SHADAC Analysis of the

Survey of Health Insurance and Program Participation

March 21, 2013

Michel Boudreaux
Brett Fried
Joanna Turner
Kathleen Thiede Call

Report to: Prepared by:

www.census.gov www.shadac.org

Acknowledgements: We are grateful to Brett O'Hara, Joanne Pascale, Amy Steinweg, and analysts at the U.S. Census Bureau for involving us in this important project, as well as their guidance and analytical support.

Executive Summary

Under a contract with the U.S. Census Bureau, researchers at the University of Minnesota's State Health Access Data Assistance Center (SHADAC) collaborated with Census Bureau experts to evaluate possible improvements to the measurement of health insurance coverage in the Annual Social and Economic Supplement to the Current Population Survey (CPS). The experimental (EXP) questions were fielded and contrasted with existing CPS and American Community Survey (ACS) questions in a test called the Survey of Health Insurance and Program Participation (SHIPP). The experiment relied on data from two frames: a nationally representative random-digit-dial (RDD) frame and a stratified sample drawn from Medicare enrollment files. This report is a deliverable under contract # 000000033114. Chapter 1 provides details on the background and variable validation of the SHIPP, Chapter 2 provides an evaluation of all (past) year coverage from the CPS and Experimental (EXP) arms, and Chapter 3 provides an evaluation of point-in-time coverage from the ACS and EXP arms.

Chapter 1: Background and Variable Validation

- The first task included consideration of the need to weight the SHIPP data and development of a strategy for creating survey weights. This task was driven by larger than expected differences between treatment arms on key demographic variables. After investigating the strengths and weaknesses of weighting the SHIPP data, SHADAC concluded that weighting provided little benefit with a split-ballot test as the goal was to test differences between treatment arms, not to generalize to the larger population. Weighting would add variance to the estimates, decreasing the power to detect differences. However, weighting the RDD sample for future face validation of the EXP duration measures may have benefits.
- Validation of programming revealed two major problems in the CATI programing.
    - The EXP grid indicators that were intended to capture coverage at any point in the 2009 calendar year actually included coverage from January 2009 to the date of the 2010 interview. The inclusion of months from 2010 resulted in an over count in "all-year" coverage for the EXP arm in the grid variables compared to the raw source variables. In addition we also found that the month-person-level grid variables did not always match the source variables.
    - The EXP instrument did not force policyholders to be included in the list of "followers" defined in the COVWHO variable series (the first person associated with a plan is termed a "leader" and all subsequent people associated with the plan are termed "followers" and listed in COVHWO). This lead to two problems:
        - The overall count of people with coverage at any point from January 2009 to the interview date is biased downward if policyholders are not explicitly included in the plan.
        - Months of coverage are not obtained for policyholders that are not "leaders" and are not explicitly identified by respondents in COVWHO.
- As a result of parallel programming by the Census Bureau and SHADAC changes were made to the CATI specification.

- The Census Bureau decided to remove the "grid" variables (indicators generated by the CATI system) and instead use a counter variable to drive the instrument. The grid variables kept a running track of all plan types at the month-person level. The counter, by comparison, does not track the plan type but only the number of plans at the month-person level. The instrument then uses this information to determine whether questions on additional coverage should be asked. The recoding into person/month/plan type variables will be completed using the "source" variables (survey responses) in post-processing.
- In the COVWHO (the list of plan members) and OTHMEMB (an indicator of the presence of other plan members in the household) variables, if the leader is a dependent, a probe to "please include the policyholder" will be added. Hard edits were also developed so COVWHO is filled with the policyholder if the respondent does not choose him or her and OTHMEMB is filled with "yes" if not selected.

- A final issue is the treatment of missing values. Currently we cannot differentiate between missing because the respondent did not answer the question or does not know how to answer the question (should be imputed) and missing because the respondent was not in the universe (did not receive the question). This is something the Census Bureau is working to resolve.

Chapter 2: CPS vs. EXP Results

- Analysis of the CPS and EXP all year insurance measures showed little evidence of differences between the CPS and EXP instruments.
- Adjusting for differences in demographic characteristics across arms did not change the substantive conclusions of the analysis and mainly served to marginally decrease the differences in coverage between arms.
- The exception to an overall trend of non-significant differences was that the EXP measured fewer instances of double coverage in the pooled (RDD and Medicare) and RDD samples.

Chapter 3: ACS vs. EXP Results

- Analysis of the ACS and EXP point-in-time insurance measures showed evidence for differences between the instruments for some insurance measures.
- Adjusting for differences did alter some of the substantive conclusions drawn from the unadjusted analysis.
- Adjusted analyses showed that in the pooled sample the EXP arm had more employer sponsored insurance (ESI), less direct purchase, less Medicare, more private, less public, and less double coverage compared to the ACS arm.

Chapter 1: Background and Variable Validation

1.1 Introduction

Chapter 1 provides details on the background and variable validation of SHADAC's evaluation of the Survey of Health Insurance and Program Participation (SHIPP). The chapter includes background on the survey and analytic sample, discussion of the consideration of weighting, and details on the variable validation and resulting improvements to the survey instrument. The chapter concludes with suggestions and potential challenges for the Census Bureau to consider when processing the data and creating a public use file, if the experimental questions are added to the Annual Social and Economic Supplement to the Current Population Survey (CPS).

1.2 SHIPP Background

The SHIPP was a randomized split-ballot test of an experimental health insurance series designed for inclusion in the CPS. The experiment was fielded between late March of 2010 and early May of 2010.

The SHIPP sample was drawn from two frames. The first frame was a nationally representative random-digit-dial (RDD) frame from which 3,081 completed household interviews (or data on 7,493 persons) were obtained. The second frame was an address frame of households that include a Medicare beneficiary enrolled as of May 2009 provided by CMS (Telematch was used to link these addresses to telephone numbers) from which 2,295 completed household interviews were obtained (or data on 5,250 persons). The Medicare sample was stratified to over-sample the households of non-elderly disabled persons and the households of persons enrolling after January 2009 (recent and younger Medicare enrollees and their households). Because data was collected for the entire household in both frames, the Medicare sample includes household members not enrolled in Medicare.

The experiment consisted of three arms. The ACS arm included the health insurance series from the American Community Survey. The ACS measures types of health insurance coverage held at the date of the interview (point-in-time). The CPS arm included the health insurance series from the Current Population Survey (CPS). The CPS series identifies types of health insurance coverage held at any point in the previous calendar year. Finally, the EXP arm contained the experimental health insurance series being considered as a replacement for the CPS health insurance questions. The EXP series measures types of health insurance held at the date of the interview as well as each month from January of the previous calendar year to the date of interview. In each arm the health insurance series was preceded by demographic and socio-economic questions borrowed from the CPS. Additional detail on the SHIPP design and procedures can be obtained from Pascale (2012) and Pascale (2009).

Ultimately the SHIPP experiment cannot reveal if the new experimental series is valid because there is no gold-standard. The value of the SHIPP is in providing a proof-of-concept that will inform future operations. A close examination of the data helps iron-out Computer Assisted Telephone Interviewing (CATI) programing and identify potential areas of concern. Further, the SHIPP test can broadly, if not conclusively, illustrate if the new series meets our prior beliefs about how it was intended to perform. For example, years of research on the CPS health insurance series has revealed that its calendar year reference period, its “laundry list” insurance menu, and its household centered design leads to under-reports of coverage and an insurance rate that is thought to be biased downward (Pascale, 2008). Therefore, our belief prior to analyzing the SHIPP data was that the experimental arm (EXP) would produce a higher insurance rate as it avoids these measurement problems.

1.3 Analytic File

To facilitate analysis of the SHIPP data by SHADAC, the Census Bureau created a public use file that met disclosure review criteria. For disclosure, the public use file was modified in two ways as compared to the internal file. The Census Bureau created a geographic variable containing groups of states and removed health insurance information for people age 65 and over. After receipt of the file, SHADAC made two additional alterations to facilitate analysis. Missing age and gender observations were imputed and households with a person age 65 and over were removed.

Creation of Geographic Variable

Because including a full set of state indicators did not meet disclosure review, the Census Bureau and SHADAC collaborated on the creation of a broader geographic variable. This geographic variable classifies states into three categories: 1) states in which the production CPS and the production ACS did not produce a significantly different uninsurance rate in any year between 2008 and 2010; 2) states in which there was a difference in one year; and 3) states where there was a difference in two or more years. This variable was chosen to check for differences between arms due to geographic differences in respondents’ understanding and characterization of their health insurance coverage in surveys. While such geographic patterns are not fully understood, they could result from a number of factors including the complexity of the local health insurance market. Appendix A provides a list of the states grouped into the three categories and provides a discussion of the different options considered for this variable.

Removal of Information for People Age 65 and Over

The Census Bureau did not release health insurance data for people age 65 and over. Because this group is nearly universally covered by the Medicare program, it was decided that there was a high disclosure risk for uninsured elderly observations. Therefore, all health insurance data is missing for elderly people, but their demographic and economic data is included. Below we discuss how this feature of the public use data was treated for the purpose of SHADAC’s analytical work.

Imputation of Age and Gender

SHADAC analysts imputed missing age and gender values using hot deck imputation mainly to reduce sample loss associated with case-wise deletion and to facilitate possible future weighting activities.

Gender was missing for only a few cases (n=9; 0.1%). The hot deck used treatment arm to impute gender.

Roughly 1,170 (9.2%) cases were missing continuous age values. There was some minimal variation by arm, but it was not significant. In addition to continuous age, the questionnaire included a categorical age variable that classified observations as 0-14, 15-64, and 65+. Only 7 cases (0.05%) were missing from the categorical age variable which allowed for the continuous age imputation to draw values within the bounds of the reported categories.

The hot deck for children (identified using the categorical variable) used treatment arm to impute continuous age. Non-elderly adults were imputed using gender, relationship to respondent and educational attainment. Continuous age for elderly adults, who were eventually removed from the file as described below, were imputed with gender and disability status. The hot deck method imputed a single version for both age and gender.

Removal of Households from the File

Since the Census Bureau deleted health insurance data for elderly people and some of the records for elderly people also contain health insurance information about other members of the household, health insurance data was missing for some non-elderly records. In the EXP arm some household members, called “followers” in SHIPP parlance, are associated with a health insurance plan through another household member. For example, person 1 in household A might be reported to have an ESI plan. If later in the instrument person 2 is reported to be the policy holder (or dependent) in person 1’s plan then person 2 would be termed a follower. In this situation the information about person 2’s health insurance would be carried on person 1’s record. Now consider that person 1 is 66 years old and person 2 is 63 years old. Due to the suppression rules used to create the public use file, information for both person 1 and person 2 would be suppressed, even though person 2 is non-elderly. To overcome this challenge, SHADAC excluded all households that had one or more elderly persons.

In addition to removing elderly households, 10 observations that were coded with relationship values of “not in universe” (NIU) were excluded. These cases appeared to be non-interviews that were missing data for a majority of variables.

After excluding observations from households with any elderly and all 10 NIU cases, the full public use file sample was reduced from 12,743 cases to 7,461 cases. The reduction varied slightly, but significantly by treatment arm ($p < 0.051$). The selected analytic sample was 60.0% of the full CPS arm; 58.2% of the full ACS arm; and 57.5% of the full EXP arm. Given the relatively large reduction of the analytic sample on the public use file and it’s non-representativeness of the full internal data, our analytical findings of the split-ballot test may not generalize to analyses conducted by the Census Bureau on the full file.

1.4 Feasibility Analysis of Weighting

The SHIPP sample is not representative of the U.S. population or of the CPS sample. The RDD sample has coverage error because it only includes landline households. A large proportion of U.S. households are wireless only and a small portion of households lack any telephone service.[1] Not including a cell phone sample is only an issue because of potential bias that result from differences in characteristics

[1] 26.6% of households were wireless only and 2.0% lacked any phone service in the first half of 2010 per the National Health Interview Survey (http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201012_tables.htm#table1)

between people living in cell phone only households and those living in households with a landline. Adults living in cell phone only households are more likely to be young, renters, low income, Hispanic, binge drinkers, current smokers and to be uninsured (Blumberg and Luke, 2012). The Medicare sample, which consisted of over-samples of young, disabled and recently enrolled beneficiaries, is representative of a subset of all Medicare enrollees and their households. This select population has a unique profile that is unlike the profile in the general Medicare population. The Medicare sample has coverage error because individuals were only included if someone in their household had a physical address on file at CMS and had a corresponding landline number. Weighting could be used to attempt to adjust for this coverage error and to bring sample characteristics for the sample frames in-line with the general U.S. and Medicare populations.

SHADAC prepared a document (see Appendix B) providing suggestions for implementing a weighting strategy. However, after thoroughly investigating the feasibility, disadvantages and advantages of weighting the SHIPP data, SHADAC recommended against weighting the data as it would not benefit the split-ballot analysis (see Appendix C). The split-ballot was designed to maximize the internal validity of the comparison of the three health insurance question series. Deriving estimates from the SHIPP that are nationally representative was not a goal of the project. Weighting is designed to maximize generalizability (or external validity) particularly of prevalence estimates rather than regression coefficients. Weighting is not a statistically efficient method for improving internal validity. Moreover, it is unlikely that weighting would fully adjust for bias introduced due to coverage error. Sampling error would be introduced through the increase in the design effect from weighting. We expect that the design effect from weighting would be especially large in the SHIPP because of the overlapping frames (it is conceivable that cases in the RDD would be included in the Medicare frame and vice versa). It's important to keep in mind that with weighted or un-weighted data, any differences observed between the EXP and CPS arm may not be generalizable to what we would see in the actual CPS fielding environment or the general population.

Nonetheless, weighting the RDD only may have a role in future activities. For example, weighting would allow for a face validity analysis of the duration of coverage measures and the recall task in the experimental series. By weighting the RDD, we could then compare duration of coverage measures from the experimental series to an external source such as the Medical Expenditure Panel Survey Household Component (MEPS/HC) or the Survey of Income and Program Participation (SIPP).

1.5 Internal and External Validation

SHADAC's initial validation test was for internal consistency of the "grid" variables. After finding an error in the way these variables were coded, SHADAC began an in depth external validation of the health insurance variable coding based on a parallel programing process.

1.5.1 Internal Validation of the "Grid" Variables

SHADAC's first analytical task was to ascertain the quality of the grid variables generated by the SHIPP CATI. The grid variables are a set of indicators generated by the CATI system. Their inputs can come from multiple questionnaire items and are populated at the time of the interview. The grid variables serve two purposes. They facilitate the complicated skip pattern within the EXP instrument. They serve as a look-up reference so that the instrument can determine what path a respondent should follow. The second purpose of the grid variables was to provide an easily accessible set of indicators to be used by analysts when the final data file is processed. In contrast to the grid variables, the “source” variables contain the raw values generated by the respondent’s answer to each item. Preliminary SHIPP reports produced by the Census Bureau were based on the grid variables.

To validate the grid variables, SHADAC began by determining if the grid variables were internally consistent with each other. For example, if monthly coverage indicators aligned with the all year summary variables in the EXP and if the specific coverage type indicators (i.e. employer sponsored insurance (ESI), direct purchase, etc.) aligned with the aggregate coverage measures (i.e. private, public, insured). During this process we discovered that the EXP indicator that was meant to indicate coverage in any month in 2009 was programmed incorrectly. It counted coverage in any month from January 2009 to the interview month which occurred in March through May of 2010. This error caused an over-count of coverage (of any type) in the EXP treatment arm: 60 cases in the public use file analytical sample. Once this error was corrected, the proportion of cases with any indication of insurance in 2009 dropped by over 2 percent.

In order to address this problem a joint decision was reached between the Census Bureau and SHADAC to forgo further use of the grid variables in the remaining analytical tasks and to instead rely entirely on the source variables.

1.5.2 External Validation of the “Source” Variables

SHADAC’s second major analytic task was validating the recoding of the source variables from each arm into meaningful coverage indicators. The task had two goals. First, to agree on recoding rules and second to verify that recoding based on those rules was correctly implemented.

Recoding Decision Rules

Four major recoding decision rules had to be agreed upon between the Census Bureau and SHADAC: 1) developing a coverage aggregation scheme, 2) deciding how to treat missing health insurance coverage, 3) coding write-ins, and 4) deciding how to treat missing months of coverage in the EXP arm.

1) *Coverage Aggregation.* A coverage aggregation strategy was devised to place respondents with specific coverage types (i.e. ESI, direct purchase, Medicaid) into aggregated coverage groupings (i.e. private, public, uninsured). Table 1.1 describes the agreed upon aggregation scheme.

Table 1.1 Health Insurance Coverage Aggregations

Private	Public	Other	Uninsured
ESI	Medicare	A residual catch all for coverage from write-ins that could not be put into another group	A residual category for people without any coverage
Direct purchase	Medicaid/CHIP		Indian health service coverage is not considered a comprehensive insurance type
Coverage from a non-household member (“Out”)	All state and other government programs		
School	Tricare or VA		

2) *Missing Health Insurance Type.* A decision was made about how to treat missing health insurance coverage data. Differentiating missing as non-response (“Don’t Know/Refused”) versus a missing value generated by the skip pattern is relatively easy in the CPS and ACS, but much harder in the EXP. A joint decision was reached to group missing values in the insurance variables with “0 = No”.

This method simplifies the coding in the short-term, but could lead to biased results when comparing treatment arms as we cannot compare the true number of “no” responses. Implications of this coding decision on comparisons of the EXP with the CPS and ACS are discussed further in Chapters 2 and 3 (sections 2.2 and 3.2).

3) *Write-Ins.* We created a common strategy for coding write-in responses. This process involved one study team member suggesting a recode for each write-in response and then discussing any concerns among all Census Bureau and SHADAC team members and revising the recode accordingly. Members of the Census Bureau production team that frequently work with this type of coding were involved with this process. The CPS had the fewest write-ins (21), then the EXP (27), followed by the ACS (72). These frequencies are not directly comparable. Since the CPS and EXP are household-centered instruments there could be one write-in for several members. Whereas in the ACS, a person centered instrument, the same respondent would have to write-in a response for all plan members. One of the advantages of the EXP is that it provides many opportunities for respondents to identify the source of their coverage before providing a potentially ambiguous write-in response.[2] However, the drawback of this approach is that it creates more variables that an analyst must consider when defining health insurance coverage recodes.

4) *Missing Months of Coverage.* The final aim of the recoding rule task dealt with the treatment of missing months of coverage information in the EXP. Amy Steinweg of the Census Bureau developed a strategy that was agreed to by SHADAC. These rules are summarized in a table produced by Amy Steinweg and included in Appendix D. Briefly, any coverage that was reported in a section of the EXP instrument intended for gathering information about currently held coverage was assigned point-in-time coverage even if the specific months of coverage were missing. (These questions were all phrased to ask specifically about coverage NOW.) Similarly, any coverage that was reported in a section of the instrument intended for gathering information on coverage that ended sometime prior to the interview was

[2] For example, if a respondent provides a verbatim write-in response of “coverage from plan X” it is not always possible to identify if plan X is private or public coverage because that insurer (e.g., Blue Cross Blue Shield) may carry products in both private and public markets.

assigned previous year coverage (what this report terms “all-year” coverage). (These questions all asked either about PAST coverage, or else specific past months.)

Verification of Health Insurance Coding

The second major goal of the external validation task was to verify that all health insurance recoding correctly reflected the intended specifications and to identify potential areas of confusion. This section describes: 1) the validation strategy; 2) one confirmed error discovered with the CATI specification; 3) additional potential errors with the CATI specification; and 4) final Census Bureau and SHADAC validation results.

1) *Validation Strategy*. The verification was accomplished with extensive collaboration between SHADAC and Census Bureau analysts. The validation process involved multiple steps: Census Bureau and SHADAC analysts wrote parallel code for all insurance measures without knowledge of the others’ specific coding process; comparison of frequencies and discussion of potential reasons for discrepancies; fixing any errors brought to light in the comparison of frequencies; and finally, trading data files and cross tabulating the merged Census Bureau and SHADAC recodes. Agreement between Census Bureau and SHADAC values in micro-data files was used as the final criterion of validation. All Census Bureau programing was conducted in SAS and all SHADAC coding was conducted in Stata. The following health insurance recodes were verified:

- Insured at any point in the past year (CPS and EXP)
- Coverage type in the past year (CPS and EXP)
- Insured at point-in-time (ACS and EXP)
- Coverage type at point-in-time (ACS and EXP)

Validation of the CPS and ACS arms was relatively simple, but validation of the EXP was much more complex. One way to demonstrate the relative complexity of the EXP compared to the CPS is to consider the number of input variables needed to produce an estimate of a coverage type during a specific time period. In the CPS each insurance type requires roughly 13 input variables (a main question, 6 “other plan” variables and 6 “verification plan” variables). An individual in the EXP can obtain coverage from one of 9 loops and each loop contains roughly 3-5 questions where a respondent could indicate a coverage type. And unlike the CPS, the EXP gathers information about coverage in each month of the 17 month look back task. This requires an analyst to consider 3-5 additional variables for each of the 9 loops to tabulate an estimate of coverage. This roughly implies that there could be up to 225 input variables in an assignment of coverage type in a specific month. While the flexibility of the EXP instrument ensures that the vast majority of respondents will not be sent through all 225 data collection points, to correctly code the data an analyst must consider the most complicated scenario the instrument could theoretically produce.

2) *Confirmed CATI Error*. During the EXP validation SHADAC discovered that the CATI did not force policyholders to be listed as followers (when the policyholder was not a leader) in the COVWHO variable. This is demonstrated in the following question series.

- The respondent is asked a series of questions about coverage.

- If coverage type is military, ESI, or direct purchase a question is asked to identify the policyholder (this can be the leader or someone else).
- Then a question is asked to determine if anyone else in the household is also covered by this coverage type and their line number is listed in the COVWHO variable series (if they are not mentioned here by the respondent, they are not captured in COVWHO).

This CATI misspecification created two problems. First, it meant that analysts would have to explicitly consider the policyholder variables to avoid undercounting coverage. In several rounds of coding, both Census Bureau and SHADAC analysts had assumed that COVWHO captured all plan members and were therefore misclassifying coverage. For all year coverage this error caused an undercount of approximately 94 cases in the analytic sample that should have been counted as covered by either military, ESI, or directly purchased coverage.

The second major problem is that information about months of coverage is driven by the COVWHO list. The EXP questionnaire obtains months of coverage by first asking if all members of the plan (defined by COVWHO) are covered in the same months as the leader if the leader's coverage was continuous from some previous time point to the date of interview. If the leader's coverage was not continuous or if at least one follower had different months of coverage than the leader, then information about months of coverage is specifically asked of each person listed in COVWHO. Since some policyholders were incorrectly absent from the COVWHO list (depending on how the respondent interpreted COVWHO), their months of coverage were not obtained in the interview.

As a temporary fix to this problem, the Census Bureau and SHADAC agreed to assign coverage to policyholders that were associated with a plan in which any other plan member had coverage at some point during 2009 (for all year coverage) or had coverage during the interview month (for point-in-time coverage). This method may undercount coverage in the EXP. By definition, policyholders have at least as many months of coverage as their dependents, but they might have more.

3) *Unconfirmed CATI Errors.* There were a handful of other isolated instances that demonstrated a potential error in the CATI. However, because the problems only affected a small number of cases it was decided that it was out of the scope of this analysis to determine if these were real CATI errors, programming errors on the part of SHADAC, or some other anomaly. SHADAC is available to discuss these cases in further detail at the request of the Census Bureau.

- In household HSEQ=4953 there are values for both ESI and Medicaid within the same loop. SHADAC's interpretation was that it was not possible to have two plan reports in the same loop.
- Household HSEQ=4147 has multiple inconsistencies. For example, the leader indicates coverage in the VERIFY question of the LC1 loop. However, they have missing values in all LC2 variables, perhaps indicating they never were asked a follow-up question. They nonetheless end up in a months of coverage question.
- In household HSEQ=4655 person 2 is a policyholder on an LC2 Military plan. The leader is the dependent. The leader (person 1) says they have different months of coverage. They report *fewer* months for the policyholder than they do for themselves. We are unaware if there are any CATI rules that would prevent a dependent from being covered for more months than their

policyholder, but such a situation is logically impossible. HSEQ=3869 has a similar pattern of policyholder and dependent coverage.

4) *Validation Results.* In early January of 2013 the parallel programing was completed. There remained a handful of EXP cases that were not validated, but the Census Bureau and SHADAC determined that resolving these cases would not provide additional information about the validity of the EXP CATI instrument. Tables 1.2 and 1.3 describe the number of cases identified by plan-type for both the all year and point-in-time measures. Appendix E contains specific notes on each case that could not be validated and suggests reasons for why the discrepancy remains. These notes were also delivered directly to the Census Bureau's SHIPP team.

The first two columns of Tables 1.2 and 1.3 report the number of cases by plan type identified by the Census Bureau and SHADAC. The third column reports the difference in the frequencies. The final two columns report the number of cases, identified by either the Census Bureau or SHADAC that was not confirmed by the other analysis team. For example, Table 1.3 shows that the Census Bureau and SHADAC identified the same number of observations with point-in-time Military coverage. However, the Census Bureau coded 1 case to Military coverage that SHADAC did not and SHADAC coded a different case to Military that the Census Bureau did not.

Table 1.2. Validation Results of All Year Coverage from the EXP

	Number of Cases		Difference in Count	Number of Cases Not Validated	
	Census Bureau	SHADAC		Census Bureau	SHADAC
Insured	2,100	2,101	-1	0	1
ESI	1,527	1,528	-1	0	1
Direct Purchase	173	173	0	0	0
Military	89	89	0	0	0
Medicare	282	282	0	0	0
Medicaid and Other Means-Tested	295	295	0	0	0
Other	12	2	10	10	0
Out-of-Household	51	51	0	0	0

Based on 2,365 observations in the EXP analytic sample.
"Number of Cases Not Validated" is the number of cases identified by either SHADAC or the Census Bureau that were not identified by the other.

Table 1.3. Validation Results of Point-In-Time Coverage from the EXP

	Number of Cases		Difference in Count	Number of Cases Not Validated	
	Census Bureau	SHADAC		Census Bureau	SHADAC
Insured	2,094	2,088	6	12	6
ESI	1,491	1,490	1	3	2
Direct Purchase	167	170	-3	0	3
Military	79	79	0	1	1
Medicare	294	293	1	1	0
Medicaid and Other Means-Tested	301	299	2	4	2
Other	7	1	6	6	0
Out-of-Household	43	43	0	0	0

Based on 2,365 observations in the EXP analytic sample.
"Number of Cases Not Validated" is the number of cases identified by either SHADAC or the Census Bureau that were not identified by the other.

1.6 CATI Modifications from Validation Results

Based on validation by Census Bureau and SHADAC analysts, the Census Bureau made corrections and improvements to the CATI specification before the planned 2013 field test. While not comprehensive, these changes include the following:

- The Census Bureau decided to remove the "grid" variables and instead use a counter variable to drive the instrument. Besides "too many" months being counted by the grid variable, the grid variable responses also did not always match those of the source variable. One potential reason identified by the Census Bureau is that when an interviewer had to "back-up" in the CATI to change the response to a question that had already been asked, the new "correct" response was not always reflected in the grid variable. The Census Bureau has addressed this issue by changing the grid variable so that it no longer tracks the plan type but only the number of plans at the month-person level (i.e., replaced the "grid" variable with a counter variable). The recoding into person/month/plan type variables will be completed using the "source" variables in post-processing.
- In the COVWHO and OTHMEMB[3] variables, if the leader is a dependent, a probe to "please include the policyholder" will be added. Hard edits were also developed so COVWHO is filled with the policyholder if the respondent does not choose him or her and OTHMEMB is filled with "yes" if not selected. These adjustments are for the content test in 2013 but the Census Bureau hopes to have the CATI automatically fill in COVWHO with the policyholder if the experimental CPS questions are used in 2014.
- Due to the treatment of missing values, we cannot differentiate between missing because the respondent did not answer the question or does not know how to answer the question (should be

[3] OTHMEMB is a variable that indicates if there are other household members on the same plan.

imputed) and missing because the respondent was not in the universe (did not receive the question). This is something the Census Bureau is working to resolve but the timeline is uncertain.

1.7 Suggestions for Post-Processing and Creating a General Public Use File

If implemented in the CPS, the experimental question series will provide an important new tool for monitoring and evaluating the Affordable Care Act. In addition to studying past year and point-in-time coverage the experimental questions will allow for analysis of duration of coverage and churning between coverage types. While the experimental question series allows for more analysis opportunities, the instrument specification and processing required is much more complicated than the current CPS. The evaluation of the SHIPP test by Census Bureau and SHADAC analysts has identified areas for improvement and the Census Bureau has already made improvements to the CATI instrument. Below we list potential features of a public use file that would be valuable to us as data consumers.

- To the extent possible involve outside experts and data consumers in developing logical editing, write-in coding, and allocation routines. Forming a “Technical Advisory Group” (TAG) could improve the content of editing rules given external content expertise and it would facilitate “buy-in” from the user community. In the production CPS the editing and allocation rules have a non-trivial impact on estimates. Building consensus in the user community on editing procedures would help build trust in the new instrument.
- To the extent possible make all variables on the final public use file person level. Avoid including information about one observation on another observation’s record unless that information pertains and is coded to everyone in a common unit like a family. For example, the leader/follower structure would be hard for a general data analyst to follow. However, the use of “line number” pointers can be valuable. For example we recommend including a line number variable that identifies a dependent’s policyholder.
- Provide month level variables if allowed by disclosure review or variables that allow for some level of duration and churning analysis as this is one of the major strengths of the experimental CPS question series. The TAG could help inform coding that meets disclosure review, for example selection of derived measures of duration such as “uninsured for 6 months or more” or aggregated yet time specific coverage measures, such as “uninsured in quarter 1” etc.
- Provide variables with as much detail as allowed by disclosure review so that analysts can make their own decisions, but also create recodes that the Census Bureau thinks are appropriate. For example: public, private, uninsured, ESI policyholder, ESI dependent, direct purchase, etc. Provide the recoded variables on the public use file in addition to documentation for how the variables were created. This will reduce analyst error when working with the files.
- Provide both a CSV and SAS file with variable labels or provide a .dat file with programs to read in the data.

Chapter 2: Comparison of CPS and EXP

2.1 Introduction

Chapter 2 presents a comparison of all year coverage measures from the Current Population Survey (CPS) arm to the Experimental (EXP) arm in the SHIPP survey.

2.2 Data

We use the validated SHIPP public use file to compare all year health insurance coverage measures from the CPS and EXP arms. Our analysis is based on an analytic sample of 5,001 records from the pooled RDD and Medicare frames. Households with at least one person age 65 and over are removed because there is incomplete information for these households due to disclosure review. More discussion of our sample selection criteria is given in Section 1.3. The Stata code used to create the tables for Chapters 2 and 3 was delivered to the Census Bureau in do files.

All missing values for the health insurance variables were coded to "0=No coverage". We interpret variables coded in this manner as "any indication of coverage". This approach to treating missing values might produce biased split-ballot results if the level of missing data varies between treatment arms. For example, two arms could have an equal proportion of explicit "no" responses, but an unequal proportion of missing values. Using our coding rules we would infer that the proportion of "no" responses was unequal between arms when in fact it was equal. This is particularly concerning for comparisons of uninsurance as that concept is operationally defined as the absence of coverage.

We do not report uninsurance as we are unable to differentiate between which respondents answered "No" to all the insurance questions and which respondents had missing data that was grouped with "0=No". For example, suppose we have a hypothetical data set where 6 out of 10 individuals have some type of coverage. All we know about the remaining 4 individuals is that they were not coded as "yes" in any of the coverage questions. Therefore we can say with certainty that 60% were reported as having insurance. We could not say that the other 40% were uninsured, *per se*, because one or more of them may have had coverage that was reported as "DK/Refuse." We examine 12 health insurance measures: 7 individual insurance types and 5 aggregated measures. The 7 insurance types are:

- Employer sponsored insurance (ESI)
- Direct purchase coverage (includes coverage from school)
- Military coverage (VA and Tricare)
- Medicare
- Medicaid and other means-tested coverage
- Out-of-household coverage (any ESI, direct purchase, or military coverage where the policyholder lives in a separate household)
- Other (any coverage that could not be categorized into the preceding coverage types)

The 5 aggregated coverage categories are:

- Insured
- Private coverage (includes ESI, direct purchase, and out-of-household)
- Public coverage (includes Military, Medicare, and Medicaid)
- Any two or more types (excluding out-of-household coverage as that coverage type conceptually implies ESI, direct purchase, or Military coverage)
- Private and public coverage in combination

All coverage measures refer to coverage held at any point in calendar year 2009. Respondents can be categorized into more than one insurance type.

2.3 Methods

The SHIPP test randomly assigned respondents to the EXP or CPS treatment arm. The interviewers and fielding periods of the two instruments were also randomized to mitigate interviewer and seasonality bias. Randomization should ensure that the differences we report are not confounded by any observed or unobserved factor that is correlated with treatment arm and insurance type.

To investigate the success of the randomization procedure we compared a number of demographic characteristics across treatment arm. The demographic characteristics we examine are known to be related to health insurance. The distribution of the covariates was compared across treatment arms using a chi-squared test of association. The assumption of this test is that balance between observed covariates implies balance between unobserved covariates.

The results of this comparison, in the fully pooled sample (composed of both the Medicare and RDD frame), are presented in Table 2.1. Of the eight characteristics we considered only race/ethnicity was significantly associated with treatment arm at the $p \leq 0.05$ level. The EXP arm had a higher percentage of whites and lower percentage of blacks compared to the CPS arm. These results are encouraging as they suggest that randomization was successful. Finding one significant difference among eight comparisons is likely due to chance.

Despite the lack of significant associations found in Table 2.1 we used logistic regression to compare health insurance measures across treatment arm, controlling for all the variables found in Table 2.1. Regression has two advantages: it controls for any observed characteristic that could plausibly bias our comparison and it increases the statistical precision of our analysis. We fit a total of 30 equations. We compared 10 coverage types ("out-of-household" and "other" coverage was not examined using regression due to a lack of sample) in the pooled data and in the Medicare and RDD frames separately. We present our model results using predicted values for each arm derived using average marginal effects. The standard errors we report are clustered at the household level and all comparisons of coverage were conducted using a t-test. We make several comparisons and run several models in this analysis. While we report and comment on differences that met a significance standard of $p \leq 0.1$ it should be noted that under that standard at least 1 test in 12 should be significant purely due to chance. We supply the p-values for each of our comparisons so that the reader can adjust the p-value for multiple comparisons.

2.4 Results

The top panel of Table 2.2 presents the unadjusted percent of each coverage type by treatment arm in the pooled analytical sample. The majority of the 12 comparisons failed to reach a meaningful level of statistical significance. The point estimates suggest that the EXP captured 1.1 percentage points more insured compared to the CPS. However, we are unable to reject the hypothesis that the difference in the population is 0. Having two or more types of coverage types did prove to be significant. The percent of respondents that had any two or more coverage types was 1.9 points lower in the EXP and the result was marginally significant ($p \leq 0.084$). The percent of respondents that had both private and public coverage in combination was also lower in the EXP – by 1.8 percentage points ($p \leq 0.049$).

The bottom panel of Table 2.2 presents the results of our logistic regressions expressed as adjusted percentages. Comparing the difference column for the unadjusted and adjusted results shows that the regression analysis accomplished its intended goals. The point estimate for the difference changed (largely moving closer to 0) and the standard errors also shrank indicating a gain of statistical precision. However, none of our substantive conclusions changed in the adjusted analysis. As before, all but two comparisons lacked statistical significance. The difference for any two or more coverage types and private and public coverage did not differ in the unadjusted and adjusted analyses.

Tables 2.3 and 2.4 repeat the analyses described in Table 2.2 for the Medicare Sample (Table 2.3) and the RDD sample (Table 2.4). In the Medicare sample, the EXP arm had 4.1 percentage points less Medicare coverage than the CPS and the comparison was marginally significant ($p \leq 0.084$). No other coverage type reached statistical significance. However, the difference for any two or more types was larger and the difference for private and public in combination was roughly the same in the Medicare sample compared to the pooled sample. In the RDD sample only private and public coverage in combination was significantly different across arms. The EXP arm had 1.9 percentage points less private and public coverage than the CPS. Given the overall lack of significance it is difficult to interpret the differences in the Medicare and RDD findings. In analyses not shown here we conducted Chow tests to determine the appropriateness of pooling the two samples. The Chow tests failed to reject the hypothesis that pooling was appropriate. This suggests that treatment effect was not moderated by membership in the Medicare vs. RDD sample.

Table 2.5 reports the untransformed logistic regression coefficients for the Insured equation. To improve the readability of the table we only report whether a coefficient reached statistical significance and not the actual p-value. The coefficient estimates should be interpreted as the difference in the log-odds of insurance status between the given level (e.g. EXP) and the reference level (e.g. CPS). The coefficient for the reference level is 0 by definition and we have omitted this from the table for brevity. All of the covariates had the expected signs and significance levels. Controlling for everything else in the model, adults were less likely to report coverage (compared to children) as were non-whites, men, the less educated, lower-income, and those working less than full time-full year. This pattern remained the same for both the Medicare and RDD samples.

2.5 Conclusions

Our split-ballot test of the CPS and EXP all year insurance measures showed that there was little evidence for differences between the CPS and EXP instruments. Adjusting for differences in demographic characteristics across arms did not change the results and largely served to moderately decrease the coverage differences between arms. We did find that the EXP tended to measure fewer instances of double coverage in the pooled and RDD samples. This is consistent with expectations because the EXP arm used prior information to direct respondents down different paths instead of the laundry list approach used in the CPS arm.

2.6 Discussion

Our analysis of the split-ballot results was based on a selected sample of non-elderly households and it may not generalize to the full internal 2010 SHIPP file or to the full CPS sample. We would not be surprised if the larger sample size of the full internal file or the characteristics of elderly households drove a different set of results. Therefore, it is difficult to use these results to make meaningful conclusions about the EXP arm. Readers should also be cautious in interpreting statistical significance as we made several comparisons and did not adjust the significance standards accordingly. We have supplied p-values so that the reader can compare the p-value to an adjusted significance threshold if desired.[4]

Our hypothesis was that the EXP would detect a higher level of insurance than the CPS. Our analysis did not find evidence to support that hypothesis. Determining why the EXP failed to measure a statistically meaningful higher level of coverage was beyond the scope of this project. However, over the course of our work we developed some speculative conjectures.

- As mentioned above our analysis was based on a select sample. We think there is a good chance that the findings might be different on the full file. If the findings are different it could suggest that there are important subgroup differences that should be investigated further. Presumably, the SHIPP was powered assuming that the full sample would be used and not the subset of cases that we analyze. Thus, it is possible that the magnitude of the difference would remain the same in the full data, but the precision of the estimates could increase.
- A key sub-population of interest is those individuals with unstable insurance coverage. If the SHIPP underrepresented people with unstable coverage and the “methods effect” of the EXP is concentrated in this group, then the SHIPP may not adequately reflect what will happen in the CPS production environment. In sub-group analyses (not shown here) we found that the difference in the insurance rate between the EXP and CPS arm was larger among low-income household than high-income households, larger among those with a high school education or less compared to those with some college or more, and larger among those working less than full-time full year compared to full-time full year workers. However, we did not find that the difference between arms was statistically different from zero in any sub-group. We also could not reject the hypothesis that the “methods effect” was the same between sub-groups. In other words, we could

[4] Several methods for multiple comparisons exist. The easiest and most common can be found at: http://www.fon.hum.uva.nl/praat/manual/Bonferroni_correction.html.

find no statistical evidence that the “methods effect” varied by groups we suspect to have more and less stable coverage.

- Due to the unique nature of the sample or interviewing procedures in which the SHIPP instrument was fielded, the EXP might have performed better than the CPS in the production environment. This can be thought of as a type of “house effect”. In order for this to have biased our results, the “house effect” would have had to be larger in the CPS than the EXP. That is, the EXP and/or the CPS in the SHIPP environment will be different than the EXP and/or the CPS in the production environment (as a supplement that comes later in the CPS interview).
- In our opinion the design of the EXP should remove a fair amount of the recall bias that exists in the CPS. However, the EXP design still relies on retrospective reporting. Taken at face value, the findings of this study in the context of previous research on the CPS suggest that the EXP did not reduce recall bias by a statistically meaningful level.
- One of the primary mechanisms that the EXP uses to reduce recall bias is to anchor respondents first to the date of interview and then to January 1, 2009. We think it is possible that using an initial anchoring point of January 1 2010, filing in coverage up to the interview date, then invoking a second anchoring point of January 2009 and working forward through 2009 months might have led to greater reductions in recall error.
- It should also be noted that this attempt to reduce recall bias also occurs in the context of another design feature: including a person rather than household loop to inquire about health insurance for all household members. Disentangling the impact the two design features is impossible.

We were encouraged that EXP picked up less double coverage than the CPS. While it is impossible to know, from this study, what the actual level for double coverage is, past experience with the production CPS suggests that it picks-up more double coverage than we expect exists in the population. We found no statistically significant difference in Medicaid coverage between the arms, even after controlling for demographic differences. Given that the production CPS is known to substantially undercount Medicaid coverage we are concerned by the statistically indistinguishable levels of Medicaid coverage in the EXP.

Table 2.1 Demographics by Treatment Arm, Pooled Sample

	CPS	EXP	Total	
Sample Size	2,636	2,365	5,001	
	%	%	%	**P-Value**
Frame				0.690
Medicare	31.1	31.7	31.4	
RDD	68.9	68.3	68.6	
Age				0.967
0 to 18	24.9	24.6	24.8	
19 to 25	7.4	7.5	7.5	
26 to 34	8.4	8.4	8.4	
35 to 44	13.0	12.4	12.7	
45 to 64	46.3	47.1	46.7	
Race and Ethnicity				0.001
Hispanic, Any Race	6.9	7.2	7.1	
White Alone, not Hispanic	76.3	80.2	78.1	
Black Alone, not Hispanic	11.8	7.9	10.0	
Other/Multiple Race, not Hispanic	4.4	4.4	4.4	
Unknown	0.5	0.2	0.4	
Sex				0.980
Female	50.5	50.5	50.5	
Male	49.5	49.5	49.5	
Marital Status				0.609
Not married	31.8	30.5	31.2	
Married	48.4	49.4	48.9	
Unknown or not in universe	19.8	20.1	19.9	
Educational Attainment				0.553
Less than high school graduate	10.6	9.2	9.9	
High school graduate	24.8	26.0	25.4	
Some college or associate's degree	21.3	20.6	21.0	
Bachelor's degree	14.5	14.2	14.3	
More than Bachelor's degree	7.8	8.2	8.0	
Unknown or not in universe	21.0	21.8	21.4	
Work Status				0.093
Full time all year	29.0	28.1	28.6	
Less than full time all year	22.3	25.0	23.6	
Not working	28.8	26.6	27.8	
Unknown or not in universe	20.0	20.2	20.1	
Household Income				0.219
High	62.1	64.5	63.2	
Low	34.6	32.6	33.7	
Unknown or not in universe	3.3	2.9	3.1	
State Groups by ACS vs. CPS Uninsured Rates				0.831
No difference	41.3	41.1	41.2	
Difference in 1 year	27.9	27.4	27.7	
Difference in 2 or 3 years	30.8	31.5	31.2	

Table 2.2 Coverage Type by Treatment Arm, Pooled Sample

	CPS		EXP		Difference		
	%	**SE**	**%**	**SE**	**%**	**SE**	**P-Value**
Unadjusted							
Insured	87.7	0.86	88.8	0.95	1.1	1.29	0.380
ESI	62.7	1.57	64.6	1.65	1.9	2.27	0.404
Direct Purchase	6.6	0.69	7.3	0.80	0.8	1.05	0.475
Military	3.9	0.55	3.8	0.62	-0.1	0.82	0.861
Medicare	12.6	0.72	11.9	0.72	-0.7	1.02	0.510
Medicaid and Other Means-Tested	14.3	1.06	12.5	1.08	-1.8	1.51	0.237
Other	0.0	0.04	0.1	0.06	0.0	0.07	0.510
Out-of-Household	1.9	0.36	2.2	0.38	0.3	0.52	0.570
Private	69.9	1.47	70.9	1.58	1.0	2.16	0.647
Public	27.3	1.23	25.2	1.26	-2.1	1.76	0.231
Any Two or More Types	12.9	0.80	11.0	0.76	-1.9	1.10	0.084
Private and Public	9.1	0.70	7.2	0.62	-1.8	0.93	0.049
Adjusted							
Insured	87.9	0.79	88.6	0.90	0.7	1.11	0.518
ESI	63.3	1.40	63.9	1.48	0.6	1.78	0.746
Direct Purchase	6.7	0.69	7.2	0.77	0.5	1.03	0.653
Military	3.8	0.53	3.8	0.62	0.0	0.81	0.982
Medicare	12.3	0.60	12.3	0.63	0.0	0.67	0.952
Medicaid and Other Means-Tested	13.9	0.94	12.9	0.99	-1.0	1.18	0.391
Private	70.6	1.28	70.1	1.39	-0.5	1.59	0.739
Public	26.8	1.11	25.8	1.15	-1.0	1.41	0.480
Any Two or More Types	12.8	0.77	11.0	0.73	-1.8	1.02	0.086
Private and Public	9.0	0.69	7.3	0.60	-1.8	0.89	0.047

SE is standard error clustered on household.

Shaded rows are significant at the $p < 0.1$ level.

Adjusted results are from logistic regression model controlling for covariates in Table 2.1.

Other and Out-of-Household are left out of the adjusted results due to lack of sample.

See report text for more information on the construction of insurance variables and the calculation of adjusted estimates.

Table 2.3 Coverage Type by Treatment Arm, Medicare Sample

	CPS		EXP		Difference		
	%	SE	%	SE	%	SE	P-Value
Unadjusted							
Insured	82.9	1.66	84.4	1.91	1.4	2.53	0.571
ESI	41.3	2.53	45.1	2.80	3.8	3.78	0.310
Direct Purchase	7.1	1.03	7.9	1.30	0.8	1.66	0.624
Military	4.1	0.90	5.2	1.16	1.1	1.46	0.466
Medicare	35.3	1.63	31.2	1.70	-4.1	2.36	0.084
Medicaid and Other Means-Tested	21.6	1.96	18.2	2.11	-3.4	2.88	0.239
Other	0.0	0.00	0.1	0.13	0.1	0.13	0.318
Out-of-Household	1.3	0.53	1.9	0.61	0.5	0.81	0.513
Private	49.5	2.55	52.2	2.81	2.8	3.79	0.468
Public	51.4	1.99	47.5	2.11	-3.9	2.90	0.183
Any Two or More Types	26.1	1.58	22.4	1.59	-3.6	2.24	0.105
Private and Public	17.3	1.37	15.5	1.37	-1.8	1.94	0.350
Adjusted							
Insured	83.0	1.57	84.2	1.78	1.2	2.20	0.597
ESI	42.3	2.31	43.7	2.53	1.4	3.03	0.655
Direct Purchase	7.2	1.03	7.8	1.25	0.6	1.59	0.705
Military	4.2	0.95	5.4	1.17	1.2	1.50	0.424
Medicare	34.2	1.42	32.5	1.55	-1.7	1.81	0.336
Medicaid and Other Means-Tested	20.7	1.76	19.1	1.92	-1.6	2.27	0.479
Private	50.7	2.31	50.5	2.50	-0.2	2.96	0.954
Public	50.0	1.85	49.0	1.98	-1.0	2.50	0.699
Any Two or More Types	25.6	1.50	22.9	1.55	-2.6	2.06	0.202
Private and Public	17.1	1.31	15.6	1.36	-1.5	1.83	0.410

SE is standard error clustered on household.

Shaded rows are significant at the $p < 0.1$ level.

Adjusted results are from logistic regression model controlling for covariates in Table 2.1.

Other and Out-of-Household are left out of the adjusted results due to lack of sample.

See report text for more information on the construction of insurance variables and the calculation of adjusted estimates.

Table 2.4 Coverage Type by Treatment Arm, RDD Sample

	CPS		EXP		Difference		
	%	SE	%	SE	%	SE	P-Value
Unadjusted							
Insured	89.9	0.97	90.9	1.06	1.0	1.44	0.469
ESI	72.4	1.81	73.6	1.89	1.2	2.62	0.636
Direct Purchase	6.3	0.88	7.1	1.00	0.7	1.33	0.590
Military	3.8	0.68	3.1	0.73	-0.7	1.00	0.478
Medicare	2.3	0.36	3.0	0.46	0.7	0.58	0.261
Medicaid and Other Means-Tested	11.0	1.24	9.8	1.21	-1.1	1.73	0.517
Other	0.1	0.06	0.1	0.06	0.0	0.08	0.935
Out-of-Household	2.1	0.46	2.3	0.48	0.2	0.67	0.769
Private	79.1	1.63	79.5	1.76	0.4	2.40	0.868
Public	16.4	1.38	14.9	1.38	-1.6	1.95	0.422
Any Two or More Types	6.9	0.84	5.6	0.78	-1.3	1.15	0.275
Private and Public	5.3	0.78	3.4	0.60	-1.9	0.98	0.048
Adjusted							
Insured	90.1	0.89	90.6	1.01	0.6	1.25	0.656
ESI	72.9	1.66	73.1	1.75	0.2	2.17	0.916
Direct Purchase	6.4	0.88	7.0	0.97	0.5	1.30	0.677
Military	3.7	0.64	3.2	0.76	-0.4	0.98	0.676
Medicare	3.5	0.53	4.4	0.64	0.9	0.79	0.255
Medicaid and Other Means-Tested	10.8	1.07	10.1	1.11	-0.6	1.31	0.619
Private	79.6	1.44	79.0	1.59	-0.6	1.85	0.752
Public	16.1	1.25	15.2	1.29	-0.8	1.62	0.610
Any Two or More Types	6.9	0.84	5.7	0.78	-1.3	1.13	0.268
Private and Public	5.4	0.78	3.4	0.60	-2.0	0.98	0.043

SE is standard error clustered on household.

Shaded rows are significant at the $p < 0.1$ level.

Adjusted results are from logistic regression model controlling for covariates in Table 2.1.

Other and Out-of-Household are left out of the adjusted results due to lack of sample.

See report text for more information on the construction of insurance variables and the calculation of adjusted estimates.

Table 2.5 Logistic Regressions Predicting Insured Status

	Pooled			Medicare			RDD		
	Coef	**SE**		**Coef**	**SE**		**Coef**	**SE**	
Arm (Reference Level: CPS)									
EXP	0.08	0.129		0.10	0.195		0.08	0.177	
Frame (Reference Level: Medicare)									
RDD	0.11	0.145		--	--		--	--	
Age (Reference Level: 0 to 18)									
19 to 25	-2.74	0.377	***	-2.11	0.531	***	-3.01	0.438	***
26 to 34	-2.77	0.397	***	-2.18	0.571	***	-3.11	0.458	***
35 to 44	-2.35	0.383	***	-1.50	0.554	**	-2.86	0.437	***
45 to 64	-2.00	0.381	***	-1.26	0.548	*	-2.42	0.425	***
Race and Ethnicity (Reference Level: White Alone, not Hispanic)									
Hispanic, Any Race	-0.97	0.201	***	-1.01	0.301	***	-0.91	0.261	***
Black Alone, not Hispanic	-0.55	0.169	**	-0.30	0.244		-0.70	0.242	**
Other/Multiple Race, not Hispanic	0.42	0.321		-0.16	0.415		0.99	0.541	
Unknown	-0.90	0.866		--	--		-1.3	1.010	
Sex (Reference Level: Female)									
Male	-0.28	0.092	**	-0.02	0.150		-0.52	0.125	***
Marital Status (Reference Level: Not married)									
Married	0.57	0.130	***	0.40	0.199	*	0.71	0.182	***
Unknown or not in universe	0.69	0.950		14.85	0.882	***	-0.53	1.003	
Educational Attainment (Reference Level: More than Bachelor's degree)									
Less than high school graduate	-1.73	0.352	***	-1.92	0.820	*	-1.55	0.414	***
High school graduate	-1.43	0.324	***	-1.43	0.802		-1.47	0.358	***
Some college or associate's degree	-0.97	0.324	**	-1.31	0.813		-0.76	0.361	*
Bachelor's degree	-0.18	0.379		-0.38	0.890		-0.09	0.424	
Unknown or not in universe	-1.63	0.423	***	-1.93	0.865	*	-1.45	0.553	**
Work Status (Reference Level: Full time all year)									
Less than full time all year	-0.64	0.145	***	-0.43	0.259		-0.79	0.177	***
Not working	-0.02	0.163		0.59	0.244	*	-0.70	0.210	***
Unknown or not in universe	-1.20	0.820		-14.62	0.709	***	-0.38	0.951	
Household Income (Reference Level: High)									
Low	-1.13	0.147	***	-0.97	0.227	***	-1.17	0.197	***
Unknown or not in universe	-0.71	0.307	*	-0.41	0.512		-0.83	0.394	*
State Groups by ACS vs. CPS Uninsured Rates (Reference Level: No difference)									
Difference in 1 year	0.06	0.153		-0.03	0.220		0.13	0.215	
Difference in 2 or 3 years	0.27	0.154		0.44	0.240		0.15	0.207	
Constant	5.73	0.541	***	4.64	1.025	***	6.47	0.587	***

SE is standard error clustered on household.

*p<0.05; **p<0.01; ***p<0.001

Chapter 3: Comparison of ACS and EXP

3.1 Introduction

Chapter 3 presents a comparison of point-in-time coverage measures from the American Community Survey (ACS) arm to the Experimental (EXP) arm in the SHIPP survey.

3.2 Data

We use the validated SHIPP public use file to compare point-in-time health insurance measures from the ACS and EXP arms. Our analysis is based on an analytic sample of 4,825 records from the pooled RDD and Medicare frames. Households with at least one person age 65 and over are removed because there is incomplete information for these households due to disclosure review. More discussion of our sample selection criteria is given in Section 1.3. The Stata code used to create the tables for Chapters 2 and 3 was delivered to the Census Bureau in do files.

All missing values for the health insurance variables were coded to “0=No coverage”. We interpret variables coded in this manner as “any indication of coverage”. This approach to treating missing values might produce biased split-ballot results if the level of missing data varies between treatment arms. For example, two arms could have an equal proportion of explicit “no” responses, but an unequal proportion of missing values. Using our coding rules we would infer that the proportion of “no” responses was unequal between arms when in fact it was equal. This is particularly concerning for comparisons of uninsurance as that concept is operationally defined as the absence of coverage.

We do not report uninsurance as we are unable to differentiate between which respondents answered “No” to all the insurance questions and which respondents had missing data that was grouped with “0=No”. For example, suppose we have a hypothetical data set where 6 out of 10 individuals have some type of coverage. All we know about the remaining 4 individuals is that they were not coded as “yes” in any of the coverage questions. Therefore we can say with certainty that 60% were reported as having insurance. We could not say that the other 40% were uninsured, *per se*, because one or more of them may have had coverage that was reported as “DK/Refuse.” We examine 11 health insurance measures: 6 individual insurance types and 5 aggregated measures. Since the ACS does not measure out-of-household coverage we do not include it in this analysis. The 6 insurance types are:

- Employer sponsored insurance (ESI)
- Direct purchase coverage (includes coverage from school)
- Military coverage (VA and Tricare)
- Medicare
- Medicaid and other means-tested coverage
- Other (any coverage that could not be categorized into the preceding coverage types)

There 5 aggregated coverage categories are:

- Insured
- Private coverage (includes ESI and direct purchase)
- Public coverage (includes Military, Medicare, and Medicaid)
- Any two or more types
- Private and public coverage in combination

Point-in-time coverage in the EXP arm is operationally defined as holding coverage in the month of interview. In the ACS it is defined as holding coverage at the date of interview. Respondents can be categorized into more than one insurance type.

3.3 Methods

The SHIPP test randomly assigned respondents to the EXP or ACS treatment arm. The interviewers and fielding periods of the two instruments were also randomized to mitigate interviewer and seasonality bias. Randomization should ensure that the differences we report are not confounded by any observed or unobserved factor that is correlated with treatment arm and insurance type.

To investigate the success of the randomization procedure we compared a number of demographic characteristics across treatment arms. The demographic characteristics we examine are known to be related to health insurance. The distribution of the covariates was compared across treatment arms using a chi-squared test of association. The assumption of this test is that balance between observed covariates implies balance between unobserved covariates.

The results of this comparison, in the fully pooled sample (composed of both the Medicare and RDD frame), are presented in Table 3.1. The following variables showed significant differences across treatment arm: race/ethnicity, educational attainment, and work status. Household income was also marginally significant ($p \leq 0.063$) with the EXP composed of a higher share of low income households. Of the eight characteristics we considered four were at least marginally significant. The relatively high number of significant comparisons suggests that randomization might not have resulted in a sample that was fully balanced on all unobserved characteristics. While we can control for the characteristics we observed we cannot rule out that our results partially reflect the confounding influence of unobserved characteristics.

We used logistic regression to compare health insurance measures across treatment arm, controlling for all the variables found in Table 3.1. Regression has two advantages: it controls for any observed characteristic that could plausibly bias our comparison and it increases the statistical precision of our analysis. We fit a total of 30 equations. We compared 10 coverage types ("other" coverage was not examined using regression due to a lack of sample) in the pooled data and in the Medicare and RDD frames separately. We present our model results using predicted values for each arm derived using average marginal effects. The standard errors we report are clustered at the household level and all comparisons of coverage were conducted using a t-test. While we report and comment on differences that met a significance standard of $p \leq 0.1$ it should be noted that under that standard at least 1 test in 11 should be significant purely due to chance. We supply the p-values for each of our comparisons so that the reader can adjust the p-value for multiple comparisons if desired.

3.4 Results

The top panel of Table 3.2 presents the unadjusted percent of each coverage type by treatment arm in the pooled analytical sample. The majority of the 11 comparisons failed to reach a meaningful level of statistical significance. The point estimates suggest that the EXP captured 0.5 percentage points more insured compared to the ACS. However, we failed to reject the hypothesis that the difference in the population is 0. Two coverage types did prove to be significant at the $p \leq 0.05$. The percent of respondents that had Military coverage was 3.3 points lower in the EXP and the result was significant ($p \leq 0.001$). The percent of respondents that had any two or more coverage types was also lower in the EXP by 3.7 percentage points ($p \leq 0.001$). The difference in direct purchase coverage was marginally significant ($p \leq 0.085$). The EXP had 2.0 fewer percentage points of direct purchase coverage compared to the ACS.

The bottom panel of Table 3.2 presents the results of our logistic regressions expressed as adjusted percentages. Comparing the difference column for the unadjusted and adjusted results shows that the regression analysis accomplished its intended goals. The point estimate for the difference changed and the standard errors shrank indicating a gain of statistical precision.

The regression analysis showed that some of the unadjusted findings were being driven by spurious correlations. The difference in ESI coverage between the two arms increased from 2.9 percentage points and non-significant in the unadjusted analysis to 4.0 percentage points and statistically significant in the adjusted analysis ($p \leq 0.023$). This carried over to private coverage. The adjusted difference for private coverage was 3.6 percentage points ($p \leq 0.028$). While the change in the difference for public coverage was only minimal (-2.8 in the unadjusted to -3.0 in the adjusted analysis) the gain in precision allowed us to detect the effect ($p \leq 0.035$). The comparisons of Military, direct purchase coverage, and any two or more coverage types all retained their sign and significance level in the regression models.

Tables 3.3 and 3.4 repeat the analyses described in Table 3.2 for the Medicare Sample (Table 3.3) and the RDD sample (Table 3.4). In the unadjusted analysis of the Medicare sample, the EXP arm had 3.9 fewer percentage points of any two or more coverage types ($p \leq 0.092$) and 3.4 fewer percentage points of Military coverage ($p \leq 0.065$). No other coverage type reached at least a marginal level of statistical significance. The regression analyses did not substantially change the majority of results in the Medicare sample. The difference for any two or more coverage types moved towards zero and lost statistical significance. The difference in Medicaid coverage increased to -3.9 and obtained a marginal significance level. Difference in Military coverage remained significant.

In the unadjusted analysis of the RDD sample, the EXP has 3.2 fewer percentage points of direct purchase coverage ($p \leq 0.031$), 3.2 fewer percentage points of Military coverage ($p \leq 0.004$), 3.3 fewer percentage points of any two or more coverage types ($p \leq 0.002$) and 1.5 fewer percentage points of private and public coverage in combination ($p \leq 0.070$). In the adjusted estimates, the EXP had a statistically significant higher percentage of ESI by 4.3 points ($p \leq 0.049$). The findings for direct purchase coverage, Military, any two or more coverage types, and private and public coverage did not substantially change.

Table 3.5 reports the untransformed logistic regression coefficients for the Insured equation. To improve the readability of the table we only report whether a coefficient reached statistical significance and not the actual p-value. The coefficient estimates should be interpreted as the difference in the log-odds of insurance status between the given level (e.g. EXP) and the reference level (e.g. ACS). All of the covariates had the expected signs and significance levels. Controlling for everything else in the model, adults were less likely to report coverage compared to children, as were men, the less educated, lower-income, and those working less than full time-full year. This pattern remained the same for both the Medicare and RDD samples.

3.5 Conclusions

Our split-ballot test of the ACS and EXP point-in-time insurance measures showed evidence for differences between the ACS and EXP instruments for some insurance measures. Notably, the regression analyses showed that in the pooled sample the EXP arm had more ESI, less direct purchase, less Military, more private, less public, and less double coverage compared to the ACS arm.

3.6 Discussion

Our analysis of the split-ballot results was based on a selected sample of non-elderly households and it may not generalize to the full internal 2010 SHIPP file or to the production ACS sample. We would not be surprised if the larger sample size of the full internal file or the characteristics of elderly households drove a different set of results. We also found that many of the demographic characteristics we considered varied significantly by treatment arm. While we can control for what we observe we cannot rule out that our findings are partially driven by unobserved confounders. Therefore, it is difficult to use these results to make meaningful conclusions about the EXP arm. Readers should also be cautious in interpreting significance findings as we made several comparisons and did not adjust the p-values accordingly. We have supplied p-values so that the reader can make such an adjustment if desired.

We were encouraged that EXP picked up less direct purchase coverage than the ACS. Previous experience with the production ACS suggests that it overestimates levels of direct purchase coverage. Mach and O'Hara (2011) suggest that the level of direct purchase in the ACS is tied to its tendency to estimate a large percentage of direct purchase coverage in combination with ESI. It is suspected that these doubly covered cases only have ESI and are either misreporting a single service plan as comprehensive direct purchase coverage or believe that the direct purchase response option is another accurate description of their ESI coverage.

That the lower levels of direct purchase coverage in the EXP is accompanied by lower levels of any two or more coverage types suggests that the EXP avoids some of the pitfalls of the ACS instrument design. Indeed, in further analysis not presented here we found that the EXP arm produced 1.6 fewer percentage points of direct purchase coverage in combination with ESI. We also found that if we edited the direct purchase variable such that direct purchase coverage was removed for any case that reported both direct purchase and ESI the unadjusted difference between the treatments became small (0.4 percentage points) and non-significant.

Despite these encouraging findings we also found areas of concern. The smaller percentage of public coverage should receive close attention given that all surveys are thought to underestimate public coverage (Call et al., 2012).

Table 3.1 Demographics by Treatment Arm, Pooled Sample

	ACS	EXP	Total	
Sample Size	2,460	2,365	4,825	
	%	%	%	**P-Value**
Frame				0.416
Medicare	32.8	31.7	32.2	
RDD	67.2	68.3	67.8	
Age				0.422
0 to 18	23.9	24.6	24.3	
19 to 25	6.8	7.5	7.2	
26 to 34	7.8	8.4	8.1	
35 to 44	11.8	12.4	12.1	
45 to 64	49.7	47.1	48.4	
Race and Ethnicity				0.006
Hispanic, Any Race	8.7	7.2	8.0	
White Alone, not Hispanic	76.8	80.2	78.4	
Black Alone, not Hispanic	8.3	7.9	8.1	
Other/Multiple Race, not Hispanic	5.6	4.4	5.0	
Unknown	0.7	0.2	0.5	
Sex				0.592
Female	51.3	50.5	50.9	
Male	48.7	49.5	49.1	
Marital Status				0.135
Not married	32.6	30.5	31.6	
Married	49.1	49.4	49.3	
Unknown or not in universe	18.2	20.1	19.1	
Educational Attainment				0.035
Less than high school graduate	10.4	9.2	9.8	
High school graduate	23.8	26.0	24.9	
Some college or associate's degree	23.0	20.6	21.8	
Bachelor's degree	14.6	14.2	14.4	
More than Bachelor's degree	8.9	8.2	8.6	
Unknown or not in universe	19.2	21.8	20.5	
Work Status				0.025
Full time all year	30.3	28.1	29.2	
Less than full time all year	22.4	25.0	23.7	
Not working	28.8	26.6	27.7	
Unknown or not in universe	18.5	20.2	19.3	
Household Income				0.063
High	67.6	64.5	66.1	
Low	29.9	32.6	31.2	
Unknown or not in universe	2.5	2.9	2.7	
State Groups by ACS vs. CPS Uninsured Rates				0.505
No difference	41.9	41.1	41.5	
Difference in 1 year	28.1	27.4	27.8	
Difference in 2 or 3 years	30.0	31.5	30.8	

Table 3.2 Coverage Type by Treatment Arm, Pooled Sample

	ACS		EXP		Difference		
	%	SE	%	SE	%	SE	P-Value
Unadjusted							
Insured	87.8	0.96	88.3	0.94	0.5	1.34	0.719
ESI	60.1	1.59	63.0	1.68	2.9	2.31	0.212
Direct Purchase	9.2	0.86	7.2	0.78	-2.0	1.16	0.085
Military	6.6	0.79	3.3	0.54	-3.3	0.96	0.001
Medicare	13.8	0.77	12.4	0.74	-1.4	1.07	0.194
Medicaid and Other Means-Tested	13.0	1.02	12.6	1.11	-0.4	1.50	0.787
Other	0.1	0.06	0.0	0.04	0.0	0.07	0.584
Private	67.0	1.55	69.4	1.62	2.5	2.24	0.268
Public	28.2	1.31	25.4	1.27	-2.8	1.83	0.125
Any Two or More Types	13.7	0.84	10.0	0.70	-3.7	1.10	0.001
Private and Public	7.4	0.66	6.6	0.56	-0.8	0.87	0.356
Adjusted							
Insured	87.4	0.94	88.7	0.85	1.3	1.18	0.261
ESI	59.6	1.42	63.6	1.48	4.0	1.77	0.023
Direct Purchase	9.2	0.87	7.2	0.77	-2.1	1.17	0.079
Military	6.6	0.80	3.4	0.54	-3.3	0.96	0.001
Medicare	13.5	0.65	12.7	0.64	-0.7	0.72	0.315
Medicaid and Other Means-Tested	13.3	0.94	12.4	0.97	-0.9	1.18	0.438
Private	66.4	1.38	70.0	1.39	3.6	1.63	0.028
Public	28.3	1.19	25.3	1.12	-3.0	1.41	0.035
Any Two or More Types	13.4	0.80	10.3	0.67	-3.1	0.99	0.002
Private and Public	7.2	0.64	6.8	0.55	-0.4	0.81	0.644

SE is standard error clustered on household.

Shaded rows are significant at the $p < 0.1$ level.

Adjusted results are from logistic regression model controlling for covariates in Table 3.1.

Other is left out of the adjusted results due to lack of sample.

See report text for more information on the construction of insurance variables and the calculation of adjusted estimates.

Table 3.3 Coverage Type by Treatment Arm, Medicare Sample

	ACS		EXP		Difference		
	%	SE	%	SE	%	SE	P-Value
Unadjusted							
Insured	87.7	1.25	85.0	1.77	-2.7	2.17	0.218
ESI	42.4	2.50	43.1	2.80	0.7	3.76	0.854
Direct Purchase	7.7	1.15	8.3	1.33	0.6	1.76	0.739
Military	8.6	1.43	5.2	1.12	-3.4	1.82	0.065
Medicare	35.7	1.67	32.3	1.76	-3.4	2.43	0.159
Medicaid and Other Means-Tested	22.7	2.07	19.5	2.26	-3.2	3.07	0.296
Other	--	--	--	--	--	--	--
Private	48.5	2.55	50.7	2.85	2.2	3.82	0.561
Public	53.7	2.04	49.8	2.14	-3.9	2.96	0.186
Any Two or More Types	26.4	1.60	22.6	1.63	-3.9	2.29	0.092
Private and Public	14.5	1.33	15.6	1.39	1.1	1.92	0.566
Adjusted							
Insured	87.0	1.26	85.7	1.58	-1.2	1.89	0.513
ESI	41.4	2.24	44.2	2.52	2.8	2.94	0.334
Direct Purchase	7.6	1.11	8.6	1.33	1.0	1.68	0.548
Military	8.6	1.43	5.2	1.10	-3.4	1.78	0.054
Medicare	35.2	1.50	32.9	1.56	-2.3	1.84	0.212
Medicaid and Other Means-Tested	23.1	1.93	19.2	1.93	-3.9	2.33	0.091
Private	47.4	2.28	51.9	2.48	4.5	2.86	0.119
Public	53.8	1.93	49.7	1.94	-4.1	2.50	0.104
Any Two or More Types	25.8	1.47	23.1	1.56	-2.7	1.98	0.175
Private and Public	14.2	1.20	16.0	1.36	1.9	1.68	0.271

SE is standard error clustered on household.

Shaded rows are significant at the $p < 0.1$ level.

Adjusted results are from logistic regression model controlling for covariates in Table 3.1.

Other is left out of the adjusted results due to lack of sample.

See report text for more information on the construction of insurance variables and the calculation of adjusted estimates.

Table 3.4 Coverage Type by Treatment Arm, RDD Sample

	ACS		EXP		Difference		
	%	SE	%	SE	%	SE	P-Value
Unadjusted							
Insured	87.8	1.29	89.8	1.09	1.9	1.68	0.249
ESI	68.7	1.93	72.2	1.93	3.5	2.73	0.204
Direct Purchase	9.9	1.15	6.7	0.96	-3.2	1.50	0.031
Military	5.7	0.95	2.5	0.59	-3.2	1.12	0.004
Medicare	3.1	0.46	3.2	0.47	0.1	0.66	0.912
Medicaid and Other Means-Tested	8.3	1.06	9.5	1.20	1.1	1.60	0.483
Other	0.1	0.09	0.1	0.06	-0.1	0.11	0.576
Private	75.9	1.83	78.1	1.82	2.2	2.58	0.404
Public	15.7	1.42	14.0	1.33	-1.7	1.95	0.391
Any Two or More Types	7.6	0.90	4.2	0.61	-3.3	1.09	0.002
Private and Public	3.9	0.72	2.4	0.43	-1.5	0.84	0.070
Adjusted							
Insured	87.7	1.24	89.9	1.00	2.3	1.48	0.126
ESI	68.3	1.75	72.6	1.74	4.3	2.18	0.049
Direct Purchase	10.1	1.18	6.6	0.93	-3.5	1.51	0.020
Military	5.8	0.97	2.5	0.60	-3.3	1.14	0.004
Medicare	3.1	0.44	3.1	0.43	0.0	0.56	0.947
Medicaid and Other Means-Tested	8.4	0.98	9.4	1.09	0.9	1.32	0.483
Private	75.6	1.64	78.4	1.58	2.8	1.94	0.144
Public	15.9	1.33	13.9	1.23	-2.0	1.66	0.229
Any Two or More Types	7.7	0.92	4.2	0.59	-3.5	1.09	0.001
Private and Public	4.1	0.75	2.4	0.43	-1.7	0.87	0.050

SE is standard error clustered on household.

Shaded rows are significant at the $p < 0.1$ level.

Adjusted results are from logistic regression model controlling for covariates in Table 3.1.

Other is left out of the adjusted results due to lack of sample.

See report text for more information on the construction of insurance variables and the calculation of adjusted estimates.

Table 3.5 Logistic Regressions Predicting Insured Status

	Pooled			Medicare			RDD		
	Coef	SE		Coef	SE		Coef	SE	
Arm (Reference Level: ACS)									
EXP	0.15	0.135		-0.13	0.196		0.28	0.180	
Frame (Reference Level: Medicare)									
RDD	-0.23	0.144		--	--		--	--	
Age (Reference Level: 0 to 18)									
19 to 25	-2.30	0.332	***	-2.15	0.542	***	-2.42	0.397	***
26 to 34	-2.14	0.352	***	-1.84	0.521	***	-2.36	0.435	***
35 to 44	-1.83	0.337	***	-1.37	0.575	*	-2.19	0.401	***
45 to 64	-1.41	0.321	***	-1.13	0.498	*	-1.66	0.382	***
Race and Ethnicity (Reference Level: White Alone, not Hispanic)									
Hispanic, Any Race	-0.48	0.219	*	-0.43	0.373		-0.47	0.272	
Black Alone, not Hispanic	-0.28	0.207		-0.12	0.293		-0.36	0.295	
Other/Multiple Race, not Hispanic	-0.42	0.352		0.26	0.512		-0.54	0.397	
Unknown	-0.33	0.564		--	--		-1.18	0.695	
Sex (Reference Level: Female)									
Male	-0.26	0.097	**	-0.15	0.175		-0.35	0.118	**
Marital Status (Reference Level: Not married)									
Married	0.55	0.140	***	0.37	0.211		0.64	0.196	**
Unknown or not in universe	0.90	0.889		2.33	0.959	*	0.14	1.214	
Educational Attainment (Reference Level: More than Bachelor's degree)									
Less than high school graduate	-1.66	0.361	***	-1.62	0.821	*	-1.70	0.437	***
High school graduate	-1.47	0.314	***	-1.56	0.784	*	-1.39	0.350	***
Some college or associate's degree	-1.00	0.325	**	-0.89	0.805		-1.04	0.365	**
Bachelor's degree	-0.18	0.321		-0.48	0.833		-0.09	0.358	
Unknown or not in universe	-1.75	0.448	***	-2.10	0.873	*	-1.68	0.571	**
Work Status (Reference Level: Full time all year)									
Less than full time all year	-0.89	0.149	***	-1.04	0.252	***	-0.81	0.186	***
Not working	0.06	0.167		0.56	0.273	*	-0.42	0.207	*
Unknown or not in universe	-0.95	0.832		-1.72	0.796	*	-0.47	1.243	
Household Income (Reference Level: High)									
Low	-1.30	0.153	***	-1.09	0.227	***	-1.43	0.199	***
Unknown or not in universe	-0.41	0.340		-0.75	0.530		0.00	0.414	
State Groups by ACS vs. CPS Uninsured Rates (Reference Level: No difference)									
Difference in 1 year	-0.19	0.165		0.04	0.236		-0.30	0.223	
Difference in 2 or 3 years	-0.04	0.162		0.07	0.242		-0.10	0.221	
Constant	5.56	0.484	***	5.06	0.958	***	5.69	0.560	***

SE is standard error clustered on household.

*p<0.05; **p<0.01; ***p<0.001

References (Report and Appendices)

Blumberg SJ and Luke.JV. 2012. "Wireless substitution: Early release of estimates from the National Heatlh Interview Survey, January-June 2012." National Center for Health Statistics, December 2012. Available from http://www.cdc.gov/nchs/nhis.htm.

Call KT, Davern ME, Klerman JA, Lynch V. 2012. "Comparing errors in Medicaid reporting across surveys: Evidence to date." *Health Services Research,* Jul 20. doi: 10.1111/j.1475-6773.2012.01446.x.

Call KT, Davidson G, Davern ME, Brown ER, Kincheloe J, Nelson JG. Winter 2008/2009. "Accuracy of self-reported health insurance coverage among Medicaid enrollees." *Inquiry,* 45(4):438-456.

Call KT, Davidson G, Sommers AS, Feldman R, Farseth P, Rockwood T. 2002. "Uncovering the missing Medicaid cases and assessing their bias for estimates of the uninsured." *Inquiry,* 38(4):396-408.

Cantor, J. C., A.C. Monheit, S. Brownlee, and C. Schneider. 2007. "The Adequacy of Household Survey Data for Evaluating the Nongroup Health Insurance Market." *Health Services Research* 42(4): 1739-1757.

Loomis, L. 2000. Report on Cognitive Interview Research Results for Questions on Welfare Reform Benefits and Government Health Insurance for the March 2001 Income Supplement to the CPS. Internal memorandum. Washington, D.C.: U.S. Bureau of the Census.

Mach, A and O'Hara, B. 2011. "Do people really have multiple health insurance plans? Estimates of Nongroup Health Insurance in the American Community Survey" SEHSD Working Paper Number 2011-28.

Pascale, Joanne. 2008. "Measurement Error in Health Insurance Reporting." *Inquiry*: December 2008, Vol. 45, No. 4, pp. 422-437.

Pascale, Joanne. 2009. "Findings from a Pretest of a New Approach to Measuring Health Insurance in the Current Population Survey." Proceedings of the Federal Committee on Statistical Methodology Conference, November, 2009.

Pascale, Joanne. 2012. "Findings from a Split-Ballot Experiment on a New Approach to Measuring Health Insurance in the Current Population Survey." American Society of Health Economists Conference, Minneapolis, MN, June 2012.

SNACC Phase II. 2008. Research Project to Understand the Medicaid Undercount: The University of Minnesota's State Health Access Center, the Centers for Medicare and Medicaid Services, the Department of Health and Human Services Assistant Secretary for Planning and Evaluation, The National Center for Health Statistics, and the U.S Census Bureau. Phase II Research Results: Examining Discrepancies between the National Medicaid Statistical Information System (MSIS) and the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC).

SNACC Phase IV. 2009. Research Project to Understand the Medicaid Undercount: The University of Minnesota's State Health Access Data Assistance Center, the National Center for Health Statistics, the Department of Health and Human Services Assistant Secretary for Planning and Evaluation, Centers for Medicare and Medicaid Services, and the U.S. Census Bureau. Phase IV Research Results: Estimating the

Medicaid Undercount in the National Health Interview Survey (NHIS) and Comparing False-Negative Medicaid Reporting in NHIS to the Current Population Survey (CPS).

Appendix A: State Clusters for Analysis

This appendix contains details of the state geographic variable discussed in Chapter 1.3 and included in some of the analysis tables in Chapters 2 and 3.

After consultation with the Census Bureau, we selected discordant annual CPS and ACS uninsurance rates to create a state geographic variable. This was deemed most relevant given our emphasis on uninsurance estimates. Table A1 lists how the states were grouped into three categories: 1) States in which CPS and ACS did not produce a significantly different uninsurance rate between 2008 and 2010, 2) states in which there was a difference in one year, 3) and states in which there was a difference in two or three years.

Table A1. Geographic Variable State Groupings

Group 1 – no significant difference	Group 2 – significant difference in 1 year	Group 3 – significant difference in 2 or 3 years
Arizona	Alaska	Alabama
Florida	Arkansas	California
Georgia	Colorado	Connecticut
Hawaii	Delaware	Iowa
Idaho	Indiana	Maryland
Illinois	Kentucky	New Jersey
Kansas	Massachusetts	Ohio
Maine	Mississippi	Utah
Michigan	Montana	District of Columbia
Minnesota	Nebraska	Louisiana
Missouri	Nevada	New York
New Hampshire	North Dakota	
New Mexico	Oklahoma	
North Carolina	Pennsylvania	
Oregon	South Carolina	
Rhode Island	Texas	
South Dakota	Vermont	
Tennessee	Wyoming	
Virginia		
Washington		
West Virginia		
Wisconsin		

We considered several potential variables to explore geographic variation in the measurement of health insurance coverage. They are briefly summarized here for documentation.

Public program complexity. States vary in both the number of public programs (e.g., Medicaid, CHIP and other state-specific programs) and the name of these programs (e.g., Medicaid versus Commonwealth

Care in Massachussetts; Medical Assistance in Minnesota, BadgerCare and ForwardHealth in Wisconsin). and research shows that this complexity may lead program enrollees to misreport their coverage (Call et al., 2002; Loomis, 2000). We considered a variable that would capture this complexity in public program names by grouping states into 3 categories: 1) Simple (standard federal programs), 2) Not simple (two or fewer additional programs or names), and 3) Complex (three or more programs or names). If two geographic variables were allowed on the public use SHIPP file, this would have been our selection.

Public program generosity. Research indicates greater movement between insurance status (insured or not), between types of insurance (public and private) and potentially between programs (Medicaid versus CHIP) in states with higher income thresholds for public program eligibility. Misreporting type of health insurance is also greater among those with higher incomes (Call et al., 2008; SNACC Phase II; SNACC Phase IV). We considered a variable that would group states according to their public program income eligibility limits (generosity), but decided against this grouping as it may be less directly tied to variation in survey stimulus than the variable selected.

Discordant annual CPS and ACS direct purchase rates. Concerns have been raised about measurement error in estimates of direct purchase or non-group insurance generally (Cantor et al., 2007), and discrepancies between CPS and ACS estimates of direct purchase insurance particularly (higher magnitude of direct purchase in the ACS as compared to the CPS). We considered a variable similar to what was chosen, but using direct purchase rates instead of uninsurance rates. We decided against this option given lack of variation; the majority of states reveal significant difference in multiple years.

State Medicaid undercount/uninsurance variation. The SNACC project revealed high state level variation in misreports of Medicaid enrollment and uninsurance (Phase II report). That is, Medicaid enrollees in some states were more likely to be reported as having Medicaid and were less likely to be reported as having no insurance. Drawing on the SNACC report we considered two dichotomous variables that flag states with significant misreporting of (1) Medicaid enrollment and (2) significant misreporting of uninsurance. This option was dropped as it was not considered a priority for this project.

Appendix B: SHADAC Draft Initial Weighting Strategy

This appendix contains additional thoughts on developing a weighting strategy for the SHIPP data initially discussed in Chapter 1.4. These deliberations were shared with the Census Bureau before concluding that weighting did not facilitate the goal of evaluating the different arms of the SHIPP test. We include here for documentation as weighting may be important for future analysis of the SHIPP data.

The SHIPP survey is a large-scale split-panel field test of three alternate sets of questions for obtaining information on health insurance coverage. The SHIPP sample was drawn from two sources: a Random Digit Dial (RDD) frame and Medicare enrollment files (Medicare). Analysis revealed demographic differences (e.g., larger household size in the RDD) and response rate differences across these two sample frames. In addition, within both the RDD and Medicare frames there were significant differences in the demographic profiles of respondents in each of the three independent treatment branches of the experiment (e.g., ethnicity and employment).

The goals of weighting the SHIPP data are (1) to adjust for these demographic imbalances across treatment groups and (2) to allow for merging the data from the RDD and Medicare frames to maximize the sample size across the three treatment groups.

Weighting

The aim of weighting survey data is to adjust the results to account for differences in probabilities of selection, sample coverage problems and reduce potential bias associated with differential participation in the survey. In order to accomplish this task it is necessary to understand who was excluded from each of the sample frames and screened out by the interviewers. As shown in the table below, households without an address or with a P.O. Box were excluded from the Medicare frame. Additionally, all individuals located in group quarters were screened out from both frames and only households that had landline telephone numbers (via an address look-up) were included in the Medicare sample. According to Census staff, group quarters were not excluded from the Medicare frame; however, as with the RDD frame group quarters would be excluded at the time of the interview. Further it is not clear whether Puerto Rico was excluded from the Medicare frame as was true for the RDD frame, but we assume so in this document.

Persons excluded from the SHIPP survey

	Do not speak English	Live in Group Quarters	Live in AK, HI or PR	Address is P.O. box	Address is not on CMS list	No members of HH are Medicare enrollees	HH does not have a listed LL phone
MCARE Frame	Excluded	Included	Excluded (PR?)	Excluded	Excluded	Excluded	Excluded
RDD Frame	Included	Excluded?	Excluded	Included	Included	Included	Excluded
Screened out of survey	Yes	Yes	No	No	No	No	Yes

Source: U.S. Census Bureau, Memorandum from Ruth Ann Killion, Chief Demographic Statistical Methods Division, April 5, 2010.

Accounting for varying probabilities of selection and response rates through the application of weights enables the survey responses drawn from statistical samples to be representative of the entire target population (i.e., the civilian, non-institutionalized population). Two types of weights are generated: 1) base weights and 2) post-stratification weights. Below we outline proposed steps in the weighting process for the SHIPP data, however, we suggest Census Bureau review of weighting decisions and the evaluation of the weights.

Calculating the base weights

The first step is to calculate base weights that adjust for the probability of selection separately for each sample frame (RDD and Medicare). For example, the Medicare sample was a stratified sample based on the age (over and under 65 years of age) and enrollment date (before or after January 1, 2009) of Medicare beneficiaries in the sample frame. Therefore, base weights will adjust for the unequal probability of selection for cases in each stratum compared to their representation in the Medicare enrollment file. In contrast, in the RDD sample all individuals were sampled with the same probability. The following is the proposed base weight calculation for the two different frames.

Medicare person base weight calculation for strata A – Use the same formula for strata B, C & D by replacing the subscript $_A$ with $_{B, C \text{ or } D}$.

Strata A= Under 65 that joined before 1/1/09
Strata B = 65 or over that joined before 1/1/09
Strata C = Under 65 that joined during or after 1/1/09
Strata D = 65 or over that joined during or after 1/1/09
$=1/[(S_A / U_A) \times (S_{ALL} / S_A) \times (S_{ALLNG} / S_{ALL}) \times (C_{NG}/ S_{ALLNG}) \times (1/P) \times P \times LL_N$, or
$=1/ (C_{NG}/ U_A) \times LL_N$

RDD person base weight
$=1/[(S_{LL}/U_{LL}) \times (S_{LLNG}/S_{LL}) \times (C_{NG}/S_{LLNG}) \times (1/P) \times LL_N \times P$, or
$=1/(C_{NG}/U_{LL}) \times LL_N$

S_A= sample of enrollee addresses in strata A
U_A= universe of enrollee addresses in strata in strata A
S_{ALL}= sample of enrollee addresses in strata A with LL match
S_{ALLNG}= sample of non-group quarters enrollee addresses with LL match in strata A
S_{LL}=sample of landline telephone numbers group and non-group
U_{LL}=universe of landline telephone numbers group and non-group
S_{LLNG}=sample of non-group quarters landline numbers
C_{NG}=non-group quarter hh interviews completed in strata
P=persons in hh in strata
LL_N=number of working LL numbers in hh in strata

Combining the frames, addressing the overlap and post-stratifying

We suggest post-stratifying the data to 2010 ACS control variables to adjust for differential participation and coverage. In the CPS this is normally done by raking by age, race/ethnicity and sex. We suggest also adjusting by age (below 65 and 65 and above)Xdisability status given the Medicare sample design that oversamples young disabled beneficiaries. We may also want to post-stratify by ageXeducation to help adjust for the exclusion of people in cell-phone only households in the RDD frame.

Separate frames

For evaluation purposes, it makes sense to weight both of the frames separately. That is, apply the Medicare base weights and then post-stratify to the ACS control totals and apply the RDD base weights and post-stratify to the same ACS control totals. This would allow for a comparison of results from the separate RDD and Medicare files with those from the combined RDD and Medicare file.

Appendix C: Reconsider Weighting Strategy Memorandum

This appendix contains a memo SHADAC delivered to the Census Bureau recommending against weighting the SHIPP data, as discussed in Chapter 1.4 and Appendix B.

Date: January 30, 2012
To: Brett O'Hara
From: Kathleen Call and SHADAC team
RE: SHIPP Task: Weighting

The SHIPP experiment provides a unique and rich source of data for understanding measurement error surrounding health insurance questions. Several concerns with the data are viewed as limiting its usefulness: (1) demographic imbalances across treatment groups within sample frames (RDD and MCARE), and (2) concerns about merging data across frames to maximize the sample size within the three treatment groups.

The first approach we proposed for overcoming this limitation is weighting the SHIPP data. After considerable investment in developing weights we have stepped back from this and instead suggest using logistic regression or multivariate ANOVA (MANOVA) to account for these demographic and sample design issues. Below we outline the reasoning for this recommendation.

Reasons to reconsider weights: (1) the goal of randomization is to distribute individual differences across treatment groups and remove the threat of selection bias, (2), the demographic differences between experimental arms are of concern but are not widespread (see Table below- for brevity, not included in Appendix C), (3) the sample weights will have large design effects that will decrease the statistical power that was hoped to be gained in merging the frames, and (4) the goal of the SHIPP experiment was not to generalize to the larger population (the goal of weighting) but to test differences between treatment arms.

Recommendation: Use MANOVA or logistic regression analysis to address demographic imbalances and merge frames within treatment groups:

- Adjust for demographic variables that were significantly different across arms within both frames.
- Include a sample frame indicator in all analysis
- Consider including an interaction term: treatment*frame to examine differences in response to treatments by frame

Please let us know how you wish to proceed.

Appendix D: EXP Month Determination

This table, prepared by Amy Steinweg of the Census Bureau, describes the assignment of months of coverage for the EXP questionnaire.

MONTH DETERMINATION in SHIPP 2010

For first round of dual-coded estimates 6/2012. Production data requires imputation of missings.

	Current Loop									Past Loop	Gaps		
Initialize to 0 for interview months, NIU for any months after	PUBEFORAFT	PUMNTHBEG1 & PUYEARBEG1	PUANYTHIS	PUANYLAST	PUCNTCOV	PUMNTHBEG2 & PUYEARBEG2	PUSPELLADD	PUWHATMNTHS1(mo)	PUWHATMNTHS2(mo)	PUWHATMNTHS_PST(mo)	PUGAPMNTHSP_(pnum)(mo)	PUGAPMNTHSS_(pnum)(mo)	How to Code Months:
If	1				1								Set 1 for mo=(1 to interview month)
If	2	1:17			1								Set 1 for mo=(MNTHBEG1 to interview month)
If						1:17							Set 1 for mo=(MNTHBEG2 to interview month)
If								1:17					Set 1 for all months reported in PUWHATMNTHS1(mo)
If									1:17				Set 1 for all months reported in PUWHATMNTHS2(mo)
If										1:17			Set 1 for all months reported in PUWHATMNTHS_PST(mo)
If											1:17		Set 1 for all months reported in PUGAPMNTHSS_(pnum)(mo)
If												1:17	Set 1 for all months reported in PUGAPMNTHSS_(pnum)(mo)
For Missings													
If no months have been reported for Current Loop:													Set 1 for Current Month
If no months have been reported for Past Loop:													Set 1 for any time in 09 indicator
If no months have been reported for Gaps Section:													Set 1 for any time in 09 indicator

Appendix E: Validation Notes

This appendix contains additional details about the validation of the EXP arm discussed in Chapter 1.5.2. Cases in **bold** indicate possible errors in the CATI instrument as opposed to discrepancies between SHADAC and the Census Bureau in applying the recoding rules.

All Year Validation

ESI

- In household 4994 person 1 has MILPLANLP2 = 2 and PUWHATMNTHS2LP2 = DK. SHADAC codes them as ESI in 2009 per our "don’t know past months" rules. The Census Bureau does not give them ESI.

Medicaid (we now match)

- **SHADAC had one case that was DK to PUGOVPLANFC3. Per our rules, SHADAC should have put this person in Medicaid. SHADAC fixed the code to reflect this, but we cannot find a write-in response for GOVSPECFC3. This should be investigated further as a possible error in the CATI.**

Other

- SHADAC did not figure out where the Census Bureau is getting 10 instances of other, and does not fully understand the plan type chart in relation to other. Any person that ends up in an undefinable plan bucket should face a write-in. So, the only "others" we should have are those where the write-in cannot be defined. If there is an undefinable write-in for GOVSPEC, SHADAC puts the person in Medicaid, per previous conversations. Otherwise SHADAC allows people to be other if they have a write-in that we cannot figure out. SHADAC finds 2 such write-ins in the analytical sample. The Census Bureau seems to have these cases plus 10 additional.

Point-in-time validation

Medicare

- In household 4320 person 1 says their coverage started after January 2009, has unknown months of coverage in a current loop, but also says “no” to PUANYTHIS. The Census Bureau codes to Medicare point-in-time, SHADAC does not.

Medicaid

- In household 3927 persons 2 and 3 have Medicaid from PUGOVPLANLC1. Their coverage started after January 2009, they have unknown months of coverage, and they say “yes” to ANYTHIS. SHADAC codes them as point-in-time, the Census Bureau does not.

- **Household 4147 is complicated. Person 1 ends up in VERIFY and they say "no". But they are missing for SRCEGENLC1 and LC2, and we are not sure why. However, they do have a value for GOVPLANLC1 where they say they have coverage. In the months of coverage portion they say they have had non-continuous months of coverage since January 2009. In MNTHBEG2 they say their current spell started in January 2009. (Maybe they meant 2010?). So, SHADAC assigns them coverage in every month since their "current spell" started in January 2009. They are missing values (as in . not DK/REF) for SAMEMNTS1 and SAMENTHS2, but nonetheless end up in WHATMNTHS3 where they say the two other members of the plan had coverage from November 2009 to February 2010. In the end, the Census Bureau gives point-in-time coverage to these two other members, but SHADAC does not.**
- In household 4655 the Census Bureau codes person 2 as having Medicaid. SHADAC finds no evidence of Medicaid and codes them as a 'Tricare for life' policyholder.
- In household 4661 both SHADAC and the Census Bureau code persons 1 and 2 with Medicaid. But the Census Bureau codes person 3 with point-in-time Medicaid and SHADAC does not find evidence that this person is covered in COVWHO or any other direct questions.
- **In household 4953 person 1 has reports within LC1 for both ESI and Medicaid. SHADAC codes as ESI and the Census Bureau codes as Medicaid. SHADAC changed to agree with the Census Bureau, but we are not clear how this person gets two reports within the same loop.**

ESI

- In household 3742 person 3 is a dependent on person 1's plan. They are covered by different months. From WHATMNTHS3_3 it looks like they are covered January 2009-August 2009 and no months in 2010. The Census Bureau codes person 3 as point-in-time, SHADAC does not.
- **In household 3869 both the Census Bureau and SHADAC code person 1 as ESI point-in-time. Per PUPOLHOLDER person 2 is the policyholder. SHADAC codes this person also as ESI point-in-time, but the Census Bureau does not. It looks like there may be an error in the CATI with WHATMNTHS3 as it seems possible that a dependent can be covered for more months than the policyholder. Technically, they do not have a reported month of ESI in the interview month. But, per our policyholder fix rules, SHADAC assigns point-in-time coverage because they are in a plan with a dependent that has point-in-time coverage.**
- In household 3961 person 5 has a direct ESI report in PUMILPLANLC1. They have unknown months of coverage in LC1. SHADAC assigns to point-in-time per the DK months of coverage rules. The Census Bureau does not.
- In household 4189, per WHATMNTHSLC13_3, person 3 is an ESI dependent covered in January 2009-June 2009. And does not appear to have other ESI coverage. The Census Bureau codes this person as ESI point-in-time while SHADAC does not.
- In household 5315 a person reports ESI in PUMILPLANLC1. They report coverage started after January 2009 in PUBEFORAFT, but are missing (.) in PUCNTCOVLC1. They end up in MTNHBEG1LC1 and report DK. They report Yes to PUANYTHISLC1. SHADAC assigns them ESI point-in-time per the DK months rules, but the Census Bureau does not.

Direct Purchase

- In household 4078 person 1 reports Medicare in LC1 and direct purchase in LC2. They do not know the months of coverage for the direct purchase LC2 plan. But say they had the coverage this year in PUANYTHISLC2. SHADAC codes them as direct purchase point-in-time. The Census Bureau does not.
- In household 5311 person 2 has a direct report of direct purchase in PUSCREDEPDIRLC1. They do not know the start date of the plan. They are missing (.) for the start month (MNTHBEG1, etc), but say they had the coverage this year. SHADAC codes them as direct purchase point-in-time. The Census Bureau does not.
- In household 5370 person 4 reports direct purchase in PUSCREDEPDIRLC1. Months of coverage are unknown, but they say yes to ANYTHISLC1. SHADAC assigns direct purchase point-in-time coverage. The Census Bureau does not.

Military

- **In household 4655 person 2 is a policyholder on an LC2 Military plan. The leader is the dependent. The leader (person 1) says they have different months of coverage. They report *fewer* months for the policyholder than they do for themselves. SHADAC assigns person 2 to Military point-in-time coverage. The Census Bureau does not. Technically, they do not have a reported month of Military in the interview month. But, per our policyholder fix rules, SHADAC assigns point-in-time because they are in a plan with a dependent that has point-in-time.**

Other

- The Census Bureau codes 6 additional other cases than SHADAC.