ASCEND for Veteran Suicide Prevention: Evaluating Sample Design and Response Characteristics of a National Veteran Survey Surveillance System

Evan R. Herring-Nathan[1], Christopher H. Johnson[1], Lindsey L. Monteith[2], Alexandra L. Schneider[2], Claire A. Hoffmire[2]

[1] NORC at the University of Chicago and [2] U.S. Department of Veterans Affairs, Rocky Mountain MIRECC for Veteran Suicide Prevention

Abstract

Veteran suicide rates are high relative to the general adult population, and suicide prevention is the top clinical priority for the Department of Veterans Affairs. Assessing Social and Community Environments with National Data (ASCEND) for Veteran Suicide Prevention aims to inform improved veteran suicide prevention initiatives through survey-based surveillance of non-fatal suicidal self-directed violence (i.e., suicidal ideation [SI] and suicide attempt [SA]). The ASCEND pilot study provided information necessary for optimal design of the first national survey wave, including with regards to the study sampling frame and design. The ASCEND pilot sampling frame (i.e., the USVETS database) was evaluated for its ability to reach targets for key demographic groups of interest (i.e., females, rural, and recently separated veterans). The quality of the sampling frame was examined for insight into coverage issues. Refinement of the sample design was considered, along with the anticipated effects of incorporating a secondary frame source to provide more complete coverage of the recently separated veteran population.

Keywords

Sample Design; Administrative Data; Complex Surveys; Stratified Sample

Disclaimer

This research is based on work supported by the Department of Veterans Affairs (VA) Office of Mental Health and Suicide Prevention and the VA Rocky Mountain MIRECC for Veteran Suicide Prevention, but does not necessarily represent the views or policy of the VA or the United States Government.

1. Introduction and Background

Suicide as the tenth leading cause of death in the United States and suicide rates continue to increase in the general population, according to the most recent data available, having increased 33% between 1999 and 2019 (Centers for Disease Control and Prevention, 2020). While U.S. Veterans remain at elevated risk for suicide compared to non-veteran adults, the suicide rate among veterans decreased 7.2% from 2018 to 2019 (U.S. Department of Veterans Affairs, 2021). Suicide prevention is a top clinical priority for the Department of Veteran Affairs (U.S. Department of Veterans Affairs, 2018). The Department of Veteran Affairs (VA) takes a public health approach to suicide prevention and is committed to improving suicide prevention for all Veterans, including those who do and do not use VA healthcare services.

Developing and targeting effective public health strategies to prevent suicide requires adequate surveillance of non-fatal suicidal self-directed violence (NF-SSDV; i.e., suicidal ideation [SI] and suicide attempt [SA]), which are associated with elevated risk for subsequent suicide. This is particularly important given that the lag in national mortality data precludes real-time tracking of trends in suicide rates overall and within populations at elevated risk. Longstanding surveys, such as the Behavioral Risk Factor Surveillance System (BRFSS) and National Survey on Drug Use and Health (NSDUH), offer nationally representative samples of the general U.S. population and can be used as survey-based surveillance systems for monitored conditions and behaviors. However, these surveys were not designed to collect data on a nationally representative sample of U.S. Veterans, nor do they include comprehensive assessment of NF-SSDV.

As such, a survey-based surveillance system of NF-SSDV among veterans would provide valuable, timely information to guide effective veteran suicide prevention efforts. Assessing Social and Community Environments with National Data (ASCEND) for Veteran Suicide Prevention was launched in 2018 to fill this critical need for targeted and comprehensive data on NF-SSDV among veterans. The ASCEND survey instrument was carefully constructed to produce reliable estimates for NF-SSDV; this instrument was used in the pilot study, which collected data from December 2020 – February 2021. We present here an evaluation of the ASCEND pilot sampling frame and corresponding sample design, which aimed to inform efficient approaches for sampling in the first full-scale national wave of ASCEND, which is expected to be implemented in 2022.

2. Sampling Frame and Design

A frame is a list or file of sampling units that represents the study population of interest. A comprehensive, high-quality frame is critical for a successful population-based survey. The file that comprises the frame should possess several desirable characteristics; chief among them are: (1) including as many cases of interest as possible to reduce the potential for selection bias; and (2) excluding cases that are not of interest, to reduce the cost of screening out ineligible individuals. Often, sampling frames are constructed from multiple pre-existing data sources.

To construct the ASCEND sampling frame, we began with the 2018 U.S. Veterans Eligibility Trends and Statistics (USVETS) Database. USVETS is managed by the VA Office of Enterprise Integration (OEI), National Center for Veterans Analysis and Statistics (U.S. Department of Veterans Affairs, 2020) and provides comprehensive data available

on all living US veterans by compiling information from various VA and Department of Defense (DoD) sources. Data sources which contribute to the USVETS file include, but are not limited to, the VA-DoD Identity Repository, Veterans Benefit Association (VBA) data, and Veterans Health Administration (VHA) administrative and healthcare data (Hauser., 2019). USVETS comprises more than 250 variables, including veterans' military history, demographics, socioeconomics, and utilization of VA benefits and services. Of particular relevance to ASCEND, USVETS includes information on individuals sufficient to classify them by state, urban vs. rural residence, and sex; these characteristics were used for stratification in the ASCEND pilot sample design.

The sampling frame consisted of over 18 million records whose information indicated that they were eligible for sampling. These records were believed to represent living veterans at the time of sampling using the most recently available USVETS data. Of these, a sample of 4,000 was selected, using a stratified random sample selection routine implemented in SAS (stratified on geography, sex, and rurality; described further below). Address updates were then conducted using Acxiom to ensure contact information was as up-to-date as possible prior to initiating recruitment.

For sampling design, sample sizes are driven primarily by determining the number of respondents belonging to subgroups for which precise estimates, and comparisons by those characteristics, are desired. The distribution of an unstratified sample would be expected to mirror that of the population covered by the frame, with sample proportions reflecting those of the population for various characteristics; if the proportion of the population with some characteristic is small, then the corresponding proportion of the sample is expected to be small.

Efficiency of estimation was a secondary consideration for the ASCEND survey, however, and it was desired that the sample design should allocate some minimum number to each stratum, for the sake of gaining broad experience during pilot data collection to inform the subsequent full-scale survey. Specifically, state-level estimates are ultimately of interest in the full-scale national survey, and states are known to vary greatly in their populations of civilians and veterans alike. An equal allocation across states would achieve precise estimates, but at a cost of inefficiency for the national sample as it would increase the design effect and thereby estimates' margins of error for most characteristics. Because stratifying the sample by 50 states for the pilot was not feasible, as an alternative, we grouped states by population size (number of veterans) to form strata with roughly equal population totals, using the quintiles of the ranked distribution. The first group was comprised of the states with the largest number of veterans, the fifth group was comprised of the states with the smallest number of veterans, and groups two, three, and four were intermediate. A sixth geographic group was composed of approximately 540,000 records with undetermined recency of state of residence, suggesting incomplete information in the database.

The sample design also considered sex and the urban/rural distribution of residence in allocating the sample by geographic group. Sampling strata were thus defined using the cross-classification of three dimensions; geography, sex, and rurality. The cross-classification of these three dimensions resulted in 22 strata. Because the sixth geographic group was not further broken out by rurality due to incomplete location information, a full 6 x 2 x 2 tabulation resulting in 24 total strata would include 2 empty cells.

Previous experience with similar surveys of veterans suggested that a 40% response rate (taking into account both successful contact and cooperation rates) might be achievable (Hoffmire et al., 2021). Unsure of the amount of intensive follow-up we would be able to employ as data collection proceeded, however, we more conservatively planned for a response rate of 12.5%, undifferentiated across the dimensions of sex, region, and rurality that comprised the design strata. As a further hedge against low response, we also planned and drew replicate samples – sometimes called reserve samples – to be released for data collection in the event of extreme shortfalls in the number of surveys completed. Reserve samples were not deployed.

Starting with a nonresponse-inflated initial sample of 4,000 to achieve an expected 500 completed surveys, we considered allocating the sample so as to achieve 100 survey participants in each of the five geographic strata and grouping unknown residence cases with the stratum of smallest states. We decided instead to allocate sample so as to achieve 50 completes to this sixth stratum, and 90 in each of the other five strata. When state was known but not urbanicity, we grouped the record with the rural stratum (the group to be oversampled). We then allocated the sample separately for male and female veterans, imposing constraints such that the sample was 25% female overall (compared to the 11% female distribution in the Veteran population/sampling frame) and, within sex, at least 20% rural for each geographic group (compared to the 11% rural distribution in the veteran population/sampling frame). Because of the interplay of these factors, the overall sample was approximately 30% rural overall (compared to a population proportion of approximately 21% overall, resulting in moderate oversampling of rural veterans).

3. Data Collection Protocol and Evaluation

The pilot recruitment approach was ten weeks long and consisted of an effective multi-mode strategy. Push-to-web mailings (pre-notification, invitation, and up to 3 reminders) were followed by weekly emails and telephone follow-up calls, when those contact details were available. While all 3,796 presumed veterans had a nominally valid mailing address, only 68% had a telephone number and an email address after the contact update process was complete. A paper and pencil questionnaire (PAPI, for "paper and pencil interview") was also provided near the middle of the data collection timeline (i.e., at week 6) for 50% of those who had not yet responded, providing a third and important mode of collection for veterans more comfortable with that response option.

4. Response Characteristics and Evaluation of Design

Overall sample yield for the ASCEND pilot was 14.9% (n=567), about 2 points higher than the conservative a priori estimate of 12.5%. Notable and significant differences in yield across veteran sub-populations were also observed. The initial sample of 4,000 veterans was expected to yield 500 completes. However, it was discovered during the early data collection phase that USVETS records with high rates of missingness for critical contact fields could not be located through our contact updating process. As such, the denominator for calculating yield included only the 3,796 selected Veterans with sufficient information in the sampling frame and upon updating contact information.

When comparing sub-populations of interest, yield differed by age group more than any other characteristic. Yield was lowest for the youngest veterans (18-34 years; 5.3%) and

highest for the oldest veterans (65+ years) at 22.2%. The association between age and yield was strong, significant, and somewhat expected.

Perhaps echoing the association observed with age, veterans who more recently separated from military service were less likely to respond to ASCEND (see Table 1). These veterans also tend to be younger.

The association between time since separation and response propensity is especially important because USVETS has known coverage issues for recently separated veterans. Veterans who have separated from the military within the last 1.5 years are generally not included on USVETS because of a lag in the administrative process to construct the USVETS database. To cover this gap, the VA-Department of Defense Identity Repository (VADIR), will be considered as a supplement to the USVETS frame for the full-scale national survey. Of note, including VADIR in the frame may lower overall yield due to lower response rates observed for younger Veterans.

Table 1. Yield by Recency of Military Separation

Recency Group	No. of Completes	N	Yield
Between 2 and 5 years	22	259	8.5%
Between 6 and 10 years	35	338	10.4%
Greater than 10 years	492	2,939	16.7%
No maximum separation date	18	260	6.9%
Total	567	3,796	14.9%

Conversely, nominal differences in yield were observed across the geographic strata. Importantly, however, we did observe that veterans with incomplete residence information (undetermined recency of state of residence) responded at about half the rate of veterans with more complete data. The proportion of veterans in this group is smaller than the other groups; however, they did comprise 6.8% of the sampling frame, and having an understanding of their response behavior benefits the efficient rollout of the larger scale study. Interestingly, yield trended higher in less populous states which also tend to be more rural.

Table 2. Yield by Geographic Strata

Geographic Strata in USVETS	No. of Completes	N	Yield
1 – Highest population states	104	706	14.7%
2...	100	708	14.1%
3...	111	706	15.7%
4...	114	710	16.1%
5 – Lowest population states	119	707	16.8%
6 – Undetermined recency of state of residence	19	259	7.3%
Total	567	3,796	14.9%

Research objectives for ASCEND include adequate representation of historically underrepresented groups, namely female and rural veterans. Therefore, realizing the sample targets for these groups is especially important. It is known that in many population surveys, females respond at higher rates than males (Porter and Whitcomb, 2005) and would theoretically be easier to engage despite comprising a much smaller portion of the veteran population than males. This was not observed for the ASCEND pilot, however. Female veterans responded at a slightly, but not significantly, lower rate than male veterans. These findings underscore the importance of oversampling female Veterans; while we sought to have 25% of the completes be female, we realized 22%.

Table 3. Yield by Sex

Sex	Population	Target Completes	Sample N	Actual Completes	Yield
Male	19,128,038 (89%)	373 (75%)	2,870 (76%)	443 (78%)	15.4%
Female	2,359,258 (11%)	128 (25%)	926 (24%)	124 (22%)	13.4%
Total	21,487,296	500	3,796	567	14.9%

5. USVETS Data Quality

As an administrative data source, USVETS should be approached with knowledge of its strengths and limitations. The quality of its data, in terms of completeness and accuracy, is generally sufficient for conducting this research, although it is not perfect. As noted above, missingness of critical characteristics, such as separation date and state of residence, is highly predictive of survey completion. This is not unexpected given that the level of completeness of data in an administrative data source can often be an indicator of the legitimacy and accuracy of the information for that record.

Table 4. Missingness in recency of separation field

Recency Group	No. of Completes	N	Yield
Has separation date	549	3,536	15.5%
No separation date	18	260	6.9%
Total	567	3,796	14.9%

Table 5. Presence of State of Residence Field

Geographic Strata in USVETS	No. of Completes	N	Yield
Has state of residence field	548	3,537	15.5%
Incomplete state of residence information	19	259	7.3%
Total	567	3,796	14.9%

It is also recognized that some records in USVETS may not correspond to veterans, but rather to spouses or dependents of veterans. In USVETS, records are assigned rankings denoting the degree of confidence that the record represents a veteran. While we did not remove observations with a lower likelihood of being veterans from the pilot sampling frame, in our experience, those observations were more likely to have missing or inconsistent data elements and thus lower yield, which may be explained in part by a lower likelihood of being eligible for the survey. Table 6 shows the distribution of the Veteran status confidence indicator in USVETS for the sample and the known ineligibility rate for the two classes as determined by actual survey responses.

Furthermore, records in the lowest probability level were very rare, but much more likely to be ineligible. However, only approximately 6.3% of veterans are not in the highest probability group (239 out of 3,796 in our selected sample) and the sample in this combined lower probability group generated a non-negligible number of completes (n=16). This is important because they represent records from less common data sources and, as such, represent a unique portion of the veteran population. Inclusion or exclusion of these records in sampling will have implications for coverage, bias, and the survey budget, which will need to be carefully weighed to determine the optimal approach for the full-scale national survey.

Table 6. Eligibility - Respondent-Reported Veteran status (i.e., eligibility) by USVETS Veteran status confidence indicator

Veteran Status	Ineligible	N	Ineligibility Rate
1 – Definite Veteran	57	3,557	1.6%
2 – Not Definite Veteran	16	239	6.7%
Total	73	3,796	1.9%

Table 7. Yield by USVETS Veteran status confidence indicator

Veteran Status	Completes	N	Yield
1 – Definite Veteran	551	3,557	15.5%
2 – Not Definite Veteran	16	239	6.7%
Total	567	3,796	14.9%

6. Weighting

Following data collection, analysis weights were developed to ensure unbiased, generalizable study results could be produced. Weighting took into account the stratified sample design which created variation in the design weights (or “base weights”). The strata were formed by the cross-classification of state group, sex, and rurality.

Even with targeted recruitment efforts, some under-representation by stratum was expected, but individual-level demographic characteristics available on the frame (characteristics known for respondents and nonrespondents alike) were available to develop eligibility and nonresponse weighting adjustments. For the eligibility adjustment, an adjustment factor was formed by considering the weight sums of sampled cases with known eligibility and for sample cases with unknown eligibility. The weighting adjustment was formed from collapsed stratification cells so that cell sizes would not be too small. Specifically, the second-most populous geography strata were grouped with their respective sex-strata within the most populous geography strata and the fourth-most populous geography strata were grouped with their respective sex-strata in the third-most populous geography strata. For example, rural females in the second-most populous geography stratum were grouped with rural females in the most populous geography stratum. Additionally, the undetermined recency of state of residence geography strata was grouped with rural records in the least populous geography strata.

Next, a non-response adjustment was performed using weighting cells formed by key demographic variables known to be associated with differential non-response and potentially suicide risk (Age Group, Race/Ethnicity, Recency of Separation from Military). Finally, the non-response-adjusted weights were raked to benchmark totals for

key dimensions (State Group, Gender, Rurality, and Veterans Health Administration usage).

So that standard errors of weighted results could be calculated accurately, variance estimation procedures are also recommended; stratum and cluster variables were created to be used in conjunction with analysis weights when describing the sample design to statistical software.

7. Conclusion

Pilot studies offer numerous benefits, one of which is refinement of the sampling plan and design. Developing and implementing a NF-SSDV survey-based surveillance system for veterans that comprehensively covers the population of interest, facilitates the collection of accurate and reliable data on sensitive topics, and does so in a fiscally sustainable manner is no small achievement. Testing elements of the process and evaluating results from a small-scale pilot allowed the ASCEND team to better understand how to navigate competing and sometimes prohibitively limiting constraints observed in full-scale operation. Importantly, we learned how VA administrative data (i.e., USVETS) can be used as a viable frame source to track NF-SSDV trends among veterans in key demographic groups: by geographical region, sex, and rurality. With this understanding, the team is well-positioned to implement a practical and efficient large-scale sampling process for ASCEND Wave 1 and gain critical information on a public health matter of the highest importance.

References

Centers for Disease Control and Prevention. Web-based Injury Statistics Query and Reporting System (WISQARS). (2020) Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.

Hauser, G. United States Veterans Eligibility Trends and Statistics (USVETS): A New Data Source with Socioeconomic Variables. VA HSR&D VIReC Database and Methods Seminar. May 6, 2019. Available at: https://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=3626

Hoffmire, Claire A et al. "Gender Differences in Lifetime Prevalence and Onset Timing of Suicidal Ideation and Suicide Attempt Among Post-9/11 Veterans and Nonveterans." Medical care vol. 59 (2021): S84-S91. doi:10.1097/MLR.0000000000001431.

Porter, S. R., & Whitcomb, M. E. (2005). Non-response in student surveys: The role of demographics, engagement and personality. *Research in higher education*, *46*(2), 127-152.

U.S. Department of Veterans Affairs, National Strategy for Preventing Veteran Suicide 2018–2028, Washington, D.C., September 2018. (https://www.mentalhealth.va.gov/suicide_prevention/docs/Office-of-Mental-Health-and-Suicide-Prevention-National-Strategy-for-Preventing-Veterans-Suicide.pdf)

U.S. Department of Veterans Affairs National Center for Veterans Analysis and Statistics. Available at: https://www.va.gov/vetdata/ Accessed: August 13, 2020.

U.S. Department of Veterans Affairs, 2021 National Veteran Suicide Prevention Annual Report, Washington, D.C., September 2021. (https://www.mentalhealth.va.gov/docs/data-sheets/2021/2021-National-Veteran-Suicide-Prevention-Annual-Report-FINAL-9-8-21.pdf)