the health insurance question was determined by comparing original responses to reinterview responses. The NDR for each question format represents the difference in proportions of false positives (false indications of having health coverage status/type) to false negatives (false indications of not having health coverage/type). A negative NDR indicates that the question tends to underestimate the underlying construct, health insurance coverage or status. Traditionally, questions about health coverage status and type tend to underestimate the true health insurance situation; our analysis is no different.

Table 4 presents the NDRs of health coverage status for each version. As most of the NDRs in this table are negative, it is clear that both versions tend to underestimate health coverage status. By taking the difference of the absolute values of the NDRs, a determination as to which version produces less systematic error can be made. In terms of the overall NDR of health coverage status, the forced-choice version produces a significantly smaller NDR than does the mark-all-that-apply version. Meaning, the forced-choice estimate of health coverage status presented in Table 2 is closer to the estimate of health coverage status obtained in the more detailed CFU than the mark-all-that-apply estimate.

Table 4. Health Coverage Status Content Follow-Up Comparison Statistics, Forced-Choice vs. Mark-all-that-Apply

Item	Net Difference Rate				
	Forced-Choice vs. CFU (%)	Mark-all-that-Apply vs. CFU (%)	Diff* |T|-|C| (%)	Margin of Error (%)	Significant
All persons	-1.0	-2.1	1.1	± 0.8	Yes
Age					
Under 18	-2.2	-3.7	1.6	± 2.1	No
18-24	-1.9	-2.7	0.8	± 2.5	No
25-44	-1.1	-1.6	0.5	± 1.2	No
45-64	0.0	-1.7	1.7	± 0.7	Yes
65+	-0.1	-0.5	0.4	± 0.4	Yes
Employment Status					
Employed	-0.5	-1.6	1.1	± 0.7	Yes
Unemployed	-2.5	-2.3	-0.3	± 4.0	No
Not in Labor Force	-0.8	-1.7	0.9	± 1.1	No

*Difference of the absolute values of the test and control net difference rates

In terms of accuracy of health coverage type, Table 5 compares NDRs for each version within three broad categories as well as their respective sub-categories. Almost unanimously, the forced-choice version has significantly smaller NDRs than the mark-all-that-apply version. This result indicates that the occurrences of dual coverage from the forced-choice version found in Table 3 are also more accurate than the estimates of health coverage type from the mark-all-that-apply version. Therefore, not only did the forced-choice format of this health insurance question produce more affirmative response, it also produced more accurate responses.

Table 5. Health Coverage Type Content Follow-Up Comparison Statistics, Forced-Choice vs. Mark-all-that-Apply

Item	Net Difference Rate				
	Forced-Choice vs. CFU (%)	Mark-all-that-Apply vs. CFU (%)	Diff* |T|-|C| (%)	Margin of Error (%)	Significant
Private Coverage	-1.3	-5.0	3.7	± 1.2	Yes
Employer Coverage	-2.2	-5.3	3.1	± 1.1	Yes
Purchased Coverage	-0.1	-6.9	6.8	± 1.2	Yes
Government Coverage	-2.0	-5.3	3.3	± 0.9	Yes
Medicare	-1.4	-4.2	2.8	± 0.6	Yes
Medicaid	-0.4	-2.0	1.6	± 0.9	Yes
TRICARE	-0.4	-0.7	0.3	± 0.4	No
VA	-1.0	-1.4	0.4	± 0.4	Yes
Indian Health Service	0.0	-0.3	0.2	± 0.1	Yes
Other Coverage	-0.4	-3.9	3.5	± 1.1	Yes

*Difference of the absolute values of the test and control net difference rates

6. Conclusions

In comparing the forced-choice format of the health insurance question used in the 2006 ACS Content Test to the mark-all-that-apply format of the question, we found two interesting results. One, while the forced-choice format of the question produced similar estimates of health coverage status as the mark-all-that-apply format, almost across the board, the forced-choice version produced higher percentages of health coverage type than the mark-all-that-apply version. This finding, which supports the work of Rasinski et al. (1994) and Smyth et al. (2005), indicates that the forced-choice format of the question produced more affirmative responses to the health insurance coverage type question than did the mark-all-that-apply version. The second major finding supports the work of Feindt et al. (1997) in that the forced-choice format of the health insurance question also produced more accurate responses than the mark-all-that-apply format. Therefore, this study shows that not only did a forced-choice format produce more affirmative responses, capturing more instances of dual health insurance coverage; these responses were more accurate than the responses to the mark-all-that-apply version. It is our hope that these findings add to the discussion about which question format is better to use in the administration of surveys.

7. References

Asiala M. and Navarro A. (2006). "Experimental Design for the 2006 American Community Survey Content Test," American Statistical Association 2006 Proceedings of the Section on Survey Research Methods [CD-ROM].

Biemer, P.P. and Forsman, G. (1992). "On the Quality of Reinterview Data with Application to the Current Population Survey." *Journal of the American Statistical Association*. 87: 915-923.

Biemer, P.P. and Lyberg, L.E. (2003). *Introduction to Survey Quality*. Hoboken, NJ: John Wiley & Sons, Inc.

Brick, J. M., and West, J. (1992). "Reinterview Program for the 1991 National Household Education Survey." Proceedings of the Section on Survey Research Methodology, 387-392. Alexandria, VA: American Statistical Association.

Bushery, John M., Royce, D., Kasprzyk, D., (1992) "The Schools and Staffing Survey: How reinterview measures data quality," American Statistical Association, Proceedings of the Section on Survey Research Methods, Boston, pp 458-463.

Feindt, P., Schreiner, I., and Bushery, J. (1997). "Reinterview: A tool for survey quality improvement." *Proceedings of the Section on Survey Research Methods* (pp. 105-110). Alexandria, VA: American Statistical Association.

Groves, R.M. (2004). *Survey Errors and Survey Costs*. Hoboken, NJ: John Wiley & Sons, Inc.

National Center for Health Statistics, Early Release of Selected Estimates Based on Data from the 2005 National Health Interview Survey, June 21, 2006. http://www.cdc.gov/nchs/about/major/nhis/released200606.htm#1.

Nelson, C. and Ericson, L. (2007). "Evaluation Report Covering Health Insurance." 2006 ACS Content Test Report Series. January 3, 2007. http://www.census.gov/acs/www/AdvMeth/content_test/P8_Health_Insurance.pdf .

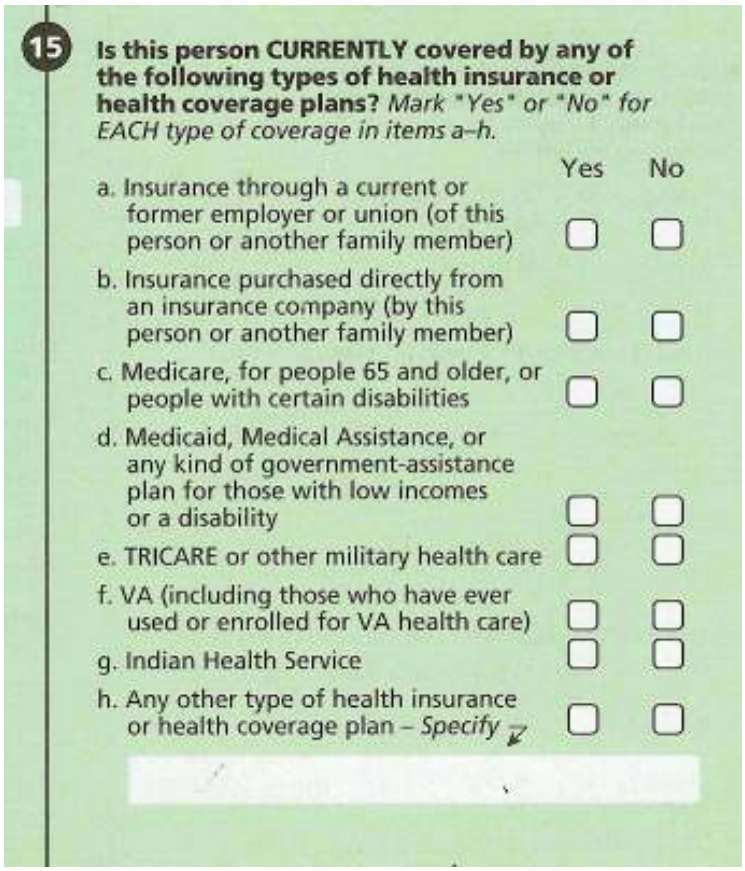
Pascale, Joanne. (2005). "American Community Survey: Cognitive Testing of Health Insurance Questions." U.S. Census Bureau Memorandum. March 25, 2005.

Rasinski, Kenneth A., David Mingay, and Norman M. Bradburn. (1994) "Do Respondents Really 'Mark All That Apply' on Self-Administered Questions?" *Public Opinion Quarterly*. 58: 400-408.

Smyth, Jolene D., Dillman, Don A., Christian, Leah Melani, and Stern, Michael J.. (2005). "Comparing Check-All and Forced-Choice Question Formats in Web Surveys: The Role of Satisficing, Depth of Processing, and Acquiescence in Explaining Differences." *Social and Economic Science Research Center Technical Report 05-029*. Washington State University: Pullman. 30pp.

WESTAT, "Cognitive Testing of Proposed Items on Veteran Status and Health Insurance Coverage for the American Community Survey." Final report prepared for the Census Bureau, May 9, 2005.

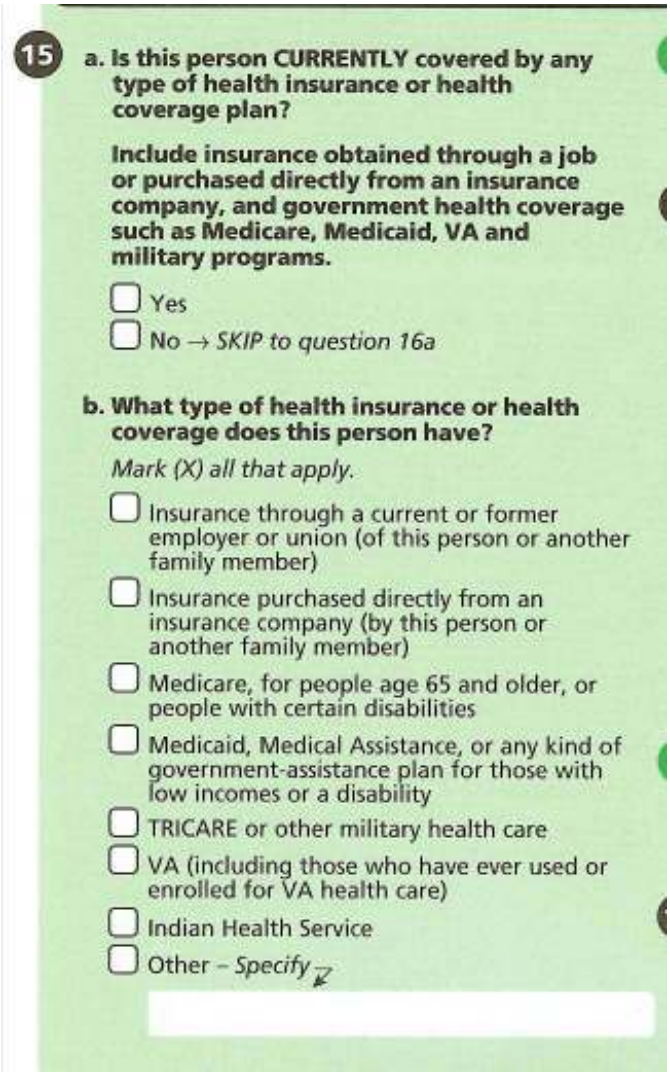
Version 1: Forced-Choice Format



15 **Is this person CURRENTLY covered by any of the following types of health insurance or health coverage plans?** *Mark "Yes" or "No" for EACH type of coverage in items a–h.*

	Yes	No
a. Insurance through a current or former employer or union (of this person or another family member)	☐	☐
b. Insurance purchased directly from an insurance company (by this person or another family member)	☐	☐
c. Medicare, for people 65 and older, or people with certain disabilities	☐	☐
d. Medicaid, Medical Assistance, or any kind of government-assistance plan for those with low incomes or a disability	☐	☐
e. TRICARE or other military health care	☐	☐
f. VA (including those who have ever used or enrolled for VA health care)	☐	☐
g. Indian Health Service	☐	☐
h. Any other type of health insurance or health coverage plan – *Specify*	☐	☐

Version 2: Mark-All-That-Apply Format



15 **a. Is this person CURRENTLY covered by any type of health insurance or health coverage plan?**

Include insurance obtained through a job or purchased directly from an insurance company, and government health coverage such as Medicare, Medicaid, VA and military programs.

- ☐ Yes
- ☐ No → *SKIP to question 16a*

b. What type of health insurance or health coverage does this person have?

Mark (X) all that apply.

- ☐ Insurance through a current or former employer or union (of this person or another family member)
- ☐ Insurance purchased directly from an insurance company (by this person or another family member)
- ☐ Medicare, for people age 65 and older, or people with certain disabilities
- ☐ Medicaid, Medical Assistance, or any kind of government-assistance plan for those with low incomes or a disability
- ☐ TRICARE or other military health care
- ☐ VA (including those who have ever used or enrolled for VA health care)
- ☐ Indian Health Service
- ☐ Other – *Specify*