COVID-19 Vaccine Survey
Sampling Methodology Report

Prepared by

Veteran Experience Office

Version 1.4

October 2020

Contents

Executive Summary **3**

Part I – Introduction **4**

A. Background 4

B. Basic Definitions 5

C. Application to Veterans Affairs 5

Part II – Methodology **5**

A. Target Population and Frame 5

B. Sample Size Determination 7

C. Data Collection Methods 10

D. Reporting 10

E. Quality Control 10

F. Quarantine Rules 11

Part III – Assumptions and Limitations **12**

Coverage Bias, and Non-Response Bias 12

Appendix 1. List of Data Extraction Variables 13

Appendix 2. Survey Questions 13

Appendix 3. References 14

Executive Summary

The COVID-19 Vaccine Survey is designed to measure the customer experience during the recruitement into the COVID-19 clinical trial and other VA Researches. VA understands that COVID-19 has significantly impacted the medium, frequency, and overall experience in which Veterans receive their healthcare. This surveys is created to understand how Veterans, ,family members, friends and other volunteers are recruited into clinical trial via various modalities such as newletters, posters, social media, targeted mails and emails, referral from a doctor or from a Veterans Service Officer (VSO) etc. Participants' experience data will be collected using an online transactional survey disseminated via an invitation email sent to selected beneficiary or via the VA.gov page questionaires (for a Landing Page Survey). Surveys invites will be sent twice per week. After the survey has been distributed, recipients have two weeks to complete the survey. Invitees will receive a reminder email after one week.

There are four surveys targeting the recruitment for Covid-19 Vaccine trial and other research.: (1) the first survey aims to understand the modality that lead the participant to the COVID-19 research landing page; (2) the second survey aims to measure the experience of participants from their interaction during the recruitment phase; (3) the third survey aims to measure the experience of participants in scheduling their appointment; (4) the fourth survey pertaints to the Study Team who enrolls volunteers at specific study sites.

The survey questionnaire is brief and contains general Likert-scale (a scale of 1-5 from Strongly Disagree to Strongly Agree) questions, checkbox and radio button questions, to assess customer experience. A qualitative study was conducted to analyze the question content and record reactions on a focus group consisting of Veterans. The sample will be distributed across the participants population defined further below.

This report describes the methodology used to conduct the COVID-19 Vaccine 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 **Veterans Experience Office** (VEO). The EMD team is tasked with conducting transactional surveys of the Veteran and Beneficiary 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 beneficiaries necessary to obtain reliable estimates. This information is subsequently used by internal stakeholders to monitor, evaluate, and improve beneficiary processes. Beneficiaries 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 beneficiary may be contacted, in order to prevent survey fatigue, across all VEO surveys.

In order to continue to provide quality service to Veterans, VEO has been commissioned to measure the satisfaction Veterans and Non Veterans, as it relates to their experience for enrolling into the the COVID-19 Vaccine research registry.

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 beneficiary interacts with the VA that impacts the beneficiary's journey and their perception of VA's effectiveness in caring for beneficiaries.
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 beneficiaries 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.

Part II – Methodology

A. Target Population and Frame

The target population of the COVID-19 Vaccine Survey is the participants who register to the volunteer list on the VA COVID-19 Research landing page for the first 3 surveys, subsequentely differentiated by other criteria; while the research admins constitute the population for the survey 4 (Study Team Survey). Survey 1 is targeting all individuals who

volunteerly register themselves on the Va.gov Covid Research website. Survey 2 is targeting the individuals who have declined to join a study after interacting with a research study team coordinator staff. Individuals who were found ineligible for a survey will be excluded from the is Survey 2 cohort. Survey 3 is targeting all the participants of the study who have scheduled an appointment. However, the survey 4 is targeting the research study team, which comprises the staff and the administrator within 3 Days of interacting with a study participant during their first appointment. See table 1 below.

Table 1. Surveys Population

	Survey	How	When	Response	Population
Survey 1	Landing page register volunteer list (assumption - survey will only be sent to people who registers on va.gov)	Va.gov	Immediately after interaction	Not Applicable	Everyone who registers on VA.gov
Survey 2	VA interaction– sign up into a research/don't sign up into a research	EMD-Email	3 days after interaction	2 weeks (one reminder after 7 days)	After volounteers register for the study, they are contacted with further information. People that do not want to move forward after being called will receive Survey 2 *(Exclude – ineligible and cohort that move forward)*
Survey 3	Appointment	EMD-Email	3 days after interaction	2 weeks (one reminder after 7 days)	Anyone who made an appointment
Survey 4	Study Team Survey	EMD-Email	3 days after interaction	2 weeks (one reminder after 7 days)	Study staff that interacted with the participants for the first Appointment.

A random sample will be drawn from the target population. The participants are the primary sampling unit and they are randomly selected from the population. Table 2 below depicts the recruitment mode and methods for each surveys.

Table 2. Survey Mode

Surveys	Recruitment Method	Time After Transaction	Recruitment Period
Landing Page Survey	Email / Social Media / Newsletter / Referral Recruitment	Within 3 days after registration	14 Days (Reminder after 7 Days)
VA interaction	Email	Within 3 days after interaction	14 Days (Reminder after 7 Days)
Appointment	Email	Within 3 days after interaction	14 Days (Reminder after 7 Days)
Study Team Survey	Email – TBD	Within 3 days after interaction	14 Days (Reminder after 7 Days)

B. Sample Size Determination

This survey aims to collect enough responses per month to begin to understand the Veterans or Beneficiaries and their families Experience during their recruitment into the COVID-19 Clincal Trial.

Since we do not currently have enough information on the target population, a sequential sampling technique can be implemented to attain the statistical significance. However, this will end up being more costly because if the required analysis between the sampling collection phases. A generic simple randomization sampling will be be adopted instead. 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

- $\boldsymbol{Z_{\alpha/2}}$ = 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.
    - 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 ECC survey the targeted margin of error is e = 0.03, or +/-3%.

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

- $\boldsymbol{n_0}$ = 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

- $\boldsymbol{Z_{\alpha/2}}$ = 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.

VEO's traditional approach of developing a probabilistic survey with a 90-95% confidence interval with a 3-5% MOE will be applied to this survey. Since we do not have enough information on the target population yet, we will be more lenient and proceed at 95% confidence interval with a 5% MOE.

Table 3 below shows the sampling numbers proposed; where the estimated response rate is suggested from past surveys and the voluntary enrolment into the trial. Survey 2 is expected to have a lower response rate because it is targeting the cohort that are declining to join the trial. We are making these recommendation because of the lack of specific information on the target population.

Table 3. Samples Recommendation:

Survey	Estimated Number in the Population	Estimated Monthly Response Needed	Estimated Response Rate	Monthly Invitation Sample Needed	Weekly Invitation Sample Needed
Survey 1	>3,000	350	~40%	875	219
Survey 2	>3000	350	~20%	1750	438
Survey 3	>3,000	350	~40%	875	219
Survey 4	<500	200	7%	2,858 or All Participating Staff	715 or All Participating Staff

C. Data Collection Methods

Recruitment occurs twice per week for all Surveys. Participants will have two weeks to complete the survey. A reminder email is sent after one week to non-respondents, to remind them that the survey is available for another week. Once the participants complete the survey, their response data is immediately available within Veterans Signals (VSignals).

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 and appropriate permissions may access the system at https://va.voice.va.gov/sso/va/pages/. The scores may be viewed by Age Group and Gender 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.

The consumption and dissemination of this data will be restricted to Authorized Users with the need to know; this will also be addressed in training and in the communication plan.

Recruitment is continuous but the results should be combined into a *monthly* data file for more precise estimates, at the call center level. Short interval estimates are less reliable for small domains, 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 are 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.

F. Quarantine Rules

VEO seeks to limit contact with Veterans as much as possible, and only as needed to achieve measurement goals. These rules 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
Repeated Sampling for the COVID-19 Vaccine Survey	Number of days between receiving one invite and receiving another for the COVID-19 Vaccione Surveys.	30 Days
Other Surveys	Veterans engaged that have recently completed other VEO surveys will not be selected for 30 days.	30 Days
Opt Outs	Persons indicating their wish to opt out of either phone or online survey will no longer be contacted.	Indefinite

Part III – Assumptions and Limitations

Coverage Bias, and Non-Response 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 beneficiaries cannot participate is because they cannot be contacted (no email address available). In both cases, the exclusion of these portions of beneficiaries 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 beneficiary 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 over-represented 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.

Appendix 1. List of Data Extraction Variables

Variables	Format
VE_MTM_KEY (provided by EMD)	Character
SURVEY_TYPE	Character
FIRST_NM	Character
LAST_NM	Character
Interaction with the VA DATEandTime	DateTime
DOB	Date
Age	Numeric
GENDER	Character
RACE/ ETHNICITY	Character
Phone	Numeric
EMAIL	Character
Zip	Numeric
State	Character
Appointment_Date (if Applicable)	DateTime
Flu shot (Yes/No)	Character
Flue Shot Date	Character
Covid (yes/no)	Character
Veteran (Y/N)	Character
SmokeorVape (Y/N)	Character
Employement Status	Character
Transportation (Y/N)	Character
Household size (with you included)	Numeric
CreateDateTime	DateTime
LatestTimeDate (if create date is blank, use this)	DateTime
receivedOtherC19Vaccine (Y/N) (Exclusion Criteria??)	Character
studyTeamEligibilityOutcome	Character
studyTeamEnrollmentStatus	Character
studyTeamName	Character
Other fields of interest	TBD

Appendix 2. Survey Questions

See insert




COVID-19 VACCINE TRIAL_VA.GOV_V5.A_TRIAL_V5.A_StudyTe COVID-19 VACCINE TRIAL_V5.A_1 COVID-19 VACCINE TRIAL_EMD 2_V5.A_1 COVID-19 VACCINE TRIAL_EMD 1_V5.A_1

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*.