



Appeals Management Office Longitudinal Survey Sampling Methodology Report

Prepared by

the

Veterans Experience Office

Version 2. December 12, 2019



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	6
C. Stratification	7
D. Data Collection Methods	7
E. Reporting	8
F. Quality Control	8
G. Sample Weighting, Coverage Bias, and Non-Response Bias	9
H. Quarantine Rules	10
Part III – Assumptions and Limitations.....	11
A. Coverage Bias	11
Appendix 1. References	12



Executive Summary

The Appeals Management Office (AMO) supports Veterans, their family members and survivors by developing policies and procedures to support the timely and accurate resolution of disagreements with Veterans Benefits Administration (VBA) decisions. The Veterans Appeals Improvement and Modernization Act of 2017 (Appeals Modernization Act), signed into law on August 23, 2017, is one of the most significant statutory changes to affect the Department of Veterans Affairs (VA) in decades. The Appeals Modernization Act (AMA) which went into effect on February 19, 2019, transforms VA's lengthy and complex legacy appeals process into one that is simple, timely, and fair to Veterans. The enhancements under the AMA are part of VA's continued commitment to improve the delivery of benefits and services to Veterans and their families. The law creates a new modernized decision review process that provides greater choice for Veterans in how they resolve disagreements with VA decisions and includes three decision review options:

- **Higher-Level Review:** A senior-level claims processor at a VA regional office will conduct a new look at a previous decision based on the evidence of record. A reviewer can overturn previous decisions based on a difference of opinion or return a decision for correction.
- **Supplemental Claim:** Veterans can submit new, relevant evidence to support their claim and claims processors at VA regional offices will assist in developing evidence.
- **Appeal:** Veterans will have the option to appeal their decisions directly to the Board of Veterans' Appeals (Board).

AMO made a commitment to Congress and the Government Accountability Office (GAO) to assess if Veterans and other beneficiaries have an improved decision review experience, how it compares to the legacy system, and how we can continue to improve the experience. VBA/AMO reporting requirements include:

- [GAO-18-352](#) VA should clearly articulate how it will “monitor and assess the new decision review process compared to the legacy process, including specifying a balanced set of goals and measures—such as...Veteran satisfaction...”
- [GAO-17-234](#) VA should “develop a strategy for assessing process reform...that ensures transparency in reporting to Congress and the public on the extent to which VA is improving Veterans' experiences with its disability appeals process.”

Without this initiative, AMO will not have the necessary information to inform process improvement and will not have the data to report to Congress and GAO. AMO has partnered with the Office of Strategic Initiatives and Collaboration (OSIC) and the Veterans Experience Office (VEO) to conduct human centered design research and use the research findings to develop three customer experience surveys around the request to file a decision review process and the two VBA managed options, the Higher-Level Review and Supplemental Claim. Veteran customer satisfaction with the Board Appeal option is being measured in a separate VSignals survey, sponsored by the Board of Veterans Appeals (Board).



The purpose of this document is to define VA’s sampling methodology for selecting potential survey respondents for this study. The sampling design intends to obtain an appropriate sample of Veterans to perform an end-to-end analysis of the Supplemental claim and Higher-Level review processes from the time a Veteran files a decision review to the time a veteran receives the VA decision regarding their decision review.



Part I – Introduction

A. Background

The **Enterprise Measurement and Design** team (EMD) in VEO is tasked with conducting transactional surveys of the Veteran population to measure their satisfaction with VA's numerous benefits and services. Thus, their mission is to empower Veterans by rapidly and discreetly collecting feedback on their interactions with such VA entities as the Veterans Health Administration (VHA), VBA, and the National Cemetery Administration (NCA). VEO surveys generally entail *probability* samples, which only contact the minimal number of Veterans necessary to obtain reliable estimates. This information is subsequently used by internal stakeholders to monitor, evaluate, and improve processes and procedures. Veterans are always able to decline participation and can 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.

Surveys issued by VEO are generally brief in nature and present a low respondent burden. Veterans or beneficiaries who have filed a request for a decision review will be contacted via email and invited to participate in a survey. A link will be enclosed so the survey may be completed using an online interface with customized Veteran information.

Most of the questions are based on human centered design (HCD) research and methodology and meets the Office of Management and Budget (OMB) A-11, Section 280 requirements of: Effectiveness/Quality, Confidence/Trust, Ease/Simplicity, Satisfaction, Equity/Transparency, Efficiency/Speed and Employee Helpfulness. The survey questions will focus on various aspects of the modernized decision review process; to include communication, filing the request for a decision review, deliberation, and receipt of the decision. Modern survey theory is used to create sample designs which are representative, statistically sound, and in accordance with OMB guidelines for federal surveys.



B. Basic Definitions

Attrition	The loss of sampled persons in a later survey period after they have responded in an earlier survey period in a longitudinal sample design.
Coverage	The percentage of the population of interest that is included in the sampling frame.
Longitudinal Sample	A sample design used to collect repeated observations within sampled persons over time.
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 <i>transaction</i> 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 <i>standard error</i> .

C. Application to Veterans Affairs

Customer experience and satisfaction are usually measured at three levels to 1) provide lines of business with the ability to track, monitor, and incentivize service quality; 2) provide service level monitoring and insights; and 3) direct point of service feedback. These measures will result in actionable data and inform process improvement activities within VBA.

Part II – Methodology

A. Target Population and Sampling Frame

The target population is all Veterans who have filed a request for decision review and received a decision from either of these reviews. There are two points in VBA's decision review process that will trigger a survey:

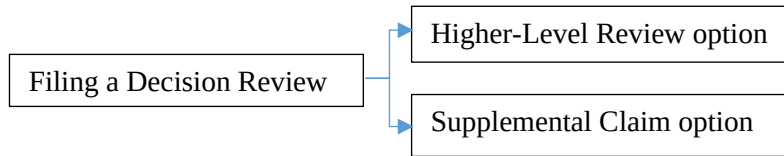
1. When the Veteran or beneficiary submits one of the below forms to VA and an end product (EP) is established.
 - a. VA Form 20-0995, *Decision Review Request: Supplemental Claim*, or
 - b. VA Form 20-0996, *Decision Review Request: Higher-Level Review*.



2. When a decision review EP is cleared:
 - 030 series – Higher-Level Review
 - 040 series –Supplemental Claim.

The sample frame contains all Veterans who have reported an email address to AMO and were recorded in Enterprise Data Warehouse (EDW) as having either filed for a review or obtained a decision on their review. The frame is created every week from data in EDW prior to sampling. For the Filing a Decision Review survey, the sampling frame includes Veterans who have filed for a Supplemental claim review or a Higher- Level review in the seven days prior to sampling. For the Decision survey, the sampling frame includes all Veterans who were sent a Filing a Decision Review notice seven to fourteen days prior to sampling¹. However, only Veterans who received a Filing survey are sent a Decision survey. Veterans who file for both a Higher-Level review and a Supplemental review in in a week will be randomly assigned to one type of review prior to sampling.²

Surveys will be sent seven days after each of the two points in VBA's decision review process. Veterans will be identified through a weekly extract from operational records stored in the VBA Enterprise Data Warehouse (EDW).



B. Sample Size Determination

The goal of the sample is to obtain end-to-end data from 300 Veterans on the Supplemental and Higher-Level review processes. Sampled Veterans receive both the Filing survey and the Decision survey to measure their experiences at the beginning of the review and the end. This type of sample design is a longitudinal sample design because multiple observations are collected from the same veterans over time. This design adds a layer of complexity in that not all veterans will provide both sets of measurements, a phenomenon known as attrition. Attrition includes nonresponse (a veteran does not respond to the Decision survey) and removal from the review process (i.e., death, halting the review process, etc.). The end effect of attrition is that fewer veterans respond to the Decision survey than the Filing survey. Consequently, the fielded sample size for the Filing survey must account for attrition in addition to the expected nonresponse for the Filed survey.

For a given margin of error and confidence level, the sample size is calculated as below (Lohr, 1999). For a population that is *large*, the equation below is used to yield a representative sample for proportions:

¹ Because the majority of decision notices are sent via mail, this timeframe was to allow Veterans time to receive their decision notice.

² One exceptional situation exists in cases in Higher-Level Review cases where a duty to assist error is found. In this situation, the Higher-Level Review end product will be cleared and a system-generated Supplemental Claim case established to address the error. When this happens, quarantine rules will keep the Veteran from being selected into the “Supplemental Claim” survey, thus AMO will be unable to obtain survey findings for Supplemental Claims in this category.



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

where

- $Z_{\alpha/2} = 1.95$, which is the critical Z score value under the normal distribution when using a 95% confidence level ($\alpha = 0.05$).
- 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 $p=0.5$, and this is often used for a conservative sample size (i.e., large enough for any proportion).
- e = the level of precision achieved with the sample. Also referred to as the margin of error (MOE).

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:

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

Where

- $Z_{\alpha/2} = 1.95$, which is the critical Z score value under the normal distribution when using a 95% confidence level ($\alpha = 0.05$).
- N = Population size.
- n = Representative sample.
- p = the estimated proportion of an attribute that is present in the population, with $q=1-p$.

Sample sizes will be targeted to achieve 300 respondents to the Decision survey. This sample size will allow for sustainable sampling from the population of veterans who have filed for a review each week by age and gender. The target sample size was expanded to 3,000 to account for nonresponse and attrition³. This sample will be drawn as a stratified systematic sample. Table 2 shows the breakout of the population eligible for sampling and the target and fielded sample sizes for each of the decision review options. Because the surveys are based on which option is chosen and the activity within that lane, this table only shows the impact of the sample on the level of precision. The MOEs were calculated for two different confidence intervals (CI) to determine the sample’s impact on the level of precision.

Table 2. Sample Targets by Activity Type

³ A response rate of 20% is used based on the usual response rates for similar VEO surveys. An attrition rate of 50% was used based on the response rates from veterans who have provided repeated responses to VEO surveys.



Activity Type	Weekly Population		Weekly Sample			Margin of Error			
	Median	Eligible for Sampling	Filed	Target for Filed	Target for Decision	95% CI		90% CI	
						Filed	Decision	Filed	Decision
HIGHER-LEVEL REVIEW	1,444	1,367	631	126	63	8.3%	12.0%	7.0%	10.1%
SUPPLEMENTAL CLAIM	6,268	5,131	2,369	474	237	4.3%	6.2%	3.6%	5.2%

C. Stratification

Stratification is used to ensure that the sample matches the population, to the extent possible, across sub-populations. The population will be stratified by type of review prior to sampling and sampled by this stratum. Additionally, the sample design will use veteran age and gender as implicit stratum to allow the age and gender distributions in the sample to match those of the population.

D. Data Collection Methods

The sample frame will be drawn from the EDW every Tuesday. The initial survey invites will be sent the following day. After 7 days a reminder invite will be sent, and the survey will close 14 days after the initial invitation. Veterans will receive Filing and Decision surveys for the type of appeal they were assigned to, unless they opted to use a different type of appeal to obtain a decision.

Table 3. Survey Mode

Mode of Data Collection	Recruitment Method	Time After Transaction	Recruitment Period
Online Survey	Email Recruitment	Within 7 days after encounter	14 Days (Reminder after 7 Days)

E. Reporting

AMO made a commitment to Congress and the Government Accountability Office (GAO) to assess if Veterans and other beneficiaries have an improved decision review experience, how it compares to the legacy system, and how we can continue to improve the experience. Based on the qualitative research that was conducted, AMO used the research to create CX surveys that will capture the overall experience of the new decision review processes and compare it to the legacy process. In addition, AMO will use the quantitative results to inform on future process improvements to processes and services provided to Veterans and their beneficiaries.

VBA will be able to use the VSignals platform for interactive reporting and data visualization. Scores can be observed for each of OMB's Section 280 A-11 drivers: Effectiveness/Quality, Confidence/Trust, Ease/Simplicity, Equity/Transparency, Efficiency/Speed and Satisfaction. The scores may also be viewed by age, gender, and race or ethnicity in various charts for other perspectives and are 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 (weekly) but the results from several weeks may be combined into a *monthly* estimate for more precise estimates, which is the recommended reporting level. Weekly estimates may include minor distortions but allow analysts to review scores more quickly and within smaller time intervals. Weekly 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 set to a 3% MOE at the 95% Confidence level. All estimates are also weighted in real time on the platform for improved representation and less bias (non-response and coverage, see section G on Sample Weighting) but the weights can introduce distortions when looking at short time windows. Quarterly estimates are the most precise, but will take the greatest amount of time to obtain (12 weeks of collection). However, Quarterly estimates are the most suitable for the analysis of small populations (e.g. Female Veterans 18-29, etc.).

F. 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. The quality control steps are as follows.

1. Population file creation
 - a. 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.
 - b. 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.
 - c. Invalid emails will be removed.
2. Loading and administration processes
 - a. The extracted sample will be reviewed for representativeness. A secondary review will be applied to the final respondent sample.
 - b. The survey load process will be rigorously tested prior to the induction of the AMO Surveys to ensure that sampled customers are not inadvertently dropped or sent multiple emails.
 - c. The email delivery process is monitored to ensure that bounce-back records will not delay the email delivery process.
3. Weighting and data management
 - a. 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.
 - b. 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.

1.

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

A final respondent sample should closely resemble the true population, in terms of the demographic distributions (e.g. age groups). One problem that arises in the survey collection process is **nonresponse**, which is defined as failure of selected persons in the sample to provide responses. This occurs in various degrees to *all* surveys, but the resulting estimates can be distorted when some groups are actually more or less prone to complete the survey. In many applications, younger people are less likely to participate than older persons. Another problem is **under-coverage**, which is the event that certain groups of interest in the population are not even included in the sampling frame. They cannot participate because they cannot be contacted: those without an email address will be excluded from sample frame. These two phenomena may cause some groups to be over- or under-represented. In such cases, when the respondent population does not match the true population, conclusions drawn from the survey data may not be reliable, and are said to be **biased**.

Survey practitioners recommend the use of sampling weighting to improve inference on the population. This will be introduced into the survey process as a tool that helps the respondent sample more closely represent the overall population. Weighting adjustments are commonly applied in surveys to correct for nonresponse bias and coverage bias. If a business rule is implemented that requires applicants to provide email address, the coverage bias for this survey would be expected to decrease. In many surveys, however, differential response rates may be observed across age groups. In the event that some age groups are more represented in the final respondent sample, the weighting application will yield somewhat smaller weights for this age group. Conversely, age groups that are underrepresented will receive larger weights. This phenomenon is termed *non-response bias correction* for a single variable. Strictly speaking, we can never know how non-respondents would have really answered the question, but the aforementioned adjustment calibrates the sample to resemble the full population – from the perspective of demographics. This may result in a substantial correction in the resulting weighting survey estimates when compared to direct estimates in the presence of non-negligible sample error (non-response bias).

Because the email population will have different demographics than the overall population, the initial sample will be selected from the frame so that the final respondent sample resembles the overall population. Stratification may also adjust for non-response (occurring when certain subpopulations are less prone to participate). Targets will be established for every permutation of the following stratification variables. As such, population values will be collected and recorded by VEO for every data collection period.

The stratification scheme above will result in a representative sample (with regard to the full population). Weighting will then be applied so that the sample is more fully matched to the population. Sample weights will be generated for Monthly, and Quarterly estimates.

Weighting will utilize subgroup weights in real time. To make this possible, targets will be based on the previous month's population. With each query on the VSignals platform for each respondent by dividing the target for a subgroup by the number of respondents in the subgroup. The weighting scheme will include, where possible, all the variables used for explicit stratification. However, cells will be collapsed if the proportion of the population is insufficient to reliably achieve a minimum of 3 completes per month. As a result, weights may be more comprehensive



for larger population segments. For instance, in the VA, women are a smaller proportion of the populations. Therefore, women will have more collapsed cells than men.

As part of the weighting validation process, the weights of persons in age and gender groups are summed and verified that they match the universe estimates (i.e., population totals). 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:

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

where

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

H. Quarantine Rules

VEO seeks to limit repeated contact with Veterans and only survey them 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 they 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 4. Quarantine Protocol

Quarantine Rule	Description	Elapsed Time
Sampling for Decision Survey	Number of days between receiving a Filing survey and a Decision survey	Varies but no earlier than 14 days
Repeated Sampling for Filing Survey	Number of days between receiving/completing the Decision survey, prior to receiving email invitation for a separate AMO Filing experience	30 Days
Other VEO Surveys	Number of days between receiving/completing online survey and becoming eligible for another VEO survey	30 Days
Prioritization	Prioritization is based on the observed sample sizes.	N/A
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. Coverage Bias

Since the VEO AMO Surveys are email only, there is a large portion of the Veteran population that cannot be reached by the survey. Veterans that lack access to the internet or do not use email may have different levels of trust and satisfaction with VA.. In order to verify whether there is a difference between Veterans who share their information and Veterans who do not share their information, VEO plans to execute a coverage bias study to assess the amount of potential coverage bias and derive adjustment factors in the presence of non-negligible bias.



Appendix 1. 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
- Kalton, G., & Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics*, 19(2), 81-97.
- Kish, L. (1992). Weighting for unequal P. *Journal of Official Statistics*, 8(2), 183-200.
- Kolenikov, S. (2014). Calibrating Survey Data Using Iterative Proportional Fitting (Raking). *The Stata Journal*, 14(1): 22–59.
- 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*.
- Wong, D.W.S. (1992) The Reliability of Using the Iterative Proportional Fitting Procedure. *The Professional Geographer*, 44 (3), 1992, pp. 340-348



Appendix 2. Survey Instruments.

