B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

Section 334 of the law requires that the VA conduct a 20-year longitudinal study on three cohorts of Veterans who began a plan of services within the VR&E program during fiscal years 2010, 2012, and 2014. This request covers the first year of data collection for this VR&E Longitudinal Study, which will include the initial survey for the Cohorts I and II. Therefore, the respondent universe for Cohort I includes all Veterans who began a plan of services in the VR&E program in FY2010. The respondent universe for Cohort II includes all Veterans who began a plan of services in the VR&E program in FY2012.

The law also requires that the study include a statistically valid sample of cohort participants. Thus, a sufficient number of Veterans should be recruited for the survey such that at the end of the 20-year study period, 1,190 Veterans remain in each cohort sample. The cohort samples selected for the study are intended to represent the cohort populations of VR&E participants for their respective fiscal years.   However, the number of Veterans who began participating in the VR&E program in FY 2010 is 10,793. Table 4 presents the potential univariate domains of interest (DOI) that may

be used for stratified sampling when selecting the random sample of Cohort I survey participants.

Table 4.   Possible univariate domains of interest to be used for stratified sampling of Cohort I for the VR&E Longitudinal Study

Tabulation variable	Value	Population prevalence[*]	Smallest population prevalence
(none)	(entire population)	100.0%	100.0%
Employment handicap	Serious	67.8%	32.2%
	Not serious	32.2%	
Gender	Male	83.4%	16.6%
	Female	16.6%	
Age	< 30	14.8%	14.8%
	30-44	41.7%	
	45-54	27.4%	
	55+	16.5%	
Track selection	Re-employment, rapid access to employment, or self-employment	6.5%	5.6%
	Employment – long term services	80.5%	
	Independent living	5.6%	
	Not selected	7.4%	
Branch of service[**]	Army	54.2%	12.9%
	Navy	17.6%	
	Air Force	13.7%	
	Marine Corps	12.9%	

[*] For FY 2011 Longitudinal Study (n = 10,793)

[**]Excludes Coast Guard, Reserves/Guard, and other

In order to ensure that a statistically valid sample remains at the end of the 20-year study period, we anticipate completing 3,500 initial surveys for each cohort. The sample will be selected such that it is representative of the respondent universe.

The sample will be organized into a main group and a reserve group which will be further partitioned into multiple release groups within each sampling stratum. We will only contact the reserve sub-sample cases if response rates are less than assumed. For Cohort I, the respondent universe will be divided

into one initial sample of 5,000, with five additional reserve sub-samples of 1,000 each, and one additional subsample of 793. The initial sample and all sub-samples will be allocated proportionately to the final strata selected from those listed in Table 4. The initial sample of 5,000 program participants will be released in the first wave of recruitment and the overall targeted response rate for the first year of data collection covered by this request, based on releasing the first wave only, is 70%. The additional samples will be held in reserve and released as necessary if the targeted response rate is not achieved.

Although a response rate of 70% in the initial year is high relative to the rates achieved on other surveys of Veterans, it is not outside the bounds of experience especially on surveys of Veteran subpopulations and Veterans who are utilizing specific benefits. For example, a recent survey of Veterans with traumatic limb loss that used mail and telephone follow-up achieved a 62% response rate[1]. Similarly, a survey on sexual assault among Veterans applying for disability achieved a 68% response rate.[2] We will utilize the reserve samples if the 70% response rate is not achieved in the initial fielding wave.

2.   Information Collection Procedures

Describe the procedures for the collection of information including:
- Statistical methodology for stratification and sample selection,
- Estimation procedure,
- Degree of accuracy needed for the purpose described in the justification,
- Unusual problems requiring specialized sampling procedures, and

---

[1] Gayle E. Reiber, Lynne V. McFarland, Sharon Hubbard, Charles Maynard, David K. Blough, Jeffrey M. Gambel, and Douglas G . Smith. Servicemembers and veterans with major traumatic limb loss from Vietnam war and OIF/OEF conflicts: Survey methods, participants, and summary findings, *Journal of Rehabilitation and Research Development*, Volume 47 Number 4, 2010, Pages 299 — 316

[2] Maureen Murdoch, Prevalende of In-Service and Post-Service Sexual Assault among Combat ad Non-Combat Neterans Applying for Department of Veteran Affairs Post Traumatic Stress Disorder Disability Benefits, *Military Medicine,* May 2004.

- Any use of periodic (less frequent than annual) data collection cycles to reduce burden.

One can use the data from a longitudinal survey to calculate estimates of <u>level</u> at individual time points or estimates of <u>change</u> between pairs of time points. For the survey variables that are being collected in the longitudinal survey of Chapter 31 benefits, there will generally be a high correlation between data collected at two different time points. Consequently, it is expected that the survey's estimates of change will be more precise than its estimates of level. As a result, to develop the proposed sample design we specify target precisions for the estimates of level, with the expectation that associated estimates of change will generally be more precise than the estimates of level.

The study will recruit three cohorts of Veterans who began a plan of services with the VR&E program in FY 2010, FY 2012, and FY 2014, respectively. In FY 2010, 10,793 Veterans began a plan of services. The number of Veterans who begin a plan of services in a given fiscal year typically ranges between 10,000 and 11,000. Not all Veterans who are invited to participate in the study will agree to participate, nor will every Veteran participant complete each annual survey for the full 20 years. Hence, in addition to specifying target precisions, it is important to specify target response rates to ensure enough Veterans are recruited to result in a statistically valid sample with sufficient power.

Target response rate. Table 5 shows the impact that attrition may have on the responses over the 20 year study period. Ultimately we need to recruit a sufficient number of Veterans for the survey such that at the end of the 20-year study period, 1,190 Veterans remain in the sample. For a 20-year longitudinal survey, a stratum's response rate is equal to $r_c(1 - r_a)^{19}$, where r_c is the recruitment rate and r_a is the annual attrition rate. As

mentioned previously, the population size for VR&E Cohort I is 10,793. Table 5 displays for various values of r_c and r_a the expected number of completed cases for Cohort I after 20 years if the entire cohort is fielded for recruiting.

The shaded cells in Table 5 indicate combinations of the recruitment rate with respect to the entire cohort and the annual attrition rate that yield at least 1,190 completed cases after 20 years. If the annual attrition rate is no more than 5%, then the shaded cells in Table 5 indicate that the lowest the initial recruitment rate with respect to the entire cohort can be in order to yield 1,190 completes after 20 years is somewhere between 25% and 40%. In particular, if the initial recruitment rate with respect to the entire cohort is 32.43%, then the number of completes in the first year would be 10,973 x 0.3243 = 3,500, and the number of completes after 20 years if the attrition rate is no more than 5% would be at least 3,500 x 0.95^{19} =1,321. Thus, an initial sample resulting from recruiting 3,500 Veterans should result in a final sample of at least 1,190 Veterans at the end of 20 years as long as the annual attrition rate is no more than 5%.

Table 5.   Expected number of completed Cohort I cases after 20 years if all cases are fielded[*]

Initial recruitment rate	Annual attrition rate				
	1%	3%	5%	7.5%	10%
75%	6,688	5,004	3,055	1,840	1,093
50%	4,458	3,336	2,036	1,227	729
40%	3,567	2,669	1,629	982	583
25%	2,229	1,668	1,018	613	364
15%	1,338	1,001	611	368	219

[*]Shaded cells are equal to greater than 1,190

NOTE: The initial recruitment rates percentages shown are based on all 10,793 initial and reserve sample members being contacted.   A 75% initial recruitment rate means that (.75*10973=) 8075 respond to the survey in the initial year.

As we recruit and enroll Veterans into the study, we will need to monitor the recruitment and enrollment process to ensure we have recruited a sufficient

number of Veterans with specific characteristics that will allow us to compare sub-populations of interest at the end of the 20-year study period. For example, the agency may be interested in comparing the outcomes of male Veterans versus female Veterans. In order to make those comparisons, we need to be sure the sample includes a sufficient number of male and female Veterans.

Target precisions. A measure of the precision of an estimate is the half-width of the 95% confidence interval about the estimate, which is referred to as the estimate's margin of error (MOE). The relative size of MOEs by DOI varies inversely with DOI sample prevalence—that is, the proportion of the sample contained in the DOI. The DOI's with large sample-prevalence values will have small MOEs, and DOIs with small sample-prevalence values will have large MOEs. In an unstratified sample in which response rates are the same across the DOIs, sample prevalence equals population prevalence. Thus, in an unstratified sample, a DOI with a small population prevalence may have a large MOE. Column 3 of Table 6 contains the MOEs for DOI estimates of p = 50% for an overall completed sample size of n = 1,190. For example, for males the expected MOE is 3.2% and for females the MOE is 7.1% if the sample is unstratified and the response rates are the same for males and females. By oversampling females, the MOE for females can be decreased, but this will increase the MOE for males. Column 4 of Table 6 contains possible maximum MOEs for a stratified sample in which the overall completed sample size is n = 1,190. Whether or not a stratified sample can be designed so the resulting MOEs are less than or equal to the target MOEs in column 4 will be determined by calculations performed in the next step of the design process.

Table 6.   MOEs for DOI estimates of p = 50% for unstratified and stratified samples yielding 1,190 completed cases and equal response rates in each DOI

Tabulation variable	Value	Expected MOE for *unstratified* sample	Possible maximum MOE for *stratified* sample
(none)	(entire population)	2.9%	4.0%

Employment handicap	Serious	3.5%	5.0%
	Not serious	5.1%	5.0%
Gender	Male	3.2%	5.0%
	Female	7.1%	6.0%
Age	< 30	7.5%	6.5%
	30-44	4.5%	6.5%
	45-54	5.5%	6.5%
	55+	7.1%	6.5%
Track selection	Re-employment, rapid access to employment, or self-employment	11.4%[*]	6.5%
	Employment – long term services	3.2%[*]	5.0%
	Independent living	12.2%[*]	6.5%
Branch of service[**]	Army	3.9%	6.5%
	Navy	6.9%	6.5%
	Air Force	7.8%	6.5%
	Marine Corps	8.1%	6.5%

[*]Ignores the completed cases for Veterans for which the rehabilitation track was initially not selected
[**]Excludes Coast Guard, Reserves/Guard, and other

Next, one performs *sample allocation*, which is a mathematical process based on population sizes that determines the number of completed cases needed in each sampling stratum so that the expected MOEs for the defined DOIs are less than or equal to the target MOEs. Another quantity that is used in the allocation calculations is a DOI's *effective sample size*, which is the sample size of a hypothetical unstratified sample that has the same precision as the actual stratified sample of completed cases in a DOI, when individual stratum variances are equal. Effective sample sizes are less than the number of completed cases because of the loss in precision in multiple-stratum estimates when individual stratum variances are equal. Table 7 displays the effective sample sizes needed to obtain various MOEs for p = 50% estimates.

Table 7.   Effective sample sizes needed to obtain specified MOEs for p = 50% estimates

MOE	Needed effective sample size
3.0%	1,111
3.5%	816
4.0%	625
4.5%	494
5.0%	400
5.5%	331
6.0%	278

Table 8 combines the information in Table 6 and Table 7. The rightmost column of Table 8 contains the result of subtracting the sum of the effective sample sizes for a set of non-overlapping DOIs from the total number completed cases, 1,190. The values in the rightmost column of Table 8 indicate that it is quite likely that a stratified sample can be developed so that the MOEs of p = 50% estimates in the proposed DOIs are less than or equal to the target MOEs listed in Column 4 of Table 6 (and Column 4 of Table 8). Thus, we can be confident that if we need to field all 10,793 Veterans in Cohort I we will still obtain an initial sample of 3,500 Veterans, which will allow for at least 1,190 Veterans in the final sample at the end of the 20-year study period. We are also confident that the sample can be allocated such that at the end of the 20 years there will be sufficient power to estimate precision levels for specific sub-populations and detect statistically significant differences on outcomes of interest for those sub-populations.

Determination of the fielded sample size. One calculates the fielded sample size for each stratum by dividing the stratum's needed number of completed cases (determined by the allocation calculations) by the stratum's expected response rate in the stratum. If a stratum's fielded sample is larger than the stratum's population size, then all of the stratum's cases are fielded, the expected number of completed cases is set equal to the product of the stratum's population size and the stratum's expected response rate, and the allocation calculation are redone, which may require increasing some target MOEs or collapsing some DOIs.

Table 8.   MOEs for DOI estimates of p = 50% for unstratified and stratified samples yielding 1,190 completed cases and equal response rates in each DOI

Tabulation variable	Value	Possible maximum MOE for stratified sample	Needed effective sample size	Total completes, 1,190, less effective sample sizes
(none)	(entire population)	4.0%	625	565
Employment handicap	Serious	5.0%	400	390
	Not serious	5.0%	400	
Gender	Male	5.0%	400	512
	Female	6.0%	278	
Age	< 30	6.5%	237	242
	30-44	6.5%	237	
	45-54	6.5%	237	
	55+	6.5%	237	
Track selection	Re-employment, rapid access to employment, or self-employment	6.5%	237	316
	Employment – long term services	5.0%	400	
	Independent living	6.5%	237	
Branch of service[*]	Army	6.5%	237	242
	Navy	6.5%	237	
	Air Force	6.5%	237	
	Marine Corps	6.5%	237	

[*]Excludes Coast Guard, Reserves/Guard, and other

Summary of sample-design strategy. In summary, we propose the following sample-design strategy:

1. Initially assume that the expected number of completed cases after 20 years will be 1,190.
2. Define DOIs, specify target MOEs, and perform the associated stratification and allocation of the expected number of completed cases after 20 years.
3. Field all 10,793 cases for recruitment for Cohort I.
4. Use the recruitment results to calculate the actual recruitment rate for each defined stratum.
5. Assuming a realistic annual attrition rate -- e.g., 5% -- calculate the fielded sample size for each stratum.

6. If a stratum's calculated fielded sample size is less than the stratum's population size, determine how many of the stratum's cases do not need to be fielded.
7. If a stratum's calculated fielded sample size exceeds the stratum's population size, then field all of the stratum's cases for recruitment, set the stratum's expected number of completed cases equal to the product of the stratum's population size and the stratum's expected response rate, and then repeat steps 2, 6 and 7 until in no stratum does the calculated fielded sample size exceed the stratum's population size. This may require decreasing the expected number of completed cases after 20 years, increasing some target MOEs, and/or collapsing some DOIs.

We will refine the sample-design strategy for Cohorts II and III based on our experience with recruiting and enrolling the Cohort I sample.

3.   Maximizing Response Rates

Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

It is becoming more difficult to conduct telephone surveys as people move from traditional landline telephones and are less willing to accept calls on their cell phones. There is increasing evidence that web and paper questionnaires, using accepted contact procedures (e.g., followup mailings; Dillman, et al., 2009) can achieve response rates as high (or higher) than a telephone survey.

A web questionnaire, by itself, will not provide the coverage that a paper-mail or telephone survey will provide. There are many households that do not have ready access to the web and other that will not take a survey via the web. Westat's recent experience on the National Survey of Veterans found that only about 15 percent or 20 percent of Veterans were willing to complete the survey on the web (Cantor, et al., 2010). Consequently, it is

imperative for the VR&E Longitudinal Study to combine the use of the web with another mode, such as a telephone or paper survey, to maximize response rates and minimize respondent burden. Therefore, the VR&E Longitudinal Study will use a multi-modal data collection approach, combining web, mail, and telephone administration.

Literature on survey research suggests that response rates in fact decrease when respondents are presented with multiple administration options at the same time. Instead, a more effective strategy is to offer respondents one survey mode at a time, and then implement a different mode with non-responders. For the VR&E Longitudinal Study, survey participants will be first asked to complete the survey online via the Internet. The initial survey invitations to complete the survey online will be mailed and emailed to sampled Veterans. An email reminder for completing the online survey will be sent to non-responders at the beginning of the 2nd and 3rd weeks of the data collection field period. The invitation letter will be printed on VBA letterhead and signed by the Director of the VR&E program. A personalized, official letter of invitation from a high-ranking executive within the VA will make the study appear to be more credible to participants, and is likely to increase participation.

A paper-pencil survey will be mailed to those who do not respond via web. In addition to the survey itself, the mailed package will also include another personalized invitation letter, a pre-addressed stamped return envelope, and possibly a pen or pencil; and will be mailed to those who did not complete the web survey at the start of week 4 of the data collection field period. Two weeks after sending out the mail surveys, trained interviewers will begin calling those who did not respond to the web or mail survey. A similar data collection and contact protocol will be used for the annual followup surveys.

Because of the longitudinal design of the survey, it is important to maintain accurate contact information and keep participants engaged throughout the study period to gain their cooperation for completing the annual surveys. To achieve a high retention rate and minimize attrition, we propose to maintain contact with participants between data collection activities. Approximately 3 months after completing an initial or annual survey, we will email a note to all participants thanking them for their participation and reminding them to provide us with any updated contact information using the toll-free study number or the study-designated email address. In the next quarter, we will use various methods to send minimally burdensome reminders to all participants to send us updated contact information; including using automated methods such as email, interactive voice response, voicemail messages, and text messages. We will send the final reminder and request for updated contact information via mail in the quarter immediately preceding the next data collection effort

An incentive payment of $20.00 will also be provided to respondents to help maximize response and retention rates. Immediately upon receipt of a completed survey, we will mail a check for $20.00 to the participant's preferred address.

Even though intensive methods will be used to increase response rates and convert non-responders, non-response bias is still a concern. High participation rates are not sufficient to rule out possible biases due to non-response. Therefore, response rates by strata will be examined and if needed, we will adjust survey weights to account for non-response.

For each cohort, the goal will be to recruit 3,500 Veterans to participate in the first year of data collection. Recruitment will begin for each cohort by fielding 5,000 invitations to participate. If the recruitment rate with respect to this initial fielding is at least 70%, the goal of recruiting 3,500 Veterans

will be achieved. If necessary, however, additional recruitment samples will be fielded in order to yield 3,500 participants in the first year of data collection for each cohort.

 If the recruitment rate with respect to all the fielded recruitment samples is less than 80%, a non-response bias study will be conducted. This study will use the VA administrative data available for Veterans enrolled in the VR&E Program. One way VA administrative data will be used is to compare the respondents with non-respondents to the invitations to participate. Another way the administrative data will be used is to estimate the bias caused by nonresponse to the invitations to participate that is present in weighted estimates computed from the administrative data. It will be possible to estimate such biases because the VA administrative data is available for both respondents and nonrespondents to the invitations to participate.

4.   Testing of Instruments

Describe whether, and to what extent, the collection of information involves the use of automated, electronic, mechanical, or other technological collection techniques or other forms of information technology, e.g., permitting electronic submission of responses, and the basis for the decision for adopting their means of collection. Also describe any consideration if using information technology to reduce burden.

No testing of procedures or methods will be undertaken for completion of this survey. In the future, survey refinement will occur based on the results of conducting the initial survey to minimize burden and improve utility wherever possible.

5.   Individuals Consulted on Statistical Issues

Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit,

contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

VR&E has contracted with Westat for conducting the survey and analyzing the data for the study.

<u>Statistical Consultation</u>:

      Richard Sigman

      Senior Statistician

      Westat

      1600 Research Boulevard

      Rockville, MD 20850

      (240) 453-2783

<u>Data Collection and Analysis</u>:

      Mustafa Karakus

      Principal Investigator

      Westat

      1600 Research Boulevard

      Rockville, MD 20850

      (301) 294-2874