2014 VHA SURVEY OF VETERAN ENROLLEES' HEALTH AND RELIANCE UPON VA

METHODOLOGICAL EXPERIMENTS AND NON-RESPONSE BIAS ANALYSIS

FINAL REPORT

— Not for Distribution —

Submitted to:

Office of the Assistant Deputy Under Secretary for Health for Policy and Planning (ADUSH/PP)

Prepared by:

126 College Street
Burlington, Vermont 05401

November 24, 2014

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. BACKGROUND

History of Survey

The Department of Veterans Affairs (VA) administers the country's largest, most comprehensive integrated health care system. More than 8 million Veterans are enrolled in the VA system and seek services ranging from specialty care to social support services to wellness maintenance. VA's authority to provide this care is regulated in part by the Veteran's Health Care Eligibility Reform Act of 1996 (Public Law 104-262). This law implements a priority-based enrollment system for Veterans and gives the Veterans Health Administration (VHA) the ability to plan to meet the needs of enrolled Veterans. Changing demographics, availability of other health care coverage, economic changes, and rising health care costs can all impact a Veteran's decision to turn to VHA for care. Understanding factors that impact Veterans' choice is critical to VA's continuous preparation and ability to meet Veterans' expectations.

The Survey of Enrollees was developed with core and supplemental groups of survey questions to gather a variety of information used to determine the relationship between utilization patterns and the demographic and socioeconomic characteristics of Veteran enrollees.

Survey of Enrollees data are used to develop health care budgets and to assist VA with its annual enrollment decisions. These data also inform the VA Enrollee Health Care Projection Model (EHCPM). Forecasts developed from this model are used for a number of purposes, such as budgeting, as well as scenario-based policy and planning analyses.

VHA has conducted twelve cycles of the Survey of Enrollees (1999, 2000, 2002, 2003, 2005, 2007, 2008, 2010, 2011, 2012, 2013, and 2014). The 2014 survey methodology can be summarized as an English-only, 15- to 20-minute survey available via Computer-Assisted Telephone Interviewing (CATI), self-administered Paper and Pencil Interviewing (PAPI), or Computer-Assisted Web Interviewing (CAWI) format, using a stratified sampling design to obtain 42,000 interviews.

ICF International, Inc. (ICF) has provided technical and data collection services to VHA in support of the Survey of Enrollees since 2005. This methodology report pertains to the 2014 data collection period from February 15 through June 30, 2014.

History of Survey of Enrollees Bias Assessments

Any information collection from the general public and conducted or sponsored by a Federal agency requires periodic Office of Management and Budget (OMB) clearance. As part of the Fiscal Year (FY) 2006 OMB clearance package, VHA was tasked with conducting a non-response bias assessment as well as examining sampling frame quality. A non-response bias assessment investigates the extent to which survey non-respondents differ from respondents in ways that may affect survey outcomes, while the examination of sample frame quality assesses the extent to which the sampling frame adequately covers, or includes all members in, the target population. In 2006, VHA and ICF met with OMB to discuss the non-response analysis and agreed to develop methods to improve the survey program. OMB granted clearance to VHA but required that VHA improve the design, starting with the 2007 survey. Since then, the Survey of Enrollees has:

- Added a pre-survey notification letter sent from the Under Secretary for Health. The letter describes the survey's purpose, explains that ICF is conducting the study on VHA's behalf, and provides a number to call with questions or to complete the survey;

- For Veterans with missing phone numbers, added a customized letter with an inbound phone number to call to complete the survey;
- Experimented with reverse phone number look-up based on address information;
- Increased the maximum number of call attempts from six to seven; and
- Improved the weighting methodology by using a propensity score adjustment based on demographics and health care utilization administrative records, as well as a post-stratification adjustment to match a consistent set of demographic control totals.

Discussion of survey bias can be organized in terms of the Total Survey Error (TSE) framework (see *Figure 1*).[1] The TSE framework divides survey error into two major sources: *errors of representation*, which are due to the systematic and random errors that influence which members of the population respond to the survey; and *errors of observation*, which are due to the systematic and random errors that influence the accuracy with which survey constructs are measured. Random error is reduced through the use of large sample sizes, such as those used in the Survey of Enrollees. On the other hand, systematic error, which is also referred to as bias, is a consistent deviation from the "true" score for a survey outcome and is not mitigated by large sample sizes. This report focuses specifically on bias in the Survey of Enrollees, both with respect to errors of observation and errors of representation.

Figure 1. Total Survey Error Analysis of the 2014 Survey of Enrollees
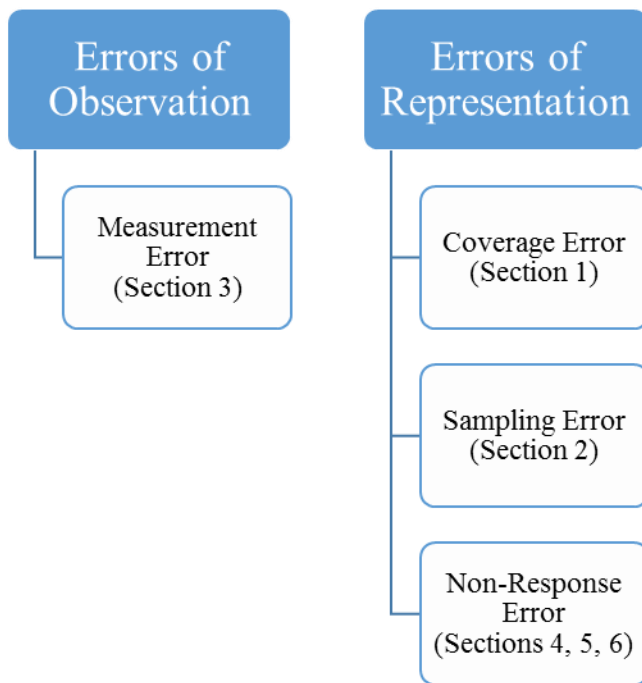
Biases in representation can arise from three major sources:
- Coverage error, due to systematic differences between enrollees included in vs. excluded from the sampling frame;
- Sampling error, due to a non-random selection mechanism or unadjusted disproportionate sampling; and
- Non-response error, due to respondents systematically differing from non-respondents with regard to survey outcomes.

---

[1] Groves Robert, Fowler Floyd, Couper Mick, Singer Eleanor, Tourangeau Roger. *Survey Methodology*. New York: Wiley; 2004.

Beginning in 2012, VHA introduced a mail mode to extend coverage to enrollees without a phone number or with a non-working number, as well as a Web survey as an alternative to mail or telephone modes. The inclusion of enrollees without a valid or working phone number in the sampling frame addressed the undercoverage of these enrollees that existed prior to 2012.[2] Beginning in 2013, the *Methodological Experiments Report* also evaluated the possibility of bias due to sampling error to verify that the random selection mechanism and subsequent design weights used to adjust for disproportionate sampling are operating as expected (see *Section 2. Sampling and Weighting Design and Bias* Evaluation). Finally, bias due to enrollee non-response continues to be evaluated by comparing responding and non-responding enrollees using available frame variables (see *Non-Response Bias Analysis*).

Biases in observation are generally due to systematic measurement error, which can arise from a variety of sources, such as question wording or item order. Since 2012, the most important potential source of systematic measurement error in the Survey of Enrollees has been the use of multiple survey modes. Although the introduction of multiple modes was needed to extend coverage to a large segment of the enrollee population, doing so necessarily introduced the possibility of mode effects. Mode effects occur when responses to survey items in one mode systematically differ from responses in other modes. Post-hoc analyses (in 2012) and a methodological experiment (in 2013) have been conducted to test for mode effects by comparing responses in the mail and CATI modes. Although some statistically significant differences have been observed, the magnitude of these differences is generally quite small. The methodological experiment was conducted again in 2014, using random assignment to survey modes as in 2013 (see *Section 3. Experiment 1 – Impact of Survey Mode on Survey Estimates*).

This 2014 report addresses sources of potential bias in both representation and observation in the 2014 Survey of Enrollees (see *Figure 1*). Following the organization of the report established in the 2013:
- Section 2 of this report evaluates the sampling and weighting processes to verify that they are unbiased.
- Sections 3, 4, and 5 report the results of the methodological experiments conducted as part of the 2014 Survey of Enrollees, including an experiment to evaluate measurement error introduced by the use of multiple survey modes and two experiments designed to reduce enrollee non-response.
- Section 6 evaluates the potential for non-response bias.

Summary of Methodological Experiments, 2006–2013

Since 2006, ICF has conducted a bias assessment and has evaluated the results of methodological experiments designed to reduce bias.

Experiments Conducted Prior to Introduction of Mixed-Mode Design

In 2006, ICF used the 2005 data to examine the survey process and potential biases resulting from missing or outdated contact information as well as survey non-response—including both the inability to make contact and the effects of respondent refusals. The report, submitted to OMB, included several recommendations to improve the research design.

One of the resulting recommendations was a propensity score weighting adjustment. This weighting adjustment, also used in 2007 and 2008, corrects for differential non-response by health care utilization and demographic information. To determine the adjustment, ICF:

---

[2] A small possibility of coverage error remains due to the frame exclusion criteria VHA applies when extracting the sampling frame from the enrollee database. Specifically, enrollees lacking a valid mailing address, living outside the U.S. or Puerto Rico, or missing one of the stratification variables are currently excluded from the sampling frame.

- Used a probability model (described below) to estimate an enrollee's individual propensity (or probability) of being in the respondent sample;
- Grouped enrollees into five equal-sized classes (or quintiles) with similar probabilities; and
- Weighted the respondents up to account for the non-respondents, using an independent adjustment for each class.

The assumption is that non-respondents would have given similar responses to the survey as the actual respondents within the quintile in which they are grouped. The accuracy of this assumption depends on the fit of the statistical model used to create these quintiles. The propensity score weighting adjustment then reduces potential bias to the extent that non-respondents and respondents with similar response probabilities are also similar with respect to the survey statistics of interest.

The 2007 Survey of Enrollees included several methodological experiments to gauge the impact of design enhancements. These experiments included sending pre-survey notification letters to potential respondents signed by the Under Secretary for Health and extending the maximum number of call attempts from six to 10.

The results of these experiments are documented in the 2007 report, *Supplementary Analysis and Technical Assistance for the 2007 Annual Survey of Veteran Enrollees Health and Reliance on VA*. The response rate among the experimental treatment group (pre-survey notification letter and 10 call attempts) more than doubled that of the control group (no pre-survey notification letters and six call attempts), at 43.3 percent vs. 21.4 percent, respectively. Based on the evidence, ICF recommended that VHA adopt both of these design enhancements for the 2008 Survey of Enrollees. VHA approved sending pre-survey notification letters and increasing the maximum call attempts to seven (concern for increased respondent burden prevented an increase to 10).

Also during the 2007 Survey of Enrollees, enrollees were sampled only from a frame of enrollees with telephone numbers. Enrollees without telephone numbers had no chance of selection—introducing a potential source of coverage bias. The 2007 survey was therefore susceptible to two major forms of bias affecting representation: coverage of enrollees with no chance of selection, and non-response bias among enrollees who did not respond. For that reason, two separate propensity score adjustments were developed: one for frame coverage and another for non-response.

In 2008, VHA approved a methodological experiment to improve sample frame coverage: utilizing reverse telephone look-up directories that used respondent addresses to obtain valid telephone numbers from a sample of 62,516 enrollees. This new process resulted in 59,426 potential respondents (95 percent coverage of the test sample), and this group yielded 12,765 completed surveys.

Since the 2008 Survey of Enrollees, the survey sample has been selected from a frame of enrollees with and without telephone numbers. Since the sample has been selected from this complete frame, coverage bias has not been a concern. However, non-response due to a variety of sources, including invalid contact information, has remained an issue. Some of these sources have been addressed through the addition of a mail survey and a Web response channel; however, some sources of potential non-response bias remain. Therefore, a single propensity score adjustment has been used to provide a general mechanism for mitigating bias due to non-specific non-response.

The 2010 Survey of Enrollees followed a methodology similar to the 2008 survey—including a reverse phone number look-up from a sample of 62,515 enrollees. Again, the results indicated that the address matching improved contact information quality, resulting in 61,376 potential respondents (98 percent coverage of the test sample). This experimental group yielded 16,851 completed surveys.

For 2011, the plan for the Survey of Enrollees also included reverse telephone look-ups. Unfortunately, this service was not implemented because the address-matching vendor was not able to comply with the project's security requirements. However, the 2011 survey did include a tailored pre-survey notification letter sent to enrollees with a known address but unknown telephone number, as listed in the database. This letter asked the enrollee to call ICF to participate in the survey. This test yielded 244 interviews from 15,339 total enrollees without phone numbers. While relatively few, these respondents represent Veterans who would not otherwise have been included in the results.

Experiments Conducted Following Introduction of Mixed-Mode Design

For 2012, two new survey modes were added to the existing telephone mode. The Survey of Enrollees had been conducted strictly as a telephone interview since its inception in 1999. Enrollees with invalid telephone numbers (e.g., missing or incorrect area code) or without a telephone were not included, and this was a source of potential coverage bias. In 2012, VHA addressed this undercoverage by developing an experimental mail survey that was sent to all enrollees without a valid telephone number. The mail survey allowed respondents to complete the survey via paper-and-pencil; the mailed materials also provided contact information if the Veteran wished to call ICF to complete a telephone interview and a link to a Web survey option. In addition, ICF conducted a follow-up mailing for phone non-respondents. Respondents in all modes also could request a mail survey at any point.

In addition to adding a mail survey, VHA offered an experimental Web option for the first time. Thirteen percent of enrollees used the Web option instead of returning a mail survey or participating in a telephone interview. Due to the cost savings on interviewer labor generated by the Web option, ICF recommended that VHA continue offering this mode.

The experimental mail survey improved the response rate and reduced bias. Counting responses via all four response channels (i.e., Web, mail, inbound CATI, and outbound CATI), the addition of a mail component (mail survey, allowing mail requests, and mail follow-up) added 10,056 interviews.

While ICF recommended that the mail mode continue to be offered, a limitation was noted in the 2012 experimental design; specifically, the confounding of survey mode with sample type meant that differences in survey responses between the survey modes could also be explained by pre-existing differences between the populations choosing to respond in each mode. ICF therefore recommended a randomized methodological experiment testing survey mode effects, which was conducted in 2013. This experiment tested for survey mode effects by randomly assigning a subset of eligible sampled enrollees to receive either the mail or CATI survey as their default mode of survey administration (i.e., the mode in which enrollees would complete the survey unless they explicitly opted to complete in a different mode). Results indicated that although survey mode (mail vs. CATI) does have a significant effect on some survey responses, the magnitude of this effect is generally quite small. ICF thus recommended continuing to administer the Survey of Enrollees in multiple modes, given the substantial increase in coverage this design affords.

A second methodological experiment was conducted in 2013 to test the effect of a second survey mailing (SSM) on response rates as part of the mail follow-up protocol for non-responding enrollees and non-working phone records. The results of this experiment indicated that, among enrollees who do not respond to the CATI survey and enrollees with non-working numbers, a second survey mailing as part of a mail follow-up protocol significantly improves response rates (by approximately seven percentage points).

Overview of Methodological Experiments, 2014

In 2014, ICF conducted three methodological experiments to investigate bias due to survey mode and enrollee non-response.

The first experiment, replicating a design first used in 2013, tested for survey mode effects by randomly assigning a subset of eligible sampled enrollees to receive either the mail or CATI survey as their default mode of survey administration (i.e., the mode in which enrollees would complete the survey unless they explicitly opted to complete in a different mode). This experiment has the potential to reveal systematic differences in survey responses due to survey mode (specifically, mail vs. CATI modes).

The second and third experiments tested the effects of "short" and "long" mail protocols on response rates among different subpopulations. In general, the "short" mail protocol involved only one survey packet mailing, whereas the "long" mail protocol involved two complete survey packet mailings.

The second experiment, also replicating a design first used in 2013, tested the effect of a second survey mailing (Second Survey Mailing/Follow-up to Phone/CATI Protocol, hereafter referred to as SSM-P) on response rates as part of the mail follow-up protocol used for CATI non-respondents and non-working phone records.

The third experiment tested the effect of a second survey mailing (Second Survey Mailing/Follow-up to Mail Protocol, hereafter referred to as SSM-M) on response rates as part of the mail survey protocol used for enrollees with only a valid mailing address. These latter two experiments continue the OMB-required research to improve response rates and to minimize non-response bias. *Figure 2* illustrates how sample was assigned to all three experiments conducted in 2014.

Figure 2. Assignment of Sample to the 2014 Methodological Experiments
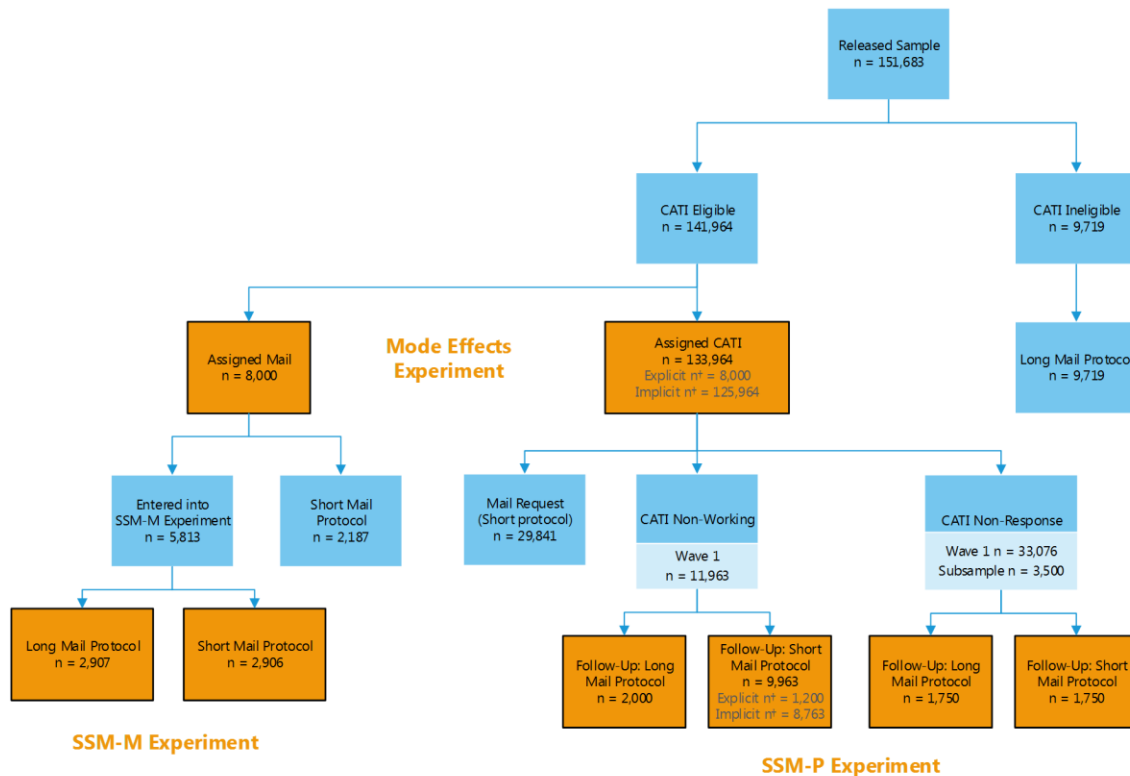
†For the Mode Effects and SSM-P experiments, sample was explicitly assigned to conditions as recommended by power analyses. In both cases, additional sample was implicitly included in the control condition due to the fact that the treatment for records not explicitly assigned to an experimental condition was identical to that of records explicitly assigned to the control condition.

2. SAMPLING AND WEIGHTING DESIGN AND BIAS EVALUATION

This section briefly presents the sampling protocol and corresponding weighting plan of the 2014 Survey of Enrollees (a detailed description of the methodology can be found in *VHA Survey of Veteran Enrollees' Health and Reliance Upon VA Methodology Report 2014*). Afterward, the bias component of the total mean squared error that can be attributed to the sampling and weighting processes is evaluated.

Sampling

Sample Stratification and Allocation

The 2014 sampling design modifies a basic framework designed to support estimates by Veterans Integrated Service Network (VISN)[3] (21 levels) and priority group[4] (eight levels) with additional stratification and oversampling by gender. In addition to this "Main" sample, an independent "Supplemental" sample was drawn of enrollees identifying as Hispanic/Latino. These modifications to the 2013 sampling design were made to increase data utility for these two emerging Veteran populations (i.e., female Veterans and Hispanic/Latino Veterans).

For the Main sample, each of the 21 VISNs was allocated 1,875 interviews as follows:

1. First, minimum sample sizes were allocated to each priority group:
   - 50 for Priority Group 7;
   - 150 for Priority Groups 4 and 6;
   - 250 for Priority Groups 1, 2, 3, and 5; and
   - 400 for Priority Group 8.
2. Second, 125 interviews were proportionally allocated to the largest priority groups within the VISN.

Within each of the 168 VISN $\times$ priority group strata, women were oversampled by allocating sample at twice the proportion of men. For example, if 10% of the stratum are women, the sample allocation would be $2\times10\% / (2\times10\% + 90\%) = 18\%$ women and 92% men.

The Supplemental sample (Hispanic/Latino) was allocated 2,625 interviews. The sample was stratified by VISN and sample was allocated in proportion to the number of Hispanics flagged on the frame. The sample selection was simple random sample drawn from each VISN's population of Hispanic/Latino enrollees.

Frame Development

VHA provided a random stratified sample of 418,832 records from its enrollee database as follows:

- VHA extracted the entire universe of enrollees who were listed as of September 30, 2013; this list included Veterans enrolled in VA health care and living in both institutionalized and non-institutionalized settings.
- VHA then eliminated all records meeting one or more of the following criteria:

---

[3] VISN is the geographic health care administration region to which each Veteran is assigned.
[4] Priority group is the patient priority group to which a Veteran was assigned at enrollment. Priority groups help VA provide health care services relative to annual funding.

- o Lacking a valid address;
- o Not living in the U.S. or Puerto Rico; or
- o Missing one of the stratification variables listed in next bullet.
- Remaining was a final file of 8,486,965 enrollees to be stratified by VISN, priority group, and gender, from which the 2014 Main sample was drawn. The 2014 Supplemental sample was also drawn from this file after filtering to include only enrollees positively identifying as Hispanic/Latino.

Sampling Process

ICF then randomly selected a subsample of these records to meet the target sample sizes in each stratum. ICF released records into the study as needed, using a random selection algorithm. To do so, ICF monitored the number of completed interviews during fielding. ICF then compared the estimated sample yield (that is, the number of completed interviews predicted from the sample at a given point in the study) to the target number required by the sampling plan. To match actual to planned performance, enrollee records were drawn and released into the study for calling/mailing randomly from the final, stratified set of records provided by VHA.

A total of 140,698 enrollees were sampled to meet these sample size requirements in all strata of the Main sample, and a total of 11,456 enrollees were sampled to meet these sample size requirements in the Supplemental sample (with an overlap of 471 enrollees).

Following data collection, ICF evaluated the Main and Supplemental samples to determine whether or not they should be combined into a single analytic dataset. Because the sample size increase gained by combining the two samples outweighed the loss of precision due to increased weighting variance, it was decided to combine the two samples (a more detailed description of this analysis can be found in *VHA Survey of Veteran Enrollees' Health and Reliance Upon VA Methodology Report 2014*). The evaluation of the weighting process will therefore focus only on the combined sample.

Weighting

The analysis weight is a product of three components:

1. A design weight that adjusts for differential selection probabilities across sampling strata and accounts for the increased probability of selection of Hispanic/Latino enrollees in the combined sample;
2. A non-response adjustment that compensates for differential response patterns across enrollee subgroups; and
3. A post-stratification adjustment that aligns weighted totals with population control totals along a set of key demographic dimensions.

Design Weight

The design weight adjusts for differential selection probabilities and accounts for overlap created by combining the Main and Supplemental samples. The Main sample was selected from the complete survey frame independently in each of the strata, which had been defined by VISN, priority, and gender. The Hispanic sample was selected from the filtered survey frame as a simple random sample (i.e., where the frame serves as the single stratum). The probability of selection for enrollees in the Main and Supplemental samples in the ith stratum is then calculated equivalently as $p_i = n_i/N_i$, where:

p_i = the probability of selection for each enrollee in the ith stratum
n_i = the number of enrollees sampled in the ith stratum
N_i = the total number of enrollees in the ith stratum

The inverse of these selection probabilities is the design weight, $w_1 = 1/p$, which is calculated for all sampled enrollees in both the Main and Supplemental samples.

In the combined sample, Hispanic/Latino enrollees received a probability of selection in both the Main and Supplemental samples. The selection probability of the combined-sample design weight was computed to account for this. Specifically, if $p_{S,j}$ represents the probability of drawing the jth enrollee from the Supplemental (Hispanic/Latino) frame, and $p_{M,j}$ represents the probability of drawing the same jth enrollee from the Main sample frame, then the correct selection probability for the jth enrollee in the combined sample is given by $p_j = p_{S,j} + p_{M,j} - (p_{S,j} \times p_{M,j})$, and the combined-sample design weight is taken as $w_1 = 1/p$. This is the delivered design weight that was used as the basis for the following non-response and post-stratification adjustments.

Non-Response Adjustment

To calculate the non-response adjustment, each sampled enrollee was classified into a non-response category (y) based on whether the attempted interview was complete or incomplete:

$$y = \begin{cases} 0 & \text{if interview is an incomplete interview} \\ 1 & \text{if interview is a complete interview} \end{cases}$$

Using logistic regression, ICF estimated the probability that an enrollee completed the interview given his or her characteristics:

$\Pr(y = 1 \mid x) = \dfrac{e^{x'\beta}}{1 + e^{x'\beta}}$, where x is a matrix of sampled enrollees and each enrollee has a set of p covariates, $\mathbf{x}'_i = (1, x_{1i}, ... x_{pi})$ for enrollee i. This set of covariates was used as explanatory (or predictor) variables, and $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)$ was a set of regression coefficients, or parameters.

The predictor variables included:

- The sample design variables (VISN, priority status, gender, and Hispanic/Latino);
- Design variables previously used for sample stratification (OEF/OIF/OND status, and enrollee type: Pre- vs. Post-enrollees);
- Seven administrative health measures (listed below);
- Demographic variables (age, urban/rural address);
- Telephone number status (valid, not valid); and
- A flag identifying whether multiple enrollees use the same telephone number.

VHA provided a file based on administrative records; the file indicated whether an enrollee had utilized any of the following VHA services in the previous year (the file did not indicate the frequency of use or amount paid for any of these benefits):
1. Received long-term care benefits,
    a. Institutional
    b. Non-institutional
2. Inpatient treatment,
    a. Mental health or substance abuse
    b. Non-mental health and non-substance abuse
3. Outpatient treatment,
    a. Mental health or substance abuse
    b. Non-mental health and non-substance abuse
4. VHA pharmacy services.

The utilization indicators have been used for weighting since the 2007 survey. From 2007–2010, the indicators were sourced from VHA workload files based on bed section and clinic stop. This categorization indicates *where* a Veteran received care. For the 2011 and 2012 survey, the indicators were based on service utilization from Health Service Categories (HSCs), indicating *what* care a Veteran received. A second change was to include institutional and non-institutional long-term care indicators as compared to 2007–2010, when a single measure of home health service was used.

The outcome of the model is the propensity score, the estimated probability that the enrollee is in the final sample of respondents given their characteristics (as defined by the list of predictor variables above).

After estimating each sampled enrollee's probability of completing an interview based on the predictor variables, respondents and non-respondents were grouped into quintiles based on their propensity score. Within each quintile, respondents were ratio-adjusted to account for non-respondents. The first quintile represents the enrollees with the lowest propensity scores; this means that these enrollees are less likely to be in the final sample—thus, they receive the largest weights. The last quintile represents the enrollees with the highest propensity scores; this means that these enrollees are more likely to be in the final sample of respondents—thus, they receive the smallest weights. See *Appendix B – Non-Response Propensity Score Quintiles* for distributions of propensity score predictors for respondents by propensity score quintiles.

Table 1. Non-Response Adjustment

Percentile	Response	Non-Response	Non-Response Adjustment (NR)
0 – <20th	211,860	1,485,526	8.01
20th – <40th	393,666	1,302,485	4.31
40th – <60th	506,344	1,192,120	3.35
60th – <80th	671,644	1,025,906	2.53
80th – <100th	804,079	893,335	2.11

To calculate the non-response adjusted weights, each respondent's design weight w_1 was multiplied by the adjustment factor NR from the quintile where he or she fell: $w_2 = w_1 \times NR$.

Post-Stratification Adjustment

Because the 2014 sample design departed from the design used in previous years, a post-stratification adjustment was included as part of the weighting to promote comparability. The primary motivation for the post-stratification adjustment is to ensure that the distribution of the weighted sample matches the distribution of the enrollee population across a stable set of dimensions, such as age and gender. Because these post-stratification dimensions are independent of the dimensions used to define sampling strata in a given year, the post-stratification adjustment facilitates flexibility in the sampling design while preserving comparability across years.

Unlike previous years, the 2014 sample stratification did not include OEF/OIF/OND status and pre/post-enrollee status. Including these dimensions in the post-stratification adjustment restores comparability to previous years.

Finally, as the enrollee age distribution is related to both of these sets of variables, as well as to reliance measures, age was included in the post-stratification. Enrollee age was categorized into seven levels: under 35; 35-44; 45-54; 55-64; 65-74; 75-84; and 85+.

The dimensions used for post-stratification in 2014 were as follows:
- Age x gender (14 levels),
- Hispanic/Latino status (two levels),
- Priority x VISN (168 levels),
- OEF/OIF/OND status (two levels), and
- Pre/Post-enrollee status (two levels).

The post-stratification adjustment was implemented via a raking, or iterative proportional fitting, algorithm. During each iteration, the non-response-adjusted weight w_2 was ratio-adjusted to match population totals along each of the above post-stratification dimensions in turn. This iterative process continues until the weighted totals match population totals along all dimensions within a specified tolerance (in this case, by less than 1.00). For the 2014 combined sample, convergence was achieved after 15 iterations, indicating a stable adjustment. The post-stratification adjustment increased the coefficient of variation of the weights (a measure of the weighting variability) from 0.78 to 0.83, indicating that only a small increase in variance was required to achieve this bias reduction. The post-stratified weight w_3 was delivered with the weighted data and should be used as the analytic weight when generating population estimates.

Survey Outcomes

Of the 418,832 records supplied by VHA, 151,683 were released into the study, resulting in 42,324 completed interviews. For the CATI treatment, 36,393 interviews were obtained with an American Association for Public Opinion Research (AAPOR) response rate (RR1) of 34 percent.[5] For the mail treatment, 5,931 interviews were obtained for an AAPOR response rate of 40 percent.

Bias Assessment

The Survey of Enrollees differs from most population-based surveys in that a considerable amount of information about the population under study is available. Specifically, seven measures of health care utilization, along with basic demographics, are present on the sampling frame, or "Universe File," for all enrollees. This allows us to compute the total mean squared error (MSE) and its components—bias and variance—for estimates of service utilization rates under different sampling and weighting schemes.

Using a resampling methodology, 400 Main and Supplemental replicate samples were drawn using the current stratification and allocation scheme. Specifically, to simulate the 2014 sampling design, each replicate involved drawing an independent Main and Supplemental sample and then combining the two samples using the combined design weight (w_1) described above. As non-response was not simulated, the non-response weight (w_2) and post-stratified weight (w_3) were not computed. For each sample replicate, each of the seven service utilization percentages (p) were computed. For each service, averaging the estimated utilization percentage across the $n = 400$ replicates approximates the expected value ($\hat{P}$) of the utilization measure produced by the sampling process:

$$\hat{P} = \frac{\sum_n p_i}{n}$$

where n is the number of sample replicates and p_i is the utilization measure from the ith sample replicate for a given service. Since the true value (P) for each utilization measure can be computed from the

[5] Documentation for these response rates is available at
http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions2&Template=/CM/ContentDisplay.cfm&ContentID=3156

sampling frame delivered by VHA, the bias in the estimate of each utilization measure can be estimated as the difference between the true value and the estimate produced by the resampling procedure:

$$\widehat{bias}(P) = \hat{P} - P$$

Bias estimates were computed for both unweighted and design-weighted data for the seven utilization measures, overall and by stratification variable categories (i.e., VISN, priority group, and gender, yielding 31 separate domains). While simple random samples were drawn within each cell defined by the crossing of all stratification variables, the disproportionate allocation means that within an overall category of a stratification variable (e.g., Priority Group 1), the sampling process did not yield a simple random sample.

Because disproportionate stratified samples are not design-unbiased, some bias in the unweighted estimates is therefore expected. This expectation is confirmed in *Table 2*, which displays both the unweighted estimated percentage ($\hat{P}_{w0}$) and the estimated bias ($\hat{P}_{w0} - P$) for each of the seven utilization measures. Negative values for bias indicate that the sample design underestimates the true value, whereas positive values indicate that the sample design overestimates the true value.

The unweighted sampling bias ranged from -3.56 percentage points to 5.48 percentage points across all seven measures and 31 domains, with a median of 0.20 and an interquartile range of 0.70. Fifty-nine of the 223 total domain bias estimates exceeded 1.00 percentage points. *Figure 3* displays the overall distribution of the unweighted bias estimates in red.

Figure 3. Distribution of Estimated Bias for 7 Utilization Percentages across 31 Domains, Unweighted (w0) and Design-Weighted (w1)
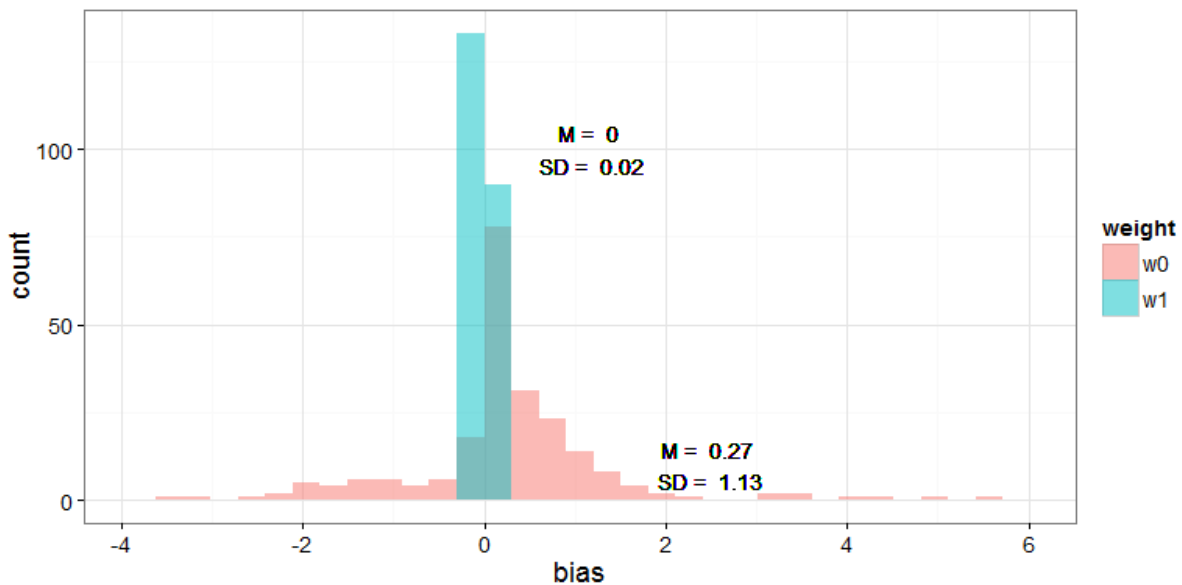
Note that the larger biases were not evenly scattered across subgroups and measures; this indicates that there are correlations between utilization rates and the characteristics used to define sampling strata, and indicates the need for weighting to reduce this bias in representation.

The design weight (w_1), computed as the inverse of combined-sample selection probabilities, compensates for the disproportionate sample allocation. The design-weighted sampling bias, distributed across estimates and stratification variables as depicted in blue in *Figure 3*, is negligible. The bias ranged

from -0.07 percentage points to 0.05 percentage points, with a median of 0.002 percentage points and an interquartile range of 0.02 percentage points. None of the design-weighted utilization measures in any stratification domains produced an expected bias above 1.00 percentage points. *Table 3* provides the design-weighted estimated percentage ($\hat{P}_{w1}$) and the estimated bias ($\hat{P}_{w1} - P$) for each of the seven utilization measures.

Overall, then, the sampling process, with proper weighting, is exhibiting minimal bias and is performing as expected.

The next sections examine the potential for bias due to other components of the survey process, particularly bias due to mode effects and to non-response.

Table 2. Sampling Process Bias Assessment, Unweighted Estimates

Stratum	Level	Inpatient Mental Health and Substance Abuse Pct. Estimate	Bias	Inpatient Non-Mental Health and Substance Abuse Pct. Estimate	Bias	Institutional Long-term Care Pct. Estimate	Bias	Non-Institutional Long-term Care Pct. Estimate	Bias	Outpatient Mental Health and Substance Abuse Pct. Estimate	Bias	Outpatient Non-Mental Health and Substance Abuse Pct. Estimate	Bias	Prescription Drug Pct. Estimate	Bias
Overall	**1**	**1.45**	**0.31**	**4.87**	**0.41**	**0.64**	**0.12**	**4.30**	**0.70**	**17.37**	**1.03**	**61.53**	**-0.67**	**53.35**	**-0.86**
VISN	1	1.89	0.47	4.08	0.43	0.67	0.05	4.04	0.73	17.42	1.36	62.71	-0.51	52.58	-0.54
VISN	2	1.53	0.34	5.54	0.86	0.67	0.16	5.17	1.05	15.34	1.86	60.27	1.64	51.56	1.27
VISN	3	1.55	0.44	4.07	0.85	0.61	0.12	5.58	1.06	17.31	3.38	53.31	4.04	45.71	3.50
VISN	4	1.48	0.31	4.16	0.55	0.81	0.17	4.31	0.89	16.83	2.31	62.79	0.99	53.42	0.79
VISN	5	1.45	0.24	4.25	0.39	0.87	0.17	4.82	0.90	14.72	-0.01	49.23	-2.11	41.05	-1.93
VISN	6	1.31	0.25	4.35	0.20	0.44	0.05	4.39	0.55	16.48	0.08	60.47	-1.45	53.81	-1.54
VISN	7	1.10	0.01	3.95	-0.11	0.27	-0.02	2.90	0.14	16.83	-1.19	59.44	-2.36	52.61	-2.49
VISN	8	1.68	0.48	6.21	0.79	0.43	0.05	5.44	1.14	22.54	4.49	73.49	4.95	64.64	5.48
VISN	9	1.69	0.41	5.70	0.22	0.44	0.05	4.13	0.38	17.43	0.13	63.77	-1.22	56.32	-1.46
VISN	10	1.44	0.19	5.31	0.32	0.77	0.12	6.14	0.81	19.51	1.37	64.32	0.19	56.33	0.05
VISN	11	1.31	0.26	4.17	0.37	0.59	0.09	4.96	0.56	16.00	1.01	63.85	0.03	55.92	-0.17
VISN	12	2.07	0.70	6.39	1.15	1.10	0.27	4.83	1.15	18.86	3.12	66.02	1.54	58.52	1.26
VISN	15	1.70	0.42	5.40	0.46	0.65	0.10	3.82	0.51	16.80	0.79	63.69	-0.89	56.13	-1.08
VISN	16	1.38	0.21	4.73	0.04	0.41	0.01	3.68	0.10	17.71	-0.16	62.27	-1.52	55.57	-1.99
VISN	17	1.37	0.17	4.12	0.18	0.42	0.00	3.17	0.30	17.06	-0.34	59.32	-1.03	52.40	-1.17
VISN	18	1.38	0.33	5.59	0.64	0.76	0.13	4.04	0.74	17.83	1.34	64.41	1.92	56.16	1.66
VISN	19	1.37	0.30	4.64	0.42	0.73	0.13	4.48	0.70	16.61	1.04	60.98	0.05	51.83	0.20
VISN	20	1.26	0.17	4.56	0.21	0.53	0.04	2.90	0.30	15.42	-0.07	60.19	-1.35	51.78	-1.37
VISN	21	1.10	0.22	5.54	0.54	0.84	0.14	3.92	0.54	17.33	0.70	61.14	0.33	51.91	0.05
VISN	22	1.18	0.23	4.75	0.27	0.61	0.06	2.72	0.22	19.03	1.70	57.68	1.48	48.66	1.16
VISN	23	1.30	0.31	4.70	0.56	0.92	0.15	4.79	0.69	14.45	1.28	66.18	-1.52	56.79	-1.00
Priority Group	1	2.38	0.26	7.66	0.03	1.34	-0.02	7.06	0.32	39.58	3.05	83.41	0.71	76.69	0.79
Priority Group	2	0.93	0.04	3.46	0.03	0.25	0.00	2.84	0.12	18.52	1.36	65.64	0.02	54.57	0.25
Priority Group	3	0.79	0.03	2.99	-0.03	0.23	-0.01	2.42	0.04	12.27	0.99	57.65	-1.11	46.05	-0.71
Priority Group	4	6.66	0.23	16.58	0.20	3.34	0.23	17.56	0.67	30.55	0.72	77.89	-0.31	73.77	-0.40
Priority Group	5	1.41	0.03	5.77	-0.06	0.42	0.00	3.80	0.12	17.32	1.18	62.52	0.81	56.92	0.45
Priority Group	6	0.30	0.01	1.05	-0.03	0.04	0.00	0.71	0.00	8.96	0.80	41.46	-1.65	30.27	-1.42
Priority Group	7	0.59	0.04	3.98	0.07	0.33	-0.04	3.43	-0.05	10.60	1.12	75.16	-0.39	62.39	0.22
Priority Group	8	0.17	0.01	1.49	0.02	0.09	0.00	1.48	0.00	4.65	0.43	47.59	-1.85	40.37	-1.88
Gender	F	1.36	0.02	3.94	-0.15	0.35	0.05	3.57	0.21	21.23	-2.20	56.66	-3.13	48.76	-3.56
Gender	M	1.47	0.35	5.08	0.59	0.70	0.16	4.47	0.84	16.51	0.72	62.62	0.23	54.37	0.01

Table 3. Sampling Process Bias Assessment, Design-Weighted Estimates

Stratum	Level	Inpatient Mental Health and Substance Abuse		Inpatient Non-Mental Health and Substance Abuse		Institutional Long-term Care		Non-Institutional Long-term Care		Outpatient Mental Health and Substance Abuse		Outpatient Non-Mental Health and Substance Abuse		Prescription Drug	
		Pct. Estimate	Bias	Pct. Estimate	Bias	Pct. Estimate	Bias	Pct. Estimate	Bias	Pct. Estimate	Bias	Pct. Estimate	Bias	Pct. Estimate	Bias
Overall	**1**	**1.14**	**0.00**	**4.46**	**0.00**	**0.52**	**0.00**	**3.61**	**0.00**	**16.34**	**-0.01**	**62.20**	**-0.01**	**54.20**	**-0.01**
VISN	1	1.44	0.01	3.66	0.01	0.63	0.00	3.32	0.01	16.04	-0.02	63.20	-0.01	53.11	-0.01
VISN	2	1.19	0.00	4.67	-0.01	0.51	0.01	4.10	-0.02	13.51	0.04	58.59	-0.04	50.24	-0.04
VISN	3	1.11	0.00	3.23	0.01	0.49	0.00	4.52	0.00	13.92	0.00	49.25	-0.02	42.20	-0.01
VISN	4	1.16	-0.01	3.61	0.01	0.63	0.00	3.42	-0.01	14.47	-0.04	61.83	0.03	52.63	0.01
VISN	5	1.20	-0.01	3.86	0.00	0.71	0.00	3.91	-0.01	14.73	0.00	51.32	-0.02	42.93	-0.05
VISN	6	1.06	0.00	4.16	0.01	0.40	0.00	3.85	0.00	16.40	0.00	61.89	-0.03	55.33	-0.01
VISN	7	1.09	0.00	4.05	-0.01	0.28	-0.01	2.76	-0.01	18.05	0.03	61.81	0.02	55.15	0.04
VISN	8	1.20	0.00	5.41	-0.01	0.38	0.00	4.31	0.01	18.02	-0.03	68.51	-0.03	59.12	-0.04
VISN	9	1.29	0.01	5.46	-0.02	0.39	0.00	3.77	0.02	17.27	-0.02	65.03	0.03	57.82	0.05
VISN	10	1.24	-0.01	4.99	0.00	0.65	0.00	5.32	0.00	18.14	0.00	64.11	-0.02	56.26	-0.01
VISN	11	1.04	0.00	3.79	-0.01	0.50	0.00	4.39	0.00	15.00	0.01	63.83	0.01	56.10	0.01
VISN	12	1.36	0.00	5.23	-0.01	0.83	-0.01	3.68	0.00	15.76	0.02	64.50	0.03	57.31	0.04
VISN	15	1.29	0.02	4.93	0.00	0.55	0.00	3.28	-0.02	16.03	0.01	64.58	0.00	57.18	-0.03
VISN	16	1.18	0.01	4.70	0.01	0.40	0.00	3.56	-0.02	17.85	-0.01	63.76	-0.03	57.54	-0.02
VISN	17	1.19	-0.01	3.93	-0.01	0.42	0.00	2.86	-0.01	17.39	-0.01	60.32	-0.03	53.53	-0.03
VISN	18	1.05	0.00	4.94	-0.01	0.63	0.01	3.34	0.03	16.48	-0.01	62.54	0.05	54.54	0.05
VISN	19	1.07	0.01	4.21	-0.01	0.60	0.00	3.79	0.00	15.60	0.03	60.93	0.00	51.59	-0.03
VISN	20	1.09	0.00	4.36	0.00	0.48	-0.01	2.63	0.02	15.46	-0.04	61.50	-0.03	53.08	-0.07
VISN	21	0.87	-0.01	5.01	0.00	0.70	0.00	3.37	-0.01	16.61	-0.02	60.80	-0.02	51.81	-0.05
VISN	22	0.96	0.01	4.48	0.00	0.55	0.00	2.51	0.00	17.35	0.02	56.19	0.00	47.52	0.01
VISN	23	0.99	0.00	4.13	-0.01	0.76	0.00	4.09	0.00	13.17	0.00	67.69	-0.01	57.81	0.02
Priority Group	1	2.11	0.00	7.63	-0.01	1.36	0.00	6.74	0.01	36.52	-0.01	82.68	-0.01	75.89	-0.01
Priority Group	2	0.89	0.00	3.44	0.01	0.26	0.00	2.72	0.01	17.17	0.00	65.64	0.02	54.34	0.02
Priority Group	3	0.76	0.00	3.01	-0.01	0.24	0.00	2.38	0.00	11.26	-0.03	58.75	-0.01	46.72	-0.03
Priority Group	4	6.46	0.03	16.39	0.01	3.12	0.01	16.89	0.00	29.85	0.02	78.17	-0.03	74.17	0.00
Priority Group	5	1.38	0.00	5.82	-0.01	0.41	0.00	3.68	0.00	16.13	0.00	61.70	0.00	56.48	0.02
Priority Group	6	0.29	0.00	1.08	0.00	0.04	0.00	0.71	0.00	8.15	-0.01	43.06	-0.06	31.65	-0.03
Priority Group	7	0.55	0.00	3.92	0.01	0.36	0.00	3.45	-0.03	9.41	-0.07	75.59	0.04	62.17	0.00
Priority Group	8	0.16	0.00	1.48	0.00	0.09	0.00	1.48	0.00	4.23	0.01	49.44	0.00	42.24	-0.01
Gender	F	1.35	0.00	4.09	0.00	0.30	0.00	3.36	0.00	23.40	-0.03	59.75	-0.03	52.32	0.00
Gender	M	1.13	0.00	4.48	0.00	0.54	0.00	3.63	0.00	15.79	0.00	62.39	0.00	54.35	-0.01

3. EXPERIMENT 1 – IMPACT OF SURVEY MODE ON SURVEY ESTIMATES

In 2013, a Mode Effects (ME) experiment tested for survey mode effects by randomly assigning enrollees for whom both a phone number and mailing address were available to one of two modes (CATI vs. mail) as the default mode of survey administration. By holding population characteristics constant in this way, any potential effects of mode on survey outcomes (including response rates and survey estimates) can be identified. Although survey mode was found to have a significant effect on some survey responses, the magnitude of these effects was generally quite small. To support trending and further increase confidence that survey mode effects are minimal, ICF recommended replicating this experiment in 2014.

Design

Given similar parameters in 2014, the power analyses conducted in 2013 were again used to determine the sample sizes required to achieve sufficient power for detecting two-way mode $\times$ stratum interactions at a 95 percent confidence level.[6] The resulting recommendation was that 7,948 records be assigned to each of the mail and CATI protocols. Ultimately, 8,000 records were assigned to receive the mail protocol.

Because the treatment of records explicitly assigned to the CATI protocol ($n = 8,000$) is functionally identical to the treatment of records eligible for the experiment (i.e., having valid mail and phone contact information) but not explicitly included in it, the size of the sample assigned to receive the CATI protocol was effectively $8,000 + 376,067 = 384,067$.[7] Power analyses based on expected response rates showed that these sample sizes would be sufficient to detect two-tailed differences in proportions between the experimental treatment groups of at least three percentage points with greater than 90 percent power, and to detect mean differences of at least 0.71 units with 80 percent power. The actual number of eligible completed surveys received from the two treatment groups (mail $n = 2,808$, CATI $n = 25,000$) resulted in over 80 percent power to detect two-tailed differences in proportions of at least three percentage points and 80 percent power to detect two-tailed mean differences of at least 0.78 units.[8]

It is important to note that this experiment manipulated the *default mode of survey administration* rather than mode of survey completion. That is, enrollees in the ME experiment received one of two versions of the pre-survey notification letter: one version indicated that the enrollee would soon receive a paper copy of the survey in the mail, and the other version indicated that the enrollee would soon receive a call to complete the survey over the phone.

The pre-survey notification letters were identical in all other respects and included a URL to complete the Web survey online, as well as a phone number that the enrollee could call to complete the survey over the phone at their convenience or to request a mailed survey (as applicable). Consequently, a sampled

---

[6] A response rate of 39 percent (observed in the 2012 Survey of Enrollees for both phone and mail) was assumed when allocating sample for this experiment; in 2013, a response rate of 40 percent was observed for mail and a response rate of 32 percent was observed for phone.

[7] The two treatment groups were not equal in size due to cost considerations. Of all enrollees sampled for this survey, 26,765 had a valid mailing address but not a valid phone number, making them ineligible for inclusion in this experiment; all other sampled enrollees had both types of contact information.

[8] The actual power of a test depends on the specific proportions being tested; proportions lower or higher than 50 percent will have less variance, and tests will therefore be more powerful than the worst-case scenario described here. In addition, power will also be affected by item-missing data; the numbers reported here assume no missing data. In addition, only enrollees who completed the survey in the same mode as that which they were assigned were eligible for analysis.

enrollee assigned to either treatment still had the choice to respond in any of the three modes offered in 2014. (The final section of this report assesses differences in response patterns due to mode of completion.)

This design choice was made to increase the response rate at the cost of complete experimental control over mode of survey response. The result is that self-selection into response mode, or "response channel," presents a threat to the randomization of the experimental design: Enrollees ultimately chose the mode of completion they preferred, regardless of the mode to which they were nominally assigned. This threat to experimental control is a limitation of the current design and, as in 2013, was considered acceptable to prevent the experiment from negatively impacting overall response rates.

The majority of respondents in the ME experiment, however, completed the survey in the default mode. Specifically, 82 percent of responding enrollees assigned to the mail mode completed a mail survey and 69 percent of respondents assigned to the CATI mode completed a CATI survey. Notably, the latter figure is lower than the 2013 rate (75%), indicating that in 2014, enrollees assigned to the CATI protocol became more likely to choose alternative modes for responding (i.e., mail or Web). In particular, compared to 2013, the use of the mail mode by this group rose from 12 percent to 16 percent, while the use of the Web mode rose from 13 percent to 15 percent.

Table 4 shows counts of completed surveys in the ME experiment broken out by default mode (i.e., the mode randomly assigned) and response channel (i.e., the mode ultimately used for response). Note that for the purposes of the experimental analyses, enrollees were grouped into treatment (mail) vs. control (CATI) conditions based on whether they completed the survey in their assigned default mode (i.e., regardless of the response channel by which they ultimately arrived there). These groups are shaded in *Table 4*. "Outbound CATI" refers to enrollees who were called by an interviewer, whereas "inbound CATI" refers to enrollees who called in to complete an interview. Note that comparisons between Web mode respondents and CATI/mail respondents are discussed later in this report.

Table 4. Survey Completes by Default Survey Mode and Response Channel

	Response Channel				
	CATI*		Mail		Web
Default Mode	Outbound CATI	Inbound CATI	Mail (Default)	Mail Request	
Mail (Treatment)	389	58	2,808	N/A	186
CATI (Control)	23,402	1,568	N/A	5,916	5,507

Note: Shaded cells indicate groups included in the ME experimental analyses, due to completing the survey in the randomly assigned mode. An additional 2,490 responding enrollees not shown in this table did not have a valid phone number, making them ineligible for inclusion in this experiment.
* The completed interviews included in the Default Mode: CATI, Response Channel: CATI groups also include completes from enrollees who requested a mail survey, did not return it, and then completed a CATI non-response follow-up survey.

Results

Health Care Coverage, Health Care Access, and Health Status

Table 5 compares the effect of default survey mode (mail vs. CATI) on selected population estimates of coverage, access, and health status. Estimates were weighted using the non-response-adjusted and post-stratified analytic weight (W3 on the data file). For each measure, the significance level (*p*) for the Rao-Scott chi-square test is reported for the comparison of mail vs. CATI estimates. Significant differences (*p* < .05) are flagged with an asterisk, and those that replicate significant effects from the 2013 ME experiment are indicated by "*Rep2013*".

Assuming minimal effects of self-selection into mode, these results suggest that survey mode does have some influence on how enrollees respond to survey items and/or some association with who chooses to respond. These mode effects were not dramatic, however, with nearly all effects creating a difference of less than five percentage points.

Of the 19 outcomes tested, 11 showed statistically significant mode effects in the 2014 survey. Nine of these significant effects replicated significant effects from the 2013 ME experiment. Of these replicated effects, the maximum difference between mail and CATI estimates was 7.50 percentage points, with a mean absolute difference of 3.91 percentage points. The following summary of findings focuses on the statistically significant mode effects that replicated this year, as these effects have the strongest evidence of being systematic.

The mail survey produced a higher estimate of the proportion of enrollees covered by Medicare, but a lower estimate of the proportion of enrollees covered by Medicaid for some health care. The mail survey produced a higher estimate of enrollees who use VA services to meet "none" of their health care needs, whereas the CATI survey produced a higher estimate of enrollees who use VA services to meet "most" of their health care needs.

The CATI survey (compared to the mail survey) produced higher estimates of enrollees being in "excellent" or "poor" health, but lower estimates of enrollees being in "good" health. This pattern may suggest that the CATI survey promotes more use of the extreme ends of response scales compared to the mail survey, which is consistent with previous findings in mixed-mode research).[9]

Only one mode effect with regard to employment status replicated; the CATI survey produced a higher estimate of unemployed enrollees ("unemployed, looking for work, or laid off") compared to the mail survey. This effect might be explained by the greater ease with which phone contacts are made with unemployed individuals.

Table 5. Comparison of Selected Coverage, Access, and Health Status Proportions by Default Survey Mode, Weighted (w3)

Survey Item	Response	Overall (%)	Mail (%)	CATI (%)	*p*
Medicare coverage	1- Yes	51.9	58.6	51.1	<.001*
		(51.1, 52.6)	(56.2, 61)	(50.3, 51.9)	*Rep2013*
Medicaid coverage for some health care	1- Yes	8.3	7.1	8.5	.034*
		(7.9, 8.7)	(5.9, 8.2)	(8.1, 8.9)	*Rep2013*
Coverage by another individual or group health plan	1- Yes	27	27.7	26.9	.532
		(26.3, 27.7)	(25.5, 29.8)	(26.2, 27.7)	
Use VA services to meet...	1- All of my health care needs	33.5	33.1	33.6	.712
		(32.8, 34.2)	(30.9, 35.3)	(32.8, 34.3)	
	2- Most of my health care needs	17.1	14.6	17.4	.002*
		(16.6, 17.7)	(13, 16.1)	(16.8, 18)	*Rep2013*
	3- Some of my health care needs	27.1	27	27.1	.900
		(26.4, 27.8)	(25, 28.9)	(26.4, 27.8)	

[9] Dillman, D., Smyth, J., & Christian, L.M. (2009). *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method*. 3rd edition. Hoboken, NJ: Wiley

Survey Item	Response	Overall (%)	Mail (%)	CATI (%)	p
	4- None of my health care needs	17.2	21.8	16.7	<.001* *Rep2013*
		(16.6, 17.8)	(19.7, 23.8)	(16.1, 17.4)	
	5- I have no health care needs	5	3.6	5.2	.030*
		(4.6, 5.4)	(2.4, 4.7)	(4.8, 5.6)	
Self-reported general health	1- Excellent	10.9	9.1	11.1	.021* *Rep2013*
		(10.4, 11.4)	(7.6, 10.6)	(10.6, 11.6)	
	2- Very good	23.7	24.4	23.6	.462
		(23, 24.3)	(22.3, 26.4)	(22.9, 24.3)	
	3- Good	30.9	36.6	30.2	<.001* *Rep2013*
		(30.2, 31.6)	(34.4, 38.9)	(29.5, 31)	
	4- Fair	23.1	21.4	23.3	.059
		(22.5, 23.8)	(19.6, 23.3)	(22.7, 24)	
	5- Poor	11.4	8.5	11.7	<.001* *Rep2013*
		(10.9, 11.9)	(7.2, 9.9)	(11.2, 12.3)	
Employment status	1- Employed full-time	23.3	19.8	23.7	.001*
		(22.6, 24)	(17.7, 21.9)	(22.9, 24.5)	
	2- Self-employed full-time	2.9	3.2	2.8	.478
		(2.6, 3.1)	(2.2, 4.2)	(2.5, 3.1)	
	3- Employed part-time	5.4	5.9	5.4	.364
		(5, 5.8)	(4.8, 7)	(5, 5.8)	
	4- Self-employed part-time	2.5	3.1	2.4	.132
		(2.2, 2.7)	(2.2, 4)	(2.1, 2.7)	
	5- Unemployed, looking for work, or laid off	7.2	4.6	7.5	<.001* *Rep2013*
		(6.7, 7.6)	(3.4, 5.8)	(7, 7.9)	
	6- Currently not employed: Either retired, a homemaker, student, etc.	58.8	63.5	58.2	<.001*
		(58, 59.5)	(61, 65.9)	(57.4, 59.1)	

Note: 95 percent confidence intervals for estimated proportions are given in parentheses.
Note: Rao-Scott chi-square tests of association were used to compare proportions for each response between ME treatment groups (mail vs. CATI).
*A *p*-value less than .05 indicates a statistically significant association between survey mode and the enrollee characteristic indicated by that response. "Rep2013" indicates a replicated finding from the 2013 ME experiment.

Key Driver Questions

For key driver questions, the respondents were read a series of statements and then asked if they: 1) completely agreed, 2) agreed, 3) neither agreed nor disagreed, 4) disagreed, or 5) completely disagreed. Mean responses to these items are presented in *Table 6*. Lower values (minimum = 1.00) indicate stronger agreement with the statement, whereas higher values (maximum = 5.00) indicate stronger disagreement with the statement.

As above, this summary will focus on significant effects in 2013 that were replicated in the 2014 ME experiment. Five of the six effects tested were replicated, although as in 2013, the magnitude of these differences was not dramatic; of the significant effects in *Table 6*, the mean absolute difference between

estimates by mode was 0.17 points on the five-point rating scale, with a maximum difference of 0.21 points.

Replicating the 2013 findings, the mail survey produced more positive opinions about VA than the CATI survey, with the ease of getting to a local VA facility showing the largest difference. The exception to this pattern was that the CATI survey produced higher estimates of how well enrollees understand how their VA health benefits work. Social desirability may explain this difference, as enrollees may be more concerned about appearing competent when being interviewed.

Table 6. Comparison of Selected Key Driver Means by Default Survey Mode, Weighted (w3)

Survey Item	Overall	Mail	CATI	p
d11c: VA offers Veterans like me the best value for our health care dollar	2.16	2.01	2.18	<.0001* *Rep2013*
	(2.15, 2.18)	(1.97, 2.06)	(2.17, 2.2)	
d12b: Veterans like me who use VA are satisfied with the health care they receive	2.26	2.14	2.28	<.0001* *Rep2013*
	(2.25, 2.28)	(2.09, 2.19)	(2.26, 2.29)	
d13b: Veterans like me can get in and out of an appointment at VA in a reasonable time	2.36	2.24	2.37	<.0001* *Rep2013*
	(2.34, 2.38)	(2.19, 2.29)	(2.36, 2.39)	
d14d: I understand how my VA health benefits works	2.42	2.59	2.40	<.0001* *Rep2013*
	(2.40, 2.44)	(2.54, 2.65)	(2.38, 2.42)	
d15f: It is easy to get to my local VA facility	2.22	2.03	2.24	<.0001* *Rep2013*
	(2.20, 2.23)	(1.98, 2.07)	(2.22, 2.26)	
d16c: I would only use VA if I did not have access to any other source of health care	2.91	2.97	2.90	<.0001*
	(2.89, 2.93)	(2.90, 3.03)	(2.89, 2.92)	

Note: 95 percent confidence intervals for estimated means are given in parentheses.
Note: Independent-samples *t*-tests were used to compare means for each response between ME treatment groups (mail vs. CATI).
*A *p*-value less than .05 indicates a statistically significant association between survey mode and the enrollee characteristic indicated by that response. "Rep2013" indicates a replicated finding from the 2013 ME experiment.

Survey Mode Effects within Strata

To explore the effects of survey mode on responses in more detail, the two ME treatment groups (mail vs. CATI) were compared within sampling strata (i.e., VISN, priority group, gender, and Hispanic identity). To simplify analyses and conserve statistical power, the 21 VISNs were collapsed into four groups according to VA area office boundaries (i.e., East, Central, South, and West).[10] The eight priority groups were collapsed into two levels (1-4 = high priority, 5-8 = low priority).

These analyses, which are equivalent to the decomposition of mode × stratum interactions, highlight outcomes where significant mode effects are observed in one level of a stratum (e.g., gender = Female) but not in another level of that stratum (e.g., gender = Male). Significant mode effects that are consistent across stratum levels are not discussed, since these are equivalent to main effects and are reflected in the discussion of the overall estimates above. Furthermore, only effects that replicated findings from the 2013 ME experiment are discussed, as these effects have the strongest evidence of being systematic. The outcome variables analyzed in this section are the same as shown in *Table 5* (coverage, access, and health status proportions) and *Table 6* (key driver questions).

---

[10] See http://www2.va.gov/directory/guide/division_flsh.asp?dnum=3

To provide a visual summary of the magnitude of the mode effects observed across domains in 2014, *Figure 4* displays the distribution of mode effects (i.e., the estimated outcome percentage from the mail survey minus the estimated outcome percentage from the CATI survey) for the 19 weighted coverage, access, and health status estimates across 14 domains (four VISN regions, high vs. low priority groups, gender, Hispanic identity, OEF/OIF/OND status, and pre- vs. post-enrollee status). The mode effects ranged from -7.80 percentage points to 12.50 percentage points across all measures and domains, with a median effect of -0.30 and an interquartile range of 4.08. Forty-six of the 266 estimated mode effects exceeded ±5.00 percentage points.

Figure 4. Distribution of Mode Effects for 19 Coverage, Access, and Health Status Estimates across 14 Domains, Weighted (w3)
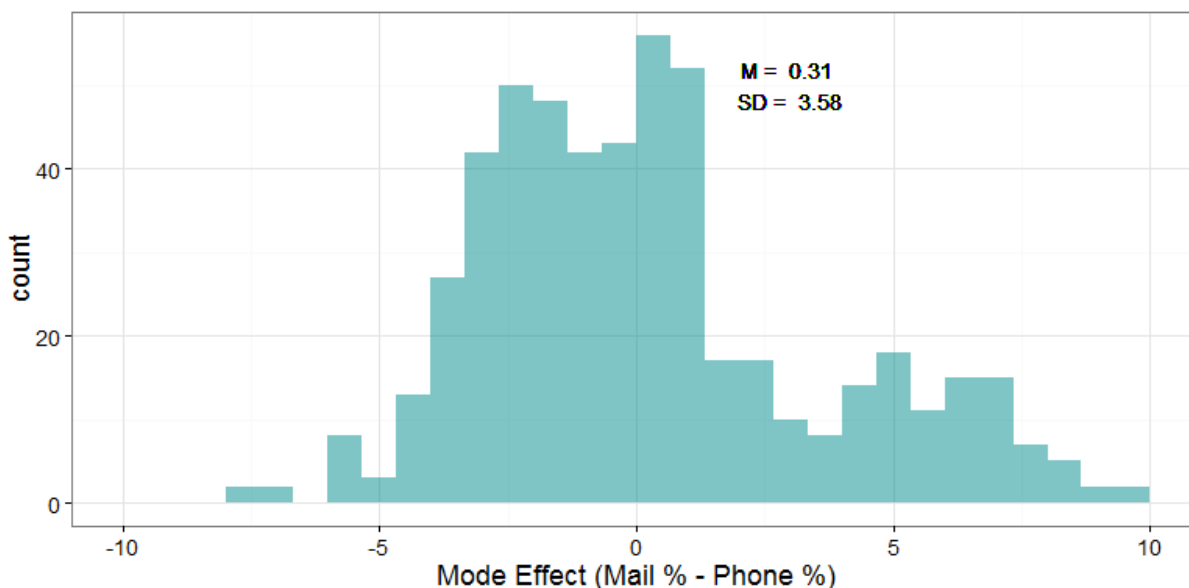
VISN Groups

As in 2013, some variability in mode effects was observed across the four geographic regions (East, Central, South, and West). The following list summarizes the replicated regional mode effects (mail vs. CATI) for the 19 measures of health care coverage, access, and health status:

- In the East region, the mail survey produced a higher estimate of enrollees having part-time employment.
- In the Central region, the mail survey produced a higher estimate of Medicare coverage; a higher estimate of enrollees who use VA services to meet "none" of their health care needs; and a lower estimate of enrollees with "poor" health.
- In the South region, the mail survey produced a lower estimate of "poor" general health.
- In the West region, the mail survey produced a higher estimate of Medicare coverage and a higher estimate of enrollees who use VA services to meet "none" of their health care needs.

None of the mode effects on key driver questions reported above varied by region.

Priority Groups

Two survey mode × priority group (high vs. low) effects were replicated. The mail survey (compared to the CATI survey) produced a lower estimate of Medicaid coverage among low-priority enrollees, whereas there was no mode effect for high-priority enrollees. The mail survey also produced a lower estimate of enrollees using VA services for "most" of their health care needs among low-priority enrollees, whereas there was no mode effect for high-priority enrollees. None of the mode effects on key driver questions reported above differed between priority groups.

Gender

Because there are many more men than women in the responding sample, statistical tests of mode effects among the male respondents have much greater power than tests among the female respondents. This difference in power leads to a higher probability of achieving statistical significance among the former subgroup even if a mode effect of the same magnitude exists in both populations. To focus on the more robust interactions between survey mode and enrollee gender, only mode effects where: a) significance was achieved in only one subgroup, and b) the absolute difference in the magnitude of the mode effects between subgroups was greater than or equal to five percentage points are discussed here.

Looking first at the measures of health care coverage, access, and health status, the mail survey (compared to the CATI survey) produced a higher estimate of Medicare coverage among men, whereas there was no significant mode effect for Medicare coverage among women. The mail survey also produced a higher estimate of "good" general health among men, whereas there was no significant mode effect for this outcome among women. Finally, the mail survey produced a lower estimate of full-time employment among men, whereas there was no significant difference for this outcome among women in the CATI mode.

None of the mode effects on key driver questions reported above varied by enrollee gender.

Hispanic/Latino Ethnicity

Due to a small proportion of the responding sample identifying as Hispanic/Latino, comparisons of mode effects between enrollees identifying as Hispanic/Latino vs. not raises the same issue of asymmetric statistical power noted with regard to enrollee gender. The same criteria used to identify the more robust interactions between survey mode and Hispanic/Latino ethnicity will be applied here.

Looking first at the measures of health care coverage, access, and health status, the mail survey (compared to the CATI survey) produced a higher estimate of Hispanic/Latino enrollees who use VA services to meet "some" of their health care needs, whereas there was no significant difference for this outcome among non-Hispanic/Latino enrollees. In addition, the mail survey produced a higher estimate of non-Hispanic/Latino enrollees who use VA services to meet "none" of their health care needs, whereas there was no significant difference for this outcome among Hispanic/Latino enrollees.

One difference in mode effects on the key driver questions was observed: Among Hispanic/Latino enrollees, the CATI survey (compared to the mail survey) created stronger agreement with the statement "I would only use VA if I did not have access to any other sources of health care," whereas there was no significant difference for this outcome among non-Hispanic/Latino enrollees.

Summary of Findings: Mode Effects Experiment

In 2012, the first year with a mail mode, an analysis of mode effects comparing enrollees responding via CATI vs. mail indicated some differences between groups. However, because enrollees were not randomly assigned to response channels, potential mode effects were confounded with pre-existing differences between the populations of enrollees who preferred to respond by CATI vs. mail. To disentangle mode effects from population differences, ICF recommended conducting a methodological experiment to randomly assign enrollees to survey modes.

Two randomized mode effects experiments have now been conducted as part of the 2013 and 2014 surveys, and the findings have been consistent: Although there are some significant differences between survey modes on key survey outcomes, the magnitude of these differences is generally small. Moreover, only 14 of the 25 overall outcomes tested (collapsing across strata) produced mode effects that replicated across years. This suggests that the mode effects observed in any given year are often not systematic.

The magnitudes of the effects that did replicate were acceptably small and do not present a substantive threat of bias to survey estimates. With regard to measures of health care coverage, access, and health status, the mean absolute difference between mail and CATI estimates was 3.91 percentage points. With regard to the key driver questions, the mean absolute difference between mail and CATI estimates was 0.17 points on the five-point rating scale.

Replicated overall mode effects indicated small differences between survey modes in estimates of enrollees covered by Medicare and Medicaid, as well as differences in estimates of general health. In addition, the mail survey appears to generate slightly more positive opinions of VA services compared to the CATI survey.

At the level of individual sampling strata, there were few replicated survey mode × stratum effects among the many that were tested. Although some effects were consistent across years, these findings were scattered across domains and outcomes, giving no indication that one mode is biasing responses in a particular direction.

Without having access to "true" values for the measures evaluated here, it is impossible to know if the mail or CATI mode (or both) is introducing measurement bias when mode effects are detected. Thus, we have no reason to assume that one mode is more accurate than the other. The current evidence justifies the recommendation to continue encouraging response in all modes, as the mixed-mode design provides a substantial reduction in undercoverage without substantially increasing measurement error due to mode effects.

4. EXPERIMENT 2 – IMPACT OF SECOND SURVEY MAILING ON RESPONSE RATES FOLLOWING CATI NON-WORKING/NON-RESPONSE

The Second Survey Mailing/Follow-Up to Phone/CATI Protocol (SSM-P) experiment tested the effect on response rates of mailing one vs. two surveys as part of the CATI survey non-response/non-working number follow-up protocols. The follow-up protocols being compared in this experiment are shown in *Table 7*. The key difference is that in the long protocol, a second complete survey is mailed following non-response to the first follow-up mail survey. This experiment is a replication of the "Second Survey Mailing" experiment reported in the 2013 *Methodological Experiments Report*.

Table 7. SSM-P Experiment Follow-Up Mail Protocols: Long (Treatment) vs. Short (Control)

Long Protocol (Treatment)	Short Protocol (Control)
1. Pre-survey Notification Letter	1. Pre-survey Notification Letter
2. 1st Survey Packet Mailing	2. 1st Survey Packet Mailing
3. Reminder Postcard	3. Reminder Postcard
4. 2nd Survey Packet Mailing	

Design

Enrollees became eligible for this experiment when their phone numbers were determined to be non-working or they were determined to be non-respondents to the CATI survey. *Table 8* shows how these records were randomly assigned to the SSM-P treatment conditions. Only a subsample of telephone non-respondents from the first wave of the sample release was entered into this experiment to receive any mail follow-up. The entire first-wave sample of dialed phone records determined to be non-working received a mail follow-up protocol (either via "explicit" assignment to the short vs. long protocols following the power analyses described below, or via "implicit" assignment to the short protocol for the balance of the first-wave non-working sample).

Table 8. SSM-P Treatment Group Sizes by CATI Non-Response/Non-working Status

SSM-P Condition	Status			
	Non-Response	Non-working (Explicit Assignment)	Non-working (Implicit Assignment)	Total
Long Protocol (Treatment)	1,750	2,000	N/A	**3,750**
Short Protocol (Control)	1,750	1,200	8,763	**11,713**

Sample sizes were determined by the decision to conduct an exact replication of the 2013 experiment. Power analyses show that for the non-working records, these sample sizes are sufficient to detect a one-sided difference in response rates (assuming a 15.4 percent response rate to the short protocol, as observed in the 2013 experiment) of at least three percentage points with over 80 percent power. Similarly, for the non-response records, these sample sizes are sufficient to detect a one-sided difference in response rates of at least three percentage points with nearly 80 percent power. Combining non-response and non-working records yields over 80 percent power for detecting a one-sided difference in response rates of at least two percentage points.

It is important to note that enrollees who entered into this experiment were allowed to complete the survey in any of the three available modes (CATI, mail, or Web). As shown in *Table 9*, the majority of enrollees entered into the SSM-P experiment and who ultimately responded did so using the mail mode, although a small number in each treatment group also completed surveys in other two modes. For the

purposes of analysis, all responses are counted toward the total response rate for each SSM-P condition regardless of the response channel, since the outcome of interest in this experiment is overall response rate improvement due to changes in follow-up protocol.

Table 9. Sampled Records and Survey Responses by Population, SSM-P Condition and Response Channel

	Sample	Responses by Response Channel				RR1	RR1 Change
Population: Non-working Phone Records							
SSM-P Condition		CATI[†]	Mail	Web	**Total**		
Long Protocol (Treatment)	2,000	3	526	7	**536**	**26.8%**	**+7.8 pts***
Short Protocol (Control)	9,963	20	1,841	35	**1,896**	**19.0%**	
Population: CATI Non-Respondents							
SSM-P Condition		CATI[†]	Mail	Web	**Total**		
Long Protocol (Treatment)	1,750	1	459	2	**462**	**26.4%**	**+4.2 pts***
Short Protocol (Control)	1,750	1	385	2	**388**	**22.2%**	
Overall (Combined Populations)							
SSM-P Condition		CATI[†]	Mail	Web	**Total**		
Long Protocol (Treatment)	3,750	4	985	9	**998**	**26.6%**	**+7.1 pts***
Short Protocol (Control)	11,713	21	2,226	37	**2,284**	**19.5%**	

*Difference is significant, $p < .05$.

[†] Inbound CATI

Total response rates (i.e., combining CATI, mail, and Web completes) for the SSM-P experiment were computed following AAPOR standards, specifically formula AAPOR RR1, which divides the number of completed interviews by the total number of attempted interviews.[11] The random assignment of records to experimental groups ensures that the expected distribution of outcome dispositions between groups is balanced, so that any differences in the number of completed interviews can be attributed to the experimental treatment (i.e., the second survey mailing).

Results

The results of the 2014 SSM-P experiment replicated the findings of the 2013 experiment in all key respects. Specifically, the long mail protocol used to follow up with phone non-working records and phone non-respondents significantly increased total response rates compared to the short protocol in both populations, as well as the overall (combined) CATI follow-up population.

As shown in *Table 9*, when looking at the population of non-working phone records, the total response rate was significantly[12] higher in the long protocol (536/2,000 = 26.8 percent) compared to the short protocol (1,896/9,963 = 19.0 percent), leading to a response rate improvement in this population of 7.8 percentage points. In 2013, a response rate improvement of 6.8 percentage points was observed due to the use of the long protocol in this population.

When looking at the population of phone non-response records, the total response rate was significantly[13] higher in the long protocol (462/1,750 = 26.4 percent) compared to the short protocol (388/1,750 = 22.2 percent), leading to a response rate improvement in this population of 4.2 percentage points. In 2013, a

---

[11] AAPOR RR1 is equivalent to the simplified, lower-bound RR3 computation used to analyze the 2013 version of this experiment, so response rates are directly comparable across replications. Documentation for response rate calculations is available at
http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions2&Template=/CM/ContentDisplay.cfm&ContentID=3156
[12] Rao-Scott $\chi^2(1) = 62.44$, $p < .0001$
[13] Rao-Scott $\chi^2(1) = 8.74$, $p < .01$

response rate improvement of 7.2 percentage points was observed due to the use of the long protocol in this population.

Finally, when looking at the overall CATI follow-up population (combining the phone non-working and non-response records), the total response rate was significantly[14] higher in the long protocol (998/3,750 = 26.6 percent) compared to the short protocol (2,284/11,713 = 19.5 percent), leading to a response rate improvement in the overall population of 7.1 percentage points. In 2013, a response rate improvement of 8.0 percentage points was observed due to the use of the long protocol in the overall population.

These results indicate that, among enrollees who do not respond to the CATI survey and enrollees with non-working numbers, a second survey mailing as part of a mail follow-up protocol significantly improves response rates. The replication of these findings across two years of the survey provides strong evidence that they are systematic. The only deviation from the 2013 results was that, in 2014, the response rate improvement due to the long protocol was higher among the phone non-working population than among the non-response population (+7.8 vs. +4.2 percentage points, respectively), whereas the opposite pattern was observed in 2013 (+6.8 vs. +7.2 percentage points, respectively). This likely reflects random variation across years and does not change the overall recommendation to use a second survey mailing as part of the CATI follow-up protocol to significantly increase response rates in both of these populations.

---

[14] Rao-Scott $\chi^2(1) = 85.97$, $p < .0001$

5. EXPERIMENT 3 – IMPACT OF SECOND SURVEY MAILING ON RESPONSE RATES AS PART OF MAIL SURVEY PROTOCOL

The Second Survey Mailing/Mail Protocol (SSM-M) experiment tested the effect on response rates of mailing one vs. two surveys as part of the mail survey protocol. The survey protocols being compared in this experiment are shown in *Table 10*. The key difference is that in the long protocol, a second complete survey is mailed to non-respondents two weeks after the first mail survey is sent out.

Table 10. SSM-M Experiment Mail Protocols: Long (Treatment) vs. Short (Control)

Long Protocol (Treatment)	Short Protocol (Control)
1. Pre-survey Notification Letter	1. Pre-survey Notification Letter
2. 1st Survey Packet Mailing	2. 1st Survey Packet Mailing
3. Reminder Postcard	3. Reminder Postcard
4. 2nd Survey Packet Mailing	4. Telephone Follow-Up
5. Telephone Follow-Up	

Design

A subsample (*n* = 5,813) of the 8,000 enrollees who were randomly assigned to receive the mail protocol as part of the ME experiment were entered into the SSM-M experiment (see *Figure 2*). Specifically, 2,907 enrollees were assigned the long mail protocol (treatment group) and 2,906 enrollees were assigned the short mail protocol (control group). Power analyses show that these sample sizes are sufficient to detect a one-sided difference in response rates (assuming a 40 percent response rate to the long protocol, as observed with the 2013 mail survey) of at least three percentage points with nearly 80 percent power.

As with the SSM-P experiment, it is important to note that enrollees entered into the SSM-M experiment were allowed to complete the survey in any of the three available modes (CATI, mail, or Web). As shown in *Table 11*, the majority of enrollees entered into the SSM-M experiment and who ultimately responded did so in the mail mode, although a small number in each treatment group also completed interviews in the other two modes. For the purposes of analysis, all responses are counted toward the total response rate for each SSM-M condition regardless of the response channel, since the outcome of interest in this experiment is overall response rate improvement due to changes in survey protocol.

Table 11. Sampled Records and Survey Responses by SSM-M Condition and Response Channel

	Sample	Responses by Response Channel				RR1	RR1 Change
SSM-M Condition		CATI[†]	Mail	Web	**Total**		
Long Protocol (Treatment)	2,907	23	825	28	**876**	**30.1%**	**-0.4 pts**
Short Protocol (Control)	2,906	362	501	24	**887**	**30.5%**	

[†] Inbound CATI

Total response rates (i.e., combining CATI, mail, and Web completes) for the SSM-M experiment were computed following AAPOR standards, specifically formula AAPOR RR1, which divides the number of completed interviews by the total number of attempted interviews.[15] The random assignment of records to experimental groups ensures that the expected distribution of outcome dispositions between groups is

---

[15] Documentation for response rate calculations is available at
http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions2&Template=/CM/ContentDisplay.cfm&ContentID=3156

balanced, so that any differences in the number of completed interviews can be attributed to the experimental treatment (i.e., the second survey mailing).

Results

As shown in *Table 11*, the difference in total response rates between the two SSM-M conditions was not significant.[16] Both the short and long mail protocols produced a total response rate of just over 30 percent. This finding indicates that sending a second survey mailing following non-response to the first mailing does not produce an advantage in total response rates given similar subsequent follow-up procedures (in this case, CATI follow-ups).

Although the total response rate did not differ between conditions, the distribution of response channels used by respondents was significantly different.[17] As shown in *Figure 5*, 94 percent of respondents in the long protocol used the mail response channel, compared to only 57 percent of respondents in the short protocol. Assuming equal rates of response to the first survey mailing (due to random assignment to conditions), and given the equivalent total response rates in *Table 11*, this finding indicates that there is a fixed number of first-mailing non-respondents who can be converted into respondents through subsequent follow-up effort. The mode used to convert these non-respondents, however, appears not to matter: If the next follow-up attempt is made in the mail mode (as in the long protocol), first-mailing non-respondents will choose to respond via mail; on the other hand, if the next follow-up attempt is made via phone (as in the short protocol), first-mailing non-respondents will choose to respond in that mode.

Figure 5. Distribution of Responses by Channel between SSM-M Experiment Conditions
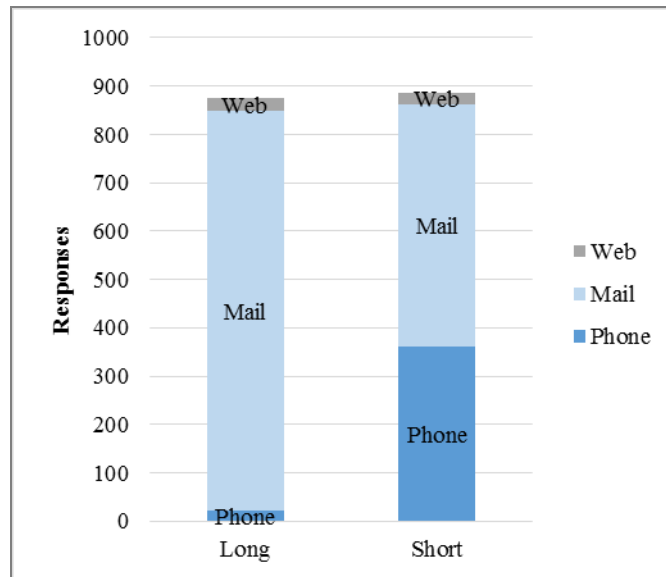
Based on these findings, and assuming that a second survey mailing has a lower cost than a phone follow-up, an initial recommendation can be made to employ the long mail protocol to reduce survey administration costs without negatively impacting response rates. In fact, if the second survey were the

---

[16] Rao-Scott $\chi^2(1) = 0.11$, $p = .744$
[17] Rao-Scott $\chi^2(2) = 416.78$, $p < .0001$

sole follow-up in the mail protocol, and phone follow-ups were eliminated, this experiment suggests that only 2.6 percent (23 / 876) of first-mailing non-respondents would fail to be converted.[18]

Given ICF's recommendation to increase the use of the mail mode in the Survey of Enrollees going forward, this experiment warrants replication to ensure that the current findings are systematic before reducing or eliminating phone follow-ups in the mail protocol.

---

[18] Of the 23 phone-channel respondents in the long protocol, 16 completed via outbound CATI and seven completed via inbound CATI. Of the 362 phone-channel respondents in the short protocol, 339 completed via outbound CATI and 23 completed via inbound CATI.

6. NON-RESPONSE BIAS ANALYSIS

Non-response bias can arise when the propensity to respond to a survey is correlated with survey outcomes. In such cases, respondents and non-respondents will be systematically different in ways that bias survey estimates. Non-response bias is typically analyzed using auxiliary variables on the sampling frame that are available for both respondents and non-respondents. In most cases, the information available from these auxiliary variables is limited; however, for the SoE, the sampling frame contains considerable administrative data about the enrollee population. This information makes it possible to estimate non-response biases with respect to enrollees' use of various VHA services described below.

This section of the report compares the utilization rate between responding and non-responding enrollees for each of these VHA services, referred to as HSCs (for details on the utilization indicators, see Appendix 1).[19] These analyses can reveal subgroups of enrollees who are less likely to respond to the survey, and may therefore benefit from more targeted survey administration efforts. For these analyses, the data are weighted to account for the differential sampling probabilities in each of the sampling strata without adjusting for non-response (i.e., using the design weight $W1$ on the survey data file).

In addition, this section of the report compares utilization rates between enrollees responding via Web and those responding via mail or CATI. Because assignment to the Web mode was not part of the 2014 Mode Effects experiment, potential differences between enrollees choosing to respond via the Web survey are examined here. For these analyses, the data are weighted using the final analysis weight ($W3$ on the survey data file), which accounts for sampling probabilities in each of the sampling strata as well as non-response and post-stratification adjustments. Because the Web mode is offered to all enrollees, utilization differences between those who responded by Web and those who responded by CATI or mail could be due to mode effects or population differences between those more likely to respond through the Web.

Past analyses have examined non-response bias for stratification variables: OEF/OIF/OND, VISN, priority group, and enrollee type (pre/post). We continue to calculate the non-response bias for these variables and also include gender (a stratification variable in 2014) and Hispanic ethnicity (oversampled in 2014).

1. Long-Term Service and Supports

A small proportion of the enrollee population receives long-term service and support (LTSS): 0.52 percent receives institutional long-term care, and 3.61 percent receive non-institutional long-term care.

Respondents vs. Non-Respondents

A significantly lower proportion of respondents (0.38 percent) compared to non-respondents (0.59 percent) receives institutional long-term care, whereas a significantly higher proportion of respondents (4.37 percent) receives non-institutional long-term care compared to non-respondents (3.23 percent). Across subgroups, the pattern is generally consistent with respondents having a lower institutional LTSS utilization rate and higher non-institutional LTSS utilization rate. Consistent with the overall pattern, responding enrollees are lower for institutional, 2.55 percent ($p = .127$) and higher for non-institutional ($p = .972$).

---

[19] Health Service Categories (HSCs) are defined as the category of care a Veteran received (Inpatient: medical, surgical, psychiatric, substance abuse, skilled nursing/extended care facility; Ambulatory care: allergy immunotherapy, allergy testing, anesthesia, cardiovascular; chiropractic, consultations, emergency room visits ,hearing/speech exams, immunizations, miscellaneous medical, office/home/urgent care visits, outpatient psychiatric, outpatient substance abuse, pathology, physical exams, physical medicine, radiology, surgery, therapeutic injections, vision exams).

Comparisons to population proportions indicate that the survey respondents under-represent the population of enrollees receiving institutional long-term care (0.52 percent of the population vs. 0.38 percent of respondents) but over-represent enrollees receiving non-institutional long-term care (3.61 percent of the population vs. 4.37 percent of respondents). After response propensity score weighting and raking, the overall LTSS utilization rate for respondents is 0.55 percent for institutional and 3.69 percent for non-institutional, not significantly different from the population values.

Web vs. Mail/CATI

Overall, enrollees responding via the Web had significantly lower utilization rates for institutional (0.29 percent) and non-institutional (2.69 percent) long-term care compared to enrollees responding via mail/CATI (0.59 percent and 3.84 percent, respectively; see *Figure 6* and *Figure 7*). This pattern was also consistent across strata, indicating that the Web mode is, in general, less likely to be used by enrollees receiving long-term care compared to the mail and CATI modes. For both of these HSC indicators, the estimated proportions among mail and CATI respondents were closer to the population values than were the proportions among Web respondents.

Figure 6. Percentage of Enrollees Receiving Institutional Long-Term Care
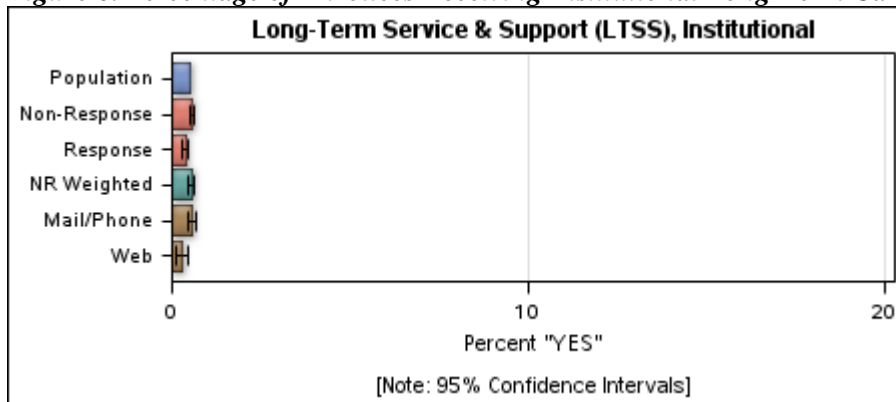
Figure 7. Percentage of Enrollees Receiving Non-Institutional Long-Term Care
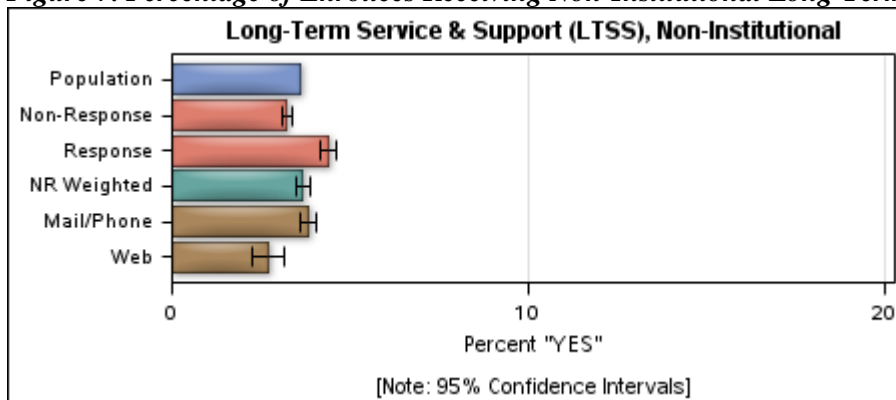
Table 12. Percentage of Enrollees Receiving Institutional Long-Term Care, by Stratum

Stratum	Level	Population %	Non-Responding Enrollees %	Responding Enrollees %	Sig.	Weighted %	Sig.	Mail/CATI %	Web %	Sig.
Overall	-	0.52	0.59	0.38	<.0001	0.55	0.62978	0.59	0.29	0.0337
Hispanic	N	0.73	0.89	0.46	<.0001	0.73	0.98348	0.78	0.41	0.0770
Hispanic	Y	0.50	0.49	0.28	0.0688	0.40	0.37509	0.45	.	
Hispanic	Unk	0.10	0.07	0.13	0.0929	0.14	0.41664	0.15	0.07	0.3041
Gender	F	0.30	0.23	0.31	0.3249	0.35	0.59062	0.38	0.16	0.2757
Gender	M	0.54	0.62	0.39	<.0001	0.56	0.67532	0.60	0.30	0.0435
OEF/OIF/OND	N	0.59	0.69	0.41	<.0001	0.63	0.57286	0.67	0.33	0.0322
OEF/OIF/OND	Y	0.05	0.04	0.02	0.5937	0.02	0.01891	0.02	.	
VISN	1	0.62	0.70	0.37	0.1773	0.61	0.95994	0.70	.	
VISN	2	0.51	0.71	0.29	0.0359	0.40	0.45046	0.42	0.30	0.6691
VISN	3	0.50	0.67	0.19	0.0126	0.36	0.52405	0.41	0.11	0.2150
VISN	4	0.63	0.82	0.58	0.3061	0.82	0.40664	0.90	0.29	0.2542
VISN	5	0.71	0.37	0.59	0.2832	0.96	0.46730	1.09	0.30	0.1977
VISN	6	0.39	0.40	0.33	0.7322	0.41	0.94944	0.46	.	
VISN	7	0.28	0.44	0.02	<.0001	0.04	0.00000	0.04	.	
VISN	8	0.38	0.40	0.17	0.1430	0.24	0.24258	0.19	0.53	0.3563
VISN	9	0.39	0.44	0.36	0.7150	0.43	0.83178	0.48	.	
VISN	10	0.65	0.78	0.58	0.4421	0.78	0.57384	0.89	.	
VISN	11	0.50	0.74	0.33	0.0482	0.42	0.60178	0.48	.	
VISN	12	0.83	1.05	0.65	0.1584	0.94	0.69536	1.04	0.42	0.3707
VISN	15	0.55	0.54	0.58	0.8894	1.01	0.18912	1.13	.	
VISN	16	0.40	0.43	0.05	<.0001	0.10	0.00000	0.11	.	
VISN	17	0.42	0.30	0.09	0.0233	0.14	0.00000	0.13	0.19	0.7453
VISN	18	0.63	0.56	0.70	0.5602	1.26	0.14274	1.41	0.40	0.0880
VISN	19	0.60	0.77	0.25	0.0213	0.29	0.03047	0.30	0.27	0.9349
VISN	20	0.48	0.77	0.63	0.6397	0.75	0.30552	0.67	1.19	0.4179
VISN	21	0.70	0.98	0.79	0.4825	0.97	0.27957	1.11	0.30	0.0600
VISN	22	0.55	0.54	0.29	0.2168	0.51	0.87301	0.42	0.97	0.4615
VISN	23	0.76	0.49	0.73	0.3045	1.08	0.34890	1.25	.	
Priority Group	1	1.37	1.87	0.74	<.0001	1.12	0.15752	1.28	0.24	0.0119
Priority Group	2	0.26	0.18	0.14	0.5415	0.14	0.02192	0.13	0.19	0.6031
Priority Group	3	0.24	0.25	0.23	0.8119	0.25	0.83976	0.29	0.08	0.0865
Priority Group	4	3.11	3.11	2.55	0.1272	4.36	0.01141	4.13	8.17	0.0544
Priority Group	5	0.41	0.39	0.46	0.4124	0.63	0.06752	0.65	0.45	0.7130

Stratum	Level	Population %	Non-Responding Enrollees %	Responding Enrollees %	Sig.	Weighted %	Sig.	Mail/CATI %	Web %	Sig.
Priority Group	6	0.04	0.02	0.05	0.2827	0.07	0.60856	0.08	.	
Priority Group	7	0.37	0.11	0.27	0.2743	0.34	0.89611	0.32	0.45	0.7833
Priority Group	8	0.09	0.09	0.06	0.3564	0.07	0.62248	0.05	0.18	0.2174
Pre/Post-Enrollee	POST	0.32	0.32	0.28	0.2687	0.37	0.19881	0.39	0.25	0.2796
Pre/Post-Enrollee	PRE	1.30	1.62	0.78	<.0001	1.20	0.53532	1.29	0.47	0.0489

Table 13. Percentage of Enrollees Receiving Non-Institutional Long-Term Care, by Stratum

Stratum	Level	Population %	Non-Responding Enrollees %	Responding Enrollees %	Sig.	Weighted %	Sig.	Mail/CATI %	Web %	Sig.
Overall	-	3.61	3.23	4.37	<.0001	3.69	0.45740	3.84	2.69	<.0001
Hispanic	N	4.91	4.61	5.33	<.0001	4.87	0.77290	5.05	3.66	0.0008
Hispanic	Y	4.59	4.24	5.04	0.0385	4.29	0.37097	4.17	5.22	0.3526
Hispanic	Unk	0.77	0.63	1.18	<.0001	0.80	0.70268	0.87	0.42	0.0507
Gender	F	3.36	2.70	4.64	<.0001	3.83	0.15171	3.89	3.43	0.5812
Gender	M	3.63	3.27	4.36	<.0001	3.67	0.66899	3.84	2.63	<.0001
OEF/OIF/OND	N	3.95	3.60	4.58	<.0001	4.00	0.65029	4.19	2.81	<.0001
OEF/OIF/OND	Y	1.33	1.17	2.05	0.0008	1.57	0.29017	1.53	1.84	0.5999
VISN	1	3.31	3.33	3.47	0.8049	3.19	0.76957	3.50	1.17	0.0662
VISN	2	4.12	3.73	5.42	0.0032	4.86	0.14568	4.68	5.73	0.4857
VISN	3	4.52	3.69	5.27	0.0034	4.01	0.21417	4.10	3.52	0.6154
VISN	4	3.43	3.16	3.62	0.3664	3.47	0.91666	3.64	2.32	0.2501
VISN	5	3.92	3.55	5.24	0.0044	3.99	0.87934	4.55	1.08	0.0015
VISN	6	3.85	3.76	4.92	0.0693	3.92	0.87785	4.15	2.29	0.1986
VISN	7	2.77	2.59	4.64	0.0003	3.70	0.05435	3.85	2.55	0.3274
VISN	8	4.30	3.59	3.98	0.4333	3.66	0.11181	3.88	2.34	0.1116
VISN	9	3.75	3.16	4.19	0.0794	3.34	0.32442	3.37	2.99	0.7730
VISN	10	5.32	4.60	6.77	0.0010	5.62	0.57855	5.92	3.48	0.1156
VISN	11	4.39	3.91	4.93	0.0905	4.20	0.67359	4.15	4.54	0.7907
VISN	12	3.68	3.23	3.38	0.7761	3.20	0.24647	3.38	2.16	0.2024
VISN	15	3.30	2.56	3.29	0.1518	2.72	0.14737	2.86	1.58	0.2747
VISN	16	3.58	3.33	4.46	0.0560	3.87	0.53730	3.94	3.32	0.6586
VISN	17	2.87	2.53	3.90	0.0118	3.19	0.46917	3.22	3.03	0.8703
VISN	18	3.31	3.22	5.03	0.0013	4.03	0.09135	4.16	3.30	0.5351
VISN	19	3.78	3.57	4.16	0.3024	3.38	0.34136	3.54	2.44	0.3256
VISN	20	2.60	2.27	3.48	0.0197	3.00	0.40888	3.10	2.49	0.5918

Stratum	Level	Population %	Non-Responding Enrollees %	Responding Enrollees %	Sig.	Weighted %	Sig.	Mail/CATI %	Web %	Sig.
VISN	21	3.38	2.92	4.09	0.0244	3.26	0.73815	3.52	1.92	0.1040
VISN	22	2.50	2.09	3.71	0.0002	2.75	0.50451	2.75	2.74	0.9933
VISN	23	4.10	3.97	5.29	0.0354	4.90	0.12319	5.28	2.36	0.0396
Priority Group	1	6.74	6.25	7.62	0.0013	7.06	0.36973	7.38	5.25	0.0162
Priority Group	2	2.72	2.47	3.30	0.0040	2.52	0.33843	2.60	2.14	0.3639
Priority Group	3	2.37	2.00	2.80	0.0024	2.13	0.19168	2.21	1.69	0.3257
Priority Group	4	16.89	16.34	17.76	0.0911	17.15	0.73687	17.40	12.98	0.1244
Priority Group	5	3.69	3.17	5.00	<.0001	4.00	0.19653	4.01	3.92	0.9213
Priority Group	6	0.71	0.58	1.27	0.0002	0.92	0.19912	0.87	1.14	0.4862
Priority Group	7	3.48	3.44	3.31	0.8551	3.21	0.62434	3.29	2.64	0.6813
Priority Group	8	1.48	1.33	1.77	0.0055	1.37	0.31760	1.47	0.77	0.0265
Pre/Post-Enrollee	POST	2.84	2.52	3.61	<.0001	2.98	0.17361	3.10	2.26	0.0038
Pre/Post-Enrollee	PRE	6.54	5.98	7.24	0.0006	6.36	0.52644	6.56	4.77	0.0340

2. Inpatient Treatment

A small proportion of the enrollee population (1.14 percent) receives inpatient treatment related to mental health or substance abuse (MHSA), and 4.46 percent receives inpatient treatment for other reasons (non-MHSA).

Respondents vs. Non-Respondents

A significantly lower proportion of respondents receives MHSA inpatient treatment (0.79 percent) compared to non-respondents (1.30 percent), whereas a significantly higher proportion of respondents receives non-MHSA inpatient treatment (5.03 percent) compared to non-respondents (4.21 percent). These differences are consistent across strata and indicate that enrollees who respond to the survey (compared to non-respondents) tend to have a lower utilization rate for MHSA inpatient treatment, but a higher utilization rate for non-MHSA inpatient treatment. After adjusting for age, the response differences still exist—those receiving MHSA inpatient treatment are less likely to respond; those receiving non-MHSA inpatient treatment are more likely to respond.

Comparison to population proportions indicates that the survey respondents under-represent the population of enrollees receiving MHSA inpatient treatment (1.14 percent of the population vs. 0.79 percent of respondents) but over-represents enrollees receiving non-MHSA inpatient treatment (4.46 percent of the population vs. 5.03 percent of respondents). Overall, the response propensity score model and raking adjustments reduce bias such that the utilization is not significantly different from the population, 1.17 percent for MHSA inpatient and 4.40 percent for non-MHSA inpatient. However, for females, the model increases the bias for MHSA inpatient treatment. This is occurring because the overall results are underestimating MHSA inpatient treatment and the non-response adjustment compensates by increasing the weights for respondents who have utilized MHSA inpatient. However, female respondents

have higher MHSA inpatient utilization than non-respondents, opposite to males.[20] Since the non-response adjustment increases the weights for those who utilized MHSA inpatient overall, the females who utilized these services are also getting increased weight, which causes the bias to increase. Note that this is the only indicator where this effect occurs for females. The non-response model reduces bias for all other indicators.

A similar effect occurs for Hispanics for the non-MHSA inpatient indicator. In this case, respondents overall overestimate the population of non-MHSA inpatient indicator so the non-response model decreases the weights for respondents who have utilized non-MHSA inpatient. The Hispanic respondents slightly overestimate the population, but the non-response adjustment overcompensates for this overestimation so the final weighted result underestimates the population.

Web vs. Mail/CATI

Overall, Web respondents had significantly lower utilization rates for MHSA-related (0.62 percent; see *Figure 8*) and non-MHSA-related (2.65 percent; see *Figure 9*) inpatient treatment compared to mail and CATI respondents (1.25 percent and 4.68 percent, respectively). This pattern was also consistent across strata, indicating that the Web mode is, in general, less likely to be used by enrollees receiving inpatient treatment compared to the mail/CATI modes. For both HSC indicators, the estimated proportions among mail/CATI respondents were substantially closer to the population values than among Web respondents.

Figure 8. Percentage of Enrollees Receiving Inpatient Treatment for MHSA
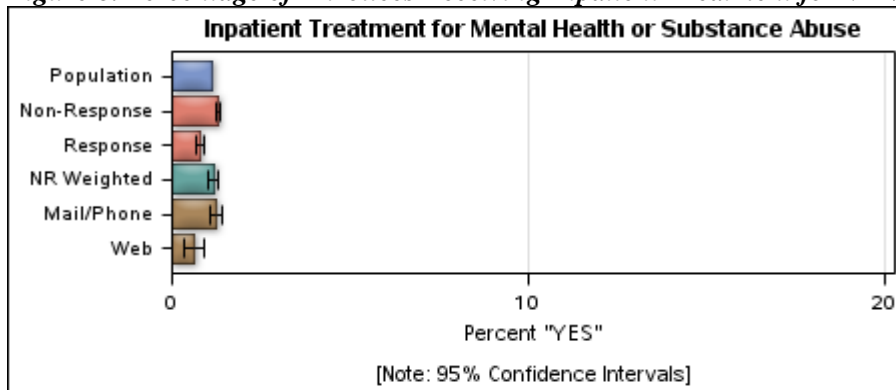
Figure 9. Percentage of Enrollees Receiving Inpatient Treatment for neither Mental Health nor Substance Abuse

---
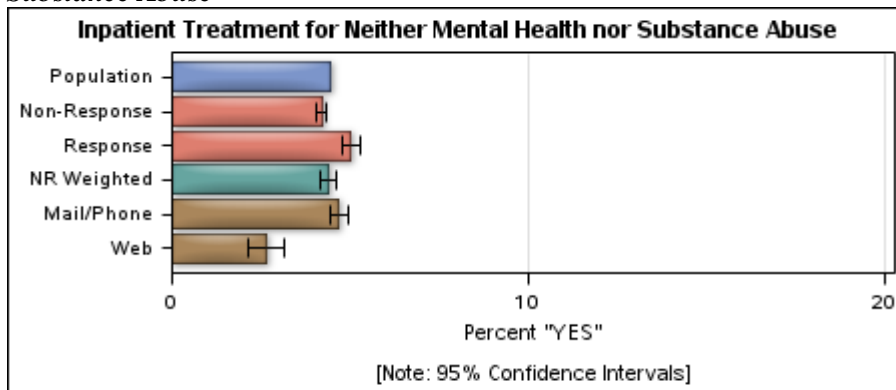
[20] Note that MHSA inpatient treatment is not a significant predictor of response among females when adjusting for age. However, it is a significant predictor of response among males even after adjusting for age.

Table 14. Percentage of Enrollees Receiving Inpatient Treatment for MHSA, by Stratum

Stratum	Level	Population %	Non-Responding Enrollees %	Responding Enrollees %	Sig.	Weighted %	Sig.	Mail/CATI %	Web %	Sig.
Overall	-	1.14	1.30	0.79	<.0001	1.17	0.73571	1.25	0.62	0.0032
Hispanic	N	1.58	1.92	0.94	<.0001	1.49	0.42868	1.60	0.78	0.0086
Hispanic	Y	1.73	1.92	1.42	0.0526	2.22	0.12989	2.28	1.79	0.6128
Hispanic	Unk	0.15	0.11	0.21	0.0522	0.24	0.21527	0.26	0.17	0.5868
Gender	F	1.35	1.22	1.42	0.3708	1.94	0.06257	2.07	1.18	0.1971
Gender	M	1.12	1.31	0.75	<.0001	1.11	0.81246	1.19	0.58	0.0067
OEF/OIF/OND	N	1.09	1.29	0.74	<.0001	1.12	0.66222	1.23	0.47	0.0005
OEF/OIF/OND	Y	1.49	1.36	1.35	0.9709	1.45	0.88895	1.41	1.76	0.6156
VISN	1	1.43	1.46	0.92	0.0947	1.46	0.93032	1.69	.	
VISN	2	1.19	1.37	0.89	0.1400	1.55	0.40973	1.51	1.76	0.8354
VISN	3	1.11	1.31	1.06	0.4122	1.62	0.17350	1.92	.	
VISN	4	1.17	1.21	0.56	0.0190	0.72	0.02557	0.83	.	
VISN	5	1.20	1.45	0.84	0.0786	1.54	0.43577	1.55	1.53	0.9890
VISN	6	1.06	1.41	0.91	0.1619	1.28	0.54349	1.44	0.12	0.0027
VISN	7	1.09	1.16	1.61	0.2457	2.01	0.04626	2.01	1.94	0.9612
VISN	8	1.21	1.53	0.84	0.0304	1.10	0.69877	1.21	0.45	0.0989
VISN	9	1.29	1.91	0.66	0.0011	0.80	0.03953	0.88	.	
VISN	10	1.26	1.39	0.64	0.0167	1.12	0.70781	1.02	1.87	0.5532
VISN	11	1.05	1.26	0.48	0.0092	0.74	0.20441	0.74	0.74	0.9962
VISN	12	1.37	1.62	0.64	0.0010	1.25	0.73777	1.46	0.04	<.0001
VISN	15	1.28	1.47	0.96	0.1304	1.35	0.82810	1.42	0.70	0.4831
VISN	16	1.17	1.11	0.89	0.4497	1.58	0.31490	1.59	1.45	0.8890
VISN	17	1.20	1.15	1.10	0.8823	1.48	0.46195	1.56	0.95	0.5128
VISN	18	1.05	1.25	0.56	0.0081	0.81	0.22310	0.83	0.69	0.8469
VISN	19	1.07	1.33	0.39	0.0018	0.54	0.00814	0.63	.	
VISN	20	1.09	1.54	0.31	0.0002	0.42	0.00004	0.51	.	
VISN	21	0.88	0.83	0.49	0.1589	0.73	0.50333	0.67	1.02	0.5898
VISN	22	0.94	0.91	0.45	0.0270	0.79	0.57687	0.89	0.29	0.1325
VISN	23	0.99	0.96	0.86	0.7158	1.34	0.31998	1.54	.	
Priority Group	1	2.12	2.41	1.57	0.0007	2.48	0.18359	2.61	1.73	0.2197
Priority Group	2	0.89	1.12	0.58	0.0032	0.80	0.59528	0.88	0.39	0.3123
Priority Group	3	0.76	0.90	0.55	0.0272	0.73	0.87626	0.84	0.16	0.0664
Priority Group	4	6.43	7.56	3.77	<.0001	6.69	0.69162	6.86	3.82	0.2876
Priority Group	5	1.38	1.54	0.95	0.0012	1.32	0.76744	1.37	0.79	0.2818

Stratum	Level	Population %	Non-Responding Enrollees %	Responding Enrollees %	Sig.	Weighted %	Sig.	Mail/CATI %	Web %	Sig.
Priority Group	6	0.29	0.27	0.04	0.0072	0.09	0.01105	0.09	0.08	0.8840
Priority Group	7	0.56	0.46	0.18	0.2056	0.37	0.47998	0.42	.	
Priority Group	8	0.16	0.19	0.10	0.1161	0.13	0.40476	0.13	0.08	0.5909
Pre/Post-Enrollee	POST	0.92	1.06	0.67	<.0001	0.96	0.61884	1.04	0.49	0.0064
Pre/Post-Enrollee	PRE	1.97	2.23	1.23	<.0001	1.94	0.89170	2.02	1.26	0.2818

Table 15. Percentage of Enrollees Receiving Inpatient Treatment for neither Mental Health nor Substance Abuse, by Stratum

Stratum	Level	Population %	Non-Responding Enrollees %	Responding Enrollees %	Sig.	Weighted %	Sig.	Mail/CATI %	Web %	Sig.
Overall	-	4.46	4.21	5.03	<.0001	4.40	0.63029	4.68	2.65	<.0001
Hispanic	N	6.15	6.15	6.11	0.8148	5.82	0.04418	6.17	3.53	<.0001
Hispanic	Y	5.75	5.45	5.88	0.3149	5.04	0.04826	5.20	3.87	0.2165
Hispanic	Unk	0.76	0.59	1.40	<.0001	0.96	0.05392	1.01	0.70	0.2757
Gender	F	4.09	3.68	4.70	0.0022	4.20	0.73790	4.35	3.31	0.1891
Gender	M	4.49	4.26	5.05	<.0001	4.42	0.57733	4.70	2.60	<.0001
OEF/OIF/OND	N	4.91	4.76	5.33	0.0004	4.87	0.73517	5.19	2.84	<.0001
OEF/OIF/OND	Y	1.45	1.20	1.64	0.0728	1.32	0.51464	1.33	1.27	0.9087
VISN	1	3.65	3.51	4.31	0.1607	3.96	0.52976	4.34	1.53	0.0417
VISN	2	4.68	4.03	5.19	0.0506	4.73	0.91871	4.90	3.90	0.5300
VISN	3	3.22	3.05	3.99	0.0562	3.74	0.27529	4.15	1.41	0.0073
VISN	4	3.60	3.92	4.36	0.4561	4.02	0.38524	4.24	2.49	0.2993
VISN	5	3.86	3.37	3.69	0.5592	3.46	0.40397	3.98	0.74	0.0093
VISN	6	4.15	3.94	4.14	0.7382	3.52	0.15948	3.78	1.59	0.1901
VISN	7	4.06	3.52	5.47	0.0029	4.85	0.17609	5.14	2.61	0.1319
VISN	8	5.42	5.17	5.11	0.9268	4.77	0.18025	5.07	2.93	0.1678
VISN	9	5.48	4.98	6.54	0.0314	5.49	0.97950	5.80	2.65	0.0621
VISN	10	4.99	4.88	5.75	0.1972	5.27	0.61998	5.44	4.04	0.5657
VISN	11	3.80	3.78	4.52	0.2216	3.72	0.85363	3.91	2.40	0.2569
VISN	12	5.24	5.17	5.10	0.9202	4.42	0.08392	4.97	1.26	0.0042
VISN	15	4.93	4.93	5.59	0.3347	5.26	0.57271	5.57	2.67	0.0630
VISN	16	4.68	4.53	4.56	0.9687	3.71	0.02948	3.99	1.53	0.0701
VISN	17	3.94	3.79	4.28	0.4308	3.54	0.38400	3.66	2.79	0.5128
VISN	18	4.95	4.72	6.48	0.0075	5.91	0.11767	6.42	2.97	0.0116
VISN	19	4.22	4.42	4.57	0.8022	3.31	0.01835	3.50	2.19	0.1748

Stratum	Level	Population %	Non-Responding Enrollees %	Responding Enrollees %	Sig.	Weighted %	Sig.	Mail/CATI %	Web %	Sig.
VISN	20	4.35	4.05	5.58	0.0273	4.53	0.72867	4.29	5.75	0.2814
VISN	21	5.00	4.72	5.92	0.0686	5.46	0.42107	5.77	3.86	0.1736
VISN	22	4.48	3.90	4.59	0.2081	4.06	0.38570	4.24	3.10	0.4407
VISN	23	4.14	3.47	5.17	0.0048	4.92	0.13952	5.30	2.40	0.0335
Priority Group	1	7.63	7.98	7.87	0.8120	7.40	0.53179	8.01	4.06	<.0001
Priority Group	2	3.43	3.02	3.65	0.0535	2.93	0.03377	3.03	2.38	0.3366
Priority Group	3	3.02	2.66	3.54	0.0030	2.81	0.37413	3.15	1.02	<.0001
Priority Group	4	16.38	16.30	16.27	0.9698	17.67	0.10869	17.38	22.42	0.1547
Priority Group	5	5.82	5.23	7.33	<.0001	6.11	0.35101	6.18	5.28	0.4599
Priority Group	6	1.08	0.88	1.50	0.0051	1.01	0.66103	0.97	1.20	0.5852
Priority Group	7	3.91	4.17	4.30	0.8739	3.84	0.90376	3.66	5.03	0.3993
Priority Group	8	1.47	1.26	1.65	0.0122	1.33	0.20983	1.45	0.60	0.0207
Pre/Post-Enrollee	POST	3.46	3.17	4.24	<.0001	3.57	0.33966	3.77	2.35	<.0001
Pre/Post-Enrollee	PRE	8.25	8.29	7.98	0.4714	7.57	0.03862	8.01	4.08	0.0001

3. Outpatient Treatment

Compared to inpatient treatment, larger proportions of the enrollee population utilize outpatient services: 16.34 percent of enrollees use outpatient treatment for MHSA and 62.21 percent use outpatient treatment for other reasons (non-MHSA).

Respondents vs. Non-Respondents

A significantly higher proportion of respondents receives MHSA outpatient treatment (17.17 percent) compared to non-respondents (16.10 percent), and a significantly higher proportion of respondents also receives non-MHSA outpatient treatment (76.22 percent) compared to non-respondents (56.26 percent). These differences are consistent across strata and indicate that enrollees who respond to the survey (compared to non-respondents) tend to have higher utilization rates for both MHSA and non-MHSA outpatient treatment. This pattern differs from that observed for inpatient treatment, where survey respondents tended to have lower utilization rates for MHSA inpatient treatment.

Comparison to population proportions indicates that the survey respondents over-represent the population of enrollees receiving MHSA outpatient treatment (16.34 percent of the population vs. 17.17 percent of respondents) and more substantially over-represent enrollees receiving non-MHSA outpatient treatment (62.21 percent of the population vs. 76.22 percent of respondents). After the response propensity score adjustment, the estimate of enrollee utilization for outpatient treatment is no longer significantly different from the population for MHSA (16.66 percent) and non-MHSA (62.09 percent).

Web vs. Mail/CATI

Overall, enrollees responding via the Web had significantly lower utilization rates for MHSA-related (12.62 percent) and non-MHSA-related (57.63 percent) outpatient treatment compared to enrollees responding via mail and CATI (17.29 percent and 62.78 percent, respectively). This pattern was also consistent across strata, indicating that the Web mode is, in general, less likely to be used by enrollees receiving outpatient treatment compared to the mail and CATI modes.

Figure 10. Percentage of Enrollees Receiving Outpatient Treatment for MHSA
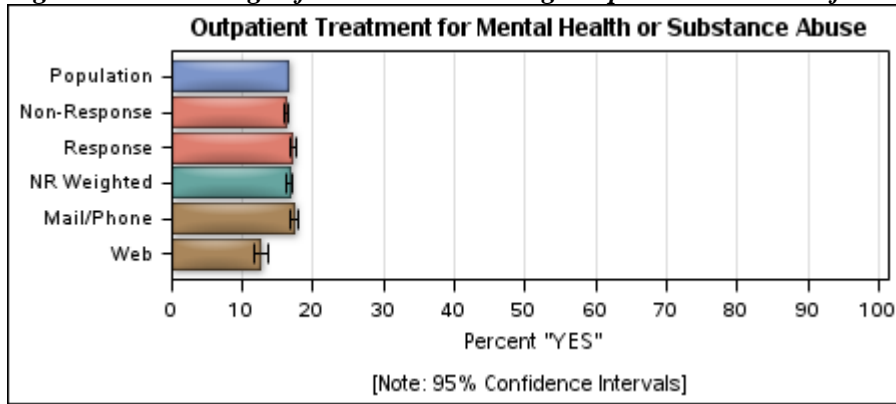
Figure 11. Percentage of Enrollees Receiving Outpatient Treatment for neither Mental Health nor Substance Abuse
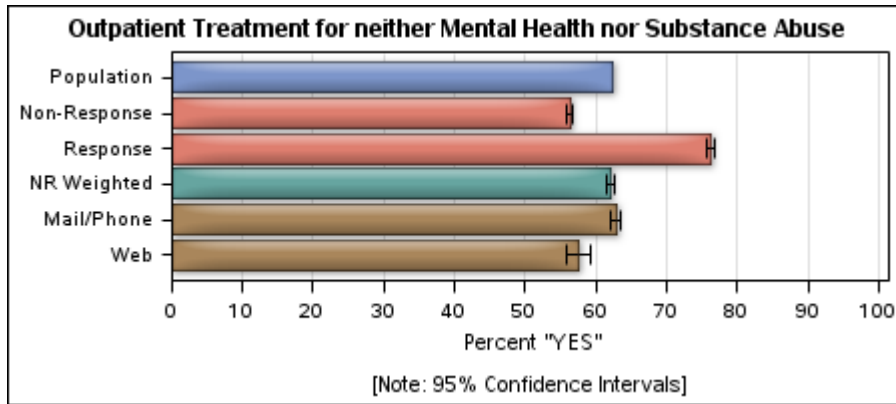
Table 16. Percentage of Enrollees Receiving Outpatient Treatment for MHSA, by Stratum

Stratum	Level	Population %	Non-Responding Enrollees %	Responding Enrollees %	Sig.	Weighted %	Sig.	Mail/CATI %	Web %	Sig.
Overall	-	16.34	16.10	17.17	<.0001	16.67	0.15171	17.29	12.62	<.0001
Hispanic	N	21.84	22.59	20.42	<.0001	21.32	0.08572	21.95	17.08	<.0001
Hispanic	Y	27.44	26.88	31.10	<.0001	30.57	0.00024	31.18	25.91	0.0407
Hispanic	Unk	3.40	3.13	4.64	<.0001	3.64	0.28097	3.93	1.96	0.0003
Gender	F	23.43	22.15	28.66	<.0001	25.36	0.00542	26.36	19.32	0.0003
Gender	M	15.79	15.58	16.49	0.0010	15.99	0.40446	16.59	12.07	<.0001
OEF/OIF/OND	N	15.37	15.17	16.03	0.0014	15.34	0.88364	15.93	11.56	<.0001
OEF/OIF/OND	Y	22.78	21.23	30.22	<.0001	25.45	0.00250	26.20	20.23	0.0159
VISN	1	16.06	15.88	15.60	0.7991	15.98	0.93363	16.45	12.95	0.1770
VISN	2	13.48	12.22	13.38	0.2226	12.96	0.53392	13.43	10.65	0.2214
VISN	3	13.93	13.62	16.05	0.0097	14.97	0.19051	15.69	10.94	0.0342
VISN	4	14.51	14.14	14.42	0.7850	14.14	0.65920	14.36	12.69	0.5109
VISN	5	14.73	13.82	16.07	0.0396	14.49	0.80272	15.28	10.39	0.0565

continued from previous page

Stratum	Level	Population %	Non-Responding Enrollees %	Responding Enrollees %	Sig.	Weighted %	Sig.	Mail/CATI %	Web %	Sig.
VISN	6	16.39	16.50	16.23	0.8211	15.25	0.24828	15.97	10.11	0.0389
VISN	7	18.02	18.13	21.29	0.0152	19.19	0.25959	19.87	13.94	0.0910
VISN	8	18.05	18.32	17.96	0.7337	19.13	0.27008	19.85	14.80	0.0516
VISN	9	17.30	16.92	18.12	0.3227	17.04	0.79344	17.12	16.25	0.7980
VISN	10	18.14	18.74	19.19	0.7005	18.69	0.59840	18.86	17.40	0.6589
VISN	11	14.99	15.11	14.93	0.8636	14.90	0.91550	15.37	11.65	0.1731
VISN	12	15.73	15.99	15.45	0.6065	16.68	0.34860	17.77	10.48	0.0111
VISN	15	16.01	14.55	18.33	0.0008	17.72	0.09282	18.59	10.22	0.0053
VISN	16	17.86	17.58	19.87	0.0584	19.08	0.25015	19.51	15.78	0.2503
VISN	17	17.40	15.47	20.28	<.0001	18.40	0.32524	18.64	16.83	0.5578
VISN	18	16.49	16.18	16.98	0.4641	15.86	0.47872	16.85	10.12	0.0087
VISN	19	15.57	15.23	14.94	0.7836	13.99	0.07405	14.58	10.45	0.1562
VISN	20	15.49	16.09	16.08	0.9906	15.50	0.99570	16.86	8.50	0.0003
VISN	21	16.63	16.51	17.64	0.3125	16.89	0.79329	17.44	14.08	0.2307
VISN	22	17.34	16.77	19.65	0.0059	18.85	0.13282	19.76	14.03	0.0225
VISN	23	13.17	13.24	12.41	0.4179	12.11	0.17500	12.66	8.52	0.0819
Priority Group	1	36.53	37.08	36.33	0.3682	36.26	0.70746	37.90	27.26	<.0001
Priority Group	2	17.17	16.74	17.07	0.6167	16.22	0.09834	16.92	12.65	0.0039
Priority Group	3	11.28	11.50	12.13	0.2581	11.52	0.62370	12.20	7.92	0.0008
Priority Group	4	29.83	30.00	28.47	0.1381	31.77	0.05033	31.87	30.13	0.6704
Priority Group	5	16.13	15.98	17.39	0.0149	17.96	0.00139	17.97	17.86	0.9542
Priority Group	6	8.16	8.25	8.60	0.5659	8.32	0.76587	8.96	5.22	0.0062
Priority Group	7	9.48	9.08	8.23	0.4443	8.55	0.30675	8.68	7.71	0.7220
Priority Group	8	4.22	4.00	4.17	0.5036	3.96	0.26013	4.26	2.26	0.0010
Pre/Post-Enrollee	POST	14.47	14.30	15.18	0.0017	14.91	0.07636	15.51	11.26	<.0001
Pre/Post-Enrollee	PRE	23.44	23.14	24.62	0.0242	23.31	0.80858	23.85	19.10	0.0078

Table 17. Percentage of Enrollees Receiving Outpatient Treatment for neither Mental Health nor Substance Abuse, by Stratum

Stratum	Level	Population %	Non-Responding Enrollees %	Responding Enrollees %	Sig.	Weighted %	Sig.	Mail/CATI %	Web %	Sig.
Overall	-	62.21	56.26	76.22	<.0001	62.09	0.71169	62.78	57.63	<.0001
Hispanic	N	80.20	76.08	87.66	<.0001	77.43	0.00000	77.93	74.07	0.0004
Hispanic	Y	76.40	73.96	85.12	<.0001	74.78	0.07710	75.06	72.67	0.3902
Hispanic	Unk	22.88	18.64	37.67	<.0001	24.07	0.01250	24.39	22.22	0.0950
Gender	F	59.79	55.06	76.26	<.0001	61.84	0.01379	62.56	57.50	0.0248
Gender	M	62.39	56.36	76.22	<.0001	62.10	0.40058	62.80	57.64	<.0001
OEF/OIF/OND	N	63.87	57.85	77.14	<.0001	63.84	0.94257	64.64	58.82	<.0001
OEF/OIF/OND	Y	51.17	47.50	65.68	<.0001	50.41	0.46674	50.59	49.23	0.6490
VISN	1	63.21	58.69	76.86	<.0001	63.80	0.67257	64.87	56.90	0.0483
VISN	2	58.63	51.25	73.73	<.0001	59.29	0.64495	59.88	56.39	0.3661
VISN	3	49.27	44.55	63.93	<.0001	50.50	0.34562	52.36	40.11	0.0009
VISN	4	61.80	55.70	76.72	<.0001	62.38	0.69415	63.42	55.47	0.0701
VISN	5	51.34	47.16	66.61	<.0001	50.52	0.55389	53.52	34.87	<.0001
VISN	6	61.92	57.13	72.85	<.0001	57.94	0.00669	59.34	47.87	0.0091
VISN	7	61.80	56.25	76.73	<.0001	62.88	0.44918	64.03	54.07	0.0257
VISN	8	68.54	62.86	81.53	<.0001	70.08	0.25469	69.93	71.00	0.7680
VISN	9	65.00	58.57	78.86	<.0001	65.40	0.77906	65.63	63.30	0.6227
VISN	10	64.13	57.85	78.50	<.0001	65.56	0.32221	65.59	65.33	0.9539
VISN	11	63.82	57.98	77.38	<.0001	63.47	0.81019	63.97	60.04	0.3464
VISN	12	64.47	58.28	77.95	<.0001	64.60	0.93095	65.05	61.99	0.4568
VISN	15	64.58	57.80	79.58	<.0001	65.97	0.34143	66.40	62.33	0.3688
VISN	16	63.79	57.92	76.53	<.0001	62.35	0.31544	63.11	56.46	0.1208
VISN	17	60.35	55.38	76.70	<.0001	61.70	0.33699	62.51	56.59	0.1491
VISN	18	62.49	57.07	76.48	<.0001	62.49	0.99901	63.37	57.38	0.1396
VISN	19	60.93	54.85	73.18	<.0001	56.94	0.00589	57.45	53.87	0.3769
VISN	20	61.54	52.87	73.94	<.0001	60.32	0.39462	60.99	56.87	0.2673
VISN	21	60.81	56.00	76.12	<.0001	61.77	0.49671	61.06	65.42	0.2522
VISN	22	56.19	51.72	69.53	<.0001	55.36	0.54803	55.53	54.49	0.7822
VISN	23	67.70	59.18	80.22	<.0001	66.70	0.48704	66.98	64.87	0.6000
Priority Group	1	82.70	79.79	88.50	<.0001	79.76	0.00014	81.18	71.99	<.0001
Priority Group	2	65.62	60.91	75.90	<.0001	62.33	0.00021	63.69	55.32	0.0003
Priority Group	3	58.76	53.04	73.02	<.0001	57.38	0.10769	58.18	53.15	0.0238
Priority Group	4	78.20	74.22	90.04	<.0001	83.60	0.00000	83.86	79.25	0.2215
Priority Group	5	61.71	55.84	79.56	<.0001	65.83	0.00000	65.59	68.42	0.3129

continued from previous page

Stratum	Level	Population %	Non-Responding Enrollees %	Responding Enrollees %	Sig.	Weighted %	Sig.	Mail/CATI %	Web %	Sig.
Priority Group	6	43.12	37.13	61.67	<.0001	42.48	0.53321	41.18	48.87	0.0031
Priority Group	7	75.55	70.27	80.69	<.0001	72.56	0.05778	72.15	75.31	0.4551
Priority Group	8	49.44	41.85	65.50	<.0001	49.00	0.47833	49.95	43.58	0.0002
Pre/Post-Enrollee	POST	59.80	53.74	74.29	<.0001	59.58	0.55104	60.17	56.01	<.0001
Pre/Post-Enrollee	PRE	71.34	66.09	83.43	<.0001	71.60	0.71026	72.39	65.34	0.0008

4. VHA Pharmacy Services

A substantial proportion (54.21 percent) of the enrollee population receives prescription drug services.

Respondents vs. Non-Respondents

A significantly higher proportion of respondents receives prescription drug services (66.79 percent) compared to non-respondents (48.91 percent). This relatively large difference is consistent across all strata and indicates that enrollees who respond to the survey (compared to non-respondents) tend to have higher utilization rates for prescription drug services.

Comparison to the population proportion indicates that survey respondents substantially over-represent the population of enrollees receiving prescription drug services (54.21 percent of the population vs. 66.79v of respondents). The propensity score model and raking adjustment reduce the utilization rate for respondents to 54.27 percent.

Web vs. Mail/CATI

Overall, enrollees responding via the Web had significantly lower utilization rates for prescription drug services (47.07 percent) compared to enrollees responding via mail and CATI (55.40 percent). This pattern was also consistent across strata, indicating that the Web mode is, in general, less likely to be used by enrollees receiving prescription drug services compared to the mail and CATI modes. This effect still exists after adjusting for age.

Figure 12. Percentage of Enrollees Receiving Prescription Drug Services
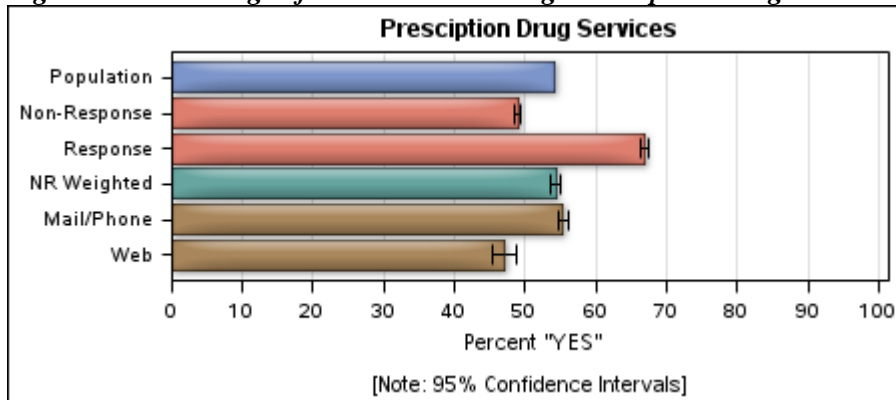
Table 18. Percentage of Enrollees Receiving Prescription Drug Services by Stratum

Stratum	Level	Population %	Non-Responding Enrollees %	Responding Enrollees %	Sig.	Weighted %	Sig.	Mail/CATI %	Web %	Sig.
Overall	-	54.21	48.91	66.79	<.0001	54.27	0.84240	55.40	47.07	<.0001
Hispanic	N	71.34	67.59	78.17	<.0001	69.01	0.00000	70.07	61.93	<.0001
Hispanic	Y	67.84	65.17	77.61	<.0001	66.90	0.32329	67.66	61.17	0.0227
Hispanic	Unk	16.76	13.50	28.14	<.0001	17.69	0.02373	18.16	14.99	0.0049
Gender	F	52.32	47.62	67.48	<.0001	54.22	0.01974	55.12	48.83	0.0050
Gender	M	54.36	49.02	66.74	<.0001	54.28	0.80910	55.42	46.92	<.0001
OEF/OIF/OND	N	56.37	51.11	67.93	<.0001	56.29	0.80057	57.51	48.51	<.0001
OEF/OIF/OND	Y	39.87	36.84	53.69	<.0001	40.90	0.29524	41.49	36.83	0.1028
VISN	1	53.12	48.11	65.45	<.0001	54.04	0.49188	55.42	45.17	0.0078
VISN	2	50.28	43.53	63.26	<.0001	50.61	0.81245	52.14	43.09	0.0120
VISN	3	42.21	38.20	55.02	<.0001	43.36	0.35783	45.40	31.93	0.0001
VISN	4	52.62	47.59	64.60	<.0001	52.20	0.76675	53.79	41.56	0.0027
VISN	5	42.98	39.82	55.86	<.0001	42.10	0.50221	45.07	26.58	<.0001
VISN	6	55.34	50.98	64.92	<.0001	52.34	0.03369	54.48	36.96	<.0001
VISN	7	55.11	50.61	68.42	<.0001	55.77	0.63235	57.24	44.53	0.0041
VISN	8	59.16	54.19	70.05	<.0001	60.27	0.40178	60.43	59.33	0.7587
VISN	9	57.77	52.22	68.77	<.0001	57.03	0.59914	57.48	52.91	0.3264
VISN	10	56.28	50.53	69.18	<.0001	57.68	0.31550	57.68	57.73	0.9906
VISN	11	56.09	51.78	68.14	<.0001	55.72	0.79476	56.70	48.97	0.0546
VISN	12	57.26	50.91	69.21	<.0001	57.14	0.92881	58.10	51.66	0.0996
VISN	15	57.21	51.58	70.56	<.0001	58.71	0.29116	60.06	47.17	0.0025
VISN	16	57.56	51.81	70.10	<.0001	57.02	0.70101	57.99	49.53	0.0450
VISN	17	53.57	48.33	67.49	<.0001	54.27	0.61199	55.50	46.45	0.0240
VISN	18	54.49	50.14	69.05	<.0001	55.83	0.32457	57.23	47.72	0.0145
VISN	19	51.63	46.33	63.19	<.0001	49.27	0.08557	50.39	42.62	0.0401
VISN	20	53.15	46.15	64.43	<.0001	52.36	0.56952	53.36	47.20	0.0895
VISN	21	51.86	47.07	65.90	<.0001	53.38	0.26351	53.25	54.07	0.8230
VISN	22	47.50	43.39	59.32	<.0001	47.20	0.81681	47.81	43.98	0.2820
VISN	23	57.79	50.45	70.24	<.0001	58.32	0.70129	59.15	52.86	0.1072
Priority Group	1	75.90	73.32	81.64	<.0001	73.76	0.00566	75.60	63.68	<.0001
Priority Group	2	54.32	50.66	62.80	<.0001	51.41	0.00061	52.92	43.70	<.0001
Priority Group	3	46.75	42.19	59.32	<.0001	46.36	0.62366	47.70	39.25	<.0001
Priority Group	4	74.17	70.63	86.10	<.0001	80.38	0.00000	80.86	72.36	0.0293
Priority Group	5	56.46	50.72	73.20	<.0001	59.93	0.00001	59.86	60.74	0.7488

continued from previous page

Stratum	Level	Population %	Non-Responding Enrollees %	Responding Enrollees %	Sig.	Weighted %	Sig.	Mail/CATI %	Web %	Sig.
Priority Group	6	31.68	26.59	46.62	<.0001	31.44	0.78227	31.03	33.47	0.2726
Priority Group	7	62.17	57.01	66.28	<.0001	59.67	0.12246	59.75	59.13	0.8905
Priority Group	8	42.25	35.58	56.15	<.0001	41.65	0.29685	43.00	33.91	<.0001
Pre/Post-Enrollee	POST	50.83	45.55	63.73	<.0001	50.82	0.98469	51.82	44.76	<.0001
Pre/Post-Enrollee	PRE	67.04	62.04	78.19	<.0001	67.36	0.63848	68.55	58.00	<.0001

7. DISCUSSION AND RECOMMENDATIONS

This is the seventh report in the Experimental Methods Series. The approach taken in the current report was to provide a comprehensive account of TSE across all aspects of the Survey of Enrollees (see *Figure 1*). The TSE framework divides survey error into two major sources: *errors of representation*, which are due to the systematic and random errors that influence which members of the population respond to the survey; and *errors of observation*, which are due to the systematic and random errors that influence the accuracy with which survey constructs are measured.

Summary of Findings

Across all areas investigated in the current report, evidence of low or no bias was found, with no evidence of major bias in any TSE domain. Where bias was detected, the survey weights were shown to effectively reduce bias in population estimates.

- The evaluation of potential bias in the sampling and weighting design revealed that the disproportionate stratified sampling plan introduces representation bias in the unweighted sample (as expected), but the design weights eliminate this bias.

- The ME experiment shows that the substantial reduction in potential coverage bias due to the introduction of a mail mode has been achieved with only minor increases in measurement error due to mode effects. The replication of conclusions drawn in 2013 greatly increases confidence that any mode effects due to the Survey of Enrollees's mixed-mode design are of small magnitude and do not threaten substantive conclusions drawn from the data.

- The SSM-P experiment found that the use of a "long" mail follow-up protocol, with two survey mailings, compared to a "short" protocol, with only one survey mailing, significantly increased response rates among both CATI survey non-respondents and enrollees with non-working phone numbers. The successful replication of 2013 experiment findings indicates that this response rate increase is a systematic effect and can be recommended for decreasing the potential for non-response bias.

- The SSM-M experiment found that, although a "long" mail protocol with two survey mailings did not increase total response rates compared to a "short" protocol with one survey mailing, the distributions of response channels used by converted non-respondents during follow-up did differ. Specifically, converted first-mailing non-respondents appeared to respond using whatever response channel was used for the initial non-response follow-up (i.e., mail in the long protocol, or phone in the short protocol). This finding suggests a potential cost savings by favoring mail follow-ups over CATI follow-ups in the mail survey protocol.

- The non-response bias analysis showed that, as in past years, although there were some differences between respondents and non-respondents with respect to health service utilization indicators, these differences were not of large magnitude and were in nearly all cases eliminated by the response propensity score and raking weight adjustments.

Thus, the general conclusions of this report are that the Survey of Enrollees is representative of the target population and that the survey instrument is accurately measuring the outcomes of interest.

Recommendations

Recommendations that have stemmed from prior annual analyses are to (parenthetical notes indicate if these recommendations were implemented and if so, when):

- Use propensity score weighting based on utilization of administrative records (**Full adoption**);
- Send a pre-survey notification letter to Veterans prior to calling (**Full adoption**);

- Increase the call attempts from six to seven **(Full adoption)**;
- Use address information to locate and update telephone numbers via database look-ups **(Mixed adopton: full adoption** in 2008 and 2010; **not implemented** in 2011 due to security and privacy concerns; **implemented sparingly** in 2012, 2013, and 2014 for seven-digit telephone numbers and invalid area codes);
- Add a mail survey **(Partial adoption** as described in the current report); and
- Add a Web survey **(Full adoption)**.

Based on the current analyses, we make the following additional recommendations:

1. **Continue to offer the mail and CATI modes**. The 2013 and 2014 mode effects experiments revealed some replicable differences in survey responses between modes, but these differences are quite small. Given that we cannot know which mode provides more accurate results (i.e., which comes closer to the "true score" for a given outcome), and given that the differences between modes were minor, the guaranteed benefit of reducing the potential for coverage bias by including all modes outweighs the introduction of small amounts of measurement error due to the use of a mixed-mode design.

2. **Implement the second survey mailing as part of CATI survey follow-up**. As part of the CATI non-response/non-working follow-up protocol, the second survey mailing raised total response rates by eight percentage points over a single follow-up mailing in 2013, and by seven percentage points in 2014. Based on this evidence of a replicable effect, we recommend broader adoption of the "long" mail protocol for following up with CATI non-respondents and non-working numbers. However, the cost implications need to be considered in more detail when deciding how to scale this protocol modification. In particular, the cost of a response generated from a second survey mailing must be compared to the cost of a response from sampling another record.

3. **Replicate the SSM-M experiment**. The SSM-M experiment conducted in 2014 found that, although a second survey mailing as part of the mail protocol did not increase total response rates compared to a single mailing, it did lead to a significant increase in the use of the mail survey by first-mailing non-respondents. Given ICF's recommendation to increase the use of the mail mode going forward, this experiment warrants replication to ensure that the current findings are systematic before reducing or eliminating CATI follow-ups in the mail protocol.

4. **Continue to offer a Web response channel**. There is some evidence that the population choosing to respond via Web differs from the populations responding via mail/CATI. Specifically, Web respondents reported lower utilization of VA health care services; continuing to offer the Web option will increase coverage of this group.

5. **Continue to investigate the potential for coverage bias.** Coverage bias arises when differences between enrollees included in vs. excluded from the sampling frame are associated with survey outcomes. Although the potential for coverage bias in the Survey of Enrollees has been greatly reduced by the introduction of the mail mode to cover enrollees without a phone number on record, the current frame development procedures still leave a small window for coverage bias due to the use of particular criteria that exclude enrollees from the sampling frame. Of particular concern is the exclusion from the sampling frame of enrollees in the VHA database who do not have a valid address on record. Through intensive efforts to match contact information to a sample of these currently excluded enrollees and obtaining completed interviews with them, a cost-benefit analysis can then be conducted to determine the extent to which the exclusion of these enrollees introduces coverage bias, and whether extending coverage to them warrants the increased cost of doing so.

APPENDIX A – UTILIZATION MEASURES

Utilization indicators based on administrative records are provided for the following services in the previous year:
- Institutional and non-institutional long-term care benefits,
- Inpatient and outpatient treatment serves, both for MHSA and non-MHSA issues, and
- Prescription drug benefits.

Based on administrative records, these measures indicate whether an enrollee had utilized any of the following services in the previous year (the file did not indicate the frequency of use or amount paid for any of these benefits):
1. Received long-term care services,
   a. Institutional
   b. Non-institutional
2. Received Inpatient treatment,
   a. MHSA
   b. Non-MHSA
3. Received Outpatient treatment,
   a. MHSA
   b. Non-MHSA
4. Received VHA pharmacy services.

Since 2007, these utilization indicators have been used in the weighting process, for bias assessment, and for assessing sample design performance.

From 2007–2010, the indicators were based on service utilization sourced from VHA workload files that were based on bed section and clinic stop. This categorization indicated where a Veteran received care.

For the 2011 survey, the indicators were based on service utilization from HSCs. This categorization indicates what care a Veteran received. A second change made in 2011 included separating long-term care in institutions and not in institutions As such, from 2007–2010, the indicator was a single measure of home health service.

APPENDIX B – NON-RESPONSE PROPENSITY SCORE QUINTILES

The following tables show the distribution of non-response propensity score model predictors for combined-sample respondents by the propensity score quintiles used to compute the non-response adjustment. For categorical variables (all dichotomous), percentages are reported, and for continuous variables, means are reported.

Table 19. Distribution of Non-Response Propensity Score Model Categorical Predictors for Combined-Sample Respondents by Propensity Score Quintiles

Predictor	1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
VISN1	9.8%	18.4%	20.4%	29.2%	22.2%
VISN2	11.1%	20.8%	20.4%	30.6%	17.0%
VISN3	24.6%	26.6%	22.9%	20.4%	5.6%
VISN4	9.5%	17.8%	17.8%	32.5%	22.3%
VISN5	24.9%	24.3%	22.6%	19.0%	9.2%
VISN6	9.0%	18.8%	19.9%	24.0%	28.2%
VISN7	8.4%	17.1%	22.2%	25.8%	26.4%
VISN8	11.1%	22.2%	27.4%	25.4%	13.9%
VISN9	6.7%	13.9%	19.4%	24.6%	35.3%
VISN10	7.5%	17.0%	21.1%	27.6%	26.8%
VISN11	7.7%	15.3%	20.1%	24.1%	32.9%
VISN12	9.3%	16.1%	18.8%	25.0%	30.7%
VISN15	6.4%	12.6%	17.1%	20.3%	43.7%
VISN16	9.5%	16.5%	23.2%	25.1%	25.8%
VISN17	13.3%	20.8%	24.9%	25.6%	15.5%
VISN18	8.8%	14.6%	22.1%	25.8%	28.8%
VISN19	9.6%	17.0%	20.3%	22.1%	30.9%
VISN20	6.3%	13.5%	21.7%	21.2%	37.2%
VISN21	11.4%	20.2%	20.6%	26.9%	20.8%
VISN22	24.2%	24.8%	25.3%	20.4%	5.3%
VISN23	5.4%	9.0%	16.2%	18.2%	51.2%
Priority Group 1	5.9%	13.0%	21.0%	30.3%	29.7%
Priority Group 2	9.8%	18.0%	20.7%	24.1%	27.5%
Priority Group 3	12.5%	18.5%	19.8%	23.7%	25.5%
Priority Group 4	12.9%	26.0%	31.9%	23.6%	5.6%
Priority Group 5	14.8%	18.0%	29.3%	25.9%	12.0%
Priority Group 6	21.0%	22.7%	13.6%	22.0%	20.6%
Priority Group 7	1.6%	7.3%	13.8%	20.0%	57.3%
Priority Group 8	10.0%	18.7%	16.9%	21.6%	33.0%
Male (vs. Female)	9.6%	16.9%	20.3%	25.2%	28.0%

continued from previous page

Predictor	1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
Has phone	11.1%	17.4%	21.3%	25.1%	25.1%
Patient (Sep13 Enrollment)	3.5%	11.1%	22.0%	31.1%	32.3%
Hispanic/Latino	24.1%	26.6%	29.4%	16.5%	3.4%
Pre-Enrollee (vs. Post-Enrollee)	9.4%	18.1%	26.0%	28.7%	17.9%
OEF/OIF/OND Yes (vs. No)	42.8%	30.4%	20.3%	5.8%	0.6%
Urban	14.8%	21.8%	23.5%	24.9%	14.9%
Rural	5.5%	12.1%	17.7%	24.1%	40.7%
Highly Rural	3.1%	6.2%	16.4%	15.3%	59.0%
Received long-term care services, Institutional	44.8%	35.2%	16.2%	3.8%	.
Received long-term care services, Non-Institutional	4.8%	13.4%	25.7%	31.1%	25.0%
Received Inpatient treatment, MHSA	44.8%	35.1%	17.8%	2.0%	0.2%
Received Inpatient treatment, Non-MHSA	7.8%	16.7%	26.1%	27.8%	21.7%
Received Outpatient treatment, MHSA	7.8%	21.6%	29.8%	26.4%	14.4%
Received Outpatient treatment, Non-MHSA	2.8%	10.7%	21.9%	31.6%	32.9%
Received VHA pharmacy services	3.0%	10.6%	21.8%	30.6%	34.0%

Table 20. Distribution of Non-Response Propensity Score Model Continuous Predictors for Combined-Sample Respondents by Propensity Score Quintiles

Predictor	1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
Age	45.8	58.0	61.7	67.3	74.8