Health Study for a New Generation of U.S. Veterans
OMB 2900-0722   VA Form 10-0523

B.   COLLECTION OF INFORMATION EMPLOYING STATISTICAL   METHODS

1.   Provide a numerical estimate of the potential respondent universe and describe any sampling or other respondent selection method to be used. Data on the number of entities (e.g., households or persons) in the universe and the corresponding sample are to be provided in tabular format for the universe as a whole and for each strata. Indicate expected response rates. If this has been conducted previously include actual response rates achieved.

This is the second iteration of the study entitled "Health Surveillance for a New Generation of U.S. Veterans." It will involve a survey of 60,000 Veterans who can respond via a paper survey or online, and will include three mailing waves. This will be followed by a telephone survey of all Veteran non-respondents and medical records retrieval of approximately 2,000 Veterans.

The response rates for the first iteration of this study are listed below, including the paper/web survey, medical records follow-up, and telephone survey of non respondents.

	Paper/Web Survey	Medical Records Follow-Up (Respondents)	Telephone Survey (Non- Respondents)
Number Contacted	60,000	1,978	36,637
Response (Number)	20,700	570	1,426
Response (Percent)	34.5%	28.8%	3.9%

2.   Describe the procedures for the collection of information, including:

- **Statistical methodology for stratification and sample selection**

Populations
        A permanent panel was selected in 2009 at the onset of this study for follow-up over a ten year period. The 30,000 Operation Enduring Freedom / Operation Iraqi Freedom (OEF/OIF) Veterans who were separated or discharged from active duty by June 30, 2008 were identified from the deployment personnel roster that Department of Defense (DOD) Defense Manpower Data Center (DMDC) shares on a monthly basis with the Department of Veterans Affairs (VA). The non-deployed control group of 30,000 was assembled from the VA/DOD Identification Repository database (VADIR) containing a population of 1.1 million. Military and demographic data for both groups of Veterans were provided by DMDC, including date of birth, gender, race, service dates, location of deployment, mailing address, rank, unit component, branch of service, education, and marital status.

Stratification

Women were oversampled to equal 20% of the entire study sample, so that each subgroup in the stratified random sampling design was adequately represented. For both the deployed and non-deployed groups, there was a separate stratification by gender, unit component (Active, Reserve, and National Guard) and service branch (Air Force, Army, Marine Corps, Navy), resulting in a total number of 20 cells for the deployed OEF/OIF group and 20 for the non-deployed controls (Marine Corps and Navy do not have National Guard units). Eligibility criteria required year of birth to be restricted to 1985 and earlier. A comprehensive tabulation, including gender, unit component, and service branch, is available on Appendix I.

Samples

To create the deployed group in the sample, a total of 30,000 (24,000 male OEF/OIF Veterans and 6,000 female OEF/OIF Veterans) were selected in a stratified random sample design. OEF/OIF Veterans were selected from the available pool of 915,965 living OEF/OIF Veterans, narrowed down to 893,939 eligible deployed Veterans after removing those born after 1985 (see Table 1).

Table 1. Population of OEF/OIF Veterans eligible for selection of study sample as of June 30, 2008, stratified by gender and service branch, with year of birth 1985 and earlier.

Branch	Male	Female	Total
Air Force	145,055	24,824	169,879
Army	416,137	54,369	470,506
Marine	107,720	3,605	111,325
Navy	124,208	18,021	142,229
Total	793,120	100,819	893,939

To create the non-deployed group, non-deployed Veterans were selected from a file sent by VADIR of N=1,932,047 deployed and non-deployed Veterans. The composition of the non-deployed comparison group mirrors the OEF/OIF group with respect to gender, branch, and unit component. Again, a stratified random sample method was used to sample an equal number of non-OEF/OIF Veterans for each Veteran cell from the eligible pool of Veterans who had served between October 1, 2001 and June 30, 2008, and who were not deployed to OEF or OIF. A tabulation of the number of deployed and non-deployed Veterans selected for this study by gender and service branch is presented in Table 2.

Table 2. Study sample by OEF/OIF or non-deployed status, stratified by gender and service branch.

Branch	OEF/OIF			Comparison Veterans		
	Male	Female	Total	Male	Female	Total
Air Force	4,378	1,446	5,824	4,378	1,446	5,824
Army	12,918	3,459	16,377	12,918	3,459	16,377
Marine	3,214	183	3,397	3,214	183	3,397
Navy	3,490	912	4,402	3,490	912	4,402
Total	24,000	6,000	30,000	24,000	6,000	30,000

Power Planning

Power calculations have suggested that 30,000 OEF/OIF Veterans and 30,000 comparison group Veterans are adequate. These sample sizes are required because of small expected frequencies for some of the medical conditions among one or more of 20 Veteran strata.

Locating Addresses

To obtain the addresses of potential participants, files of the samples of 30,000 OEF/OIF and 30,000 control Veterans were prepared for processing through an interagency agreement with the National Institute for Occupational Safety and Health (NIOSH) for the Taxpayer Retrieval System, which enables us to obtain taxpayers' last known addresses. If an address of a Veteran obtained from DMDC at the time of separation from active duty was different from the IRS address, the IRS address was tried first. For those who were missing mailing addresses from both sources, one or more of the proprietary databases such as a credit bureau (Experian, Trans Union, Equifax), National Change of Address, or Telematch were searched for alternate mailing addresses. The process of finding the most current address will be repeated for future survey mailings.

Vital Status Ascertainment

For vital status ascertainment, researchers have access to the VA Beneficiary Identification and Records Locator System (BIRLS) file through the Austin Automation Center. The Social Security Administration (SSA), under the terms of an agreement, periodically sends a computer file of deceased individuals for whom the deaths were reported to SSA (Death Master File). Researchers will search these two national data sources and those who are recorded as deceased will be deleted from the sample before follow up.

- **Estimation procedure**

Statistical power for a study of a given sample size depends on the prevalence of specific conditions among the controls (non-OEF/OIF Veterans) and the relative risk of specific conditions which one considers as important to detect. The table below describes the sample size required for each group and the statistical power of the study under various conditions. Assuming, for example, that a condition is present among 5% of non-OEF/OIF Veterans and 7.5% of OEF/OIF Veterans (Relative Risk=1.5), to establish that this difference is true with 80% power (1-β) and 5% statistical significance (α) would require a sample of 1,469 Veterans in each of the two groups. Detection of differences in rarer conditions would require larger sample sizes or vice versa.

Sample Size Required for Each Group

RR	P=0.01		P=0.05		P=0.10	
	90%	80%	90%	80%	90%	80%
1.2	57,100	42,645	10,910	8,149	5,137	3,837
1.5	10,364	7,741	1,966	1,469	916	685
2.0	3,100	2,316	581	434	266	199
2.5	1,602	1,197	296	221	133	99
3.0	1,027	767	187	140	82	62

α = 0.05, two-sided test
RR = smallest relative risk detectable
P = the prevalence rate of disease in the controls
1-β = 90% and 80% statistical power

- **Degree of accuracy needed**

The results of the study will be presented in three different ways and discussed accordingly. First, all outcome data will be included in the analyses and reported as such.

Second, analyses will be based on self-reported questionnaire data but with appropriate adjustment for reporting errors. If the accuracy of self-reported data is found to be reasonable from the validation study (Kappa value above 0.4) and misclassification of the outcome value is not significantly biased, it is theoretically possible to correct for the effects of measurement error on the magnitude of the observed association.

In practice, however, correcting estimates in this way will be seriously limited by the absence of sensitivity and specificity data on many variables. Another more practical way to correct for measurement error will be used as proposed by Green (1983). He pointed out that when binary outcomes are truly present in only a small proportion of the population and when misclassification is non-differential, the only data needed to obtain excellent corrected value for the risk ratio is an estimate of the true proportion of those having the outcome of interest among those who reported having it.

Since it is much easier to validate a relatively small number of Veterans who reported having certain adverse outcomes than investigating the entire group, this method of adjustment will be used for most outcomes.

- **Unusual problems requiring specialized sampling procedures**

Women Veterans will be over-sampled to ensure that there will be adequate numbers in this study, to include 20 percent of the study sample.

- **Any use of less frequent than annual data collection to reduce burden**

  Data will be collected every three years.

3. **Describe methods to maximize response rate 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 population studied.**

The Dillman method, which incorporates four contacts by first class mail with an additional contact by telephone to increase the response rate for self-reported survey data collection, was the proven model for satisfactory response rates until proliferation of cell phone usage. We plan to adopt a mixed-mode of postal survey, Web-based survey, and Computer Assisted Telephone Interviewing (CATI) telephone survey methods.

As in the previous iteration of the survey, all Veterans will initially be sent an advance letter request for participation with an invitation to complete the survey on the Web. Then they will be sent a reminder letter. Next, Veterans will receive up to three mailing waves of envelopes containing paper questionnaires and the suggestion to compete the survey online, each followed by three thank you/reminder postcards. All Veterans who do not respond to the three waves of paper survey mailings nor to the Web survey by either survey submission or refusal will be contacted by telephone to complete the survey. As a check of accuracy and reliability, we will request medical records from approximately 2,000 respondents and compare the participants' responses to the reason for one health care visit to their medical records.

In the past iteration of the study, giving participants the option of completing the survey online helped to maximize response rates. The response rates for Web and paper were very similar (49.7% submitted via Web, 43.9% via the paper questionnaire, and 6.4% CATI), confirming that this population was willing and interested in adopting Web-based survey methods.

We are offering a $10 check to all potential participants in an effort to maximize response rates. In our pilot of this study, we saw that a pre-paid incentive made a difference. For the pilot study, a sample of 3,000 Veterans were assigned to one of three incentive groups: one group was pre-paid a $5 check with their survey; one group was promised a $5 check after survey completion; and one group received no incentive. The response rates were 25.2% for the pre-paid group, 22.3% for the promised group, and 16.9% for the no incentive group. Based on the pilot study, we decided to mail a $10 check with each initial survey packet for the main study to maximize response rates. We also decided to mail a $10 check to those that completed the CATI interview.

A field test of eight Veterans and the pilot study was also used to identify and minimize any potential problems with the survey content and administration. For the field test, the study team noted the time it took to complete the survey, evaluated the appropriateness of the survey responses, and asked participants if they had difficulty understanding any of the questions. We incorporated participants' feedback from the field test and pilot study into the final survey design.

Participation rates in epidemiological studies have declined dramatically in the past decade. Reasons for decreasing participation rates include (1) survey fatigue due to proliferation of research studies with resulting increased number of requests to potential subjects; (2) a decrease in volunteerism in the U.S. including willingness to participate in research studies; (3) relative importance of prospective study to one's own life; and (4) invasive demand on study subjects to participate in survey assessment, biologic sampling, requests for long-term follow-up, and lengthy consent forms written at inappropriately high reading levels (Galea and Tracy, 2007 Annals of Epidemiology). We have attended to these reasons for a drop in participation rates in the design of our studying in the following manner.
(1) Survey fatigue—Veterans are informed that participation will be requested every three years - a not too burdensome frequency. In addition, Veterans have the novel option to complete the survey online.
(2) Drop in volunteerism— To increase participation, a $10 incentive check is provided to all Veterans invited to complete the survey.
(3) Importance to one's own life—Participation can ultimately affect care provided or benefits received from VA. Veterans are told in an introductory letter that there are no direct benefits to participation, but are also told in this letter and the informed consent that participation will improve VA's understanding of the services that Veterans need.
(4) Demand on participants—

- Invasive demand – Veterans are informed in the introductory letter that they can skip questions which they consider sensitive, yet still continue to participate in the study;

- Biologic sampling—There will be no biologic sampling as a part of this study;

- Consent form written at inappropriate reading level— A simple, one-page consent form with appropriate reading level was designed and is in compliance with IRB requirements.

4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions of 10 or more individuals.

We will be comparing self-reports versus medical records for approximately 2,000 Veterans as a validation test.

5. 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.

Principal Investigator

Aaron Schneiderman, Ph.D., M.P.H., R.N.
Acting Director
Epidemiology Program
Office of Public Health
Department of Veterans Affairs
Washington, D.C. 20420
Tel: (202)266-4695

Co-Investigators

Han Kang, Dr.P.H.
Senior Scientist
Epidemiology Program
Office of Public Health
Department of Veterans Affairs
Washington, D.C. 20420
Tel: (202)266-4695

Clare M. Mahan, Ph.D.
Biostatistician/ Health Science Specialist
Epidemiology Program
Office of Public Health
Department of Veterans Affairs
Washington, D.C. 20420
Tel: (202)266-4695

Steven Coughlin, Ph.D.
Senior Epidemiologist
Epidemiology Program
Office of Public Health
Department of Veterans Affairs
Washington, D.C. 20420
Tel: (202)266-4695

Stephanie Eber, M.P.H.
Health Science Specialist
Epidemiology Program
Office of Public Health
Department of Veterans Affairs
Washington, D.C. 20420
Tel: (202)266-4695

Shannon Barth, M.P.H.
Statistician/ Health Science Specialist
Epidemiology Program
Office of Public Health
Department of Veterans Affairs
Washington, D.C. 20420
Tel: (202)266-4695

Matthew Reinhard, Psy.D.
Director
War Related Illness and Injury Study Center
50 Irving Street, N.W.
Washington, DC 20422
Tel: (202) 745-8249

Data Collection Support Contractor
To be selected.

Appendix I

Cell Sample Sizes
Deployed/Separated OEF/OIF Sample, by Gender, Unit Component, and Service Branch*

Service Branch	Air Force	Army	Marines	Navy	Total
Males n=24,000					
Unit - Active					
Frequency	67,502	161,275	81,026	101,706	411,509
Percent	16.4	39.19	19.69	24.72	100
Sample Size	1,575	3,762	1,890	2,373	9,600
Unit – Reserve					
Frequency	26,767	85,269	26,694	22,502	161,232
Percent	16.6	52.89	16.56	13.96	100
Sample Size	1,328	4,231	1,324	1,117	8,000
Unit - Guard					
Frequency	50,786	169,593	0	0	220,379
Percent	23.04	76.96	0	0	100
Sample Size	1,475	4,925	0	0	6,400
Total male sample	4,378	12,918	3,214	3,490	24,000
Eligible Population	145,055	416,137	107,720	124,208	793,120
Females n=6,000					
Unit - Active					
Frequency	14,129	22,823	2,949	14,707	54,608
Percent	25.87	41.79	5.4	26.93	100
Sample Size	621	1,003	130	646	2,400
Unit - Reserve					
Frequency	4,188	16,771	656	3,314	24,929
Percent	16.8	67.28	2.63	13.29	100
Sample Size	336	1,345	53	266	2,000
Unit - Guard					
Frequency	6,507	14,775	0	0	21,282
Percent	30.58	69.42	0	0	100
Sample Size	489	1,111	0	0	1,600
Total female sample	1,446	3,459	183	912	6,000
Eligible Population	24,824	54,369	3,605	18,021	100,819

*N=893,939
Stratified by gender, unit component, and branch
Year of birth 1985 and earlier

Cell Sample Sizes
Non-Deployed Sample, by Gender, Unit Component, and Service Branch *

Service Branch	Air Force	Army	Marines	Navy	Total

Males n=24,000					
Unit - Active					
Sample Size	1,575	3,762	1,890	2,373	9,600
Unit – Reserve					
Sample Size	1,328	4,231	1,324	1,117	8,000
Unit - Guard					
Sample Size	1,475	4,925	0	0	6,400
Total male sample	4,378	12,918	3,214	3,490	24,000
Females n=6,000					
Unit - Active					
Sample Size	621	1,003	130	646	2,400
Unit - Reserve					
Sample Size	336	1,345	53	266	2,000
Unit - Guard					
Sample Size	489	1,111	0	0	1,600
Total female sample	1,446	3,459	183	912	6,000

*Set to mirror cell sizes for deployed
  Stratified by gender, unit component, and branch
  Year of birth 1985 and earlier