Service Level Measurements – Same Day Services

Sampling Methodology Report

Prepared by

Veteran Experience Office

Version 1

June 2020

Contents

Executive Summary

The Same Day Services Survey is designed to measure customer experience in obtaining and receiving non-emergency VHA services on the same day that they requested the services. The more specific goal of this survey is to measure customer experience in for Veterans that received same day services in the following categories:

- Getting an appointment that day (same-day appointments)
- Obtaining a medication refill through outpatient pharmacy
- Evaluation of a Veteran's medication renewal request
- Nurse visit
- Issue with a medical device/equipment
- Walk-in vaccinations
- Traveling Veterans reaching a Traveling Veteran Coordinator
- Scheduling a future appointment

Veterans experience data is collected by using an online transactional survey disseminated via an invitation email sent to randomly selected beneficiary. The data collection occurs **once** per week with invitation being sent out within **8 days** of their same day interaction**. The** questionnaire is brief and contains general Likert-scale (a scale of 1-5 from Strongly Disagree to Strongly Agree) questions to assess customer satisfaction as well as questions assessing the knowledge, speed, and manner of the interaction. After the survey has been distributed, recipients have two weeks to complete the survey and will receive a reminder email after one week.

The overall sample size for the Same Day Service Survey population is determined so that, where sufficient sample is available, the reliability of monthly survey estimates is at 3% Margin of Error at a 95% Confidence Level for each of the seven categories of care outlined above. The survey will be sent to a representative sample of Veterans who received the service. Once data collection is completed, the participant responses in the online survey will be weighted to reflect the proportions found in the same day service population.

This report describes the methodology used to conduct the Same Day Service Survey. Information about quality assurance protocols, as well as limitations of the survey methodology, is also included in this report.

Part I – Introduction

A. Background

The **Enterprise Measurement and Design** team (EMD) is part of the **Insights and Analytics** (I&A) division within the **Veterans Experience Office** (VEO). The EMD team is tasked with conducting transactional surveys of the Veteran population to measure their satisfaction with the Department of Veterans Affairs (VA) numerous benefit services. Thus, their mission is to empower Veterans by rapidly and discreetly collecting feedback on their interactions with such VA entities as NCA, VHA, and VBA. VEO surveys generally entail *probability* samples which only contact minimal numbers of Veterans necessary to obtain reliable estimates. This information is subsequently used by internal stakeholders to monitor, evaluate, and improve beneficiary processes. Veterans are always able to decline participation and have the ability to opt out of future invitations. A *quarantine* protocol is maintained to limit the number of times a Veteran may be contacted, in order to prevent survey fatigue, across all VEO surveys.

The VHA is dedicated to offering quality services to Veterans as quickly as possible and with minimal barriers to care. When possible, this means being able to serve the Veteran's with the convenience of same day services.

In order to continue to provide quality services to Veterans, VEO has been commissioned to measure the satisfaction of Veterans with the VHA's same day services. To complete this goal, VEO proposed to conduct a brief transactional survey on randomly selected Veterans who had received such services. Three surveys were developed with three multiple choice screener questions and 10 Customer Experience (CX) questions centered around a Veteran's recent encounter pertaining to the notions of Trust/Confidence, Satisfaction, Quality, Ease/Simplicity, Efficiency/Speed, Equity/Transparency, and Employee Helpfulness. These Likert-scale (a scale of 1-5) questions are designed through extensive Veteran input and recommendations from subject matter experts within the VA.

Where sample is available to meet the targeted accuracy of the survey, Veterans are randomly selected to participate in the survey. Otherwise all Veterans that used a particular same day service will be selected. Invitation to the survey will be sent via an invitation email. A link is enclosed so the survey may be completed using an online interface, with customized participant information. The data is collected on a **weekly basis** and the survey is reported on a monthly basis. The purpose of this document is to outline the planned sample design and provide a description of the data collection and sample sizes necessary for proper reporting.

B. Basic Definitions

Coverage	**The percentage of the population of interest that is included in the sampling frame.**
Measurement Error	The difference between the response coded and the true value of the characteristic being studied for a respondent.
Non-Response	Failure of some respondents in the sample to provide responses in the survey.
Transaction	A *transaction* refers to the specific time a Veteran interacts with the VA that impacts the Veteran's journey and their perception of VA's effectiveness in caring for Veterans.
Response Rate	The ratio of participating persons to the number of contacted persons. This is one of the basic indicators of survey quality.
Sample	In statistics, a data sample is a set of data collected and/or selected from a statistical population by a defined procedure.
Sampling Error	Error due to taking a particular sample instead of measuring every unit in the population.
Sampling Frame	A list of units in the population from which a sample may be selected.
Reliability	The consistency or dependability of a measure. Also referred to as *standard error*.

C. Application to Veterans Affairs

Customer experience and satisfaction are usually measured at three levels to: 1) provide enterprises the ability to track, monitor, and incentivize service quality; 2) provide service level monitoring and insights; and 3) give direct point-of-service feedback. This measurement may bring insights and value to all stakeholders at VA. Front-line VA leaders can resolve individual feedback from Veterans and take steps to improve the customer experience; meanwhile VA executives can receive real-time updates on systematic trends that allow them to make changes.

1) To collect continuous customer experience data
2) To help field staff and the national office identify areas of improvement.
3) To understand emerging drivers and detractors of customer experience.

Part II – Methodology

A. Target Population, Frame, and Stratification

The target population of the Same Day Service Survey is defined as any Veterans who in the previous week has obtained and received non-emergency VHA services on the same day that they requested the services.

The sample frame is prepared by extracting population information directly from the VHA's Corporate Data Warehouse (CDW). These extracts are also used to obtain population figures for the sample weighting process. The Veteran is the primary sampling unit and is randomly selected from the population according to a stratified design with allocation proportional to the true population. The sample will be stratified by the type of same day interaction and balanced to reflect the age, gender, and district distribution of the population.

B. Sample Size Determination

To achieve a certain level of reliability, the sample size for a given level of reliability is calculated below (Lohr, 1999):

For a population that is *large*, the equation below is used to yield a representative sample for proportions:

$$n_0 = \frac{Z_{\alpha/2}^2 pq}{e^2}$$

where

- $Z_{\alpha/2}$= is the critical Z score which is 1.96 under the normal distribution when using a 95% confidence level (α = 0.05).
- p = the estimated proportion of an attribute that is present in the population, with q=1-p.
  - o Note that pq attains its maximum when value p=0.5 or 50%. This is what is typically reported in surveys where multiple measures are of interest. When examining measures closer to 100% or 0% less sample is needed to achieve the same margin of error.
- e = the desired level of precision or margin of error. For example, for the Same Day Service Survey the targeted margin of error is e = 0.03, or +/-3% where sufficient sample is available.

For a population that is relatively *small*, the finite population correction is used to yield a representative sample for proportions:

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

Where

- n_0 = Representative sample for proportions when the population is large.
- N = Population size.

The margin of error surrounding the baseline proportion is calculated as:

$$Margin\ of\ Error = z_{\alpha/2} \sqrt{\frac{N-n}{N-1}} \sqrt{\frac{p(1-p)}{n}}$$

Where

- $Z_{\alpha/2}$= 1.96, which is the critical Z score value under the normal distribution when using a 95% confidence level (α = 0.05).
- N = Population size.
- n = Representative sample.
- p = the estimated proportion of an attribute that is present in the population, with q=1-p.

Estimates from the population files drawn for the first part days of 2020 through March 15 indicate that, in the average month, 690,000 same day encounters with the VHA occur. Table 1A indicates the population figures based on numbers from that period, as well as estimated population with email addresses on file and the proportion that is likely to be usable after removing duplicates and quarantine rules across VEO surveys.

For this survey, the goal is to reach a +/- 3.0 margin of error with 95% confidence for each type of same day service where sufficient sample is available. Where sample is not sufficient, the sample plan proposes that all available sample be used. Currently we are unable to identify sample sources for some of the encounter types. If data becomes available to reach Veterans that experience these interactions, all sample will be used unless the sample exceed the threshold to reach the +/-3.0% target using a sample[1].

Table 1A. Target Population Figures, Sample Size, and Email Contacts

Type of Same Day Service	Estimated Monthly Callers	Estimated Monthly Callers w/ Email Addresses	Estimated Monthly Callers w/ Email Addresses Available After Exclusion Rules and Deduplication	Target MOE	Confidence	Minimum Monthly Responses Needed	Response Rates	Minimum Monthly Sample Needed
Getting an appointment that day	7,809	6,166	1,855	4.55%	95%	371	20%	1,855
Obtaining an RX refill thru Outpat. Pharmacy	144,115	81,526	36,818	3.00%	95%	1,058	20%	5,288
Evaluation of a RX renewal request	Currently Unavailable							
Nurse visit	156,387	102,694	62,810	3.00%	95%	1,070	20%	5,352
Issue with a medical device/equipment	671	607	441	9.34%	95%	88	20%	441
Walk-in vaccinations	15,513	9,377	2,378	4.02%	95%	476	20%	2,378
Traveling Veteran Coordinator	Currently Unavailable							
Scheduling a future appointment	366,585	207,044	134,076	3.00%	95%	1,080	20%	5,401

[1] The threshold for reaching the +/-3.0% target is currently 1,300 qualifying veterans per week.

Table 1B shows the estimated sample frame and minimum target sample size on a weekly basis. Minimum targets are rounded upward to assure the prescribed accuracy is achieved.

Table 1B shows the weekly sample availability and sample needs.

	Estimated Weekly Callers w/ Email Addresses Available After Exclusion Rules and Deduplication	Minimum weekly sample needed	Rounded weekly sample targets	Sampling Rate
Getting an appointment that day	427	427	427	100.0%
Obtaining an RX refill thru Outpatient Pharmacy	8,473	1,217	1,300	15.3%
Evaluation of a RX renewal request	Currently Unavailable			
Nurse visit	14,455	1,232	1,300	9.0%
Issue with a medical device/equipment	102	102	102	100%
Walk-in vaccinations	547	547	547	100%
Traveling Veteran Coordinator	Currently Unavailable			
Scheduling a future appointment	30,856	1,243	1,300	4.2%

The sample will be drawn using a systematic sampling methodology. This statistical valid approach allows the team to balance the sample across several variables such as age, gender, and district. These balancing variables are often referred to as implicit strata. Use of implicit strata has been proven to improve the accuracy of estimates, stabilize weights, and reduce the variability that make trends difficult to interpret.

Email addresses will be acquired by matching Veteran ID numbers to the VHA's Corporate Data Warehouse (CDW). Each email address encountered is validated in several ways:

- Validation that the email address has a valid structure
- Comparison with a database of bad domains
- Correction of common domain misspellings
- Comparison of a database of bad emails including
  - Opt outs
  - Email held by multiple veterans
- Comparison to a database of valid TDLs (e.g. ".com", ".edu")

C. Data Collection Methods

Invitations will be sent out each week to assure that initial invites are sent within nine days of their same day service interaction. Information for qualifying veterans will be extracted weekly from

VHA database resource: the VHA's Corporate Data Warehouse (CDW). Invitation will be sent on Tuesdays. Invitees that have not completed the survey will receive a reminder after one week. The survey will remain open for a total of two weeks. Survey responses are immediately available within VSignals as soon as feedback is submitted.

D. Reporting

Researchers will be able to use the Veteran Signals (VSignals) system for interactive reporting and data visualization. VA employees with a PIV card may access the system at https://va.voice.medallia.com/sso/va/. The scores may be viewed by demographic (e.g. Age Group, Gender, and Race/Ethnicity) in various charts for different perspective. They are also depicted within time series plots to investigate trends. Finally, filter options are available to assess scores at varying time periods and within the context of other collected variable information.

Recruitment is continuous but the results should be combined into a *monthly* data file for more precise estimates, at the interaction type level. Short interval estimates are less reliable for small domains, (i.e., VAMC-level) and should only be considered for aggregated populations. Monthly estimates will have larger sample sizes, and therefore higher reliability. Estimates over longer periods are the most precise but will take the greatest amount of time to obtain and are less dynamic in that trends and short-term fluctuation in service delivery may be missed. Users examining subpopulation should be particularly diligent in assuring that insights stem from analysis with sufficient sample in the subpopulations being examined or compared.

E. Quality Control

To ensure the prevention of errors and inconsistencies in the data and the analysis, quality control procedures will be instituted in several steps of the survey process. Records will undergo a cleaning during the population file creation. The quality control steps are as follows.

1. Records will be reviewed for missing sampling and weighting variable data. When records with missing data are discovered, they will be either excluded from the population file or put into separate strata upon discussion with subject matter experts.
2. Any duplicate records will be removed from the population file to both maintain the probabilities of selection and prevent the double sampling of the same Veteran.
3. Invalid emails will be removed.

The survey sample loading and administration processes will have quality control measures built into them.

1. The survey load process will be rigorously tested prior to the induction of the survey to ensure that sampled customers is not inadvertently dropped or sent multiple emails.
2. The email delivery process is monitored to ensure that bounce-back records will not hold up the email delivery process.

The weighting and data management quality control checks are as follows:

1. The sum of the weighted respondents will be compared to the overall population count to confirm that the records are being properly weighted. When the sum does not match the population count, weighting classes will be collapsed to correct this issue.

2. The unequal weighting effect will be used to identify potential issues in the weighting process. Large unequal weighting effects indicate a problem with the weighting classes, such as a record receiving a large weight to compensate for nonresponse or coverage bias.

F. Sample Weighting, Coverage Bias, and Non-Response Bias

Weighting is commonly applied in surveys to adjust for nonresponse bias and/or coverage bias. Nonresponse is defined as failure of selected persons in the sample to provide responses. This is observed virtually in all surveys, in that some groups are more or less prone to complete the survey. The nonresponse issue may cause some groups to be over- or under-represented. Coverage bias is another common survey problem in which certain groups of interest in the population are not included in the sampling frame. The reason that these Veterans cannot participate is because they cannot be contacted (no email address available). In both cases, the exclusion of these portions of Veterans from the survey contributes to the measurement error. The extent that the final survey estimates are skewed depends on the nature of the data collection processes within an individual line of business and the potential alignment between veteran sentiment and their likelihood to respond.

Survey practitioners recommend the use of sample weighting to improve inference on the population so that the final respondent sample more closely resembles the true population. It is likely that differential response rates may be observed across different age and gender groups. Weighting can help adjust for the demographic representation by assigning larger weights to underrepresented group and smaller weights to overrepresented group. Stratification can also be used to adjust for nonresponse by oversampling the subgroups with lower response rates. In both ways of adjustments, weighting may result in substantial correction in the final survey estimates when compared to direct estimates in the presence of non-negligible sample error.

Weights are updated live within the VSignals reporting platform[2]. Proportions are set based on the monthly distribution of the previous month.[3]

If we let w_{ij} denote the sample weight for the ith person in group j (j=1, 2, and 3), then the CW formula is:

$$w_{ij} = \frac{\% \, Veterans \in population \in group \; j}{¿ \, Veterans \in group \; j \in the \, sample}$$

As part of the weighting validation process, the weights of persons in an age and gender group are summed and verified that they match the universe estimates (i.e., population proportion). Additionally, we calculate the *unequal weighting effect*, or UWE (see Kish, 1992; Liu et al., 2002). This statistic is an indication of the amount of variation that may be expected due to the inclusion of weighting. The unequal weighting effect estimates the percent increase in the variance of the final estimate due to the presence of weights and is calculated as:

$$UWE = 1 + cv^2_{weights} = \left(\frac{s}{w}\right)^2$$

where

---

[2] Realtime weighting may cause some distortions at the beginning of each cycle due to empty cells or random variance in small sample distributions.

[3] Using previous months data is a design option for handling the problem of setting targets prior to fielding each month. An alternative design is to set targets off annualized estimates to create more stability month to month. If the population is known to fluctuate from month to month, past month population estimates may not be the optimal solution.

- **cv** = coefficient of variation for all weights w_{ij}.
- **s** = sample standard deviation of weights.
- $\overline{w}$ = sample mean of weights, $\overline{w} = \dfrac{1}{n}\sum_{ij} w_{ij}$.

G. Quarantine Rules

VEO seeks to limit contact with Veterans as much as possible, and only as needed to achieve measurement goals. Quarantine rules, therefore, are enacted to prevent excessive recruitment attempts upon Veterans. VEO also monitors Veteran participation within other surveys, to ensure Veterans do not experience survey fatigue. All VEO surveys offer options for respondents to opt out, and ensure they are no longer contacted for a specific survey.

Table 5. Proposed Quarantine Protocol

Quarantine Rule	Description	Elapsed Time
Past waves	Number of days between completing online survey any VEO survey and receiving another invitation.	30 Days
Active Waves	Number of days between receiving an invitation to a VEO survey and receiving another invitation.	14 Days
Opt Outs	Persons indicating their wish to opt out of either phone or online survey will no longer be contacted.	N/A

Part III – Assumptions and Limitations

A Veterans Only

At the onset of the Same Day Service Survey, email addresses are only available for Veterans and not other beneficiaries. Since Veteran attitudes may differ from those of non-Veterans, the exclusion of non-Veterans from the survey may contribute bias to the survey estimates.

B Coverage Bias due to Email-Only Data Collection

Since the Same Day Service Survey is email-only, there is a segment of the population qualifying veterans that cannot be reached by the survey. This will correspond to persons that lack access to the internet, and those who do not have an email address, or elect to not share their email address with the VA. Such beneficiaries may have different levels of general satisfaction with their service they received.

Appendix 1. List of Data Extraction Variables

Survey Variables
Survey Person ID

Agent ID
Date Time Call
Call Center
Phone Number
Coach
Full Name
Service Request Action
Caller Relation to Veteran
Has eBenefit Account
Credit Level
Call Type
Sub Type
NCC Start Date
Age
Gender
Period of Service
Veterans Email
Veteran ID # (MVI)

Appendix 2. Survey Questions

Appendix 3. References

Choi, N.G. & Dinitto, D.M. (2013). Internet Use Among Older Adults: Association with Health Needs, Psychological Capital, and Social Capital. *Journal of Medical Internet Research*, 15(5), e97

Kish, L. (1992). Weighting for unequal P. *Journal of Official Statistics*, 8(2), 183-200.

Lohr, S. (1999). *Sampling: Design and Analysis* (Ed.). Boston, MA: Cengage Learning.

Liu, J., Iannacchione, V., & Byron, M. (2002). Decomposing design effects for stratified sampling. *Proceedings of the American Statistical Association's Section on Survey Research Methods*.