AD_____

Award Number: DAMD17-98-1-8662

TITLE: Relationships of Stress Exposures to Health in Gulf War Veterans

PRINCIPAL INVESTIGATOR: John A. Fairbank, Ph.D.

CONTRACTING ORGANIZATION: Duke University Medical Center
Durham, NC 27710

REPORT DATE: October 2003

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 2003	Annual (1 Oct 2002 – 30 Sep 2003)

4. TITLE AND SUBTITLE

Relationships of Stress Exposures to Health in Gulf War Veterans

5. FUNDING NUMBERS

DAMD17-98-1-8662

6. AUTHOR(S)

John A. Fairbank, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Duke University Medical Center
Durham, NC 27710

E-Mail: jaf@psych.mc.duke.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

Research on Gulf War (GW) illnesses leaves many questions unanswered about diagnostic syndromes of GW illnesses, dimensions of stressor exposures encountered by GW veterans, relations among stressor exposures and GW syndromes or symptoms, and factors that may mediate these relationships. The proposed study has five key aims intended to address these gaps and enhance understanding of illnesses reported by GW veterans:
(1) to identify and examine dimensions of illnesses and health problems commonly reported by GW veterans;
(2) to assess exposures to environmental toxins and psychosocial stresses comprehensively and to identify the dimensions of these exposures;
(3) to examine the extent to which particular types of dimensions of exposure experienced during deployment and participation in Operation Desert Shield/Desert Storm (ODS/S) are associated with the commonly reported and undefined post-war health problems of GW veterans;
(4) to clarify how premilitary and predeployment adversities, risk factors, and protective factors affect GW illness outcomes;
(5) to examine the mediating role of post-ODS/S factors (e.g., instrumental and emotional social support, general psychological functioning, specific comorbid psychiatric disorders, such as posttraumatic stress disorder and depression) on GW illness outcomes.
These aims will be achieved by obtaining in-depth data from a national probability sample of GW veterans, including men and women in active and reserve components.

14. SUBJECT TERMS

NO SUBJECT TERM PROVIDED

15. NUMBER OF PAGES

38

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Tenth Anniversary Gulf War Veterans Health Study

Progress Report: 1 October 2002 – 30 September 2003

INTRODUCTION: Brief Narrative of Subject, Purpose, and Scope of Research

Study Aims. Research on Gulf War (GW) illnesses leaves many questions unanswered about diagnostic syndromes of GW illnesses, dimensions of stressor exposures encountered by GW veterans, relations among stressor exposures and GW syndromes or symptoms, and factors that may mediate these relationships. This study has five key aims intended to address these gaps and enhance understanding of illnesses reported by GW veterans. These aims are:

(1) to identify and examine dimensions of illnesses and health problems commonly reported by GW veterans;
(2) to assess exposures to environmental toxins and psychosocial stressors comprehensively and to identify the dimensions of these exposures;
(3) to examine the extent to which particular types of dimensions of exposure experienced during deployment and participation in Operation Desert Shield/Desert Storm (ODS/S) are associated with the commonly reported and undefined post-war health problems of GW veterans;
(4) to clarify how premilitary and predeployment adversities, risk factors, and protective factors affect GW illness outcomes;
(5) to examine the mediating role of post-ODS/S factors (e.g., instrumental and emotional social support, general psychological functioning, specific comorbid psychiatric disorders, such as posttraumatic stress disorder and depression) on GW illness outcomes.

Original Scope of Work. Originally, the aims of the Tenth Anniversary Gulf War Veterans Health Study were to be achieved using a two-phase study design. Phase 1 of the study was a mail survey of a national probability sample of GW veterans, including both men and women in active and reserve components. The goal of this survey was to provide comprehensive probability-based data about the dimensions of Gulf War illnesses, dimensions of exposure to environmental toxins and psychosocial stressors encountered by GW veterans, and relationships between these exposures and GW illness syndromes. These issues, as well as factors that may mediate relationships between stressor exposures and GW illnesses, were to be examined further in Phase 2 of the study, a computer-assisted telephone interview (CATI) conducted with a subsample of mail survey respondents with and without symptoms of GW illnesses.

Revised Scope of Work. A revised scope of work was executed in DAMD17-98-8662, Mod P0002 on 6/26/2001. The revised protocol involved:

(1) modification of the study's incentive structure and procedures to improve response rate; and
(2) incorporating into the mail survey the key aspects of the CATI telephone interview formerly referred to as Phase 2.

Despite efforts to enhance the response rate to the Tenth Anniversary Gulf War Veterans Health Survey using state-of-the-science methods, we received completed surveys from only about 5182 veterans for a response rate of approximately 54%. This response rate was higher than that typically achieved for mail surveys of military populations of comparable length (e.g., the 1998 Department

of Defense sponsored Total Forces study obtained a response rate of 38%; the Naval Health and Research Center's 1995 Perceptions of Wellness and Readiness Study obtained a response rate of 36%). Nonetheless, the scientific and public controversy regarding the nature and etiology of Gulf War illnesses is such that we felt that it was critical for us to be able to empirically address questions about the effects of nonresponse on findings from this study. Therefore, we conducted a nonresponse follow-up with a subgroup of the sample for the Tenth Anniversary Gulf War Veterans Health Study that was nonresponsive to the initial mail survey that enabled the computation of an overall weighted response rate of 70.5%.

PROGRESS REPORT (BODY OF REPORT)

This section describes research activities and accomplishments toward achieving the approved statement of work.

Data Collection and Related Activities.

Tenth Anniversary Gulf War Veterans Mail Survey

Data collection for the Tenth Anniversary Gulf War Veterans Health Survey was completed in March 2002. The survey was initially implemented as a mail survey in 2001. We achieved a response rate of 54.6 percent (using the AAPOR RR3 definition) after three mailings of the instrument, as well as a reminder post card, and a reminder telephone call.

Table 1. *Status of sample upon completion of mail survey*

Strata	Sample Size	Known Inelig	# of Elig Resps	Non Resps	Add. Est. # Eligs	Response Rate[1]
Active Males	4,488	160	2,099	2,229	2,045	50.6%
Active Females	2,312	151	1,115	1,046	915	54.9%
Reserve Males	2,310	67	1,282	961	908	58.5%
Reserve Females	1,191	39	686	466	440	60.9%
Males	6,798	227	3,381	3,190	2,954	53.4%
Females	3,503	190	1,801	1,512	1,355	57.1%
Active Duty	6,800	311	3,214	3,275	2,960	52.1%
Reserves	3,501	106	1,968	1,427	1,349	59.3%
CCEP	4,274	59	2,499	1,716	1,677	59.8%
Non CCEP	6,027	358	2,683	2,986	2,632	50.5%
Overall	**10,301**	**417**	**5,182**	**4,702**	**4,309**	**54.6%**

Gulf War Veterans Non-Response Follow-up Study

A key aspect of this research was the inclusion of a national probability sample of Gulf War veterans. Because of the lower than expected response rate to the Tenth Anniversary Gulf War

[1] The (unweighted) response rates are calculated using AAPOR RR3 formula.

Veterans Health Survey we felt that it was imperative to make every effort to determine the representativeness of the study sample. Therefore, after careful consideration, we decided to conduct a nonresponse follow-up with approximately 1000 sample members who did not respond to the initial mail survey.

Data collection for the Gulf War Non-Response Follow-up Study began on 6/12/02 and continued through 10/4/02. We had originally planned on an eight to ten week data collection period; however, locating accurate telephone numbers for nonresponse sample members proved to be more difficult than we had expected, resulting in delays due to the need for intensive tracing on an ongoing basis. Because of particular difficulties locating Active Duty sample members, we also requested an updated locator file from DMDC, after which we pursued new leads via telephone, US mail, and e-mail. Table 2 provides a summary of case status for the nonresponse sample at the end of data collection. According to these data, 55.5% of nonresponse sample members who were eligible for participation completed the follow-up CATI or a hard copy questionnaire.

Table 2. *Status of non-response sample upon completion of Gulf War Veterans Health Follow-up Study*

	Telephone Follow-Up					
Strata	**Sample Size**	**Known Inelig**	**# of Elig Resps**	**Nonresp**	**Add. Est. # Eligs**	**Response Rate[2]**
Active Males	472	8	239	225	215	52.7%
Active Females	218	6	113	99	91	55.4%
Reserve Males	208	8	114	86	77	59.6%
Reserve Females	102	2	61	39	37	62.2%
Males	680	16	353	311	292	54.7%
Females	320	8	174	138	128	57.6%
Active Duty	690	14	352	324	306	53.5%
Reserves	310	10	175	125	114	60.5%
CCEP	372	4	211	157	153	58.0%
Non CCEP	628	20	316	292	268	54.2%
Easy Follow-up	400	4	281	115	113	71.4%
Difficult Follow-up	600	20	246	334	307	44.4%
Overall	**1,000**	**24**	**527**	**449**	**420**	**55.5%**

[2] The (unweighted) response rates are calculated using AAPOR RR3 formula.

Effective Response Rate

A total of 5,709 eligible sample members responded to either the mail survey or the telephone follow-up. The overall weighted response rate[3] (a.k.a. the *effective response rate)* among eligible sample members for the combined mail survey and telephone follow-up was 70.5 percent with a 95 percent confidence interval of +/- 3.3 percent. The weighted response rate can be thought of as a population parameter. That is, it is the response rate that would be achieved if everyone on the sampling frame had been selected for the survey. The estimated weighted response rates for key subpopulations are shown in Table 3.

Table 3. *Weighted Response Rates*

	Mail Survey		Follow-up Survey		Overall	
Strata	**Rate**	**95% CI**	**Rate**	**95% CI**	**Rate**	**95% CI**
Active Males	43.1%	± 1.8%	46.6%	± 5.1%	69.3%	± 4.0%
Active Females	48.9%	± 3.3%	45.7%	± 8.0%	70.9%	± 6.1%
Reserve Males	51.8%	± 2.5%	51.7%	± 7.8%	75.8%	± 5.1%
Reserve Females	58.2%	± 3.4%	57.2%	± 9.0%	80%	± 6.5%
Males	44.4%	± 1.6%	48.1%	± 4.3%	70.3%	± 3.5%
Females	51.7%	± 2.6%	49.3%	± 6.2%	73.7%	± 4.7%
Active Duty	43.5%	± 1.7%	46.3%	± 4.3%	69.4%	± 3.8%
Reserves	52.6%	± 2.2%	53.5%	± 6.0%	76.3%	± 4.5%
CCEP	57.7%	± 1.2%	53%	± 5.4%	78.7%	± 3.3%
Not on CCEP	44.2%	± 1.6%	45.8%	± 4.5%	70%	± 3.5%
Easy Follow-up	*na*	*na*	71.2%	± 4.5%	*na*	*Na*
Hard Follow-up	*na*	*na*	42.1%	± 4.3%	*na*	*Na*
Overall	**44.9%**	**± 1.5%**	**48.5%**	**± 3.5%**	**70.5%**	**± 3.3%**

Data Cleaning/Editing

Intensive efforts have focused on cleaning data, combining information from the original mail survey (Tenth Anniversary Gulf War Veterans Health Survey) and the non-response telephone follow-up survey (Gulf War Veterans Follow-Up Survey), writing and implementing scoring algorithms for established scales, and developing created/summary variables resulting in the creation of an analysis dataset along with accompanying documentation describing study procedures and data editing decisions. Among the data cleaning activities carried out in this reporting period are:

[3] We calculated the weighted response rate as weighted proportions of a zero/one response indicator to each of the 10,301 sample members using design weights as the weights. Details on how weights were derived is found in Appendix A.

- Careful review of data associated with skip patterns. Some subjects did not appear to correctly follow the skip pattern for one item. Corrections were made after thoroughly reviewing the data associated with this item.

- Examination of missing responses. A review of patterns of missing data indicated that in some longer scales, respondents marked positive but not negative responses. Rules were developed for recoding missing items in these cases to negative responses.

  Information from DMDC data files was used to provide information when data were missing on date of birth, rank, branch of service, and type of unit (active duty/reserve/national guard).

- Reconciling inconsistencies between self-report data and DMDC data. In some instances there were inconsistencies between self-report data and DMDC data for gender, active versus reserve/national guard status, and branch of service. Resolving conflicting reports of gender involved comparing DMDC data with the respondents name and/or responses to female-only health questions. To resolve conflicting reports of duty status and branch of service, information was compared across multiple survey items that asked about duty status conflicts on this variable.

- Scoring algorithms were developed for all established measures that were used in the study (e.g. Hopkins Symptom Checklist-25; SF-36; Alcohol Use Disorders Identification Test). The corresponding program code for all scale and sub-scale scores has been written and variables have been created.

- Recoding rules were written to address cases where veterans' reports of the location in which they served in the Gulf were missing and or conflicted with reports of the area in which they spent the longest time.

Weight Derivation

Analysis weights appropriate for the complex survey design of the 10th Anniversary Gulf War Veterans Health Survey have been developed. Statisticians developed and implemented an innovative methodology for constructing the analysis; weights incorporated information from the 2-phase design to adjust for non-response bias and utilize a jackknife approach for computation of design-consistence variance estimates. Appendix A is a detailed report of the methodology for using this dual-frame estimation together with sampling weight calibration to develop composite weights.

Data Analysis

Based on a recent thorough review of the literature, an analytic plan for initial analyses has been developed to (1) determine whether patterns of symptoms and illnesses can identify meaningful sub-groups of Gulf War veterans and (2) examine the relationship of subgroup membership to demographic characteristics, measured exposures, and functional outcomes.

We plan to use non-hierarchical crisp clustering methods (e.g. k-means) to accomplish the first analytic task of identifying sub-groups. In order to test the stability and reproducibility of the identified subgroup structure, we will develop the initial clusters in a 'calibration' sample and then examine whether similar results are obtain when the same methods are applied to a 'validation'

sample. These calibration and validation samples will be obtained by splitting the total study sample into two-random sub-samples, balanced on sampling stratum to ensure equal numbers of male/females, registry/non-registry, active/guard & reserve personnel in each group. Generalized linear models (e.g. multinomial logistic regression) will be used to examine the relationship of subgroup membership with demographic factors (e.g. age, sex, rank, marital status, branch, unit), measured exposures (e.g. area served, receipt of vaccines, warzone stress), and functional outcomes (e.g. employment, physical impairment). Through the use of these models we hope to determine whether the subgroups identified in the cluster analysis are meaningful in terms of risk or prognosis factors.

Schedule Projections.

9/29/03 – Submitted request for no cost extension.

Activity	Months
Obtain study sample file from the Defense Manpower Data Center (DMDC; request for sample submitted on 12 March 2001)	30 completed
Apply for and obtain clearance for revision to study incentive procedures and protocol from the USAMRMC Human Subjects Committee and Contract Specialist.	30-31 completed
Apply for and obtain clearances for revisions to study incentive procedures and protocol from the DUMC and RTI IRBs	30-31 completed
Submit request for current addresses of study sample to National Institute of Occupational Safety and Health (NIOSH); obtain address file from NIOSH.	30-31 completed
Edit address file and send to National Computer Systems to format for mailing of survey and related materials.	32 completed
Conduct and complete survey data collection activities	33-38 completed
Plan and develop procedures for data editing and data analyses	33-38 completed
Edit survey data and develop analysis weights	39-48 completed
Construct analytic variables and conduct preliminary analyses.	48-60 completed
Conduct main analyses of survey data	56-66 in progress
Present study findings at scientific meetings and prepare article-length manuscripts.	66-72 in progress

KEY RESEARCH ACCOMPLISHMENTS

Main analyses are currently underway. We plan to develop and submit a manuscript to a peer-reviewed journal based on the results of this analysis. In addition, a manuscript will also be completed based on the development of the analysis weights for this study which employed novel methods for optimizing the use of non-response follow-up sample information in addressing issues of bias and precision.

REPORTABLE OUTCOMES

Preliminary estimates of the prevalence of certain health and mental health problems have been computed. These can be found in Appendix B.

APPENDICES

This report includes two appendices— Appendix A: 10th Anniversary Gulf War Veterans Health Survey: Weighting and Sampling Report; Appendix B: 10th Anniversary Gulf War Veterans Health Survey: Preliminary Estimates of Certain Physical and Mental Health Problems

APPENDIX A

10th Anniversary Gulf War Veterans Health Survey

Sampling and Weighting Report

PREPARED BY:
Vincent G. Iannacchione
Jill A. Dever
Avinash C. Singh
Shelton M. Jones

October 19, 2003

TABLE OF CONTENTS

1.0 Sampling Design

1.1 Target Population

The *target population* for the Tenth Anniversary Gulf War Veterans Health Survey is the estimated 689,183 men and women who served in all branches of the U.S. Armed Forces during the 1991 Persian Gulf War. The estimated distribution of the target population is shown for key attributes in **Table 1.**

The frame counts in **Table 1** are taken from the Defense Manpower Data Center (DMDC) Desert Shield/Storm File. This file contains a listing of all service personnel in the military during the time of the Kuwaiti Theatre of Operations – August 2, 1990 through July 31, 1991. Records for study-ineligible personnel such as Standby Reserves and Retired Reserves were excluded before determining the relevant counts. The counts for the "estimated number of eligibles" and the associated 95% confidence interval half-widths were calculated using the eligibility rates identified in the GWHS. Ineligibles are classified as individuals listed on the DMDC file who stated that they were never deployed to the Persian Gulf War.

Table 1. Estimated Distribution of the Target Population

Key Attribute	Frame Counts	Estimated # Eligible	95% CI[1]
Overall	689,183	627,883	± 40,876
Army	347,254	313,029	± 36,770
Navy + CG	155,911	144,755	± 27,763
Marines	102,876	93,678	± 22,366
Air Force	82,538	75,817	± 15,580
Enlisted	613,161	560,910	± 40,635
Officers	63,856	56,055	± 8,367
NH Black	156,555	142,933	± 24,723
NH White	468,368	422,771	± 32,808
Hispanic	40,961	40,106	± 11,588
Active Duty	578,616	526,868	± 38,532
AGR/TARs	61,515	58,448	± 10,888
Males	639,745	585,027	± 40,707
Females	49,438	42,856	± 3,704
CCEP	34,716	34,003	± 3,857
Not on CCEP	654,467	593,880	± 40,694
Active Male, CCEP	28,483	27,934	± 3,828
Active Female, CCEP	3,140	3,095	± 410
Active Male, nonCCEP	515,140	469,576	± 38,193
Active Female, nonCCEP	31,853	26,263	± 3,352
Reserve Male, CCEP	2,492	2,389	± 237
Reserve Female, CCEP	601	586	± 29
Reserve Male, nonCCEP	93,616	85,114	± 13,555
Reserve Female, nonCCEP	13,858	12,927	± 1,501

Source: Defense Manpower Data Center (DMDC) Desert Shield/Storm File.

1.2 Allocation and Selection of the Mail Survey

The primary objectives of the study are to provide national estimates of Gulf War veterans who report significant health concerns and to model the key correlates of those health concerns. Other objectives include comparisons between active-duty military and reservists, and the development of separate explanatory models for the occurrence of health concerns in male and female veterans. The objective of

[1] Confidence interval half-widths.

the sample design for this study was the selection of a probability sample of veterans from the target population of sufficient size to support these analytic objectives.

Our statistical power analysis indicated that a sample of approximately 850 active males, 450 active-duty females, 400 reserve males and 220 reserve females, all with health concerns, would be of sufficient size for most statistical comparisons. For example, a sample that includes 450 active female veterans with health concerns will detect differences between those with and without health concerns of 10 percent with 80 percent power at the five percent level of significance (two-tailed test). Given these requirements, the problem confronting the sample design was the identification of veterans in the population with health concerns.

Without prior knowledge of the health status of all veterans in the population, screening would be required to identify veterans with health concerns. If the screening were done at random, we estimated that an initial sample of approximately 17,440 veterans would be needed to achieve the desired number of veterans with health concerns (assuming a 15 percent overall incidence of health concerns and a 75 percent response rate). Instead, we over-sampled veterans who had registered with Department of Defense's Gulf War Comprehensive Clinical Evaluation Program (CCEP) and received a medical diagnosis (based on International Classification of Diseases, 9th Revision[2]) to obtain a sufficient number of veterans reporting significant health concerns with a smaller sample size. Our analysis indicated that approximately 78 percent of the CCEP registrants exhibit health concerns. As a result, veterans listed on the CCEP account for only five percent of the population of Gulf War veterans but account for approximately 26 percent of those with health concerns.

We capitalized on this disproportionate incidence of health concerns by sampling veterans listed on the CCEP at a much higher rate than other veterans. (We determined the actual sampling rate with a design optimization.) This approach allowed us to reduce the sample size needed to find the desired number of veterans to 10,301. **Table 2** shows that the use of the CCEP enabled us to increase the effective number of respondents with health concerns by between 20 percent and 35 percent depending on the reporting domain.

[2] CCEP registrants accounted for 34,506 (5 percent) of the veterans on the sampling frame. Approximately 78 percent of Gulf War veterans evaluated by the CCEP exhibited significant health concerns compared to an estimated 15 percent overall incidence rate among all Gulf War veterans. As a result, we estimate that veterans listed on the CCEP account for 26 percent of all Gulf War veterans exhibiting significant health concerns.

Table 2. Expected sample sizes for sample designs with and without the CCEP registry

Primary Stratum	Sample Design Without CCEP			Sample Design With CCEP		
	Initial Sample Size	Effective Respondents[4]		Initial Sample Size	Effective Respondents[4]	
		High SS[2]	Low SS[2]		High SS[2]	Low SS[2]
Active						
Males	4,488	673	3,815	4,488	852	2,514
Females	2,312	347	1,965	2,312	471	1,263
Reserve						
Males	2,310	347	1,963	2,310	416	1,317
Females	1,191	179	1,012	1,191	221	672
Total	**10,301**			**10,301**		

[1] Assumes random sampling within each stratum and a 15 percent incidence of health concerns.
[2] Sample allocation with CCEP maximizes the effective number of respondents with health concerns in each stratum.
[3] Assumes a 75% response rate.
[4] Number of respondents (assuming a 75% response rate) adjusted for design effects caused by over-sampling from the CCEP. Effective sample sizes are not additive across strata.

We selected a stratified systematic sample of 10,301 veterans from a sampling frame of 685,074 veterans. We defined eight sampling strata by subdividing veterans in each of the four primary strata into CCEP and non-CCEP categories. Prior to selection, we sorted the file by race/ethnicity within each stratum to implicitly stratify the sample distribution with respect to race/ethnicity.

1.3 Allocation and Selection of the Telephone Follow-up

The 54.4 percent response rate to the mail survey was 20 percentage points lower than expected. In an effort to reduce the potential bias associated with nonresponse to the mail survey, the project team decided to conduct a telephone follow-up of a sub-sample of nonrespondents to the mail survey. We based the follow-up sub-sample size of 1,000 mail non-respondents (about one-fifth of all mail respondents) on funding available to the study. The one-in-five sub-sampling fraction applied to the mail nonrespondents was less than optimal[3]. As a result, the reduction in bias obtained from the follow-up was adversely affected by the increase in sampling variance due to the increased variability in the sampling weights of the combined sample.

We allocated the follow-up sample inversely proportional to the mail response rates of each stratum. Prior to selection, each mail nonrespondent was classified as probable 'easy' or 'difficult' to contact based on whether an interviewer had made contact with someone in the veteran's household during calls

[3] In retrospect, an optimal sub-sampling fraction of x in x mail nonrespondents would have been needed assuming the actual response rate to the mail survey, the ratio of the cost of the telephone survey to the mail survey, and the estimated differences in the outcomes of mail and telephone respondents. See Cochran 1977, pps 367-371 for details.

made to prompt the return of the mail survey. Mail nonrespondents classified as 'easy to contact' were over-sampled to increase the expected effective sample size of the follow-up. To decrease response burden, the telephone follow-up obtained information on 69 of the 151 questions included in the mail survey. A total of 220 'hard refusals' to the mail survey and 264 'unable to locate' cases[4] were excluded from selection for the telephone follow-up.

2.0 Survey Response Rates

2.1 Response to the Mail Survey

The survey was initially implemented as a mail survey in 2001. We achieved a response rate of 54.6 percent (using the AAPOR RR3 definition) after three mailings of the instrument, as well as a reminder post card, and a reminder telephone call. As **Table 3** shows, the response rates to the mail survey were highest among females, reservists, and those who had been evaluated by the Gulf War CCEP.

Table 3. Response Rates to the Mail Survey

Strata	Sample Size	Known Inelig	# of Elig Resps	Non Resps	Add. Est. # Eligs	Response Rate[5]
Active Males	4,488	160	2,099	2,229	2,045	50.6%
Active Females	2,312	151	1,115	1,046	915	54.9%
Reserve Males	2,310	67	1,282	961	908	58.5%
Reserve Females	1,191	39	686	466	440	60.9%
Males	6,798	227	3,381	3,190	2,954	53.4%
Females	3,503	190	1,801	1,512	1,355	57.1%
Active Duty	6,800	311	3,214	3,275	2,960	52.1%
Reserves	3,501	106	1,968	1,427	1,349	59.3%
CCEP	4,274	59	2,499	1,716	1,677	59.8%
Non CCEP	6,027	358	2,683	2,986	2,632	50.5%
Overall	**10,301**	**417**	**5,182**	**4,702**	**4,309**	**54.6%**

[4] All 'unable to locate' cases received intensive tracing. The available information suggested that the majority were military personnel who were deployed overseas (e.g., in Afghanistan).
[5] The (unweighted) response rates are calculated using AAPOR RR3 formula.

2.2 Response to the Telephone Follow-Up

As **Table 4** shows, we achieved a 55.5 percent response rate (AAPOR RR3 definition) among the 1,000 mail nonrespondents selected for telephone follow-up. The response patterns for the follow-up were similar to the mail survey although the largest increase in response rate occurred among active-duty males not evaluated by the CCEP.

Table 4. Response Rates to the Telephone Follow-Up

	Telephone Follow-Up					
Strata	**Sample Size**	**Known Inelig**	**# of Elig Resps**	**Nonresp**	**Add. Est. # Eligs**	**Response Rate[6]**
Active Males	472	8	239	225	215	52.7%
Active Females	218	6	113	99	91	55.4%
Reserve Males	208	8	114	86	77	59.6%
Reserve Females	102	2	61	39	37	62.2%
Males	680	16	353	311	292	54.7%
Females	320	8	174	138	128	57.6%
Active Duty	690	14	352	324	306	53.5%
Reserves	310	10	175	125	114	60.5%
CCEP	372	4	211	157	153	58.0%
Non CCEP	628	20	316	292	268	54.2%
Easy Follow-up	400	4	281	115	113	71.4%
Difficult Follow-up	600	20	246	334	307	44.4%
Overall	**1,000**	**24**	**527**	**449**	**420**	**55.5%**

2.3 Weighted Response Rates

A total of 5,709 eligible sample members responded to either the mail survey or the telephone follow-up. The overall weighted response rate[7] (a.k.a. the *effective response rate)* among eligible sample members for the combined mail survey and telephone follow-up was 70.5 percent with a 95 percent confidence interval of +/- 3.3 percent. The weighted response rate can be thought of as a population parameter. That

[6] The (unweighted) response rates are calculated using AAPOR RR3 formula.

[7] We calculated the weighted response rate as weighted proportions of a zero/one response indicator to each of the 10,301 sample members using the design weights (defined as d_{kb} in Step 2 of weight calculations in the next section) as the weights. The standard errors of the response rates were calculated using the variance replicates described in Step 4 of the weight calculations.

is, it is the response rate that would be achieved if everyone on the sampling frame had been selected for the survey. The estimated weighted response rates for key subpopulations are shown in **Table 5.**

Table 5. Weighted Response Rates

	Mail Survey		Follow-up Survey		Overall	
Strata	**Rate**	**95% CI**	**Rate**	**95% CI**	**Rate**	**95% CI**
Active Males	43.1%	± 1.8%	46.6%	± 5.1%	69.3%	± 4.0%
Active Females	48.9%	± 3.3%	45.7%	± 8.0%	70.9%	± 6.1%
Reserve Males	51.8%	± 2.5%	51.7%	± 7.8%	75.8%	± 5.1%
Reserve Females	58.2%	± 3.4%	57.2%	± 9.0%	80%	± 6.5%
Males	44.4%	± 1.6%	48.1%	± 4.3%	70.3%	± 3.5%
Females	51.7%	± 2.6%	49.3%	± 6.2%	73.7%	± 4.7%
Active Duty	43.5%	± 1.7%	46.3%	± 4.3%	69.4%	± 3.8%
Reserves	52.6%	± 2.2%	53.5%	± 6.0%	76.3%	± 4.5%
CCEP	57.7%	± 1.2%	53%	± 5.4%	78.7%	± 3.3%
Not on CCEP	44.2%	± 1.6%	45.8%	± 4.5%	70%	± 3.5%
Easy Follow-up	*na*	*na*	71.2%	± 4.5%	*na*	*Na*
Hard Follow-up	*na*	*na*	42.1%	± 4.3%	*na*	*Na*
Overall	**44.9%**	**± 1.5%**	**48.5%**	**± 3.5%**	**70.5%**	**± 3.3%**

3.0 Analysis Weights

We assigned analysis weights to each of the 5,709 survey respondents to the mail survey and telephone follow-up. Analysis weights enable design-consistent estimation of population parameters by scaling the disproportionalities between the sample and the population. These weights may be viewed as inflation factors that account for the number of veterans in the target population that a sample member represents. The initial component of an analysis weight is the inverse of the selection probability that is specified by the sample design. Adjustments are then made to the weights to compensate for potential biases attributable to differential response and coverage among sample members based upon relevant characteristics of the sample members.

RTI statisticians developed and implemented an innovative methodology for constructing the analysis weights for the Gulf War Veterans Health Survey. In this section of the report, we summarize the development of the analysis weights, and demonstrate how the use of the weights improves the efficiency of the survey estimates. A detailed, theoretical treatment of the methodology will soon be available (Singh et al 2003).

3.1 Motivation for the Use of Dual-Frame Calibration

In surveys where initial response rates are low, a follow-up survey of nonrespondents may be used to reduce the nonresponse bias associated with survey estimates. However, when cost considerations require that the follow-up sample size be small, the reduction in bias obtained from the follow-up may be negated by the increase in sampling variance due to highly unequal selection probabilities in the combined sample. We developed and implemented a weight calibration method that can improve the efficiency of survey estimates while preserving the known population totals for key variables. Our method is motivated by analogy with small-area estimation which balances the variance of an unbiased but unstable design-based estimator with a biased but stable model-based estimator. We show that the ideas underlying dual-frame estimation together with sampling weight calibration can be used to develop composite weights to produce estimates that strike a balance between variance and bias.

Variability in the sampling weights often is attributable to extreme weights that result from differential sampling of sub-populations, adjustments for differential response rates, or adjustments for post-

stratification. Procedures that trim extreme weights[8] (Potter 1990) may be used to reduce the mean square errors (MSEs) associated with key survey estimates. The goal of weight trimming is to reduce sampling variance of an estimate more than enough to compensate for the possible increase in bias caused by the use of trimmed weights. This can be accomplished by ensuring known population controls are met while controlling for extreme weight values. This idea is incorporated in the generalized exponential model (GEM) (Folsom & Singh 2000), a general unified weight-calibration[9] model that controls for extreme weights by incorporating pre-specified upper- and lower-bounds (separately for initially identified extreme and non-extreme weights) into the adjustment factors made for nonresponse and post-stratification.

We used GEM calibration to develop efficient estimators for surveys with nonresponse follow-up while keeping bias under control. In addition, we borrowed ideas from composite estimation for dual-frame surveys to extract more information from the data. In our case, the composite estimator is a weighted combination of two correlated estimators:

1. The first estimator ignores the follow-up and uses only the mail (i.e., first-phase) sample (adjusted for unit nonresponse). This estimator is expected to be stable but potentially biased especially if the response rate is low. This can be considered as a quasi-model based estimator because modeling for nonresponse adjustment plays a major role.

2. The second estimator combines the first-phase sample with the follow-up (or second-phase) sample and is expected to be relatively unbiased but unstable, especially if the follow-up sample is small. Typically, another model is needed to adjust for nonresponse in the follow-up. This can be considered as a quasi-design based estimator because here modeling for nonresponse adjustment plays a minor role.

Our proposed calibration method combines these two estimators in a way that is similar to small-area estimation (SAE) in that it attempts to achieve a balance between bias and variance. However, unlike SAE, here the combined estimators are in the spirit of design-based estimates because all variance calculations are based on large sample considerations.

[8] Weight trimming (a.k.a. winsorization) involves the truncation of that part of a weight that exceeds a specified limit. The truncated portion of the weight is re-distributed to the untrimmed weights.
[9] Weight calibration is a term used for weight adjustments that meet specified constraints for weighted sample totals.

The two estimators described above can be viewed as estimates based on samples selected from two completely overlapping frames. We use the method of dual-frame calibration (DFC) introduced by Singh and Wu (1996, 2003) for combining independent estimates from two frames for our purpose with suitable modifications for dependence between samples. As a result, the calibration problem can be recast as a problem in dual-frame estimation except that we need to account for the dependence between the two samples.

We used a modified version of GEM to develop a single set of final calibrated weights that enable the combined estimator to be expressed as an expansion estimator, which, in turn, provides convenience in constructing estimates for any study variable. We provide details of the construction of the calibrated analysis weights in **Appendix A**.

3.2 Unequal Weighting Effects

The combination of a one-in-five sub-sampling rate for the telephone follow-up and a 55 percent response rate to the follow-up resulted in analysis weights for follow-up respondents that were approximately ten times as large as those for mail respondents. As a result, the reduction in bias obtained from the follow-up was adversely affected by the increase in sampling variance that resulted from the increased variability in the sampling weights of the combined sample.

Table 6 shows that, before DFC, the effective sample size actually decrease significantly from 1,672 to 535 when the follow-up respondents are included in the analysis. In other words, the precision associated with estimates based on the overall sample are larger than those based on only on the mail portion of the survey. After DFC, the effective sample sizes of the overall sample exceed those of the mail survey for every major reporting domain.

Table 6. Comparison of Unequal Weighting Effects (UWE) and Effective Sample Sizes (Eff. n's) Before and After Dual-Frame Calibration (DFC)

Domain	Respondents		Before DFC				After DFC			
			Mail-Only		Mail & FUs		Mail-Only		Mail & FU	
	Mail Only	Mail & Follow-Up	UWE	Eff. n's	UWE	Eff. n's	UWE	Eff. n's	UWE	Eff. n's
Overall	5,182	5,709	3.10	1,672	10.67	535	3.11	1,666	3.09	1,850
Active Duty	3,214	3,566	2.62	1,227	9.01	396	2.63	1,223	2.62	1,362
Reserves	1,968	2,143	2.32	848	7.50	286	2.33	845	2.33	919
Males	3,382	3,735	2.31	1,464	7.98	468	2.32	1,458	2.31	1,620
Females	1,800	1,974	1.76	1,021	6.51	303	1.77	1,014	1.79	1,100
Active Duty, Males	2,100	2,339	1.91	1,101	6.58	355	1.91	1,097	1.91	1,223
Active Duty, Females	1,114	1,227	1.72	649	6.60	186	1.73	645	1.75	700
Reserve, Males	1,282	1,396	1.93	666	6.26	223	1.93	664	1.94	721
Reserve, Females	686	747	1.79	383	5.81	129	1.80	381	1.82	411

3.3 Effects of Dual-Frame Calibration on Survey Estimates

Table 7 presents survey estimates and corresponding sampling errors before and after dual-frame calibration. Two sets of survey outcomes are presented. The first set includes the ten key outcomes that comprise the *z*-vector of 'zero controls.' After DFC, the difference between these estimates for these variables using w_{KA} and w_{kB} is zero. Note that equal variances would be obtained for each sample estimator if the variances are adjusted for calibration through the use of Taylor-series linearization or a replication method. The second set of 'other' outcomes illustrates the effects of the DFC procedure on outcomes that are not explicitly part of the calibration procedure. For these outcomes, the DFC estimator is the composite of the w_{KA} and w_{kB} estimates using the combining factor of $\xi_A = 0.82$.

Table 7. Effects of the DFC Procedure on Survey Outcomes

Survey Outcomes	Before DFC					After DFC					
	Sample A w^1_{kA}		Sample B w^1_{kB}			Sample A w_{kA}		Sample B w_{kB}		DFC w^*_k	
	Mean	SE	Mean	SE	Diff	Mean	SE	Mean	SE	Mean	SE
Key Outcomes (z- variables):											
% w/CDC Mulitsymptom Illness	68	1.16	63.6	1.91	4.5 *	67.3	1.05	67.3	1.05	67.3	1.05
% w/PTSD Indicator	7.6	0.63	8.5	1.17	-0.9	7.8	0.59	7.8	0.59	7.8	0.59
Chronic Fatigue Indicator	10.3	0.64	11.7	1.49	-1.3	10.5	0.65	10.5	0.65	10.5	0.65
13-Item Chalder Fatigue Score	60.5	1.26	57.9	2.05	2.6	60.1	1.16	60.1	1.16	60.1	1.16
% w/Sexual Discomfort	10.1	0.77	9.5	1.44	0.5	10.1	0.75	10.1	0.75	10.1	0.75
% Current Smoker	24.3	1.16	28.6	1.93	-4.3 *	25	1.05	25	1.05	25	1.05
% Drinking (Mult X/Week)	31.3	1.20	32.1	2.05	-0.8	31.4	1.10	31.4	1.10	31.4	1.10
% Married/Cohabiting	76.4	0.96	73.8	1.94	2.6 *	76.1	0.92	76.1	0.92	76.1	0.92
Depression Subscale (Continuous)	1.7	0.02	1.7	0.03	0	1.7	0.01	1.7	0.01	1.7	0.01
Physical Impairment Score (~Continuous)	7	0.04	7	0.07	0.1	7	0.04	7	0.04	7	0.04
Other Outcomes:											
% w/General Health Status=Excellent	8.8	0.63	11.7	1.53	-3	8.8	0.60	11.3	1.50	9.2	0.61
% w/General Health Status=Very Good	29.4	1.14	27.1	1.74	2.3	29.5	1.11	26.8	1.77	29	1.05
% w/General Health Status=Good	38.5	1.21	39.2	2.31	-0.7	38.3	1.20	40.7	2.18	38.7	1.18
% w/General Health Status=Fair	20	1.13	18.4	1.69	1.5	20	1.09	17.9	1.50	19.7	1.05
% w/General Health Status=Poor	3.3	0.42	3.5	1.07	-0.2	3.4	0.44	3.3	0.96	3.4	0.42
% w/Reduce Time Spent on Work	16.5	0.92	18.6	1.70	-2.1	16.7	0.96	17.2	1.19	16.8	0.93
% w/Limit Type of Activities	22.9	1.14	24.5	2.09	-1.6 *	23.1	1.11	23	1.37	23.1	1.10
% w/Loss of Interest in Usual Activities	57.7	1.27	61.5	1.93	-3.8 *	57.7	1.19	61.3	1.73	58.4	1.18

* Difference significant at 0.05.

4.0 Variance Estimation

For variance estimation, we created 294 variance replicates (a.k.a. random groups) that enable us to combine the data obtained from the mail survey with that obtained from the telephone follow-up and then estimate the variances of survey outcomes in a design-consistent fashion. Within each of the eight first-phase strata, we randomly assigned 35 sample members to each replicate with the requirement that each replicate have approximately equal numbers of mail respondents and at least one follow-up respondent.

The primary advantage of random groups is that standard survey software packages (e.g., SUDAAN®) can be used to analyze the data. In fact, the variance estimates that we obtained for outcomes for mail respondents (i.e., excluding the follow-up) using random groups are only slightly conservative compared to the usual variances obtained for (single-phase) stratified designs. For two-phase variance estimation, we calculated 294 sets of replicate weights for use with the 'delete one' Jackknife method of variance estimation (Lohr 1999, p 298). We constructed each set of replicate weights by serially deleting one replicate from the sample and then adjusting the DFC weights to account for the deleted replicate.

An example of Jackknife variance estimation in SUDAAN® is presented in **Appendix B**.

5.0 References

American Association for Public Opinion Research. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 2000.

Folsom, R.E. Jr. and Singh, A.C. (2000). "A Generalized Exponential Model for Sampling Weight Calibration for a Unified Approach to Nonresponse, Post-stratification, and Extreme Weight Adjustments." *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 598-603.

Frankel, Lester R., "The Report of the CASRO Task Force on Response Rates," *in Improving Data Quality in a Sample Survey*, edited by Frederick Wiseman. Cambridge, MA: Marketing Science Institute, 1983.

Lohr, S.L. (1999). *Sampling: Design and Analysis*, Duxbury Press, New York.

Potter, F. (1990). "A Study of Procedures to Identify and Trim Extreme Sampling Weights." *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 225-230.

Research Triangle Institute (2002). *SUDAAN User's Manual, Release 8.0*. Research Triangle Park, NC: Research Triangle Institute.

Singh, A.C. and S. Wu (2003). "An Extension of Generalized Regression Estimator for Dual-Frame Surveys." *Proceedings of the American Statistical Association, Section on Survey Research Methods* (in print).

Singh, A.C., V.G. Iannacchione, and J.A. Dever (2003). "Efficient Estimation for Surveys with Nonresponse Follow-Up Using Dual-Frame Calibration." *Proceedings of the American Statistical Association, Section on Survey Research Methods* (in print).

Appendix A. Dual-Frame Calibration (DFC) Weight Calculation

We applied the DFC methodology to compute the analysis weights for the Gulf War Veterans Health Survey as follows:

1. **Define two over-lapping samples.**

   s_A* = 10,301 veterans initially selected for the survey; and,

   s_B* = 5,182 mail respondents plus the 1,000 mail nonrespondents selected for follow-up.

   Note that s_B* is a proper subset of s_A*.

2. **Assign the design weights.**

   d_{kA} = the inverse of the selection probability assigned to the k^{th} sample member.

   d_{kB} = d_{kA} if the k^{th} sample member responded to the mail survey. For the 1,000 mail nonrespondents selected for the follow-up, d_{kB} equals d_{kA} times the inverse of the follow-up selection probability

   Note that $\Sigma_{sA*} d_{kA} = \Sigma_{sB*} d_{kB}$ = 685,074 veterans on the sampling frame.

3. **Define two over-lapping samples of respondents.**

   s_A = 5,182 respondents to the mail survey; and,

   s_B = 5,182 respondents to the mail survey plus the 527 follow-up respondents.

   Note that s_A is a proper subset of s_B.

4. **Construct variance replicates.**

   We created 294 variance replicates (a.k.a. random groups) that enable us to combine the data obtained from the mail survey with that obtained from the telephone and then use the jackknife method to estimate the variances of survey outcomes in a design-consistent fashion. Within each of the eight first-phase strata, we randomly assigned 35 sample members to each replicate with the requirement that each replicate have approximately equal numbers of mail respondents and at least one follow-up respondent.

5. **Calculate nonresponse adjustment factors.**

   For s_A*, the weighting classes corresponded to the eight first-phase sampling strata. For s_B*, the weighting classes corresponded to the eight first-phase strata for mail respondents, and to the

sixteen strata for the 1,000 mail non-respondents selected for the follow-up. The adjustment factors $a_{kA,nr}$ and $a_{kB,nr}$ equal the inverse of the weighted response rates within each weighting class. Note that $\Sigma_{sA} d_{kA} a_{kA,nr} = \Sigma_{sB} d_{kB} a_{kB,nr} = 685{,}074$ veterans on the sampling frame.

6. Obtain post-stratification totals.

The control totals T_x corresponded to the following 17 counts:

- First-stage strata (8): Gender x Component x CCEP evaluation
- Branch of Service (4): Army, Navy and Coast Guard, Marine Corps, Air Force
- Race/ethnicity (3): White, Black, Other
- Military rank group (2): Officer, Enlisted

Note that the post-stratification totals sum to 689,183 veterans and reflect slightly more complete totals than those for the sampling frame.

7. Calculate post-stratification adjustment factors.

We used GEM to calculate post-stratification adjustment factors $a_{kA,ps}$ and $a_{kB,ps}$ that were applied to the nonresponse-adjusted design weights to force them to sum to the 17 control totals. The design weights with adjustments for nonresponse and post-stratification are written as

$w_{kA}^{(1)} = d_{kA}\, a_{kA,nr}\, a_{kA,ps}$, and, $w_{kB}^{(1)} = d_{kB}\, a_{kB,nr}\, a_{kB,ps}$.

Note that $\sum_{sA} w_{kA}^{(1)} = \sum_{sB} w_{kB}^{(1)} = 689{,}183$ veterans.

Also note that these weights are used as input weights to the DFC procedure.

8. Obtain key outcome variables.

The following ten survey outcomes comprise the z-vector of 'zero controls':

- CDC Multi-Symptom Illness (CDCMULTI)	2 levels
- Post-traumatic Stress Disorder (PTSD)	2 levels
- Chronic Fatigue Indicator (CFATIGUE)	2 levels
- SF 36 Impairment Score (SF36_RP)	Continuous
- Hopkins Symptom Depression (HSCL_DEP)	Continuous
- Chalder 13-Item Fatigue Score (FATIGUE13)	2 levels
- Partner has Discomfort during Sex (SEX_DISC)	3 levels
- Current Smoking status (SMOKE)	3 levels
- Current Drinking status (DRINK)	5 levels
- Current Marital Status (MARIT)	2 levels

9. Construct replicate sets of DFC weights.

We constructed 294 replicated sets of DFC weights for use in the jackknife method of variance estimation. Each set of weights was constructed by excluding the respondents of a variance replicate and then applying the DFC methodology to the remaining respondents. This process was repeated until each of the 294 variance replicates had been excluded from one replicated set of weights.

For each replicated set of DFC weights, we used GEM to calculate DFC adjustment factors $a_{kA,DFC}$ and $a_{kB,DFC}$ that were applied to the nonresponse-adjusted and post-stratified design weights so that the differences between the ten key outcomes using w_{KA} and w_{kB} were zero while maintaining the 17 control totals. Using a grid search, we determined that a scaling constant of $\eta_A = 0.80$ minimized the sum of the variances of the key outcome variables. We used the scaling constant to calculate the DFC weights w_{KA} and w_{kB} and then a combining factor of $\xi_A = 0.82$ to calculate w^*_k, the set of final DFC weights for the full respondent sample s_B.

Appendix B. Example SUDAAN® Program

```
TITLE1 "GULF WAR VETERANS HEALTH STUDY";
FOOTNOTE9 "Program: SUDAAN_Phase1-2.SAS";

*****************************************************************************;
** Input and Output Data Sets.                                              **;
*****************************************************************************;
** Input Data Files **;                          %let datain =ETA80_AFL;
** Output Data Files **;                         %let dataout=;

*****************************************************************************;
** Variable Formats.                                                        **;
*****************************************************************************;
FILENAME fmts "C:\GWHS\FORMATS.TXT";
%inc fmts;

*****************************************************************************;
** Process Data File for SUDAAN Runs.                                       **;
*****************************************************************************;
DATA A_FILE;
  SET &datain.(KEEP=ELIG PH12_AWT1-PH12_AWT294 SF36_RP COMPONENT GENDER);
RUN;

*****************************************************************************;
** SUDAAN Analyses - Jack-knife Variances.                                  **;
*****************************************************************************;
PROC DESCRIPT DATA=A_FILE DESIGN=JACKKNIFE;
  SUBPOPN (ELIG=1);
  JACKWGTS PH12_AWT1-PH12_AWT294;
  JACKMULT 53*0.9811 28*0.9643 75*0.9867 38*0.9737 27*0.9630
           14*0.9286 39*0.9744 20*0.9500;
  SUBGROUP COMPONENT GENDER _one_;
  LEVELS   2 2 1;
  VAR SF36_RP;
  TABLE COMPONENT*GENDER;
  RFORMAT COMPONENT comp_.;
  RFORMAT GENDER gen_.;
  RTITLE "GWHS - SF36 (PTSD)";
RUN;
```

APPENDIX B

10th Anniversary Gulf War Veterans Health Study

Preliminary Estimates of Certain Health and Mental Health Problems

October 28 2003

1.0 Purpose

This report provides preliminary estimates on rates of health concerns among members of the U.S. armed forces who were deployed to the Persian Gulf in support of Operations Desert Shield and Desert Storm.

2.0 Methods

2.1 Sample Design

The target population for the 10th Anniversary Gulf War Veterans Health Survey was defined as military personnel (active duty, reserve, and national guard) who were deployed to the Persian Gulf between August 1, 1990 and July 31, 1991. Sampling was based on a stratified single-stage systematic sample of deployed personnel. The sampling frame was created from eligible personnel on the Operation Desert Shield/Storm database maintained by the Defense Manpower Data Center (DMD). Eight levels of design strata were defined based on a combination of three factors: (1) military status at the time of the Persian Gulf War (active versus reserves/national guard), (2) gender, and (3) risk status. Risk status refers to the likelihood of individuals having health concerns subsequent to serving in the Gulf (Table 2.1a).

Table 2.1a Distribution of GW veteran population over sampling strata

		High Risk	Not High Risk	Total
Active	Males	28,388	513,419	541,807
	Females	3,068	31,122	34,190
Reserve				
	Males	2,460	92,411	94,871
	Females	590	13,616	14,206
Total	Males	30,848	605,830	636,678
	Females	3,658	44,738	48,396
			0	
Pop. Count		34,506	650,568	685,074

A sample of 10,300 veterans was optimally allocated over the eight design strata. Military personnel who participated in the Comprehensive Clinical Evaluation Program (CCEP) were assumed to be at elevated risk for GW illnesses and were over-sampled relative to the GW military-population. Women and reservists were also over sampled because prior research

suggested that veterans in these two sub-populations may be at increased risk for reporting health problems following GW service.

Based on their experience with the Operation Desert Shield/Storm database, DMD personnel estimated that 2-3% of the cases in the database would not meet the study's inclusion criteria (i.e., individuals who were not in the Persian Gulf Theater between August 1, 1990 and July 31, 1991). To compensate for potentially ineligible subjects, the initially planned sample of 10,000 veterans was increased to 10,301. Rates of eligible responses, ineligible responses, refusals and non-response for the mail survey are shown in the table below.

Table 2.1b Rates of eligible responses, ineligible responses, refusals and non-response for GWVHS mail survey

	Freq	%	Cum Freq	Cum %
Eligible and useable survey	5198	50.5	5198	50.5
Ineligible (gw0=1)	199	1.9	5397	52.5%
Ineligible (note)	9	0.1	5406	52.9%
Refuse (blank)	30	0.3	5436	52.9%
Refuse (note)	3	0.03	5439	52.9%
Not useable (missing)	3	0.03	5442	52.9%
No survey returned	4844	47.1	10286	100%

2.1.1. Non-Response Sample

Despite efforts throughout the study to enhance the response rate to the GWVHS, approximately 4,844 veterans (47% of sample) did not return a survey. This response rate is higher than that typically achieved for mail surveys of similar length in a military population (e.g., the 1998 Department of Defense sponsored Total Forces study obtained a response rate of 38%; the Naval Health and Research Center's 1995 Perceptions of Wellness and Readiness Study obtained a response rate of 36%). Nonetheless, the scientific and political controversy regarding the nature and etiology of Gulf War illness is such that it was important to ascertain the potential effects of nonresponse on findings from this study. Therefore, a sample of 1,000 people who did not respond to the initial mail survey were selected to take part in telephone interviews. Useable data was obtained from 512 of these individuals.

2.1.2 Effective Response Rate

Analysis weights appropriate for the complex survey design were developed that incorporate information from the 2-phase design to adjust for non-response bias and which utilize a jackknife approach for computation of design-consistence variance estimates. Applying these weights, the overall weighted response rate (a.k.a. the *effective response rate)* among

eligible sample members for the combined mail survey and telephone follow-up was 70.5 percent with a 95 percent confidence interval of +/- 3.3 percent. The weighted response rate can be thought of as a population parameter. That is, it is the response rate that would be achieved if everyone on the sampling frame had been selected for the survey.

2.2 Questionnaire Items

A prototype version of the GWVHS questionnaire was pilot tested in a convenience sample of 41 Gulf War veterans. The questionnaire was administered in two group sessions followed by individual and small group debriefing. The questionnaire was shortened based on feedback from pilot participants. Administration time for the final version of the survey was estimated to be approximately one hour. A subset of items from the mail survey questionnaire was developed into a Computer Assisted Telephone Interview (CATI) and administered to a subsample of GWVHS sample members who did not respond to the mail survey.

The majority of items in the GWVHS were taken from previously published instruments, sometimes with adaptation. Sources of the survey items used in producing this report include

Alcohol Use Disorders Identification Test (AUDIT)

Reinert, Duane F; Allen, John P. (2002) The Alcohol Use Disorders Identification Test (AUDIT): A review of recent research. Alcoholism: Clinical & Experimental Research. 26: 272-279.

CDC Chronic Multisymptom Illness

Fukuda, K., R. Nisenbaum, *et al. (1998)*. Chronic multisymptom illness affecting Air Force veterans of the Gulf War. Journal of the American Medical Association 280: 981-988.

Chalder Fatigue Scale

Chalder, T. Berelowitz G., Pawlikowski, T., et al. (1978). Development of a fatigue scale. Journal of Psychosomatic Research, 37:147-153

Hopkins Symptom Checklist-25

Derogatis, L.R., Lipman, R.S., Rickels, K., Uhlenhuth, E.H., & Covi, L. (1974) The Hopkins Symptom Checklist (HSCL): A self-report inventory. Behavioral Science, 19:1-15.

Iowa Persian Gulf War Study

Iowa Persian Gulf War Study Group. (1997). Self-reported illness and health status among Gulf War veterans. Journal of the American Medical Association, **277**: 238-245.

PTSD Checklist

Weathers, F.W., Huska, J.A., & Keane, T.M. (1991). The PTSD Checklist-Military Version (PCL-M). Boston: National Center for PTSD

2.3 Procedures

2.3.1. Recruitment and Incentives

Sample members received an information pamphlet describing the GWVHS in advance of the first mailing of survey questionnaires. There were three waves of mailings that occurred on August 31, 2001, November 5, 2001, and December 31, 2001. In the first wave, questionnaires were mailed to all sample members. A souvenir pencil was included in the mailing along a $5.00 check. Checks, however, were only sent to veterans who were not on active duty since military regulations prohibit active duty military personnel from receiving compensation for participating in government-funded research unless it involves blood draws. A second questionnaire was mailed to wave 1 non-respondents. Reminder calls were made to a subset of individuals who had not returned the survey by one month after the second mailing. Because there is little data from which to determine if telephone calls are a cost effective means of increasing the response rate to a mail survey, only a subset of non-responders received calls. This allowed for the systematic evaluation of the effectiveness of this approach. The third and final mailing included a check for $20.00 (for non-activity duty personnel).

2.3.2 Scanning

Questionnaires were optically scanned employing a method that is sensitive to non-traditional marks such as check marks or 'X's – a method known as the 'check box' approach. Basically, the way the check box works is that when there are multiple marks for a given item, the darkness of the two or more marks is graded. If the differential between the two marks exceeds a specified level on a darkness scale, then the mark that scores higher on the darkness scale is treated as the intended response to the item. The darkness scaling depends on both the position of the mark and overall darkness. When compared to decisions made by human examiners, the error rate was minimal.

3.0 Preliminary Estimates of Health Problems

3.1 Health Problems in Past Six Months

Table 3.1 provides preliminary estimates for self-reports of health problems in the six months prior to veterans completing the GWVHS questionnaire. The most common health complaints were allergies/hay fever, migraine headaches, joint diseases, and high blood pressure. Rates of critical low prevalence health problems include reported cancer among an estimated 1.69% of the population and aplastic anemia among 0.74%.

Table 3.1. Estimated Prevalence of Health Problems in Past 6 Months (Weighted)

Health Problem	*%*	*Health Problem*	*%*
Allergies or hay fever	37.50	Heart attack or other serious heart trouble	3.41
Migraine headaches	28.84	Fibromyalgia, fibrostitis or myofascial pain syndrome	3.32
Arthritis, rheumatism, other bone or joint diseases	26.82	Disease of the genital organs	3.02
High blood pressure/hypertension	23.54	Severe hernia or rupture	2.85
Other major health problems	14.95	Emphysema or other lung diseases	2.46
Chronic stomach or gallbladder trouble	14.16	Multiple sclerosis, epilepsy, or other neurological disorders	2.38
Eczema, psoriasis or dermatitis	13.57	Lupus, thyroid disease, or other autoimmune disorders	2.36
Bronchitis	12.50	Severe kidney or liver disease	1.86
Ear infections	11.82	Other cancer or a malignant tumor of any kind	1.69
Blindness or severe visual or hearing impairment	7.71	Aplastic anemia	0.74
Asthma	7.01	Stroke	0.64
Diabetes or high blood sugar	4.70	Leukemia	0.35
Ulcer	4.02	AIDS	0.29

3.2 Persistent Health Problems

Table 3.2 reports preliminary prevalence estimates for persistent health problems which according to veterans first occurred subsequent to serving in the Persian Gulf. Among the most prevalent of these is the array of symptoms (i.e., joint pain, fatigue, numbness, memory loss, negative affective states) characteristic of the multi-symptom condition referred to as Gulf War Illness.

Table 3.2. Estimated Prevalence of Health Problems in Past 6 Months (Weighted)

Persistent Problem	*%*	*Persistent Problem*	*%*
Pain or aches in more than one joint	45.97	Nausea or an upset stomach	18.34
Feeling unrefreshed after sleep	42.42	Frequent diarrhea	14.91
Joint stiffness	42.29	Cough	14.90
Muscle tension, aches, soreness, or stiffness	39.71	Sore throat or irritation	13.84
Numbness or tingling in parts of your body	39.62	Abdominal pain	13.59
Back pain	39.29	Tender, painful or swollen lymph glands	12.99
Feeling moody or irritable	37.18	Inflammation of eyes	12.13
Reflux, heartburn, gas or indigestion	36.83	Tremors or shaking	11.72
Feeling weak in parts of your body	32.43	Wheezing	11.28
Feeling depressed	30.59	Wounds that are slow to heal	11.09
Feeling anxious or nervous	29.20	Unexpected hair loss	10.84
Trouble finding words	28.12	Mouth score	10.13
Any headaches	26.90	Swelling of both feet or both ankles	9.86
Loss of hearing or ringing in your ears	26.65	Constipation	9.51
Pains in heart or chest	24.62	Frequent or painful urination	9.13
Faintness, lightheadedness, dizziness or trouble maintaining balance	24.23	Unusual discomfort in genitals during or after intercourse	9.00
Dryness or scaling of your skin	24.16	Eruption of hives or welts on your skin	8.78
Hot or cold spells, fever, night sweats, chills	22.81	Trouble swallowing	8.68
Shortness of breath	22.48	Tendency to bruise or bleed easily	7.83
Feeling of bodily discomfort after exertion	21.81	Unintended weight loss > 10#	5.52
Unintended weight gain > 10 #	21.70	Vomiting	5.13
Persistent sensation of itching on your skin	21.56	Blisters, open sores, or skin ulcers	5.08
Heart palpitations, pounding or racing	21.50	Double vision	3.38
Sinus problem/rhinitis	21.43	Seizures or convulsions	0.98
Skin redness or rash	19.85		

3.3 Probable Post-Traumatic Stress Disorder

Veterans were asked to report separately on symptoms of post-traumatic stress associated with stressful 'Gulf War experiences' and with 'another stressful experience from the past'. The rates of probable post-traumatic stress disorder associated with stressful Gulf War experiences was marginally higher than the rate associated with other past experiences.

Table 3.3. Estimate of Prevlance of Probable Post-Traumatic Stress Disorder (Weighted)

Measure	%
PTSD Checklist (Gulf War)	7.81
PTSD Checklist (Other Past Experiences)	7.25

4.0 Analyses Currently Underway

Analyses are underway to identify sub-groups of veterans that are meaningful in terms of risk exposure and prognostic factors using non-hierarchical crisp clustering methods (e.g. k-means). In order to test the stability and reproducibility of the identified subgroup structure, the initial clusters will be developed in a 'calibration' sample and then examine whether similar results are obtain when the same methods are applied to a 'validation' sample. Generalized linear models (e.g. multinomial logistic regression) will be used to examine the relationship of subgroup membership with demographic factors (e.g. age, sex, rank, marital status, branch, unit), measured exposures (e.g. area served, receipt of vaccines, warzone stress), and functional outcomes (e.g. employment, physical impairment).