B.   COLLECTIONS 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.

Phase I will include a mail survey of 60,000 veterans during a 19-month period that will include three mailing waves. This will be followed by a telephone survey of 2,000 veteran non-respondents and medical records retrieval of 1,000 veterans over a 26-month time period.

Except for the initial number of veterans to be contacted, these numbers are estimates. The final numbers will be determined by the response rate and the number of surveys returned undeliverable during each of the mailing waves in Phase I.

	Mail/Web Survey	Medical Records Follow-Up (Respondents)	Telephone Survey (Non- Respondents)
To Be Contacted	60,000	1,000	2,000
Expected Response (Number)	36,000	700	1600
Expected Response (Percent)	60%	70%	80%

Expected response rates are 60 percent, based on our prior experience with surveys among Gulf War veterans. However, there is a trend toward decreasing response rates to research surveys in the United States, so that the actual response rate may be lower, which may require offering a small incentive or enhancement to the mail package (see more detailed description in later section).

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

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

     All potential study subjects and comparison group veterans in this study will be identified by the Department of Defense (DOD) Defense Manpower Data Center

(DMDC). The OIF/OEF veterans who were separated or discharged from active duty by December 2007 will be identified from the deployment personnel roster that DMDC shares with VA on a monthly basis. A total of 30,000 (24,000 male OIF/OEF veterans and 6,000 female OIF/OEF veterans) will be randomly selected from the available pool of 710,470 male and 89,321female OIF/OEF veterans, respectively. Women veterans will be oversampled by the stratified random sampling method to ensure that each subgroup (gender by branch) such as "Female-Army" will be adequately represented in both the OIF/OEF and veteran comparison groups. Women represent about 11% of total OIF/OEF veterans who are potentially eligible for the study, and we will double their representation in the study to 20%. There is no need for a separate stratification by unit component because the Active Duty component and Reserve component (National Guards and Reservists) are almost equally distributed in the eligible pool (49% Active vs. 51% Reserve component).

Power calculations have suggested that 30,000 OIF/OEF veterans and 30,000 comparison group veterans should be adequate. These sample sizes are required because of small expected frequencies for some of the medical conditions among one or more of eight veteran strata (gender x branch). A breakdown of the number of OIF/OEF veterans to be included in this study by gender and service branch (Air Force, Army, Marine Corps, and Navy) can be seen in the table below. The composition of the non-OIF/OEF veteran comparison group will mirror the OIF/OEF study group with respect to gender and branch. Again, a stratified random sample method will be used to sample an equal number of non-OIF/OEF veterans for each veteran stratum from the available pool of veterans who were separated from active duty anytime between October 1, 2001 and December 31, 2007, and who were not deployed to OIF or OEF. Military and demographic data for both groups of veterans will be provided by DMDC: DOB, gender, race, service dates, location of deployment, mailing address, rank, unit component, branch of service, education, marital status, etc.

Files of the samples of 30,000 OIF/OEF and 30,000 control veterans will be 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 is different from the IRS address, the IRS address will be tried first. For those who are 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, Telematch, etc will be searched for alternate mailing addresses.

Vital status ascertainment

We have access to the VA BIRLS file through the Austin Automation Center and through our Target system on site. The Social Security Administration (SSA), under the terms of an agreement, periodically sends us a computer file of deceased individuals for whom the deaths were reported to SSA (Death Master File). We will search these two

national data sources and those who are recorded as deceased will be deleted from the sample.

Table 1. Population of OIF/OEF veterans available for selection of study sample as of 4th Quarter of Fiscal Year 2007

Branch	Male	Female	Total
Air Force	129,638	21,449	151,087
Army	378,268	49,466	427,734
Coast Guard*	717	60	777
Marine	96,337	3,143	99,480
Navy	105,510	15,203	120,713
Total	710,470	89,321	799,791

*Coast Guard personnel will be combined with the Navy personnel in the study sample.

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

	OIF/OEF			Comparison veterans		
Branch	Male	Female	Total	Male	Female	Total
Air Force	4,320	1,440	5,760	4,320	1,440	5,760
Army	12,720	3,300	16,020	12,720	3,300	16,020
Marine	3,360	240	3,600	3,360	240	3,600
Navy/Coast Guard	3,600	1,020	4,620	3,600	1,020	4,620
Total	24,000	6,000	30,000	24,000	6,000	30,000

- **Estimation procedure**

Statistical power for a study of a given sample size depends on the prevalence of specific conditions among the controls (non-OIF/OEF 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-OIF/OEF veterans and 7.5% of OIF/OEF 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

$\alpha = 0.05$, two-sided test
RR = smallest relative risk detectable
P = the prevalence rate of disease in the controls
$1-\beta$ = 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 CATI telephone survey methods.

Participation rates in epidemiological studies have declined dramatically over the past three decades. 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 shall attend to these reasons for a drop in participation rates in the design of our studying in the following manner.
  (1) Survey fatigue—Veterans will be informed that participation will be requested every three years - a not too burdensome frequency. In addition, veterans will have the novel option to complete the survey on-line.
  (2) Drop in volunteerism—A small incentive may be considered to increase participation.
  (3) Importance to one's own life—Participation can ultimately affect care provided or benefits received from VA. Veterans who screen positive for various conditions (such as depression or PTSD) according to the survey will be notified and given information on how to access VA healthcare.
  (4) Demand on participants—
      - Invasive demand – Veterans will be 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 in the baseline 2008 health survey;

- Consent form written at inappropriate reading level— A simple, short consent form with appropriate reading level will be designed which is still in compliance with IRB requirements.

The specific methods for this study are as follows: In the Wave 1 mailing, the researchers will send out pre-mailing introductory letters, initial questionnaires, and postcard reminders to the veterans. The letters will include a "do not forward address corrections to sender" request. The postcard reminders will be sent two weeks after the questionnaires. Subsequent questionnaire mailings (Waves 2 and 3) will require address location efforts. For those envelopes returned by the post office with a "forwarding order expired" sticker, the specific post office the in ZIP Code will be contacted by letter on VA letterhead to obtain the proper address (we were told that Federal agencies have this prerogative). The initial letter or questionnaire packet will be re-mailed within two weeks of receipt rather than being held for a later mass mailing. After Wave 1, an independent address location effort will take place for all non-respondents, to either confirm or update the addresses on file.

All participants, including those who have received a paper questionnaire, will have the option of answering the questionnaire online. The Web site will be built with a firewall and will be password-protected to ensure participant confidentiality. Maintenance of the Web-based survey will include procedures to disable or eliminate duplicate responses, aggregate data collected online with data collected from mailed responses, and allow for the identification of online respondents who completed the Web-based questionnaire.

Two-thousand non-respondents will be contacted to receive a full CATI health interview. One-thousand respondents will be contacted to retrieve information necessary for medical record retrieval. Permission will be obtained from veterans for the retrieval of medical records and medical records will be obtained from medical care providers for the contacted respondents and non-respondents.

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 1,000 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
Epidemiologist
Han K. Kang, Dr.P.H.
Director
Environmental Epidemiology Service
Department of Veterans Affairs
Washington, D.C. 20420
Tel: (202)254-0370

Co-Investigators

Biostatistician
Clare M. Mahan, Ph.D.
War-Related Illness and Injury Study Center
Veterans Affairs Medical Center
Washington DC, 20422
Tel: (202)254-0367

Epidemiologist
Elisa R. Braver, PhD
Associate Professor, Epidemiology and Preventive Medicine
National Study Center for Trauma & EMS
University of Maryland School of Medicine
701 W. Pratt St., Rm. 526
Baltimore, MD 21201
Tel: (410)328-7491

Data Collection Support Contractor
To be selected