AD_____

Award Number: DAMD-17-03-1-0236

TITLE: Prostate Cancer Mortality and Herbicide Exposure in Vietnam Veterans

PRINCIPAL INVESTIGATOR: Steven D. Stellman, Ph.D.

CONTRACTING ORGANIZATION: Columbia University
New York, NY 10025

REPORT DATE: Apr 2006

TYPE OF REPORT: Final

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-04-2006	Final	1 APR 2003 - 31 MAR 2006

4. TITLE AND SUBTITLE

Prostate Cancer Mortality and Herbicide Exposure in Vietnam Veterans

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DAMD-17-03-1-0236

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Steven D. Stellman, Ph.D.

E-Mail: sds91@aol.com

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

Columbia University
New York, NY 10025

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. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
AIMS. The association between exposure to military herbicides and subsequent death from prostate cancer was investigated in a cohort of approx. 100,000 Vietnam veterans who applied to the Agent Orange Veteran Payment Program (AOVPP), 1985-1994. Exposure opportunity was assessed via a geographical information system (GIS) and herbicide and troop location databases previously developed by the investigators. Cause of death was ascertained via National Death Index (NDI) searches. Epidemiological analyses include comparison of age-specific prostate cancer death rates, proportional mortality ratios (PMR), standardized mortality ratios (SMR), and logistic regression methods, using unexposed veterans as a reference population. RESULTS. NDI queries yielded causes of death for claimants who had died by 12 December 1999. Overall mortality was elevated for malignancies which the Institute of Medicine deemed "sufficient evidence" of association with herbicides. Microfilm record abstraction of military unit histories was completed for prostate cancer cases and controls. Exposure assessment for veterans serving in stable Army and Air Force units was completed. Preliminary risk estimates do not show a significant relationship between death from prostate cancer and exposure to dioxin-contaminated phenoxy or other herbicides in Vietnam.

15. SUBJECT TERMS
Prostate cancer, Vietnam veterans, Agent Orange

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	45	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

INTRODUCTION

Prostate cancer is the most common cancer diagnosed among males in the US and other Western countries[1], yet little is known about its causes. Elevated prostate cancer risks have been associated with occupational exposure to phenoxy herbicides and other agricultural chemicals[2,3]. One large group of men who may be at increased risk from exposure to phenoxy herbicides consists of veterans of the Vietnam War (1962-1975), during which the US military sprayed over 19 million gallons of phenoxy and other herbicides in the Republic of Vietnam (RVN). Many of the 3.2 million servicemen stationed in RVN were assigned to military duties which involved exposure to these herbicides and to 2,3,7,8-tetrachloro-*p*-dibenzodioxin (TCDD or dioxin), a carcinogenic contaminant of Agent Orange.[4] An expert Committee of the Institute of Medicine (IOM) concluded that there is "limited/suggestive evidence" of association with prostate cancer.[5,6,7,8] but this conclusion is based on studies of non-veteran populations, or on very small veteran studies.[9] The purpose of this study is to evaluate the association between exposure to military herbicides and subsequent prostate cancer mortality in a cohort of about 100,000 claimants to the Agent Orange Veteran Payment Program (AOVPP). The AOVPP was created by the Eastern US Federal District Court (New York) to administer the settlement of a class-action lawsuit of veterans against chemical manufacturers. Fact and cause of death among the 26,000 expected deaths in this cohort are being ascertained via Social Security and National Death Index searches. Exposure is being assessed from military records abstracted from the extensive AOVPP microfilm files, using our geographic information system (GIS) database of military unit locations throughout the War, and an exposure assessment methodology that we developed under a contract from the National Academy of Sciences.[10,11] Measures of association include comparisons of age-specific rates with unexposed veterans and the general population, and comparison of proportional mortality ratios, standardized mortality ratios, and odds ratios generated from logistic regression models among subgroups with different estimated levels of exposure opportunity.

BODY

The overall project tasks outlined in the approved Statement of Work (SOW) are shown in Table 1, below.

Table 1. Statement of Work Tasks and Status

***Task 1*. To obtain the causes of death for known decedents.**	**Status**
a. Initiate application process for National Death Index and vital status searches.	Completed
b. Obtain cause-of-death file for known decedents from National Death Index NDI-Plus and incorporate into AOVPP database; resolve ambiguities.	Completed
c. Obtain vital status for applicants who were alive at time of filing claims.	Completed
d. Obtain cause-of-death data from NDI-Plus for approximately 10,000 applicants reported deceased in vital status searches.	Completed

***Task 2*. To obtain Agent Orange Exposure Opportunity Scores for AOVPP cohort members**	**Status**
a. Code military unit with UIC code based upon information in microfilm copy of military records.	Completed for prostate cancer
b. Refer to primary reference matter for supplementary military history data.	Completed
c. Use GIS to obtain Exposure Opportunity Scores from UICs.	Completed

***Task 3*. To analyze mortality rates in relation to Agent Orange exposure scores and develop final report**	**Status**
a. Overall and disease-specific mortality analysis of cohort Proportional mortality ratios (PMR) Standardized mortality ratios (SMR)	Completed
b. Mortality in relation to exposure (OR)	Completed

Synopsis. During Project Year 1 (01 April 2003 to 31 March 2004) most of **Task 1** was completed, as will be described. Project Year 2 (01 April 2004 to 31 March 2005) was devoted primarily to completing **Task 1**, to accomplishing much of **Task 2** including developing and testing software tools, and to beginning **Task 3**, which was completed during the approved no-cost extension period (01 April 2005 to 31 March 2006).

The body of this Final Report is in five sections. The first four present results of major project tasks, while the last summarizes the final conclusions.

Section 1. Vital Status and Cause of Death Ascertainment
Section 2. Selection of cases and controls for nested case-control study
Section 3. Exposure Assessment
Section 4. Data Analysis
Section 5. Conclusions

Section 1. Vital Status and Cause of Death Ascertainment

The study population consists of approximately 16,000 deceased veterans on whose behalf claims to the AOVPP were filed by survivors, and an additional 85,000 veterans who filed claims while still alive. Cause of death information was obtained for the decedent population during Year 1 via National Death Index searches. Matches were obtained for 15,962 veterans. As noted in the Year 1 Annual Report, extensive quality assurance testing was done to assure ourselves that the death records were obtained for the right veterans. Also in Year 1, vital status searches were made to ascertain which of the non-deceased veterans had died subsequent to filing his AOVPP application. A total of 14,683 veterans who had died were identified in this manner. Vital records searches indicated that 9,805 veterans died between the time they filed their claims and December 31, 1999. A second National Death Index search was submitted in April, 2004, to determine causes of death for this group. This cutoff date was determined by budgetary limitations.

The final cohort consists of all applicants to the Agent Orange Veteran Payment Program who were active duty members of the US Armed Forces between 28 February 1961 and 7 May, 1975, and who were discharged alive and were alive on 1 January 1, 1979, subject to follow-up date restrictions noted above. A further restriction was that deceased veterans on whose behalf claims were filed by survivors were excluded from the cohort if the cause of death was not from "natural causes," i.e., accident, homicide, or suicide, because these individuals were automatically ineligible for compensation and their microfilmed records may not have been complete. Finally, we created an analysis dataset for preliminary mortality studies which restricted deaths to those occurring through 31 December 1998, the last day on which the ICD-9 coding system was in use. We did this in order to avoid having to deal with minor incompatibilities with the ICD-10 system which went into effect 1 January 1999 during exploratory analyses; the great majority of deaths occurred prior to the switchover to ICD-10.

The breakdown of veterans in the analysis dataset is shown in Table 2.

Table 2a. Vital status of all claimants at time of filing of claim

Vital status at time of filing	No. of claimants
Deceased	15,962
Alive	84,910

Table 2b. Vital status of claimants who filed while still alive, as of 31 December, 1998

Vital status on 31 December, 1998	No. of claimants
Deceased	7,633
Alive	77,277
Total	84,910

Section 2. Selection of Cases and Controls for Nested Case-Control Study

A major analytical goal is estimation of relative risks for prostate cancer in relation to herbicide exposure. Because of the effort required to extract from the microfilm archives the military records needed for exposure, it is not practical to do an exposure assessment for the entire cohort. Instead, we planned a nested case-control study using all prostate cancer cases, with a sample of controls from each of the sub-cohorts used as the comparison groups.

An important principle of control selection in case-control studies is to avoid choosing as controls persons with diseases potentially related to the exposure under study[12]. We took as a guide to potentially related diseases the Institute of Medicine's listings in its biennial reviews of health effects of Agent Orange,[13] specifically those disease which the IOM regards as having "Sufficient Evidence" of an association as well as those for which "Limited/Suggestive Evidence" exists. The "Sufficient Evidence" category includes soft-tissue sarcoma, chronic lymphocytic leukemia, Hodgkin disease, and non-Hodgkin's lymphoma. The latter category includes prostate and respiratory cancers, multiple myeloma, and diabetes.

After the causes of death were known from the NDI results, we extracted controls from the pool of claimants not dying from the above diseases. We used a ratio of five controls per case, frequency matched on year of birth and state of residence (or residence of payee if that of veteran was not available.) We used this relatively high control:case matching ratio in order to make allowances for veterans with missing or incomplete records. With experience we found that approximately 24% of veterans fell in this category: 12% had only partial data, 8% had no usable data, and 4% were ineligible. Table 3 shows the final numbers of cases and controls for the nested study.

Table 3. Numbers of cases and controls ascertained for nested prostate cancer case-control study

	Cases	Controls	Total	Percent
All Vietnam tour records available	161	761	922	71.0
Partial Vietnam tour records only	43	191	234	18.0
No usable data	18	75	93	7.2
Ineligible	13	36	49	3.8
Total SSNs	**235**	**1063**	**1298**	**100.0**

Section 3. Exposure Assessment

Exposure metrics in this study are expressed via a set of exposure opportunity indexes (EOIs) which we previously developed and validated under a contract from the National Academy of Sciences. The concepts and methods have been published[10,11]. To summarize, we developed a geographic information system (GIS) for Vietnam, whose layers include databases of all known herbicide spraying (the so-called HERBS file, which we extensively cleaned and

edited) as well as other types of military and civilian facilities. We have also developed an extensive set of location history databases for military units. Exposure assessment for an individual cohort member consists of the following steps:

1. Extract his military records from the AOVPP microfilm archives
2. Code his military unit history and military occupation specialties for each unit in which he served during his tour or tours of duty in Vietnam
3. Create a chronological list of unit location records (the "location history) by referencing the unit location databases
4. Apply our EOI algorithms to his location history records to obtain a set of exposure metrics (direct hits within 0.5 km, 1 km, 2 km, 5 km, and E4 score weighted by time and distance).

Details of these four steps have been presented in previous Reports. We will summarize them here and refer to Appendices to provide detailed documentation.

Step 1: Extract military records. This procedure is described in Appendix A. Military records are abstracted from the microfilm document archives of the AOVPP using a Minolta MS-6000 scanner. The operator of the scanner is prompted by a Microsoft Access database which has been programmed with the locations of desired documents (located on over 850 reels of film). The process has been optimized to the extent possible by retrieving documents in physical order rather than veteran order. The documents images in pdf format are placed in holding folders and periodically reorganized as each veteran's dossier is completed.

In all, we identified 7,127 documents in the microfilm archives potentially relevant to assessment of exposure in the 1,298 prostate cancer cases and associated controls. All documents were retrieved and reviewed, and 6,065 were scanned as pdf image files (on examination the remainder was deemed not relevant to exposure, blank, of poor quality, or for the wrong veteran). For veterans with at least one document on file, an average of 5.15 documents was identified and an average of 4.42 was scanned.

Step 2: Code military unit history. This task consists of abstracting military unit histories from the microfilm records that were scanned in Step 1. Procedural details are provided in Appendices B-D. We previously created a Unit Identification Code (UIC) for each of about 7,000 military units (to the company level) that served in Vietnam for the four service branches plus Coast Guard. Unit history coding is done by Ms. Francine Benjamin, the military records specialist. Veterans have been coded to over 2,500 distinct military units. The median start date for the first tour of duty was 30 December, 1967. The distribution of first-tour start dates is shown in Figure 1 below and illustrates that the great proportion of veterans were stationed in Vietnam during the peak years of herbicide spraying.

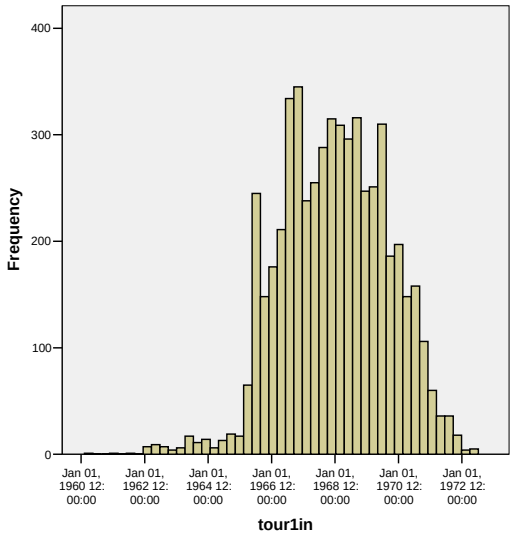


Figure 1. Frequency distribution of first tour of duty

Step 3. Once we have a listing of the military units in which a veteran served we create a chronological list of locations for each of those units in a format suitable for referencing the so-called "Tracking Databases" which contain location histories for many units. The format of these tracking databases is somewhat different from that required by the exposure assessment software. Therefore, a special set of software – "Location History Generator" or LHG – was developed to extract all available unit history data from the unit coding database and re-format it as a set of sequential records suitable for input to our exposure assessment program. Details are provided in Appendix E.

Step 4. We apply our EOI algorithms to the location history records for each study subject to obtain a set of exposure metrics (direct hits within 0.5 km, 1 km, 2 km, 5 km, and E4 score weighted by time and distance). This is the final step in preparing data for exposure analysis. The software – "Herbicide Exposure Assessment – Vietnam" – has been previously developed and validated and runs essentially instantaneously. This has been described in previous reports. Our specially developed software, Herbicide Exposure Assessment – Vietnam (HEA-V) was developed under contract to the National Academy of Sciences. A committee of the NAS Institute of Medicine ("Making Best Use of the Agent Orange Exposure Reconstruction Model") is currently tasked with proposing wider applications of our overall methodology to Vietnam veteran health – see the section on Research Accomplishments below.

Additional sources of exposure data.
Each applicant to AOVPP was given the opportunity to provide a description of personal exposure to herbicides in his Exposure Information Form, which we capture from the microfilm archives. After evaluating a large number of these forms we have concluded that the information contained in them is of limited use. We believe that it can reasonably be used to confirm high exposure scores as determined from objective EOI measures, but is subject to random misclassification that is difficult to classify either as differential or non-differential.

We also recorded military occupation specialty (MOS) codes wherever they were available in the military records. While these provide some information about other potential exposures (e.g., engine exhaust among garage mechanics), they do not necessarily provide reliable data on herbicide exposure except in a small number of instances. We hope to use these data in future analyses where control for such exposures may be important.

Section 4. Data Analysis

Task 3. To analyze mortality rates and risks in relation to Agent Orange exposure scores and develop final report.

1. Cohort mortality studies. We sought to characterize the mortality patterns in comparison to the general US population. An important aim was to see whether there was an elevated risk for illnesses for which the Institute of Medicine has concluded that there is "sufficient evidence" for an association with herbicide exposure, and to do the same for those illnesses for which it concluded "limited/suggestive evidence" exists.

We used Life Table Analysis Software (LTAS) distributed by the National Institute of Occupational Safety and Health, because it has built-in rate tables for most of the diseases of interest. Risks are reported by LTAS in the form of either standardized mortality ratios (SMRs) or proportional mortality ratios (PMRs), each with 95% confidence intervals. PMRs are easily converted to PCMRs using the method of Breslow and Day[14]. For diseases not covered by LTAS (histology-specific leukemias), and as a check, we computed an independent set of rates, using our observed deaths as numerators and official population figures as denominators. We calculated SMRs and PCMRs using standard statistical and life table methods.

Acquisition of population databases. We obtained mortality detail files for every year from 1966 through 2001 from the Office of Analysis and Epidemiology, National Center for Health Statistics. The CD's for 1966-1988 are public use files and were sent to us immediately upon request. Files for later years are subject to approval. On 28 June, 2004, we filed the appropriate applications for Compressed Mortality Files for 1989-98 and for 1999-2001. The requests were approved and the CD's were sent soon thereafter.

The population of eligible veterans was treated as two separate cohorts for the purpose of analysis. The deceased cohort consists of those veterans who were deceased at the time of filing. Analysis of deaths in this cohort is limited to PCMR analysis. Claimants who were alive when their claims were filed are termed the living cohort, and their deaths are studied using SMR methods. Approximately 40% of the living cohort were highly disabled when they filed their claims; it is well known that persons with severe disabilities have a considerably higher death rate than the general population[15]. This implies that population rates will underestimate mortality even if there is no impact of herbicide exposure. This is evidenced by an all-cause SMR of 1.93. To offset this potential bias we used the relative standard mortality ratio (RSMR), which is simply the cause-specific SMR divided by the SMR for the entire population. While this does not remove the potential bias due to high baseline mortality rates (most causes were greater than in the general population), it does permit one to assess whether specific causes are elevated relative to each other and to benchmark causes such as coronary heart disease.

The SMR results for the living cohort are shown in Table 4 and the PCMR results for the deceased cohort are shown in Table 5. The causes of death which are designated "Sufficient" by the Institute of Medicine were elevated in both cohorts. Prostate cancer was elevated in the living cohort with an SMR of 1.78. However, its RSMR was 0.92 (95% confidence interval 0.72 – 1.12).

One shortcoming of the microfilm records is that they rarely contain information about the race or ethnic background of the participant. This makes it impossible to apply race-specific mortality rates. Since prostate cancer rates are higher in blacks than in whites, failure to adjust for race could lead to biased risk estimates. We examined this issue in a sensitivity analysis in which we used US Census data for 1990 to determine the racial makeup of the zip code in which the claimant had lived (or the payee if the claimant's address was not available). We repeated the mortality analyses using only those individuals who resided in a zip code which was either at least 90% white or 90% non-white, using US white and non-white death rates, respectively, as references, and obtained results which differed very little from those in Tables 4 and 5.

The mortality analysis methods which we use provide outcome data for all causes, not just those of primary interest to the parent study. It was of additional interest to determine whether other causes were also elevated, both as a way of probing potential selection biases as well as to gather data which may be of importance for veteran health in general. Noting recent recommendations of the Institute of Medicine regarding neurological diseases, we examined mortality for two neurodegenerative diseases, amyotrophic lateral sclerosis (ALS) and multiple sclerosis (MS), both of which we found to be elevated. In the deceased cohort, the PMR for ALS was 1.23 (0.90 – 1.62) and for MS was 1.37 (1.00 – 1.80). In the living cohort the RSMR for ALS was 1.49 (.95 – 2.03) and for MS was 2.92 (1.98 – 3.86). These findings were reported at the 2005 meeting of the Society for Epidemiologic Research in Toronto.

Table 4. Standardized Mortality Ratios (SMRs) for selected causes among 84,010 disability claimants to the Agent Orange Veteran Payment Program, followed up 1989 – 1998.

IOM DESIGNATION[a]	CAUSE OF DEATH	OBSERVED DEATHS	EXPECTED DEATHS	SMR[b]	RELATIVE SMR	95% CONFIDENCE INTERVAL
Sufficient	Chronic lymphocytic leukemia	29	6.30	4.60	2.39	1.52 – 3.25
	Connective tissue cancer	40	8.75	4.57	2.37	1.63 – 3.10
	Non-Hodgkin lymphoma	219	49.53	4.42	2.29	1.99 – 2.59
	Hodgkin's disease	32	4.38	7.31	3.79	2.47 – 5.10
Limited/Suggestive	Prostate cancer	82	46.03	1.78	0.92	0.72 – 1.12
	Respiratory cancers[c]	756	385.88	1.96	1.02	0.94 – 1.09
	Multiple Myeloma	88	18.24	4.82	2.50	1.98 – 3.02
	Diabetes Mellitus	176	100.47	1.75	0.91	0.77 – 1.04
None	Esophageal cancer	71	41.63	1.71	0.88	0.68 – 1.09
	Stomach cancer	42	32.01	1.31	0.68	0.47 – 0.89
	Colorectal cancer	174	101.98	1.71	0.88	0.75 – 1.02
	Acute Myeloid Leukemia	22	12.13	1.81	0.94	0.55 – 1.33
	Heart disease[d]	1828	1133.54	1.61	0.84	0.80 – 0.87
	All Cancers	2330	1062.75	2.19	1.14	1.09 – 1.18
	All Deaths	7633	3964.25	1.93	1.00	-

[a] Sufficient = IOM category “Sufficient Evidence of an Association.” Limited/Suggestive" = IOM category “Limited/Suggestive Evidence of an Association”

[b] Adjusted for age and calendar-year using US Males as reference

[c] IOM category is larynx, trachea, bronchus, and lung combined, ICD-9 codes 161-162

[d] ICD-9 codes 390-398, 402-404, 410-414, and 420-429. This grouping includes ischemic heart disease, rheumatic heart disease, hypertension with heart disease, and endocardial diseases.

Table 5. Proportional cancer mortality ratios (PCMRs) in the deceased cohort.

IOM Designation[a]	Cause of death	Observed Deaths	PCMR[b]	95% Confidence Interval.
Sufficient	Chronic Lymphocytic Leukemia	54	1.41	1.06 – 1.81
	Connective Tissue	149	1.39	1.18 – 1.63
	Non-Hodgkin's Lymphoma	742	1.55	1.44 – 1.67
	Hodgkin's Disease	173	1.49	1.28 – 1.72
Limited/Suggestive	Prostate Cancer	111	0.60	0.49 – 0.71
	Respiratory Cancers[c]	2840	0.94	0.90 – 0.97
	Multiple Myeloma	191	1.53	1.32 – 1.76
None	Esophagus	217	0.72	0.63 – 0.82
	Stomach	270	0.88	0.78 – 0.99
	Colorectal	709	0.86	0.80 – 0.92
	Acute Myeloid Leukemia	171	1.26	1.08 – 1.46

[a] Sufficient = IOM category “Sufficient Evidence of an Association.” Limited/Suggestive" = IOM category “Limited/Suggestive Evidence of an Association”

[b] Adjusted for age and calendar-year using US Males as reference

[c] IOM category is larynx, trachea, bronchus, and lung combined, ICD-9 codes 161-162

2. Case-control analysis with exposure

As explained above and in our published work, exposure assessment is done by comparing the locations of a veteran's military units to the database of spray locations in the HERBS file using software designed for this purpose (Herbicide Exposure Assessment – Vietnam or HEA-V). We compute two types of scores: counts of "hits" within 0.5, 1,2, and 5 km, and an E4 continuous exposure score which takes into account environmental decay including that from past spraying.

This process requires a detailed inventory of actual locations of military units. As noted in previous reports, we have developed several such inventories over the years for different types of military units. The most complete database is for Army units at the company level designated as "stable," that is, those units which moved infrequently and whose locations are known very well. This covers approximately 80% of all Army personnel who served in Vietnam. The great majority of Air Force personnel were assigned to about 13 fixed air bases whose coordinates are also well known. Hence, we regard people belonging to Army "stable" units and all Air Force claimants to have known location histories which can be evaluated for exposure using our software. The great majority of Naval personnel were stationed offshore and were never exposed to herbicides (herbicides were delivered to Vietnam in civilian or merchant marine transports and not by military vessels). At present we are undertaking research at the National Archives and Records Administration to fill in locations for "mobile" Army units and for Marine units, which cover collectively about 20% of claimants.

Preliminary case-control data

As noted above, we are continuing to research locations of military units at the National Archives and Records Administration Vietnam collection in College Park, MD. Most of the units for which we do not yet have complete location data are for "mobile" Army units, such as Artillery and Infantry, which were often more heavily involved in the theater of military operations than so-called "stable" units. Herbicide spraying was generally done as tactical support for such units, so that their exposure are likely to be higher than for "stable" units. As these location databases are filled in we anticipate identifying more heavily exposed veterans.

Since these investigations are still in progress, we provide only a preliminary analysis, using simple tabulations and logistic regression. In future analyses we will use standard proportional hazards methods to take account more fully for survival time. However, even this preliminary report provides support for the feasibility of more detailed studies.

We calculated exposures for 104 prostate cancer cases and 547 controls, distributed between the two cohorts as follows:

Table 6. AOVPP Disability and Survivor claimants by case-control status.

		Claim_Type		Total
		D Disability	S Survivor	
Disease	0 Control	222	325	547
	1 Case	47	57	104
Total		269	382	651

The following Table (Table 7) shows the percentage of Stable Army and Air Force claimants in the case-control analysis. It demonstrates that over 15% of these AOVPP claimants were within 500 meters of at least one herbicide spray, and nearly three-quarters had a non-zero continuous E4 score.

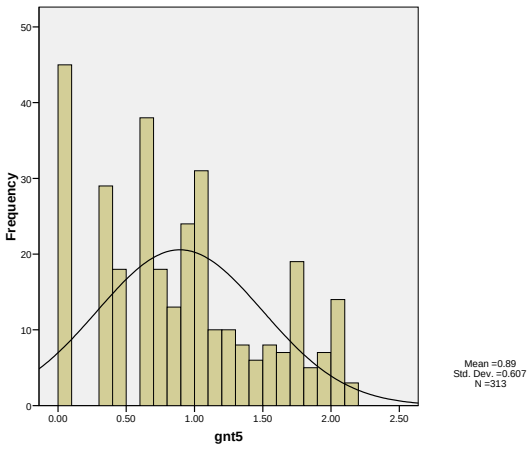
Table 7. Exposure opportunity – Stable Army and Air Force Claimants with non-zero scores

Exposure type	Disability cohort	Survivor cohort
Hits within 0.5 km	15.6%	17.8%
Hits within 1 km	21.9	22.5
Hits within 2 km	32.7	27.7
Hits within 5 km	46.8	70.1
E4 score	73.5	70.1

Figures 2a and 2b below show the distribution of the log of NT5 (hits within 5 km) and the log of the continuous E4 score for the study population, with normal curves superimposed. The Figures demonstrate that exposure ranged over several orders of magnitude and that the cohort thus includes persons with very high as well as very low exposures.

Figure 2. Histograms of log (hits within 5 km) and log (E4). Claimants with zero scores are excluded to avoid skewing which would obscure the appearance of the graphs.

2a. Log (hits within 5 km)



2b. Log (E4)

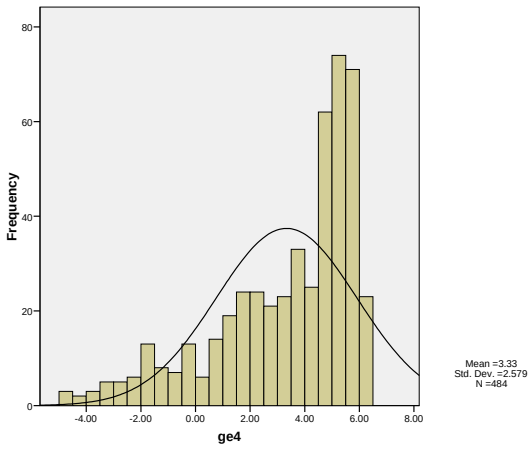


Table 8 shows the results of a logistic regression analysis, done separately for the disability and survival cohorts, to evaluate the odds ratio for prostate cancer in relation to the log of the E4 exposure score. Year of birth is included as a covariate.

Table 8. Logistic regression results for prostate cancer risk in relation to log (E4) controlling for year of birth.

Variables in the Equation

Claim_Type			B	S.E.	Wald	df	Sig.	Exp(B)
D Disability	Step 1[a]	e4k	.811	.447	3.288	1	.070	2.251
		birthyear	-.050	.024	4.279	1	.039	.951
		Constant	-.537	.828	.420	1	.517	.584
S Survivor	Step 1[a]	e4k	-.349	.316	1.219	1	.270	.705
		birthyear	-.026	.022	1.389	1	.239	.974
		Constant	-.641	.763	.706	1	.401	.527

a. Variable(s) entered on step 1: e4k, birthyear.

In the disability cohort the odds ratio estimate is 2.25 (95% confidence interval (0, 3.52). In the survivor cohort the odds ratio estimate is 0.705 (95% confidence interval (0, 0.97). Neither estimate is significantly different from the null value of 1.0. However, considering that we have not yet included veterans who are likely to have higher exposures than those observed so far, our judgment is that it is worth pursuing this hypothesis with additional subject whose exposures can be determined by archival research that we have demonstrated is feasible.

KEY RESEARCH ACCOMPLISHMENTS

- Completed vital ascertainment of over 84,000 members of the living cohort through 1999
- Obtained causes of death for all 7,633 known decedents in the living cohort through the National Death Index, including special processing for those dying in New York City
- Identified 235 cases and 1,149 controls for nested prostate cancer case-control studies
- Identified 556 cases and 2,731 controls for comparison studies of soft-tissue sarcoma, chronic lymphocytic leukemia, and acute myelogenous leukemia
- Identified and examined 7,127 documents in the microfilm archives potentially relevant to prostate cancer case-control study
- Scanned, abstracted, and coded tours of duty and military units records from 6,065 documents in the prostate cancer case-control study and began records abstraction for comparison case-control studies
- Carried out standardized mortality analysis of living cohort
- Carried out proportional cancer mortality analysis of decedent cohort
- Carried out mortality analysis for amyotrophic lateral sclerosis and multiple sclerosis
- Completed exposure assessment phase for nested case-control studies
- Carried out risk regression analysis for prostate cancer in relation to herbicide exposure, showing an elevated but non-significant odds ratio in relation to the continuous E4 exposure score within the disability cohort but not the survivor cohort.

REPORTABLE OUTCOMES

A manuscript is in preparation which presents the findings shown in Tables 6 and 7. A preliminary version of these data was presented at the 2005 joint meeting of the Society for Epidemiologic Research and the Canadian Society for Epidemiology and Biostatistics in Toronto. In addition to the cancer data shown in the Tables, we observed and reported increased proportional excesses for two neurodegenerative diseases: amyotrophic lateral sclerosis and multiple sclerosis, as described above.

PRESENTATIONS AT MEETINGS

Cancer Mortality in a Cohort Of 100,000 Vietnam Veterans. S.D. Stellman, C. Tomasallo, J.M. Stellman. Society for Epidemiologic Research, Toronto, June, 2005.

Neurodegenerative Disease Mortality in Vietnam Veterans Exposed to Phenoxy Herbicides. C.D. Tomasallo, S.D. Stellman, J.M. Stellman. Society for Epidemiologic Research, Toronto, June, 2005.

Heart disease in Vietnam-Era veterans: a 14-year follow-up of American Legionnaires. CD Tomasallo, SD Stellman, JF Sommer, Jr., JM Stellman. Society for Epidemiologic Research, Toronto, June, 2007.

CONCLUSIONS

The first year of this study was devoted to establishing procedures for efficient identification and abstraction of military and medical records from the large microfilm archives (860 reels, over 1.2 million documents), and to initiation of vital status ascertainment and cause of death determination.

In the second year vital status ascertainment was completed, the nested case-control studies of prostate cancer and several other comparison endpoints was begun, and preliminary mortality analyses were carried out for decedent veterans whose claims were filed by next of kin, and for veterans who filed claims while still alive. In both subcohorts we determined that four malignancies were elevated for which the Institute of Medicine determined that "sufficient evidence" of an association with herbicide exposure exists: soft-tissue sarcoma, non-Hodgkin lymphoma, Hodgkin disease, and chronic lymphocytic leukemia. We made a similar finding for multiple myeloma, whose evidence the IOM classifies as "suggestive/limited." Prostate cancer was significantly elevated in the living cohort (SMR = 1.78, 95% confidence interval 1.42 – 2.19), but was not significant after adjustment for overall cohort mortality (RSMR = 0.92; 0.72 – 1.12). We also found elevated mortality from ALS and MS.

Our preliminary findings are consistent with the conclusions of the Institute of Medicine regarding cancers associated with exposure to military herbicides in Vietnam, especially those such as Agent Orange which were contaminated with dioxin, at least for "sufficient evidence" cancers. For "limited/suggestive evidence" cancers, including prostate, overall mortality is not elevated.

Exposure assessment was completed during the no-cost extension Year 03, along with analyses of cancer risk in relation to exposure opportunity. Preliminary risk analyses of veterans in stable Army and Air Force units do not show a significant relationship between death from prostate cancer and exposure to dioxin-contaminated phenoxy or other herbicides in Vietnam, but the risk in the disability cohort was 2.25 (95% confidence interval 0, 3.52).

Military records, when available, provided highly reliable information on the military unit assignments and tour of duty dates for veterans. However, these were available for only 71% of all veterans, with slightly fewer being available for cases (68.5%) than for controls (71.6%). In future studies it would be advisable to find supplemental sources of military records, such as Unit Rosters, Morning Reports, and other rosters that can often be found in at the National Archives or in private archives.

It is to be emphasized that while none of the individual procedures for data location extraction is technologically complex, these procedures are part of an overall process

which so far as we know is unique and was, until now, untested. The many steps embody a smooth data stream from historical records to continually updated relational databases and involves human judgment at critical points. There were several points at which we needed to develop special-purpose file management macros under the Windows XP operating system in order to reorganize record images for individual veterans. This unavoidably added time to the overall project but minimized human errors and increased confidence in data quality. Because of the unique nature of the process, we have included key specification documents as Appendices. Appendix 1 is a set of functional specifications for the program to compare military unit assignments for individual veterans with existing military unit location databases, for the purpose of generating a file that can be read by the exposure software. Appendices 2 and 3 are, respectively, instructions for scanning indexed microfilm documents and for abstracting military unit assignments from those scanned microfilms.

LIST OF PERSONNEL RECEIVING PAY FROM THIS RESEARCH EFFORT

Steven D. Stellman, PhD MPH	Principal Investigator
Jeanne M. Stellman, PhD	Co-Investigator
Francine Benjamin	Military records expert
Daniel Kabat	Research assistant
Nicholas Marian	Research assistant
Nori Murakami	Research assistant
Carrie Tomasallo	Research assistant
Nick Wada	Research assistant
Stephen Yang	Research assistant

REFERENCES

1 Hsing AW, Tsao L, Devesa SS. International trends in prostate cancer incidence and mortality. Int J Cancer 2000; 85:60-67.

[2] Blair A, Zahm SH. Cancer among farmers. Occup Med 1991; 6:335-354.
[3] van der Gulden JWJ, Kolk JJ, Verbeek ALM. Prostate cancer and work environment. Am J Occup Med 1992; 34:402-409.
[4] Stellman JM, Stellman SD, Christian R, Weber T, Tomasallo C. The extent and patterns of usage of Agent Orange and other herbicides in Vietnam. Nature 2003;422:681-7.
[5] Institute of Medicine 1994. Committee to Review the Health Effects in Vietnam Veterans of Exposure to Herbicides. Veterans and Agent Orange: Health Effects of Herbicides Used in Vietnam. Washington, DC: National Academy of Sciences Press.
[6] Institute of Medicine. 1996. Veterans and Agent Orange: Update. Washington, DC: National Academy of Sciences Press.
[7]Institute of Medicine. 1998. Veterans and Agent Orange: Update. Washington, DC: National Academy of Sciences Press.
[8]Institute of Medicine. 2001. Veterans and Agent Orange: Update 2000. Washington, DC: National Academy Press.
[9] Giri VN, Cassidy AE, Beebe-Dimmer J, Smith DC, Bock CH, Cooney KA. Association between Agent Orange and prostate cancer: a pilot case-control study. Urology. 2004 Apr;63(4):757-60; discussion 760-1.
[10] Stellman JM, Stellman SD, Weber T, Tomasallo C, Stellman AB, Christian R, Jr. A geographic information system for characterizing exposure to Agent Orange and other herbicides in Vietnam. Environ Health Perspect 2003;111:321-328.
[11] Stellman SD, Stellman JM. Exposure opportunity models for Agent Orange, dioxin, and other military herbicides used in Vietnam, 1961-1971. J. Expos. Analysis Environ. Epidemiol. 14: 354-362, 2004.
[12] Rothman KJ, Greenland S. Modern Epidemiology, 2nd Ed (1998). Philadelphia: Lippincott-Raven.
13 Institute of Medicine. Veterans and Agent Orange: Update, 2002. Washington, DC: National Academies Press, 2003.
[14] Breslow N E, Day N E. Statistical Methods in Cancer Research, Volume II. The Design and Analysis of Cohort Studies (IARC Scientific Publications No. 82). Lyon, International Agency for Research on Cancer, 1987.
[15] Social Security Administration Office of Disability Programs. "Disability Evaluation Under Social Security." SSA Pub. No. 64-039. Washington, DC, January, 2005.

APPENDICES

Appendix 1. Specification for Location History Generator

Appendix 2. Instructions for Form A

Appendix 3. Instructions for Form B

Appendix 1. Specification for Location History Generator

Program input: list of military unit assignments for individual veterans and military unit location databases

Program output: location history file formatted for use by Herbicide Exposure Assessment – Vietnam exposure software

This file is c:\Orange\Location History Generator\Specification for Location History Generator.doc. It was copied into the current folder on May 9, 2005 and replaces all other versions.

1. Introduction

1.1 Purpose. The purpose of this document is to present to the developer and tester the specification for a Location History Generator (LHG).

1.2. Scope. This document contains a complete description of the functionality of LHG.

1.3. System Overview. The primary function of this program is to transform a veteran's military unit assignment history into a location history that is acceptable as valid input to Herbicide Exposure Assessment – Vietnam (HEA-V).
The primary user of LHG will be a researcher who wishes to conduct a herbicide exposure assessment for one or more US Vietnam veterans. The exposure assessment process consists of four steps:

Task	Software used	Person
a. List in chronological order all military units to which the veteran was assigned while in Vietnam. This list is the military unit assignment history.	Scanning software Form B	Francine Benjamin
b. Compare that list against a master list of all military units, their locations, and the dates during which they occupied those locations.	LHG	Carrie Tomasallo
c. Using the results of that comparison, construct a chronological list of the locations that the veteran occupied during his stay in Vietnam, and the dates he occupied each location. These are the location history records.	LHG	Carrie Tomasallo
d. Use the location history records generated in step (c) as input to the exposure assessment software HEA-V, to produce an exposure vector for the veteran.	HEA-V	Jeanne Stellman Steven Stellman Carrie Tomasallo

2. Definitions

2.1 Glossary

AOVPP stands for Agent Orange Veteran Payment Program. The AOVPP was created by the US Federal District Court in 1985 as a mechanism for handling claims for compensation filed by veterans as the result of a lawsuit against chemical manufacturers of military herbicides used in Vietnam. The program was terminated in 1994.

Date systems. Two equivalent date systems are used by this system: ordinary dates and day numbers. Every day of the Vietnam War period is assigned a number beginning with 1 (January 1, 1961) and ending with 4,017 (December 31, 1971), the last day of the final year of herbicide use. Dates and day numbers may be converted back and forth using the DATES table (defined later in this document).

An entity is an individual, military unit, or location for which an exposure assessment is to be carried out.

The Exposure Information Form (EIF) is a four-page form that was completed by claimants to the AOVPP. It contains a list of military units and the dates the veteran was assigned to them, a narrative section concerning possible herbicide exposure, and a checklist of Vietnam locations and dates. Every veteran should have an EIF in the microfilm archives.

HEA-V (Herbicide Exposure Assessment – Vietnam) is a software system created by Stellman and Greene Consulting which reads location history records as input and produces exposure records as output.

A location history record is a record which describes a single location of one entity during one time interval. The time interval may be as short as one day or as long as many years. An entity may require several location history records to describe its movement over time. It consists of the following information:

- ID (text)
- Longitude (up to 5 decimal places)
- Latitude (up to 5 decimal places)
- Date-in
- Date-out

The microfilm archives is a set of 865 cartridges (reels) of 35mm microfilm on which all paper documents relating to claims filed with the AOVPP reside. Each cartridge contains up to 6,000 frames. Individual documents may occupy several frames. There are approximately 1.2 million documents in the archives. A computerized index lists each document by SSN, cartridge, frame, and document type.

A military record for the purpose of this document is an archival record, usually on microfilm, which contains information about an individual's assignments and duties while in the US Armed Forces. Military records are usually official documents issued by the Dept. of Defense. Each branch of the military has its own set of documents with characteristic document identifiers and formats. A description of common document types is given in the ..\..\Epidemiologist Guide\Epidemiologist guide to military records_final.doc.

Mobility factor. One of three possible designations for a military unit that describes the movement of its location over time:

Mobility factor	Definition
Stable	Did not move from base camp (or did so rarely, and usually to a new permanent quarters)
Stable with mobile elements	Otherwise stable units which included individuals who traveled away from regularly assigned base camps in order to accomplish the unit mission
Mobile	Routinely left base camp to assigned field destinations, carried out mission, which generally included movement through the field, and then returned to base camp, landing zone or fire support base to await next mission

PointID Gridpoint identification number. A 6-digit integer that describes a unique grid within the grid system set up for Vietnam in this project. For a more complete description see ..\..\Systems Requirements Document\SRS Specification Document\SRS_April_8_2003.doc.

Unit Identification Code (UIC). A 7-digit number which uniquely identifies a military unit that was assigned to Vietnam. The first digit is the branch of service. The second digit may be 0, 1, or 2 for Army units. It is zero for for all other branches. The lowest level unit which has a UIC is a company or its equivalent. There are UICs for units up to the size of a division.

First digit	Branch
1	Army
2	Marines
3	Air Force Units
4	Air Force Bases
5	Navy
6	Coast Guard
7	Vietnam Names

Unit coding is the action of determining for an individual veteran the UIC or UICs for all units to which he or she belonged while in Vietnam. It is usually performed by the military records expert, Francine Benjamin, with the collaboration of the senior military expert, Lt. Col. Richard S. Christian (USAR, ret.).

2.2 Input Tables

All input data that are required for this program (LHG) are in two Microsoft Access databases:

Name of Database	Nickname	Contents of database
Stable_Database.mdb	ULDB	Locations of all Stable units over time
Unit_Assignments_AOVPP.mdb	VetUA	Unit assignments for individual veterans

[This was previously called aovpp table for carrie 110103.mdb]

2.2.1 Description of databases

Database 1: Stable_Database.mdb (nickname ULDB)

Until further notice, this program deals only with Stable units. The database named above covers Stable units only. It contains three tables:

Table name	No. records	Description
Stable data_condensed 062603	10,481	Chronological listing of all known locations for each stable unit.
UICLIST	4,778	Definitive listing of all Army units with a description of each entry.
UIC_BNCODE	4,778	Same as UICLIST but defined at the battalion level.

Layout of tables
Stable data_condensed 062603

Field	Content
UIC	Unit Identification Code

UTM	UTM coordinate of the unit
dayin	day number on which the unit first arrived at that location*
dayout	day number on which the unit left that location*
Lat	Latitude of the location
Long	Longitude of the location
PointID	PointID of the location*
location name	Name of location, nearest town, or other nearby landmark – for identification only

* For additional description see ..\..\Systems Requirements Document\SRS Specification Document\SRS_April_8_2003.doc

Sample excerpt from Stable data_condensed 062603 for one particular UIC (1002003):

uic	utm	dayin	dayout	Lat	Long	PointID	location name
1002003	BR470455	1705	2099	13.96858	108.65811	125466	AN KHE
1002003	BR471445	2100	2403	13.95955	108.65912	125246	AN KHE
1002003	BR470445	2404	2434	13.95955	108.6582	125246	AN KHE
1002003	BR471445	2435	2891	13.95955	108.65912	125246	AN KHE
1002003	XT960490	2892	3771	11.29319	106.79568	59140	PHUOC VINH

Database 2: Unit_Assignments_AOVPP.mdb (nickname VetUA)

This database has one table:

Unit assignment data abstraction table
[A previous version had two tables: Unit assignment data abstraction table and Unit assignment data abstraction table_Units 9-16. The first table contained data for units 1-8. These two tables have been condensed into a single table. Two tables were used rather than one in order as a convenience to the military records expert who must abstract data from microfilm copies of the military records. The first table contained data on up to eight units. If a veteran's records indicated that he or she belonged to more than eight units, data for the ninth and above were entered into the second table in a format identical to the first. Both tables could be linked to form a single master table for all units for all veterans.]

This Table has 35,206 records, which is the total number of AOVPP veterans classified as deceased. Each record contains military unit assignment data for a single veteran. The table is continually being updated, with the ultimate goal of recording a complete military unit assignment history for all deceased claimants.

Database 2 has room to record up to sixteen military units per veteran. There are four groups of variables:

- Identification fields
- Tour of duty fields
- UIC fields
- Veteran data status fields

Identification fields:

Field name	Content
Date-modified	Last date this record was updated
SSN	Social Security Number
Vet Name	Name of veteran
Vet DOB	Date of birth as mm/dd/yy
Diff_SSN	Corrected or alternative Social Security Number
Diff_DOB	Corrected or alternative date of birth
Diff_Name	Corrected or alternative name
DOB_verified	Checked if DOB in original database has been verified

Tour of duty fields.
Up to four tours of duty can be recorded for a veteran. For each tour, there are two checkboxes. The first is checked if the date the tour began could not be found in the veteran's record but was imputed from other information available. The second is checked if the date the tour ended was imputed.

Field name	Content
tour1In	Date first tour of duty began
tour1out	Date first tour of duty ended
t1_DayIn_imputed?	Checked if dayin is imputed
t1_DayOut_imputed?	Checked if dayout is imputed
tour2in	Date second tour of duty began
tour2out	Date second tour of duty ended
t2_DayIn_imputed?	Checked if dayin is imputed
t2_DayOut_imputed?	Checked if dayout is imputed
tour3in	Date third tour of duty began
tour3out	Date third tour of duty ended
t3_DayIn_imputed?	Checked if dayin is imputed
t3_DayOut_imputed?	Checked if dayout is imputed
tour4in	Date fourth tour of duty began
tour4out	Date fourth tour of duty ended
t4_DayIn_imputed?	Checked if dayin is imputed
t4_DayOut_imputed?	Checked if dayout is imputed

UIC fields.

Data on between 1 and 8 military units (UICs) can be recorded. A veteran can belong to the same UIC at two different time intervals. Each UIC has 8 data fields, for a total of 64 UIC fields. There are 8 sets of fields like the following, one per unit.

Field name	Content
UIC1	UIC code for the first unit to which veteran was assigned
Unit1in	Date veteran was assigned to unit
Unit1out	Date veteran left unit
Day1In_imputed?	Checked if the day of the month was imputed (e.g., the 1st or the 15th) from other information
Date1In_imputed?	Checked if the entire date the unit assignment began was imputed
Day1Out_imputed?	Checked if the day of the month was imputed (e.g., the 1st or the 15th) from other information
Date1Out_imputed?	Checked if the entire date the unit assignment ended was imputed
MOSs_U1	This is a free-text field which may contain data on one or more military occupation specialties (MOSs) listed in the veteran's record for this unit assignment

Veteran data status fields:

Field name	Content
Notes	Free-text field for coder's remarks
More Units?	Checked if data are recorded for more than eight units
Data Source	This field can take one of three values: "Mil Record" Data were abstracted from a military record "EIF" Data were abstracted only from Exposure Information Form "BothMil Rec+EIF" Data were abstracted from both sources
Done-with-veteran	Checked if all available data for this veteran have been entered
Partial	Checked if only a partial military history is available from records
No Usable Data	Checked if there is no useable data for this veteran

DATES table

Every day of the Vietnam War period is assigned a number beginning with 1 (January 1, 1961) and ending with 4,017 (December 31, 1971), the last day of the final year of herbicide use. Dates and day numbers may be converted back and forth using the DATES table.

The DATES table is used to convert between day number and calendar date. This table is needed because most queries will have residence dates connected with them. The user supplies the residence dates in calendar format (e.g., 5/27/74) but calculations of E4 use day numbers. To convert calendar dates to Day values that can be used with the Exposure_Master table, retrieve the row where the Date column matches the date in question. Day values can be converted back to dates by looking up the row with the Day value in question. This table can convert Week values as well.

Name	Type	Description
Date	Date/Time	Calendar date. Every day between 1/1/61 and 3/31/73 is represented
Day	Integer	Day number. Every day is numbered consecutively from 1/1/61 = 1 through 3/31/73 = 4473
Week	Integer	Week number. Consecutive from 1 through 639
Month	Integer	Month number. Consecutive from 1 through 147
Year	Integer	Year number. Consecutive from 1 through 13

<table>
<tr><td>Name</td><td>FR-DR1: Input Data Verification</td></tr>
<tr><td>Summary</td><td>The system must verify imported data.</td></tr>
<tr><td>Rationale</td><td>A rigid and consistent set of rules is required which governs flagging and handling of potential date conflicts in tours of duty and unit assignment histories for an individual veteran.</td></tr>
<tr><td>Requirements</td><td>When location data are imported from the Form B database, a set of validation rules must be enforced. When user input violates these rules, the record the user is attempting to input may not be saved until the problem is resolved or the operation is cancelled. When data being imported violate these rules, the offending records are to be skipped and displayed to the user after the import is finished.

This database has one table:
Unit assignment data abstraction table
This table has 35,206 records, which is the total number of AOVPP veterans classified as deceased. Each record contains military unit assignment data for a single veteran. The tables are continually being updated, with the ultimate goal of recording a complete military unit assignment history for all deceased claimants. The database has room to record up to sixteen military units per veteran.

Identification Fields
SSN — Every record must have a valid 9-digit Social Security Number [A valid Social Security Number aways consists of 9 numeric digits, which may contain leading zeros.]
Vet Name — Every record must have a name
Vet DOB — Every record must have a valid birth date that falls between January 1, 1901 and December 31, 1960.

Tour of duty fields
Tour of duty fields may contain data or they may all be blank.
If any tour of duty fields are non-blank, they must obey the following rules.
Up to four tours of duty may be recorded. If more than one tour of duty is recorded, they must all be listed in chronological order by in-date.
There may be no gaps in the data fields between tours. In other words, if two tours are listed, data for the second must begin with the field tour2in, the third must begin with the field tour3in, and the fourth must begin with the field tour4in.
If a tour exists, it must have a valid in-date and a valid out-date. The out-date cannot precede the in-date.
No tour of duty may begin earlier than January 1,1960.
No tour of duty may end later than December 31, 1975.
If there is more than one tour, it is permissible for the out-date of one tour and the in-date of the next tour to be the same day. However, two tours may not overlap. That is, the in-date of one tour may not come before the out-date of the preceding tour.

UIC fields
A "block" of UIC data consists of the following eight data items:

<table>
<tr><th>Data item</th><th>Data type</th><th>Sample field name</th></tr>
<tr><td>UIC</td><td>Number</td><td>UIC1</td></tr>
<tr><td>in-date</td><td>Date/Time</td><td>Unit1in</td></tr>
</table>
</td></tr>
</table>

		in-day-imputed	Yes/No	Day1In_imputed?
		in-date-imputed	Yes/No	Date1In_imputed?
		out-day-imputed	Yes/No	Day1Out_imputed?
		out-date-imputed	Yes/No	Date1Out_imputed?
		MOSs	Text	MOSs_U1
	Up to sixteen UIC blocks may exist in the database for one veteran. They must fill consecutive blocks in the veteran's data record in chronological order by in-date. There may be no gaps in the data fields between UICs. In other words, if two UICs are listed, data for the second must begin with the field UIC2. Every non-blank UIC block must contain a valid UIC. A valid UIC is a seven-digit integer whose first digit is 1-5 and whose second digit is zero. Every non-blank UIC block must contain a valid in-date and a valid out-date. A valid in-date must be on or after January 1, 1961. A valid out-date must not fall after May 7, 1975. Within every UIC block, the in-date may not come after the out-date. However, the in-date and out-date may be the same. In other words, it is possible for a unit assignment to last only one day. If there is more than one UIC, it is permissible for the out-date of one UIC block and the in-date of the next UIC block to be the same day. However, two UIC time intervals may not overlap. That is, the in-date of one UIC may not come before the out-date of the preceding UIC. The in-date for the first UIC may not precede the in-date for the first tour of duty. The out-date for the last UIC may not come after the out-date for the final tour of duty.			
References	FR-DR2: Dataset Import Error Handling			
Revision history	11/22/03 Created (Steven Stellman) 11/26/03 Revised (Steven Stellman)			

Name	UC-AR1: Import unit assignment data
Summary	The researcher imports unit assignment data from an external database
Rationale	The main purpose of the LHG software is to transform unit assignment data for one or more veterans into a set of location histories. This use case describes the process whereby the researcher initiates and carries out the input of unit assignments.
Basic Course of Events	1. This use case begins as soon as the system has been started. 2. The researcher indicates that an external dataset containing unit assignments is to be read in. 3. The system responds by asking the user to provide a dataset name. 4. The user enters a name. 5. The system prompts the user to select an input database. 6. The researcher selects a Microsoft Access file containing the desired input. 7. The system displays the available Tables within the selected Microsoft Access file and prompts the user to select a Table. 8. The user selects a Table from the Microsoft Access file. 9. The system responds by reading data from the selected table and subjecting its records to a set of validation rules (FR-DR1). 10. After the data have been read, if there were no input errors or violation of validation rules, the system so informs the researcher.

	11. The system stores the data under the dataset name provided by the user in Step 4.
Alternative Paths	If any records contain violations of validation rules given in (FR-DR1), the system displays a message that indicates the number of invalid input records. It writes those records to a text file which contains the invalid data and a description of the violation. It displays the name of that invalid-data text file so that the researcher can find and open it. The system then offers the user the option of either aborting further action or proceeding with creation of location history records for valid input data.
Exceptions	None
Trigger	Initiation of program
Assumptions	None
Preconditions	
Postconditions	
Functional Requirements	FR-DR1: Input Data Verification
Nonfunctional Requirements	
Revision History	11/22/03 Create (Steven Stellman) 11/26/03 Revised (Steven Stellman)

Name	UC-AR2: Generate location history records
Summary	Location history records are generated using unit assignment input data
Rationale	The main purpose of the LHG software is to transform unit assignment data for one or more veterans into a set of location histories. This use case describes the process whereby the system generates those records from unit assignment data that was previously read in
Basic Course of Events	1. This use case begins when the user indicates that location history records are to be generated. 2. The system prompts the user to select an existing dataset. 3. The researcher indicates the name of a dataset to be used to generate location history records. 4. The system generates the location history records using the procedure described in FR-DR2. When finished it indicates to the user that the records have been created and asks whether the user wishes to view the records 5. The user indicates that the location history records should be viewed 6. The system displays the location history records in SSN order, sorted by UIC in-date within SSN. 7. The system provides the user with the opportunity to save the location history records. 8. The user indicates that the location history records are to be saved 9. The system allows the user to indicate a filename for the location history records and a folder in which those records are to be saved 10. The user enters a filename and indicates a folder
Alternative Paths	5. The user may decline to view the location history records. If the user declines, the records are not displayed.
Exceptions	None
Trigger	A dataset with unit assignment histories has been read in (UC-AR1)
Assumptions	None
Preconditions	
Postconditions	
Functional Requirements	FR-DR2
Nonfunctional Requirements	None
Revision History	11/22/03 Create (Steven Stellman) 11/26/03 Revised (Steven Stellman)

Name	FR-DR2: Procedure for creating location history records
Summary	Location history records are generated using this procedure
Rationale	The main purpose of the LHG software is to transform unit assignment data for one or more veterans into a set of location histories. This is done by extracting location data from a master database of unit location histories, using as a lookup key the UIC code or codes for each unit to which the veteran belonged.
	Input This procedure uses information contained in two databases: Database 1: Master unit location database (ULDB) Database 2: Veteran unit assignment database (VetUA) Sample Database — Contents of database Database 1: Stable_Database.mdb — Nickname ULDB: Locations of all Stable units over time – each record contains the location of one unit (UIC) during one time interval Database 2: Unit_Assignments_AOVPP.mdb — Nickname VetUA: Unit assignments for individual veterans – each record pertains to a single veterans and contains a list of 1 to 16 units (UIC) and the time intervals of assignment Output The output of this procedure consists of two tables: Table 1: Location history records (variable number per veteran) SSN Tour number (1, 2, 3, etc.) UIC number (1, 2, 3, etc.) Day in Day out Latitude Longitude Table 2: Veteran information records (one per veteran) SSN Veteran name: First – MI – Last Veteran date of birth Number of tours of duty served Whether or not any tour of duty dates were imputed (yes/no) Number of units veteran served in Whether or not any UIC dates were imputed (yes/no) Data source: Records only, EIF only, or both Whether only a partial military history is available from records (yes/no) Percent coverage = 100 x number of days for which location data exist in the unit location database divided by the total number of days in the veteran's service history, rounded to a whole number

This procedure creates two sets of output, one containing the desired location history records and one containing some process information, such as imputed or missing data flags.

Procedure for creating location histories. The end product of this procedure will be a set of output records which reflect chronologically the locations of each UIC to which a veteran was assigned. The program will go through VetUA and process one veteran at a time, and when processing a veteran the program will proceed one unit at a time. For each unit to which the veteran belonged, the UIC in VetUA is used to extract all records in ULDB with the same UIC and which include the interval in VetUA for that UIC. There can be many matching records in the ULDB, depending on how frequently that unit moved. Selected fields from those matching records are written to the output file. The dates in the output records must be edited so as not to predate or postdate the veteran's UIC dates.

Example.

Input comes from a veteran who served in only one unit. His VetUA unit assignment record contains the following data:

SSN	111223333
UIC1	1002003
Unit1in	12/15/1966
Unit1out	6/30/1967

The veteran's unit was the headquarters and headquarters battalion of the 1st Cavalry Division. In the Stable Unit database this unit was listed in five locations between 1965 and 1971. The final location (Phuoc Vinh) is a considerable distance from An Khe. The entire representation of this unit in the ULDB Stable Unit database is:

uic	utm	datin	dayout	Lat	Long	PointID	location name
1002003	BR470455	9/1/1965	9/30/1966	13.96858	108.65811	125466	AN KHE
1002003	BR471445	10/1/1966	7/31/1967	13.95955	108.65912	125246	AN KHE
1002003	BR470445	8/1/1967	8/31/1967	13.95955	108.65811	125246	AN KHE
1002003	BR471445	9/1/1967	11/30/1968	13.95955	108.65912	125246	AN KHE
1002003	XT960490	12/1/1968	4/29/1971	11.29319	106.79568	59140	PHUOC VINH

The correct output is:

SSN	uic	datin	dayout	Lat	Long
111223333	1002003	12/15/1966	6/30/1967	13.95955	108.65912

These are the rules for constructing the location history output.

Identify every record in the ULDB for which at least one day is included in the VetUA interval. Sort them by day-in. For each ULDB record identified, create a location history output record with the following fields:

	VetUA Input ULDB Output SSN SSN UIC UIC day-in VetUA: day-in or ULDB day-in, whichever is later day-out VetUA: day-out or ULDB day-out, whichever is earlier Latitude Longitude If the veteran's UIC in the VetUA does not exist in the ULDB, there will be no location history record output.
Exceptions	None
Assumptions	None
Preconditions	All unit assignment records are assumed to have been checked using the rules stated in FR-DR1: Input Data Verification. It is also assumed that there are no date overlaps for any UIC in the ULDB.
Postconditions	
Functional Requirements	FR-DR1 FR-DR2
Nonfunctional Requirements	None
Revision History	11/22/03 Created (Steven Stellman) 11/26/03 Revised (Steven Stellman)

Appendix 2. Instructions for Form A

Procedure for locating and scanning military and medical documents in Agent Orange Veteran Payment Program microfilm archives

Instructions for use of Form A
Summary: This document explains how to use Form A to capture document images.

Where to turn for help

1. Dr. Stellman – Room 726
email sds91@columbia.edu
Tel. 212-305-4911

2. Carrie Tomasallo - offsite
email ct178@columbia.edu
Tel. 520-319-8011

3. Francine Benjamin - offsite
email fb98@columbia.edu
Tel. 718-474-1955

Background: The prostate cancer study requires us to determine an exposure score for each member of the Agent Orange Veteran Payment Program (AOVPP) cohort. This is done in three stages:

(1) determine the veteran's military unit assignments
(2) construct a location history for the veteran by linking his military unit assignment chronology to the military unit location history database
(3) use the veteran's location history as input to the exposure assessment software HEA-V

Stage 1 is composed of two specific steps that utilize the AOVPP microfilm database

(a) gather all of the veteran's documents that can be use to deduce his units (this is done using Form A as described below)
(b) extract the unit information from the collection of documents (Form B)

Scope: This document deals exclusively with Form A: determination of the veteran's military unit assignment history. A separate document deals with Form B.

Resources used:
Microsoft Access database: Scanners_NDI_case-control study_071304.mdb (actual date in filename will be current with latest date of access)

Table: Docs_all_decedents

This Table contains: A list of every veteran who applied, or whose family applied, to the AOVPP, and the microfilm location of every document that was submitted to Aetna. There are 99 different types of documents, each with a numerical code.

We are only interested in scanning the following eight types:

Doc Type	Name of Document	Condition / Requirement	%
04	Exposure Information Form	This is a 4-page, double sided form. Only scan documents which have information on at least one military unit with DATES. There are many blank forms. Do not scan completely blank forms. However, if a form has any information at all, scan all four pages.	21.4
05	Death Certificate	Usually a 1-page document.	17.6
30	DD214 – discharge document	Usually a 1-page document.	21.6
31	DD215 – discharge document	Usually a 1-page document.	0.3
32	Other military documentation	DA-20, AF-7, etc. Some can be very long and detailed.	13.0
40	Attending Physician Statement	This is a 4-page, double sided form. There are many blank forms. Do not scan completely blank forms. However, if a form has any information at all, scan all four pages.	7.7
45	Response to operator #9	This occasionally has military records. It must be checked when it appears.	8.7
99	Miscellaneous	Can be almost anything. This code was not consistently applied by Aetna.	9.7

Queries:

docs sorted by veteran
docs_sorted_by_cart_frame
Reason-not-scanned

Form:

FORM-Docs_sorted_by_cart_frame (also called Form A)

General instructions
You will try to scan every document whose type is shown in the Table above, provided it has "useful" information for us. You will learn how to decide whether a document is useful or not. When in doubt about whether a given document has useful data, scan it.

Detailed instructions:

1. Every microfilm operator has been assigned a folder corresponding to his/her name.
   Example: c:\yang.

2. Before beginning a scanning session, create inside your folder a new sub-folder with today's date and username as part of its name.

   Example: c:\Nikolas_Marian\2005_May_31_Nick.

   This is where you will store all the documents that you scan in the current session.

3. The master database for microfilm work is on the desktop. Its name reflects the most recent date that it was used.

   Example: Master_Scanning_Database_11Jan2005.mdb

4. When beginning work for the day, rename the master database to show the current date.

   Example: Master_Scanning_Database_12May2005.mdb becomes

   Master_Scanning_Database_13May2005.mdb

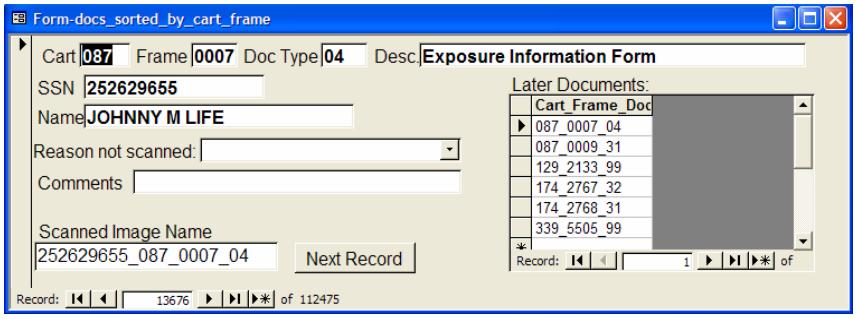
5. After renaming the database, open it into Microsoft Access directly from the desktop by double-clicking it.

6. Open the form FORM-Docs_sorted_by_cart_frame

7. Open Adobe Acrobat and arrange the Access and Acrobat windows one atop the other.

8. The form will display the first un-scanned record in the database. Ordinarily that will be your starting point. If you wish to start beyond that point, you should have recorded the record number of the last record you processed. Use the record selector guide at the bottom of the Form window to find the next record after the one that you last worked on. Alternatively, you may position the cursor in the Cart field and use ctrl-F (find) to find a particular cartridge, then do the same in the Frame window. This is now the target document.

   Form A: Looks like this:



9. Find the document you need on the microfilm by positioning the desired Cartridge at the indicated Frame. The frame number is only approximate. If the vertical scroll wheel is adjusted after arriving at a frame, it will add 4 to the frame number despite

whether the frame has changed or not. You will nearly always have to search for the document in nearby Frames.

10. Most documents need to be centered or adjusted for clarity on the microfilm. The knob on the bottom right corner of the monitor rotates the image. The knob under "SCAN" pans horizontally. The scroll wheel adjusts the image vertically. The blue cogwheel below the monitor adjusts the zoom, while the larger gray cogwheel below it adjusts the focus.

11. Open the "Minolta Series Scanner Setup" window by clicking on File, Import, Scan, OK or use Alt-F-M-N-S, then scan the document into Adobe Acrobat. Acrobat knows that some documents have more than one page. After each frame is scanned it will ask whether you are finished or wish to scan an additional page. Make sure you capture every page of a given document.

12. If the system is left idle for too long the settings in the "Minolta Series Scanner Setup" window may reset. Most often, the Resolution (DPI) will revert to 400, however the Film Type, Brightness and Contrast, or position of the image may change as well. The microfilm making a different sound during scanning indicates these changes. To delete a page use Ctrl-Shift-D.

13. If you decide not to scan a document, enter the "Reason not scanned" into the reason box. There is a drop-down menu. If the reason does not appear on the list, type in a new reason. This new reason will permanently appear in the list beginning with the next time you use Access to open the Form. Try to use existing reasons as much as possible.

14. Enter any additional comments in the Comments box.

15. Tab the cursor to the "Scanned Image Name" box and highlight the entire contents of the box. This will be a 21-character string consisting of the SSN, Cartridge, Frame, and Document Type separated by underscores. Copy this string into the buffer using Ctrl-C.

16. Return to Adobe Acrobat and use Save As to save the document into the working folder under the proposed name, by pasting it into the File Name field using Ctrl-V. The working folder is the new folder that you created at the beginning of today's session (step 2 above). For example, if the Scanned Image Name contains the string 252629655_087_0009_31, the document will be saved under the name 252629655_087_0009_31.pdf.

17. Return to the Form window and click on the "Next Record" button.

18. When a scanning session is finished, make a note of the record number of the last document scanned. Close Acrobat and Access.

19. Daily closeout procedure:

    a. Locate on the desktop the icon for the master database that you have been working with.

    b. Move the mouse directly over the icon and right-click. This will bring up a menu of actions beginning with Open and ending with Properties. One of the choices is WinZip (probably the fourth choice).

    c. Hold the cursor over the WinZip choice for a moment. A new menu will appear which contains six choices from Add to Zip File … to Configure. The second choice will be:

        Add to Master_Scanning_Database_31May2005.mdb

d. Left click that choice to start the WinZip program. Click on Use Evaluation Version.

e. WinZip will zip up the file and then terminate. The zipped file will be visible on the desktop as a new icon, with the appropriate zip-file name
Example: Master_Scanning_Database_31May2005.zip

f. Move the zip-file you just created from the desktop into the folder with today's work. Today's folder should now contain all of today's scans plus a zipped version of the current master scanning database.

g. Log onto your account on Dixie and upload today's folder in its entirety.

h. Send an e-mail to Carrie Tomasallo that you have done so.

Appendix 3. Instructions for Form B

Instructions for transcribing military unit assignment data from microfilmed military records

Instructions for use of Form B

Summary: This document explains how to use Form B to enter the military unit assignments for individual veterans.

Background. One component of the Army prostate cancer study is to determine an exposure score for each member of the Agent Orange Veteran Payment Program (AOVPP) cohort. This is done in three stages: (1) determine the veteran's military unit assignments, (2) use our extensive library of military locations in Vietnam to link each assigned unit to its location history during the time the veteran belonged to that unit, and (3) use the location history as input to the exposure assessment software HEA-V.

To carry out stage 1 for a veteran requires two specific steps:

(a) gather together all of the veteran's documents that can be use to deduce his units

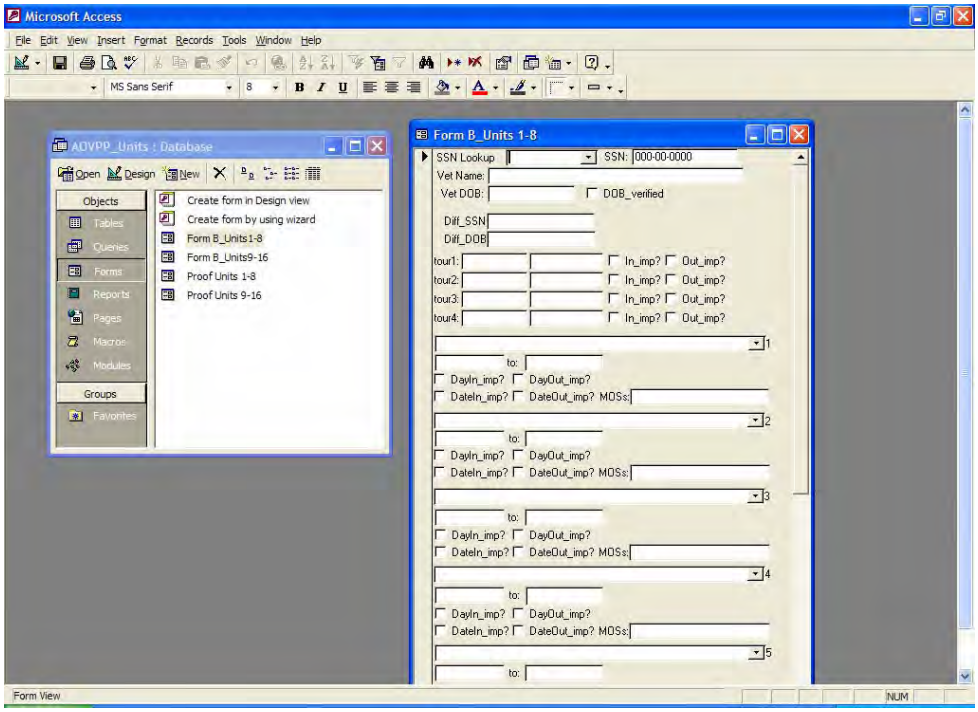
(b) extract the unit information from the collection of documents.

To carry out these steps requires use of a microfilm database and two forms called Form A and Form B.

Scope: This document deals exclusively with Form B: determination of the veteran's military unit assignment history. A separate document deals with Form A.

Resources used:
Microsoft Access database: AOVPP_Units.mdb

1. Open the database – AOVPP_Units.mdb
2. Choose the Form button from the list of database objects
   Open Form B_Units1-8 – the screen should look something like this:



3. Enter the SSN of the first veteran to be investigated into the first field of the form, labeled "SSN lookup". The veteran's name and DOB will appear. Confirm that this is the correct veteran. If the

form has a date of birth, compare it with the DOB entry in the form. If they match, check the DOB box .

If you determine that the SSN or DOB are incorrect, enter the correct data in the blank fields Diff_SSN or Diff_DOB as appropriate.

4. Enter the start and finish dates for up to four tours of duty. If dates are not completely known use the following imputation rules.

All date fields require a complete date: mm/dd/yy (2-digit year).

Tour of duty dates:

Only enter a date for tours 1-4 if you know at least the month and year.

You may impute the day of the month as 1 (in) or 28-31 (out). If you impute the day of the month, check the appropriate imputation box.

Unit assignment dates:

If you know the month and year but not the day of the month, you may impute the day of the month as 1 (in) or 28-31 (out). If you impute the day of the month, check DayIn_Imp and/or DayOut_Imp as appropriate.

If you impute the entire date from the tour of duty or other data, enter a complete date (mm/dd/yy) and check DateIn_Imp and/or DateOut_Imp as appropriate.

5. Enter any available MOS codes in the MOSs box. Enter as many MOS's as available for given unit, separated by commas.

6. Repeat through unit 8 if necessary.

7. If the veteran is done, check the “Done-with-veteran” box. Then click on the “Next Record” button.

8. If there are more than 8 units, check the “More Units?” box, then go to the Microsoft Access database object selector. Click on Forms, then click on FORM_B_UNITS_9-16 to bring up the window for Units 9 through 16.

9. Enter Units 9 and higher. When done click the “Next record” button. (There is no “Done-with-veteran” box on this form.) Then close the Units 9-16 Form (click on little x in upper right-hand corner) and return to the Units 1-8 Form.
10. If the veteran is done, check the “Done-with-veteran” box. Then click on the “Next Record” button.
11. If the veteran is not done, simply click on the “Next Record” button.
12. Enter the SSN of the next veteran to be entered into the SSN box. You will have to over-write the SSN of the veteran you just finished.

When you are finished entering a day’s worth of veterans, proceed to the proof-reading phase by carrying out the following steps.

1. Close the Units 1-8 Form.
2. In the Microsoft Access objects window, click the Module tab to bring up a list of Modules.
3. Select Module_Open and Run Step1 Queries in design view. Do not double-click it.

4. Select all the lines of code in the Module. You can do this either by pressing Ctrl-A or by clicking Edit – Select All.

5. Run the Module. You can do this in any of three equivalent ways:
   Click on the small triangle that appears in the menu line near the top of the screen
   Click the sequence Run – Run Sub-User Form
   Press the F5 function key
6. Press the enter key every time a prompt appears (32 times) until there are no more prompts
7. Close the Visual Basic Module window by clicking on the X in the upper right-hand corner of the window
8. In the Microsoft Access objects window, click the Queries tab to bring up a list of Queries. Open Step2_Making Report for Proofing Data in design view.
9. The very first column in this Query contains the field “date-modified.” A date or a range of dates appears in the row labeled “Criteria.” Change this field to reflect the date or range of dates on which data were entered which need to be proofed.

The correct format for a single date is: #7/23/2003#
The correct format for a range of dates is: >= #7/23/2003# And <=#7/30/2003#

10. Save the query by clicking on the diskette icon on the menu line. Close the query by clicking on the little x in the upper right-hand corner of the window, but don't run it yet.
11. In the Microsoft Access objects window, click the Reports tab to bring up a list of Reports. Click once on Report for Proofing Data. Then press the enter button to run the report. You can modify the view with the zoom control. Verify that the report covers the dates desired. Print it out and proofread the data. Enter corrections on the paper copy, which will be saved in a notebook.

When you are finished proofreading a set of data, make corrections as follows.

1. In the Microsoft Access objects window, click the Forms tab to bring up a list of Forms. Double-click on Proof Units 1-8 and make the necessary corrections. When done, click the “Next Record” button. If there are more than 8 units, return to the Forms window and double-click on Proof Units 9-16.

2. Notice that the report form condenses data to save space. Check-box data are represented by a 0 for No and a –1 for a check mark (same as Yes).