

Methodologic issues in a population-based health survey of Gulf War veterans

Bradley N. Doebbeling^{a,b,c,*}, Martha F. Jones^d, Daniel B. Hall^e, William R. Clarke^d,
Robert F. Woolson^d, James C. Torner^b, Leon F. Burmeister^d, Terri Snyders-Crumley^a,
Drue H. Barrett^f, Kenneth H. Falter^f, James A. Merchant^{a,g}, Sarah Nusser^h,
Dianne Anderson^h, David A. Schwartz^{a,c,g}

^aDepartment of Internal Medicine, The University of Iowa College of Medicine, Iowa City, Iowa, USA

^bDepartment of Epidemiology, The University of Iowa College of Public Health, Iowa City, Iowa, USA

^cVeterans' Affairs Medical Center, Iowa City, Iowa, USA

^dDepartment of Biostatistics, The University of Iowa College of Public Health, Iowa City, Iowa, USA

^eDepartment of Statistics, University of Georgia, Athens, Georgia, USA

^fDivision of Environmental Hazards and Health Effects, National Center for Environmental Health,
Centers for Disease Control and Prevention, Atlanta, Georgia, USA

^gDepartment of Occupational and Environmental Health, The University of Iowa College of Public Health, Iowa City, Iowa, USA

^hStatistical Laboratory Survey Section, Department of Statistics, Iowa State University, Ames, Iowa, USA

Received 16 October 2000; received in revised form 12 November 2001; accepted 21 November 2001

Abstract

This report describes the principal methods used in the development, conduct, and analysis of the research study “Health Assessment of Persian Gulf War Veterans from Iowa” (Iowa Gulf War Study). The methods presented include an outline of the organizational structure, study timeline, hypotheses, outcome definitions, and study design. Adhering to a strict timeline, the study protocol and instruments were developed, and a stratified sample of 3,695 military personnel (76% participation) was located and surveyed by structured telephone interview. The study tracked personnel from all service branches residing nationally and internationally, including those discharged from service. This study required development and implementation of methods appropriate to analysis of data collected in a complex sampling framework and methodological procedures to ensure scientific rigor in a highly public and politicized environment. Statistical analyses were conducted on a priori health outcomes and required development of methods to compute Cochran-Mantel-Haenszel adjusted rate differences. This environment facilitated rapid implementation, critique by scientific and public advisors, a high participation rate, and rapid publication. © 2002 Elsevier Science Inc. All rights reserved.

1. Introduction

Operations Desert Shield and Desert Storm involved over 750,000 participating U.S., British, and Canadian military personnel during 1990–1991. In 1992, widespread speculation began to arise regarding instances of unexpected illnesses among U.S. military personnel involved in the Gulf War [1–3]. Shortly after returning home, some individuals reported a variety of symptoms, including fatigue, myalgias, and neurocognitive and mood problems.

Gulf War veterans have reported multiple chronic symptoms, and multiple studies have found a higher prevalence

of symptoms among Gulf War veterans compared with military controls [4–9]. Data from several large, population-based studies with appropriate controls demonstrate that the pattern of symptoms is comparable in Gulf War deployed and Gulf War era personnel [6,10–12]. Clinical evaluations have identified medical or psychological conditions among many individuals reporting for Gulf War Registries, although symptomatology remains unexplained in approximately one fifth [13,14]. Many potential etiologies for these problems have been raised [15]. A broad range of self-reported exposures during the period has been associated with reported symptoms [5,8,9]. Expert scientific panels have reviewed the evidence, but none have yet provided a definitive explanation [1,16–18].

To investigate these problems, we conducted a population-based telephone survey of the health of Gulf War veterans and a similar control population, the “Health Assess-

* Corresponding author. Department of Internal Medicine, SE 625 GH, The University of Iowa College of Medicine, 200 Hawkins Drive, Iowa City, IA 52242. Tel.: (319) 356-8556; fax: (319) 356-3086.

E-mail address: brad-doebbeling@uiowa.edu (B.N. Doebbeling).

ment of Persian Gulf War Veterans from Iowa” research study (Iowa Gulf War Study) [5,11,19–22]. Notably, the participation rate obtained was one of the highest published to date [23]. Another important strength of the study was its evaluation of personnel from all service branches, both those remaining on active duty and those discharged from service [9]. The Institute of Medicine recently described the Iowa Gulf War Study as “perhaps the strongest study on Gulf War veterans’ experience of symptoms related to deployment” [23].

Due to widespread interest in these methods, we have summarized key methodologic issues and lessons learned. These include (i) the development and organization of the study, (ii) the strict timeline under which the study was carried out, (iii) study aims and hypotheses, (iv) locating strategies, (v) health outcome development, (vi) the sampling methods and instrument development, and (vii) statistical analysis methods. We consider the use of a small research working group, a Public Advisory Committee (PAC), and our subject tracking methods and statistical power and analysis methods particularly innovative and successful strategies.

2. Materials and methods

2.1. Study objectives

The primary objective of the Iowa Gulf War Study was to compare the self-reported health status of military personnel who served in the Gulf War theater (Gulf War group) to that of military personnel serving at the time who were not deployed to the Gulf (Gulf War-era group). Additional objectives were comparisons of (i) the health status of those who were “active duty” military personnel (hereafter referred to as “regular military”) with members of either the National Guard or Reserves (NG/R) who were “activated” as a result of the Gulf War military operations, (ii) the Gulf War with the Gulf War-era group within the regular military, and (iii) the Gulf War with the Gulf War-era group within the NG/R.

2.2. Organization

In November 1994, the Iowa Department of Public Health (IDPH) was awarded a cooperative agreement from the Centers for Disease Control and Prevention (CDC) for the purpose of conducting the Iowa Gulf War study. Fig. 1 presents the organizational structure of the collaborators and participating organizations for the study.

In collaboration with the CDC and IDPH, The University of Iowa (UI) was responsible for the overall design and conduct of this 2-year research study. We formed a small interdisciplinary research work group of experienced clinical researchers and internists (B.N.D. and D.A.S.), biostatisticians (R.F.W., W.C., and D.H.), and study coordinators (M.J. and T.S.C.) who met in work sessions several times per week. Working under a tight timetable, the work group planned and reviewed every component of the study and subsequent analy-

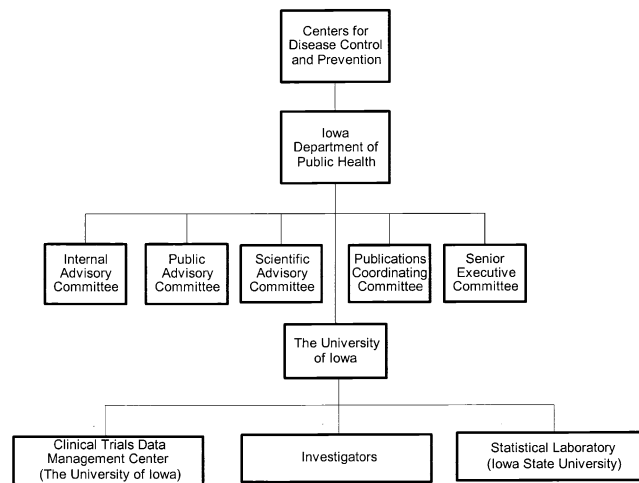


Fig. 1. Organizational structure for Iowa Gulf War study.

ses. Personnel resources were substantial. One of the Co-Principal Investigators (B.N.D.) devoted considerable effort (40%) to the project. Investigators from the CDC and IDPH participated in an Internal Advisory Committee (IAC) with other key UI investigators, which met bimonthly with additional conference calls as needed. Over 20 investigators devoted effort to the study. Development and revision of the study instruments, study correspondence, subject locating, and tracking required 1.5 full-time equivalents (FTE) for 2 years. Database development, data cleaning, planning and programming analyses, and developing documentation and reports required approximately 4.0 FTE for 2 years. At ISU, development and testing of the structured interview required 1.0 FTE for 6 months. Approximately 30 part-time interviewers (15 FTE) conducted the interviews over a 9-month period. Supervisory personnel to conduct monitoring, retraining, soft refusal conversions, and data editing required approximately 3 FTE for 1 year. When early evidence of delays occurred, additional personnel were hired on a temporary basis. The Department of Statistics Statistical Laboratory (Stat Lab) at Iowa State University (ISU) developed, tested and administered the subject interviews.

2.3. Public advisory, scientific oversight, and communications committees

Several committees (Fig. 1) were established to provide scientific guidance and public oversight. The Senior Executive Committee, consisting of one senior official from UI, CDC, and IDPH, was established to resolve outstanding issues related to interpretation and reporting of study results, authorship, or other key research issues. The Publications Coordinating Committee, consisting of three senior investigators from UI, a representative from CDC, and a representative from IDPH, formulated and implemented operational procedures for developing, organizing, and managing the preparation of manuscripts and presentations resulting from this

study. A Scientific Advisory Committee (SAC) of six leading scientists in clinical and research methods also provided advice regarding methodologic issues, review of the study instruments, review of the a priori outcome definitions, and study results.

A unique feature of this study was the establishment of the PAC, which was initially established to provide coordination, liaison, and open communication with veterans, veterans' service organizations, affected parties, and legislative bodies [23,24]. The committee also participated in the revision of the study questionnaire, review of confidentiality concerns, and the development of publicity releases for interview initiation. Members of the committee included representatives of local veterans' service organizations, Gulf War veterans, and spouses of Gulf War veterans. The PAC met twice annually with the principal investigators during the study period. It provided advice and comment regarding subject-locating strategies, review of results, and liaison with veterans' organizations. The input of the PAC helped us communicate to veterans that the study was being done with their interests at heart and may have contributed to the high response rate [23,24].

2.4. Study timeline

The study was funded and conducted over a 2-year period from December 1994 through November 1996. During the first 6 months, the study organizational structure was established, the study protocol was developed, survey instruments were constructed and pilot tested, and study approval was received from the UI and CDC Institutional Review Boards.

Because the study was federally funded and involved the use of normal population controls, protocol and survey instruments were submitted to the Office of Management and Budget (OMB) in June 1995. OMB was interested in assessing subject burden and asked for a series of revisions to make some items comparable between the Iowa and VA National Health Surveys [25]. OMB approval was received, and interviewing began in September 1995.

The interview phase lasted approximately 32 weeks. Data were then edited and analyzed with the initial review of the study results conducted by a limited group of UI investigators at the end of June 1996. The preliminary results were then presented at a combined meeting of study investigators and the SAC in July 1996. Initial results were published in JAMA in January 1997 [5].

2.5. Health outcomes

Due to the paucity of basic epidemiologic data at its inception, this study was formulated to assess self-reported health status in regard to both primary and additional outcomes in the study population. The primary health outcomes of interest were symptoms of depression, posttraumatic stress disorder (PTSD), chronic fatigue, respiratory disease, and cognitive dysfunction. Other outcomes of interest were

general health symptomatology, health-related quality of life, reproductive health, adverse birth outcomes, sexual dysfunction, cancer, fibromyalgia, multiple chemical sensitivity, substance abuse, anxiety, injuries, and health care utilization. These specific outcomes were investigated based on previous reports in self-selected populations of Gulf War veterans that suggested that these outcomes might be elevated in Gulf War veterans [1,17,26].

2.6. Study design and methods

This project was conducted as a population-based epidemiologic study, surveying (cross-sectionally via telephone interview) a sample of military personnel who served during the Gulf War period. Two criteria were used to define the military personnel eligible for inclusion: (i) listing Iowa as the home of record at enlistment and (ii) service as regular military or activated NG/R between 2 August 1990 and 31 July 1991—the period of the Persian Gulf military operations. Based on these criteria, the Defense Manpower Data Center (DMDC), operated by the Department of Defense, identified 29,010 military personnel from their Desert Shield/Storm File and Active Duty Military Master files eligible for inclusion. Provided in electronic format, these personnel records included demographic data, military information, and personal identifiers.

2.7. Sample size

Study sample sizes were determined on the basis of the primary comparison of interest, namely, the comparison of the Gulf War groups to the non-Gulf War group. As shown in Table 1, there are two groups within which a Gulf War/Gulf War-era comparison of health outcome rates were made, the regular military and NG/R. In planning the sample size, it was anticipated that a Cochran-Mantel-Haenszel (CMH) statistic would be used to compare the Gulf War and non-Gulf War groups [27]. To arrive at an adequately powered comparison of the Gulf War-deployed and Gulf War-era within each of the two strata individually, NG/R were oversampled. The dependent variables used in sample size calculations focused on symptoms of fibromyalgia, chronic fatigue, airways disease, or depression. The existing literature suggested that base rates between 5% and 10% would be conservative estimates for several of these outcomes [28–30].

Sample size calculations utilized the methods provided by Woolson et al. [31], who described generation of sample size for specified significance level, power, and odds ratios in a stratified CMH framework. The health outcome rates of

Table 1
Study domains

| | Gulf War | Gulf War-era |
|------------------------|----------|--------------|
| Regular military | Domain 1 | Domain 2 |
| National guard/reserve | Domain 3 | Domain 4 |

the Gulf War-deployed and Gulf War-era groups were to be compared both within and across the regular military versus NG/R strata combined, utilizing the CMH statistic. If, in fact, the regular military and NG/R difference did not differ appreciably across the two strata, then such a CMH analysis would be a powerful way to compare the Gulf War-deployed and Gulf War-era groups. Sample size was calculated to achieve these objectives.

Table 2 is a summary of the sample size calculations generated for this particular investigation. Sample size figures were generated under the assumption that two-tailed significance tests would be conducted.

Because the calculations assumed an equal number of individuals in each of the four domains defined by military and exposure status, 1,364 would be the number required in each domain if the Gulf War and Gulf War-era difference varied by domain. The sample size of 1,364 would yield 80% power to detect a 50% increase. If the Gulf War/Gulf War-era difference was homogeneous across the two subgroups and the base rate was 5%, then 80% power to detect a 50% increase would be achieved with 2,946 completed interviews. Hence, a total of 3,000 individuals interviewed equally divided into the four domains (750 per domain) would yield reasonable power to detect meaningful differ-

ences for each comparison of interest over a broad range of base rates. The target sample for this investigation was selected to be 3,000 (750 per domain). Our predicted base rates for the conditions of primary interest were conservative; thus, the study power was greater than expected for most comparisons [5].

2.8. Sampling procedures

Using the data files provided by the DMDC, we developed and conducted the sampling procedures for the full study. Each individual in the study population was placed into one of four domains. Within each domain, the population was stratified by the combinations of the levels of five stratification variables displayed in Table 3. Although 29,010 military personnel records were originally identified as eligible for inclusion, 42 of these records were not included for several reasons. The most common exclusions were (i) retirees, due to their small numbers and dissimilarity with the major study domains ($n = 14$) and (ii) having participated in the pilot study ($n = 24$). The Coast Guard was combined with the Navy due to their small numbers. The final target study population consisted of 28,968 persons.

To obtain approximately 750 interviews in each of the four domains, the sampling was carried out as a two-step process. In the first step, approximately 750 individuals were selected from each of the four domains using stratified random sampling with proportional allocation with two exceptions. If any stratum contained only one subject, then that subject was selected into the sample. In strata where more than one subject was available to be sampled, at least two subjects were selected.

A computer program using computer-generated random numbers yielded simple random samples of size n_{hj} from the N_{hj} subjects in stratum h ($h = 1, \dots, 64$) of domain j ($j = 1, 2, 3, 4$). Apart from the noted exceptions, the proportional allocation procedure dictated sample sizes

$$n_{hj} = 750 \frac{N_{hj}}{N_j},$$

where

$$N_j = \sum_{h=1}^{64} N_{hj}$$

is the total number of veterans in domain j , for all h and j .

This first step of sampling produced 3,075 subjects for interview. However, it was understood that some proportion of the subjects selected in the first step of sampling would not be located or would refuse to participate. Because this proportion could not be accurately estimated at the beginning of the study, a second sample was planned in which additional subjects would be selected from each stratum. The actual number selected in this second step was to be determined based on estimated response rates from the first 1,600 subjects processed from the first step. Response rate estimation was delayed as long as possible between step one

Table 2

Total sample size required utilizing the Cochran-Mantel-Haenszel Test Statistic to compare Gulf War and Gulf War-era with two strata, regular military and National Guard/Reserves (two-tailed tests)^a

| Proportion of control with Dx = P _c = 0.10 | | | Study sample size $\alpha = 0.05$ Power = | | |
|---|----------------|------|--|--------|--------|
| P _c | P _e | OR | 0.8 | 0.9 | 0.95 |
| 0.10 | 0.125 | 1.29 | 4,877 | 6,528 | 8,072 |
| 0.10 | 0.15 | 1.59 | 1,364 | 1,825 | 2,256 |
| 0.10 | 0.20 | 2.25 | 398 | 532 | 657 |
| Proportion of control with Dx = P _c = 0.075 | | | Study sample size $\alpha = 0.05$ Power = | | |
| P _c | P _e | OR | 0.8 | 0.9 | 0.95 |
| 0.075 | 0.09375 | 1.28 | 6,708 | 8,980 | 11,104 |
| 0.075 | 0.1125 | 1.56 | 1,915 | 2,562 | 3,168 |
| 0.075 | 0.150 | 2.18 | 552 | 738 | 912 |
| Proportion of control with Dx = P _c = 0.05 | | | Study sample size $\alpha = 0.05$ Power = | | |
| P _c | P _e | OR | 0.8 | 0.9 | 0.95 |
| 0.05 | 0.0625 | 1.27 | 10,419 | 13,948 | 17,248 |
| 0.05 | 0.075 | 1.54 | 2,946 | 3,943 | 4,876 |
| 0.05 | 0.100 | 2.11 | 870 | 1,164 | 1,439 |

Abbreviation: OR, odds ratio.

^aValue tabulated is the total number of individuals to be interviewed. It is assumed that the total number of individuals is divided by four to give the number of individuals to be interviewed in the Gulf War regular military, the Gulf War reserve/guard, the Gulf War-era regular military, and the Gulf War-era reserve/guard.

In the table, the quantity labeled P_c is the base rate assumed in the control group. P_e is the rate we wish to detect in the Gulf War-era and the OR is defined as $\frac{P_e(1 - P_c)}{P_c(1 - P_e)}$.

Table 3
Distribution of the study population over levels of the stratification variables

| Stratification variables | Gulf War regular military <i>n</i> (%) | Gulf War-era regular military <i>n</i> (%) | Gulf War NG/R <i>n</i> (%) | Gulf War-era NG/R <i>n</i> (%) | Total <i>n</i> (%) |
|--------------------------|--|--|----------------------------|--------------------------------|--------------------|
| Branch of service | | | | | |
| Army | 2,935 (44.5) | 5,721 (30.5) | 1,161 (77.7) | 1,690 (80.8) | 11,507 (39.7) |
| Air Force | 779 (11.8) | 5,776 (30.7) | 5 (0.3) | 200 (9.6) | 6,760 (23.3) |
| Marines | 1,264 (19.2) | 1,846 (9.8) | 234 (15.7) | 129 (6.2) | 3,473 (12.0) |
| Navy/Coast Guard | 1,616 (24.5) | 5,445 (29.0) | 95 (6.4) | 72 (3.4) | 7,228 (25.0) |
| Rank | | | | | |
| Enlisted | 6,090 (92.4) | 16,949 (90.2) | 1,396 (93.4) | 1,889 (90.3) | 26,324 (90.9) |
| Officer | 504 (7.6) | 1,839 (9.8) | 99 (6.6) | 202 (9.7) | 2,644 (9.1) |
| Gender | | | | | |
| Male | 6,298 (95.5) | 16,692 (88.8) | 1,372 (91.8) | 1,930 (92.3) | 26,292 (90.8) |
| Female | 296 (4.5) | 2,096 (11.2) | 123 (8.2) | 161 (7.7) | 2,676 (9.2) |
| Race | | | | | |
| White | 6,279 (95.2) | 18,034 (96.0) | 1,441 (96.4) | 2,031 (97.1) | 27,785 (95.9) |
| Black/other/unknown | 315 (4.8) | 754 (4.0) | 54 (3.6) | 60 (2.9) | 1,183 (4.1) |
| Age | | | | | |
| ≤ 25 | 4,049 (61.4) | 9,914 (52.8) | 775 (51.8) | 1,032 (49.4) | 15,770 (54.4) |
| > 25 | 2,545 (38.6) | 8,874 (47.2) | 720 (48.2) | 1,059 (50.7) | 13,198 (45.6) |
| Total | 6,594 (22.8) | 18,788 (64.9) | 1,495 (5.2) | 2,091 (7.2) | 28,968 |

Abbreviation: NG/R, National Guard/Reserves.

and step two, achieving relatively precise estimates of success rates without interrupting the interviewing of subjects.

By mid-October 1995, the initial 1,600 subjects from the first phase-sample had been processed for location of valid phone numbers. Based on the locating success and response rate experience obtained, it was projected that, across strata, approximately 58–69% of the first sample would provide completed interviews. Using these estimates, it was decided that to reach the goal of 750 completed interviews in each domain by the end of February 1996, an additional 1,800 subjects should be selected in the second phase. Furthermore, because the locating success and response rates were somewhat different in the domains, it was determined that 28% of the second sample should be allocated to each of the two regular military domains and 22% allocated to each of the two NG/R domains. Thus, using the first-step sampling methods and sampling fractions, a total of 1,857 subjects were selected in the second step. The additional 57 subjects over and above the sampling goal of 1,800 was a result of oversampling small strata as described previously. These step-two subjects then became part of the total sample of 4,932 subjects and were processed in the same manner as those selected in the first step.

2.9. Instrument development

The structured interview was developed to assess a broad array of health concerns and to determine the prevalence of symptoms of specific medical conditions. Study investigators, working in small groups according to their areas of interest and expertise, were responsible for developing the survey questions to address study hypotheses. The symptom items selected were primarily based on published peer-reviewed data, personal interviews of Department of Veter-

ans Affairs Gulf War Registry participants, pilot studies, and input from the PAC and SAC [32]. Items included symptoms most often reported by Gulf War veterans, across each organ system of interest (constitutional, respiratory, cardiovascular, gastrointestinal, dermatologic, endocrine, and neurologic) [11]. Specifically, the item pool contained multiple items related to disorders hypothesized by the investigators to potentially occur at an increased prevalence in this population. These included asthma, chronic bronchitis, dermatitis, cognitive dysfunction, fibromyalgia, chronic fatigue, substance abuse, major depression and dysthymia, panic disorder, and PTSD, among others.

Emphasis was placed on using standardized, validated, and well recognized instruments whenever possible. Sources used in their entirety for valid scales or subscales included the PRIME-MD [33,34], the CAGE Questionnaire [35], the PTSD Checklist - Military (PCL-M) [36], the Fatigue Scale [37], the Brief Sexual Function Questionnaire for Men [38], and the Marlow-Crowne Social Desirability Scale [39]. Other specific subscales or groups of items were incorporated from other widely used sources and instruments to enhance comparability with other datasets: the CDC Chronic Fatigue Syndrome Questionnaire [40], National Health Interview Survey [41], the Behavioral Risk Factor Surveillance Survey [42], the Health Status of Vietnam Veterans Telephone Survey [43–46], the National Medical Expenditures Survey [47], the Brief Symptom Inventory [48], the American Thoracic Society Questionnaire [49], the Sickness Impact Profile [50], the Agricultural Health Study [51], and a series of questions to assess fibromyalgia [52–55] and military exposures [56]. In addition, the Medical Outcomes Study Short Form-36 (SF-36) was used in its entirety to assess health-related quality of life [57,58].

Subjects were asked whether any of 78 symptoms had

been persistent or recurrent in the past year. If present, they rated how much it bothered them: a little bit, moderately, quite a bit, or extremely, coded from 0 (none) to 4 (extremely). The response format was dichotomous for 35 additional symptom items from standardized instruments and 24 other medical problems. Declined responses were coded as missing and “don’t know” responses as negative. For symptoms present in the past year, subjects were asked if onset was before, during, or after the Gulf War. Missing data on symptoms, for example, were rare ($n = 59$, 1.6%).

The telephone instrument was split into two sections for ease of administration and to facilitate scheduling of interviews and participation. A short Introductory Interview (mean, 10 minutes; range, 3–30 minutes) obtained subject consent, collected general military and demographic information, and allowed the interviewer to schedule an appointment (if needed) to conduct the Main Interview. The Main Interview (mean, 60 minutes; range 28–185 minutes) included the questions intended to gather the health and exposure data necessary to assess the identified study outcomes. The Introductory and Main Interviews were pilot-tested on a group of individuals selected from both the study population and military personnel outside it. Based on the pilot information, the Introductory and Main Interviews were refined into a final form. The survey was further reviewed and revised based on input from CDC, IDPH, and the IAC, PAC, and SACs.

In addition to the two-part survey instrument, Proxy and Reliability Interviews were developed. The Proxy Interview provided an abbreviated assessment of the health and exposure experience of subjects who were deceased or who were mentally or physically incapable of participating themselves. Additionally, for those subjects who were deceased, the date, place, and cause of death were included in the surrogate interview. No validation of the proxy interview was done because very few subjects were unable to participate due to mortality, institutionalization, or severe disability. Because of the small numbers of proxy interviews, the results were not used in analyses.

The Reliability Interview was developed to assess test-retest reliability. However, the reliability interview contained only a subset of items from both the Introductory and Main Interviews. Approximately 5% of the sample was administered this follow-up interview 2–4 weeks after completion of the Main interview.

2.10. Locating strategies

Extensive subject database searches and locating measures were used to locate, recruit, and interview all subjects. A subject-tracking database was developed to assist in tracking potential new addresses and contact information. However, no specific financial incentives were provided. Although data provided by the DMDC included the name, address, social security number, and date of birth for study subjects, many of the names and addresses from this data-

base were subsequently discovered to be inaccurate and/or incomplete. Consequently, it was necessary to develop a multi-phased locating process. This process included mailing “preview” letters to subjects to inform them that they had been selected for participation, explain the procedure for selection and purpose of the study, update address information and telephone numbers, and prepare them for a telephone call. These letters noted that the study was being conducted by the UI, acknowledged funding from and collaboration with the CDC and IDPH, and included an explanation of the study’s purpose, procedures for contact, interview, and subsequent plans.

All letters were mailed with address and forwarding services requested. Included with each letter was a toll-free telephone number and a return card with a postage-paid envelope through which the subject could furnish updated address and/or telephone information. The mailed letter, toll-free number, and return postcard for updated information were particularly effective locating measures.

Directory assistance was then used to ascertain telephone numbers for addresses identified. Military base locators were used to identify telephone numbers for active duty personnel and were particularly helpful for those overseas.

When a valid address could not be obtained, multiple search strategies were employed. Telephone CD-ROMs and multiple low- or no-cost web-based person-finder searches were used to search for individuals using names, prior geographic locations, and any other identifying information. Other sources for updated address information included the Iowa Department of Motor Vehicles records, updated DMDC data files, referrals from family members or friends located through the previously mentioned search strategies, the DEERS database of spouse contact information, and the Internal Revenue Service (IRS) database. Three commercial vendors were used in tracking study subjects nationally: Equifax, Inc. (Fairfax, VA), Telematch (Springfield, VA), and the Trans Union Corporation (Chicago, IL). Credit agency searches were performed. Because the DEERS and IRS sources were accessed following several credit agency searches, they provided less than 5% of new locating information. The combination of these multiple strategies resulted in an overall location rate of 84% of the sampled subjects, and 91% of those subjects participated.

A nationwide death certificate search was conducted. Only 19 of 4,889 selected subjects had died. They were equally distributed between deployed veterans ($n = 9$) and nondeployed controls ($n = 10$) [11]. The final disposition of all subjects is displayed in Fig. 2.

2.11. Confidentiality

The Institutional Review Boards approved the study procedures and instruments. Contact letters and information summaries included an explanation of the study’s purpose, procedures for contact, interview and protection of data security and confidentiality, and subsequent plans. A Public Health Service Certificate of Confidentiality was obtained to

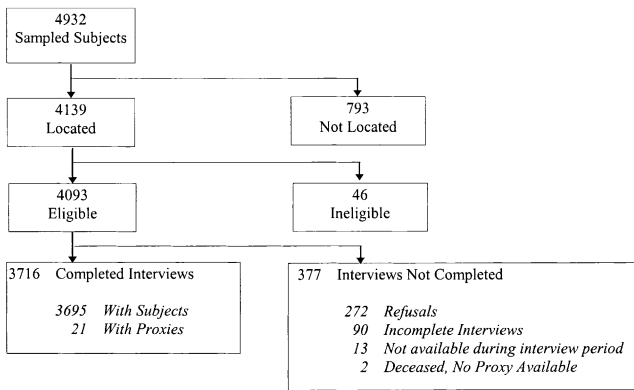


Fig. 2. Final disposition of sampled subjects.

protect confidentiality of the data to subpoena and encourage participation. Other measures to protect subject identity and survey data were explained to subjects in the information summary during recruitment.

2.12. Structured interviews

An experienced survey research group, the ISU Stat Lab, conducted the telephone interviews from September 1995 through May 1996 using experienced interviewers specifically trained for the study. Training included 20 hours of instruction for each interviewer on the principles and procedures of research interviewing, project specifics, and the computer-assisted telephone interviewing (CATI) software use. The CATI system was used to display survey questions and potential responses on a video screen, allowing interviewers to accurately code responses. Because of the length and branching complexity of the survey instrument, interviewer and respondent burden were significantly reduced with the use of this system. In addition, programmed skip patterns and range and consistency checks minimized missing data. Interview conduct was monitored on an ongoing basis, and all completed interviews were subject to supervisory review. The Computer-Assisted Survey Execution System is a CATI software package developed by the Computer-Assisted Methods Program (CSM), a widely recognized software package used by many survey research organizations [59].

Telephone contact was scheduled on a standard daytime, evening, and weekend rotating schedule, covering all time zones, including those overseas. A minimum of 12 calls were made to a given telephone number before it was classified as “unable to reach.” Most subjects were contacted within eight phone calls, although a few cases required more attempts. Because the cohort was young and typically employed, active, and mobile, they often used answering machines and were frequently not at home. If there was repeatedly no answer, those cases received additional call attempts. Difficult-to-engage subjects or those providing “soft” refusals were contacted by more experienced interviewers, with an additional yield of approximately 10%.

2.13. Participation

Interviews were completed on 3,695 of 4,886 eligible subjects (76% participation; 91% of those contacted). Participants included 1,896 deployed (78%) and 1,799 individuals not deployed (73%) to the Gulf, representing 889 deployed and 893 nondeployed units [11]. Participation rates were comparable across strata. Deployed study participants were more often younger, male, enlisted, Army or Marines, and less educated (Table 3). There were no differences in the deployed and nondeployed in terms of race, active duty versus NG/R, income, smoking, or unemployment [11].

2.14. Data analysis

Before the completion of the study interviews and before the initiation of any data analyses, study investigators developed operational definitions for each of the identified health outcomes. Beginning in June 1995, UI study investigators, working in small groups according to their areas of expertise, developed initial operational definitions for each of the health outcomes. An internal study group led by the UI principal investigator or co-principal investigator participated in all working group meetings to provide oversight, continuity, and documentation. All definitions were then assembled and distributed for review to study investigators. A second round of working group investigator meetings was held in August 1995 to discuss and update these operational definitions. In October 1995, these definitions were reviewed during a joint meeting of the UI investigators with the SAC, CDC, and IDPH. The last round of working group meetings was held in December 1995, during which final revisions were made and covariates for analysis identified. The study investigators had no access to accumulating interview data while developing these definitions.

The outcome definitions developed by the investigators were primarily based on the instruments and scales contained in the study survey that had been previously validated and/or published. Most instruments were used intact, and many outcomes were developed using previously validated coding. For example, most of the mental health outcomes used algorithms from the PRIME-MD.

Whereas investigators defined a few outcomes from a single question, most were constructed as composites of multiple questions. These composite outcomes combined component questions by requiring that (i) all of several criteria be true, (ii) n or more of several criteria be true, or (iii) some combination of conditions of types 1 and/or 2 hold.

Typically, these outcomes took on positive, negative, or missing values corresponding to the situations in which data were sufficient to conclude that the adverse health experience occurred, data were sufficient to conclude that the adverse health experience did not occur, or data were insufficient to conclude whether or not the adverse health experience occurred during the period of inquiry. Because the period of inquiry was typically currently or recently (e.g., during the

year previous to the interview), such outcomes were appropriate for the estimation of prevalence rather than incidence. Prevalence was defined as the proportion of positive outcomes among all positive or negative outcomes. That is, outcomes taking on missing values were excluded from the denominator of prevalence measures.

Concurrent with development of the health outcome definitions, investigators developed an analysis plan to meet primary and secondary study goals. Primary goals were to obtain point estimates and standard errors of population prevalence and incidence rates, and prevalence differences and odds ratios, within the four study domains. Secondary goals were to assess the reliability and validity of the study data, summarize nonresponse patterns, account for the effects of nonparticipation due to death or physical or mental incapacity, and summarize a variety of covariates that were not study outcomes but for which data were collected.

2.15. Validity analyses

Birth defect and cancer outcomes from the Iowa Gulf War Study were validated by a comparison of study outcomes with information available from the Iowa State Health Registry. To assess the level of response bias, the 10-item short version of the Marlow-Crowne Social Desirability Scale was administered as part of the main interview [60].

Content validity was demonstrated by the incorporation of sections of our instrument in other investigators' mail-out surveys [9,61] and independent expert review [23]. The SF-36, a well-validated instrument assessing eight domains of functioning, was used to assess general health status [57]. Symptom scales showed evidence of convergent and construct validity with the SF-36 [11]. The SF-36 demonstrated excellent reliability and construct validity [22]. Analyses of military, behavioral, environmental, and other health factors were associated in expected directions with postdeployment SF-36 scores [22]. Analyses with other investigator-derived outcomes also appeared reliable and valid [19,21,62].

As in any survey, an unevenly distributed pattern of nonresponse is a potential source of bias. To assess the degree to which such a pattern occurred, response rates were reported by each of the stratification variables separately, as well as for all combinations of the stratification variables (that is, by the seven-way cross-classification of exposure status, NG/R versus regular military status, age, gender, race, rank, and branch of service). This plan and a corresponding set of proposed statistical methods were reviewed and revised by study investigators and the SAC before implementation.

2.16. Statistical methods

Although most study outcomes were dichotomous (health experience occurred/did not occur), there were several continuously distributed and a few polytomous outcomes defined in the study. In addition, covariates were measured that were taken into account in the analysis of some study outcomes. This combination of several outcome

scales and the presence or absence of covariates, depending on the outcome, required that several different statistical techniques be used to meet the primary goals outlined above. Although all of these techniques are standard in the context of simple random sampling, the stratified sampling design of the study required that adjustments to these methods be utilized. Although these adjustments were available for most methods, it was necessary to develop methods to compute the CMH adjusted rate differences [63,64]. All programming pertinent to implementation and completion of these goals was completed before June 1996, when final data from the interviews were received.

Alpha was set at 0.05, and all P values were two-tailed. There was no adjustment for multiple comparisons. The investigators considered it more important to identify potential associations of interest, given the nature of the illness under study, rather than obscure potential differences if they were observed through adjustment of alpha.

To account for the nonsimple random sampling design of the study, the statistical analyses required sampling weights. For each of the H combinations of stratification variables, a sampling weight w_h was computed to be the ratio of the population size to the sample size. Sampling weights were set equal to 0 for strata in which the sample size equaled 1.

2.16.1. Stratification variables

For analysis purposes, the values of the stratification variables (age, gender, race, rank, branch of service) were assigned based on responses to items from the study questionnaire. Any missing values were replaced by the strata information provided by the DMDC files. Stratum-specific population sizes were adjusted to account for any discrepancies that were observed between subjects' self-reported strata data and the DMDC records. Stratum-specific population sizes were also adjusted by removing ineligible subjects discovered in the interviewing process.

2.16.2. Dichotomous outcomes

The most common outcome type in the study was a dichotomous response variable. Three statistical analysis techniques were used for such variables. To test the study hypotheses, CMH tests [27,65] were performed. In addition to CMH tests, CMH rate difference estimates [66] were computed for each of the comparisons corresponding to the study hypotheses. CMH estimates, associated standard errors, and 95% confidence intervals were reported. Although a CMH rate difference provides a measure of the excess prevalence in one of the populations, it is also of interest to estimate the prevalence to which this rate difference is relative. For this reason, prevalence estimates and associated standard errors were computed for each of the four domains defined in Table 1 as well as for the two collapsed populations consisting of all Gulf War subjects and all Gulf War-era subjects.

2.16.3. Ordered polytomous outcomes

Another variable type encountered among the study outcomes was a polytomous variable with ordered response

categories. For these outcomes, the proportion of responses in each of the possible response categories was reported by domain. In addition, CMH rate difference estimators were computed, and generalized CMH [67] tests and mean score tests were completed. The CMH rate difference estimators differ slightly in this case from the estimators described above for dichotomous variables in that in the polytomous case rate differences were estimated for each of the C response categories rather than for just the condition-present category. The two types of generalized CMH tests that were performed are alternative generalizations of the CMH test for $H 2 \times 2$ tables to the situation in which the data can be described by $H 2 \times C$ tables.

2.16.4. Continuous outcomes

A few outcomes considered in the study were continuous in scale. For these variables, domain-specific means and standard errors were reported. In addition, the study hypotheses were addressed by fitting four separate regression models corresponding to the comparisons of interest for each continuous variable. The explanatory variables in each regression model included the stratification variables, as well as an indicator variable defining the populations to be compared (e.g., regular military Gulf War versus regular military Gulf War-era). The coefficients of the population-defining indicator variables provided estimates of the mean difference in the outcomes between the populations. Ninety-five percent confidence intervals were computed for these coefficients, providing tests of the hypotheses of no mean difference. Because these confidence intervals are based on an assumption of normality of the responses, a test of normality was provided for each regression model. For models based on 2,000 or fewer observations, this normality test was the Shapiro-Wilk test [68]; otherwise, the normality test was the Kolmogorov test [69].

3. Summary

Our purpose has been to present the principal methods used in the development, conduct, and analysis of the Iowa Gulf War Study. This information should be useful to researchers faced with designing and administering complex sample surveys.

Acknowledgments

This work was partially supported by CDC Cooperative Agreement, U50/CCU711513, and Department of Defense Grant #DAMD17-97-1. We were very fortunate in enlisting the support and participation of many individuals throughout this study. Investigators at the University of Iowa, Iowa Department of Public Health, and Centers for Disease Control and Prevention include (in addition to the authors): D.W. Black, K.G. Saag, P.S. Thorne, Z.K. Ballas, J. Barrash, L.P. Boss, T.B. Casale, E. Chrischilles, H. Falk, M.D. Hansen,

R.G. Kathol, J.A. Kelly, B.M. Pfohl, W.W. Piette, P. Quinlisk, J.E. Rohrer, N.L. Sprince, P. Tallboy, and C. Zwerling). We also acknowledge the Iowa State University Statistical Laboratory (Toni Genalo, Andrew Williams, and Jan Larson) who assisted in the structured telephone interview and supervised the telephone survey. The Scientific Advisory Committee (Chair: Jonathan Samet, MD, MS [Johns Hopkins School of Hygiene and Public Health]; Mark Cullen, MD [Yale University School of Medicine]; Pierce Gardner, MD [SUNY at Stony Brook]; Michele Marcus, PhD [Emory University]; Earl Pollack, ScD [The Center to Protect Worker's Rights]; and Janet B.W. Williams, DSW [Columbia University/New York State Psychiatric Institute]) provided advice, guidance, and critique throughout all phases of this investigation. The Public Advisory Committee (Co-chairs: Tim Striley and John Kelly [Iowa Department of Public Health], Oscar Ballard [Paralyzed Vets of America], Randy Brown [Commission of Veterans Affairs], Sgt. Scott Clyce, Dennis W. Gooden [Disabled American Veterans], Mark Heiderscheit, Richard Hodges [VAMC, Des Moines], M. Terry Lipovac [Veterans of Foreign Wars], Lydia Siefken, Robert O. Steben [AmVets], Phillip Talboy [CDC], Annie Tuttle [VAMC, Iowa City], Tully Walker [VAMC, Des Moines], Larry Wisnosky [VA Regional Office], and Colonel Mark Zirkelbach [Iowa National Guard, Camp Dodge]) established an open dialogue between our study and veterans, veteran organizations, and affected individuals.

References

- [1] NIH Technology Assessment Workshop Panel. The Persian Gulf experience and health. *JAMA* 1994;273:391–5.
- [2] Evans S. Veterans' groups call for study of possible Desert Storm illnesses. *Washington Post* April 29, 1992.
- [3] What happened over there? [video recording] ABC News 20/20. Denver, CO: Journal Graphics Inc. August 14, 1992.
- [4] Stretch RH, Bliese PD, Marlowe DH. Physical health symptomatology of Gulf War-era service personnel from the states of Pennsylvania and Hawaii. *Mil Med* 1995;160:131–36.
- [5] The Iowa Persian Gulf Study Group. Self-reported illness and health status among Gulf War veterans. A population-based study. *JAMA* 1997;277:238–45.
- [6] Fukuda K, Nisenbaum R, Stewart G, Thompson WW, Robin L, Washko RM, et al. Chronic multisymptom illness affecting Air Force veterans of the Gulf War. *JAMA* 1998;280:981–88.
- [7] Gray GC, Hawksworth AW, Smith TC, Kang HK, Knoke JD, Gackstetter GD. Gulf War veterans' health registries: who is most likely to seek evaluation? *Am J Epidemiol* 1998;148:343–49.
- [8] Proctor SP, Heeren T, White RF, Wolfe J, Borgos MS, Davis JD, et al. Health status of Persian Gulf War veterans: self-reported symptoms, environmental exposures and the effect of stress. *Int J Epidemiol* 1998;27:1000–10.
- [9] Unwin C, Blatchley N, Coker W, Ferry S, Hotopf M, Hull L, et al. Health of UK servicemen who served in Persian Gulf War. *Lancet* 1999;353:169–78.
- [10] Ismail K, Everitt B, Blatchley N, Hull L, Unwin C, David A, et al. Is there a Gulf War syndrome? *Lancet* 1999;353:179–82.
- [11] Doebbeling BN, Clarke WR, Watson D, Torner JC, Woolson RB, Voelker MD, et al. Is there a Persian Gulf War syndrome? Evidence from a large population-based survey. *Am J Med* 2000;108:695–704.

- [12] Knoke JD, Smith TC, Gray GC, Kaiser KS, Hawksworth AW. Factor analysis of self-reported symptoms: does it identify a Gulf War syndrome? *Am J Epidemiol* 2000;152:379–88.
- [13] Coker WJ. A review of Gulf War illness. *J R Nav Med Serv* 1996;82: 141–46.
- [14] Joseph SC. Comprehensive Clinical Evaluation Program Evaluation Team. A comprehensive clinical evaluation of 20,000 Persian Gulf War veterans. *Mil Med* 1997;162:149–55.
- [15] Presidential Advisory Committee on Gulf War Veterans' Illnesses. Final report. Washington, D.C.: Government Printing Office; 1996.
- [16] Defense Science Board Task Force on Persian Gulf War health effects. Defense Science Board. Final report. Washington, D.C.: Office of the Undersecretary of Defense for Acquisition and Technology; 1994.
- [17] Institute of Medicine. Health consequences of service during the Persian Gulf War: initial findings and recommendations for immediate action. Washington, D.C.: National Academy Press; 1995.
- [18] Committee on the DoD Persian Gulf Syndrome Comprehensive Clinical Evaluation Program. Institute of Medicine Division of Health Promotion and Disease Prevention. Evaluation of the U.S. Department of Defense Persian Gulf comprehensive clinical evaluation program. Washington, D.C.: National Academy Press; 1995.
- [19] Black DW, Doebbeling BN, Voelker MD, Clarke WR, Woolson RF, Barrett DH, et al. Quality of life and health-services utilization in a population-based sample of military personnel reporting multiple chemical sensitivities. *J Occup Environ Med* 1999;41:928–33.
- [20] Black DW, Doebbeling BN, Voelker MD, Clarke WR, Woolson RF, Barrett DH, et al. Multiple chemical sensitivity syndrome: symptom prevalence and risk factors in a military population. *Arch Intern Med* 2000;160:1169–76.
- [21] Zwerling C, Torner JC, Clarke WR, Voelker MD, Doebbeling BN, Barrett DH, et al. Self-reported injuries among Gulf War veterans. *Public Health Rep* 2000;115:346–49.
- [22] Voelker MD, Saag KG, Schwartz DA, Chrischilles E, Clarke WR, Woolson RF, et al. Health-related quality of life among Gulf War era veterans. *Am J Epidemiol* 2001; in press.
- [23] Committee on Measuring the Health of Gulf War Veterans. Institute of Medicine. Gulf War veterans: measuring health. Washington, D.C.: National Academy Press; 1999.
- [24] Presidential Advisory Committee on Gulf War Veterans' Illnesses. Interim report. Washington, D.C.: Government Printing Office; 1996.
- [25] Kang HK, Mahan CM, Lee KY, Magee CA, Murphy FM. Illnesses among United States veterans of the Gulf War: a population-based survey of 30,000 veterans. *J Occup Environ Med* 2000;42:491–501.
- [26] Office of the Under Secretary of Defense for Acquisition and Technology. Report of the Defense Science Board Task Force on Persian Gulf War health effects. Washington, D.C.: Department of Defense; 1994.
- [27] Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;22:719–48.
- [28] Bates DW, Schmitt W, Buchwald D, Ware NC, Lee J, Thoyer E, et al. Prevalence of fatigue and Chronic Fatigue Syndrome in a primary care practice. *Arch Intern Med* 1993;153:2759–65.
- [29] US Dept of Health and Human Services. Clinician's handbook of preventive services. Washington, D.C.: Public Health Service; 1994.
- [30] Weiss EB, Segal MS. Bronchial asthma. Boston, MA: Little Brown; 1976.
- [31] Woolson RF, Bean JA, Rojas PB. Sample size for case control studies using Cochran's statistic. *Biometrics* 1986;42:927–32.
- [32] Kreutzer RD, Neutra RR, Lashuay N. Prevalence of people reporting sensitivities to chemicals in a population-based survey. *Am J Epidemiol* 1999;150:1–12.
- [33] Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy III FV, Hahn SR, et al. Utility of a new procedure for diagnosing mental disorders in primary care. *JAMA* 1994;272:1749–56.
- [34] Spitzer RL, Kroenke K, Linzer M, Hahn SR, Williams JBW, deGruy FV, et al. Health-related quality of life in primary care patients with mental disorders: results from the PRIME-MD 1000 study. *JAMA* 1995;274:1511–17.
- [35] Ewing JA. Detecting alcoholism: the CAGE questionnaire. *JAMA* 1984;252:1905–7.
- [36] Weathers F, Litz BT, Heran DS, Huska JA, Kean TM. The PTSD Checklist (PCL): reliability, validity, and diagnostic utility. San Antonio, TX: International Society of Traumatic Stress Studies; 1993.
- [37] Chalder T, Berelowitz G, Pawlikowska T. Development of a fatigue scale. *J Psychosom Res* 1993;37:147–53.
- [38] Reynolds CF, Frank E, Thase MD, Houck PR, Jennings JR, Howell JR, et al. Assessment of sexual function in depressed, impotent, and healthy men: factor analysis of a Brief Sexual Function Questionnaire for men. *Psychiatry Res* 1988;24:231–50.
- [39] Strahan R, Gerbaski KC. Short, homogenous versions of the Marlow-Crowne Social Desirability Scale. *J Clin Psychol* 1972;28:191–93.
- [40] Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A, et al. The Chronic Fatigue Syndrome: a comprehensive approach to its definition and study. *Ann Intern Med* 1994;121:953–59.
- [41] Chyba MM, Washington LR. Questionnaires from the National Health Interview Survey, 1985–89. *Vital Health Stat* 1. 1993; 1–412.
- [42] Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance Survey (BRFSS). Atlanta, GA: Centers for Disease Control and Prevention; 1995.
- [43] Centers for Disease Control. Postservice mortality among Vietnam veterans: The Centers for Disease Control Vietnam Experience Study. *JAMA* 1987;257:790–5.
- [44] Centers for Disease Control. Health status of Vietnam veterans. I. psychosocial characteristics: The Centers for Disease Control Vietnam Experience Study. *JAMA* 1988;259:2701–7.
- [45] Centers for Disease Control. Health status of Vietnam veterans. II. physical health. The Centers for Disease Control Vietnam Experience Study. *JAMA* 1988;259:2708–14.
- [46] Centers for Disease Control. Health status of Vietnam veterans. III. reproductive outcomes and child health. The Centers for Disease Control Vietnam Experience Study. *JAMA* 1988;259:2715–19.
- [47] National Center for Health Services Research and Health Care Technology Assessment. National medical expenditure survey: questionnaires and data collection methods for the household survey and the survey of American Indians and Alaska natives. Rockville, MD: National Center for Health Services Research and Health Care Technology Assessment; 1989.
- [48] Boulet J, Boss MW. Reliability and validity of the Brief Symptom Inventory. *Psychol Assess* 1991;3:433–37.
- [49] Ferris BG. Epidemiology standardization project. *Am Rev Respir Dis* 1978;119:1–120.
- [50] Bergner M, Bobbitt RA, Kressel S, Pollard WE, Gilson BS, Morris JR. The Sickness Impact Profile: conceptual formulation and methodology for the development of a health status measure. *Int J Health Serv* 1976;6:393–415.
- [51] The University of Iowa. Female and family health questionnaire: the agricultural health study. Iowa City, IA: University of Iowa; 1993.
- [52] Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia. *Arthritis Rheum* 1990;33: 160–72.
- [53] Wolfe F. When to diagnose fibromyalgia. *Diagn Issues* 1994;20: 485–501.
- [54] Wolfe F, Ross K, Anderson J, Russell IJ. Aspects of fibromyalgia in the general population: sex, pain threshold, and fibromyalgia symptoms. *J Rheumatol* 1995;22:151–56.
- [55] Wolfe F. The fibromyalgia syndrome: a consensus report on fibromyalgia and disability. *J Rheumatol* 1996;23:534–39.
- [56] National Centers for Infectious Disease. Centers for Disease Control and Prevention. Illness in Persian Gulf Veterans, Lebanon, Pennsylvania (Questionnaire). Atlanta, GA: National Centers for Infectious Disease, Centers for Disease Control and Prevention; 1994.
- [57] Ware J. Appendix C: script for personal interview SF-36 administration. Boston, MA: Nimrod Press; 1993.

- [58] Stewart AL, Hays RD, Ware Jr. JE. The MOS short-form general health survey: reliability and validity in a patient population. *Med Care* 1988;26:724–35.
- [59] Computer-Assisted Survey Execution System (CASES). Computer-Assisted Survey Methods Program (CSM). Berkeley, CA: University of California; 1994.
- [60] Barthel CE. The need for approval, task categorization, and perceptual defense. *J Consult Clin Psych* 1926;26.
- [61] Gilroy G. Health study of Canadian Forces Personnel involved in the 1991 conflict in the Persian Gulf. Vol. I. Ottawa, Ontario: Goss Gilroy; 1998.
- [62] Jones MF, Doebbeling BN, Hall DB, Snyders TL, Barrett DH, Williams AN, et al. Methodologic issues in a population-based health survey of Gulf War veterans. Iowa City, IA: Department of Preventive Medicine and Environmental Health, The University of Iowa; 1998.
- [63] Hall DB, Woolson RF, Clarke WR, Jones MF. Health assessment of Persian Gulf War veterans from Iowa statistical techniques. Iowa City, IA: Department of Preventive Medicine and Environmental Health, The University of Iowa; 1996.
- [64] Hall DB, Woolson RF, Clarke WR, Jones MF. Cochran-Mantel-Haenszel techniques: applications involving epidemiologic survey data. In: Sen PK, Rao CR, editors. *Handbook of statistics*. Amsterdam: Elsevier; 2000.
- [65] Cochran WG. Some methods of strengthening the common chi squared tests. *Biometrics* 1994;10:417–51.
- [66] Greenland S. Interpretation and estimation of summary ratios under heterogeneity. *Stat Med* 1982;1:217–27.
- [67] Landis JR, Heyman ER, Koch GG. Average partial association in three-way contingency tables: a review and discussion of alternative tests. *Int Stat Rev* 1978;46:237–54.
- [68] Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika* 1965;52:591–611.
- [69] Kolmogorov AN. Sulla determinazione empirica di una legge di distribuzione. *G Inst Ital Attuari* 1933;4:83–91.