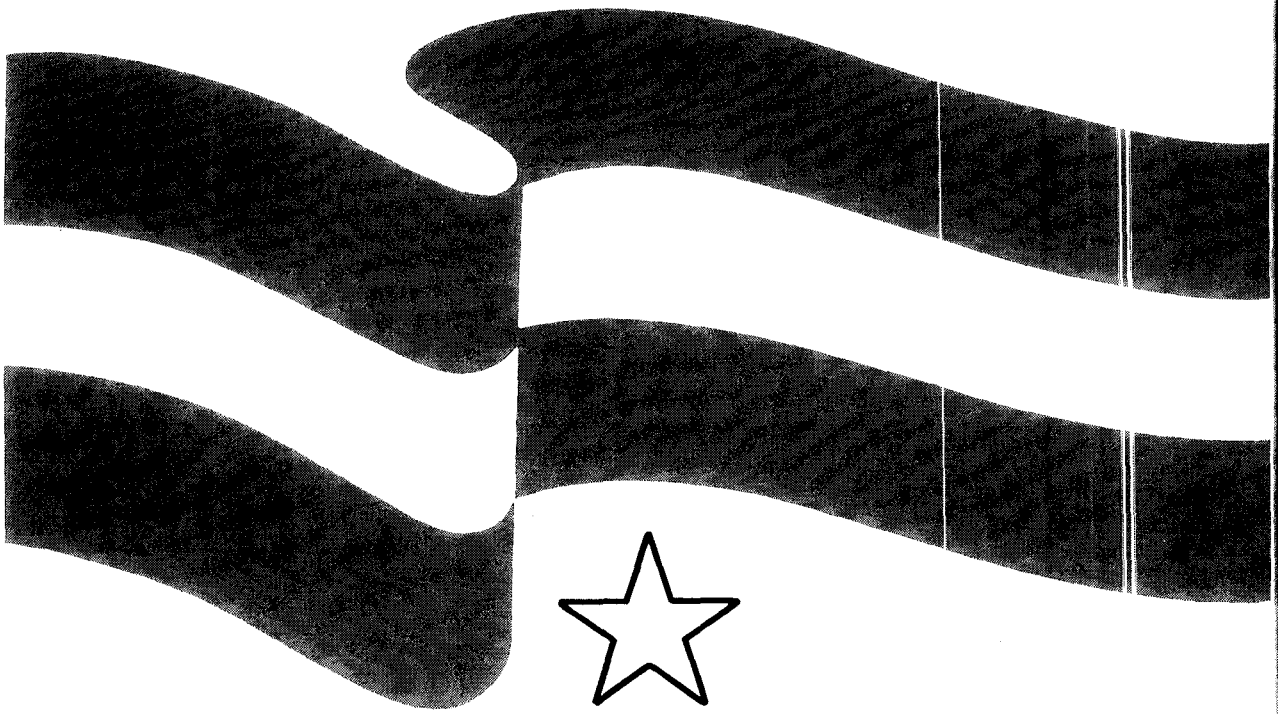


HEALTH STATUS OF VIETNAM VETERANS

SUPPLEMENT B
MEDICAL AND PSYCHOLOGICAL DATA QUALITY

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
Centers for Disease Control

HEALTH STATUS OF VIETNAM VETERANS

SUPPLEMENT B
MEDICAL AND PSYCHOLOGICAL DATA QUALITY

The Centers for Disease Control
Vietnam Experience Study
January 1989

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
Centers for Disease Control
Center for Environmental Health and Injury Control
Atlanta, Georgia 30333

Use of trade names is for identification only and does not constitute endorsement by the Public Health Service or by the U.S. Department of Health and Human Services.

ACKNOWLEDGMENTS

This supplement was prepared by Andrew L. Baughman, Robert R. German, Nancy E. Stroup, and Joseph L. Annest.

TABLE OF CONTENTS

SUMMARY

We used various methods to assure the good quality of data collected for the medical and psychological examination components of the Vietnam Experience Study (VES). These methods included using standardized tests and procedures, performing repeat testing, and making extensive checks on the quality of the data. In this supplement we focus primarily on three potential sources of error: observer (physician, technician, or interviewer), date of examination, and test order. In addition, we tested selected semen analysis measures for the effects of the magnification setting (of the ocular lens) at which semen specimens were recorded on video tape. For selected items in each medical examination, we analyzed interobserver variability, results of repeat tests, and temporal trends, and, for selected items in each psychological examination, we analyzed interobserver variability, temporal trends, and test order. Results of these analyses indicate that interobserver variability did not introduce confounding or effect modification into the analyses of cohort differences for any of the medical or psychological outcomes evaluated. The reliability of measurements for individual veterans was generally good for all repeated items from the medical examinations. Slightly more Vietnam veterans than non-Vietnam veterans participated in the study during the early phase. However, the results of our comparison of Vietnam and non-Vietnam cohorts did not vary by time period for any of the medical or psychological outcomes evaluated. The Combat Exposure Index scores of Vietnam veterans varied over time, as did the proportion of veterans who met criteria for Post-traumatic Stress Disorder (PTSD). This suggests that Vietnam veterans with higher levels of self-reported combat exposure or with PTSD were more likely to participate at the beginning of the study. The order in which the psychological tests were administered does not appear to have confounded or modified the association between Vietnam service and test results. The magnification setting for the video recording of semen specimens did not affect the results of the comparison between cohorts for sperm concentration or motility.

1. INTRODUCTION

This report is one of three supplements to the five-volume monograph, *Health Status of Vietnam Veterans*. In Volumes III and IV, titled *Medical Examination* and *Psychological and Neuropsychological Evaluation*, we summarize the physical health findings and the psychological health findings, respectively. In Supplement A we describe laboratory methods and quality control. Supplement C includes the medical and psychological procedures manuals and the data collection forms used in the Vietnam Experience Study (VES).

In assuring the good quality of data collected during the medical and psychological examination components of the VES (see Volumes III and IV), we used specific methods. These methods included standardized tests and procedures; laboratory statistical quality control; repeat tests; interobserver variability studies; site visits and data collection and processing procedures to find and correct errors in the data. These methods were designed to minimize various sources of systematic and random error. Systematic error may result from differences in measured outcomes among examiners or among dates of the examination. For instance, differences between Vietnam veterans and comparison veterans may be due, at least in part, to the examiner or the date of the examination. For the psychological examinations, differences between cohorts may also be due to the order in which the tests were administered. Random error reflects sampling variability and may have occurred at any stage in the data collection process.

To determine the extent of these potential sources of error in the medical and psychological examination components of the VES, we conducted numerous statistical analyses. These analyses were aimed at assessing whether the interobserver variability, date of examination, or test order (for psychological tests) influenced associations between Vietnam service and examination findings reported in Volumes III and IV. Throughout the rest of this supplement, the term observer refers to either physician, technician, or interviewer, depending on the examination component being addressed.

2. METHODS

We analyzed interobserver variability, temporal trends, and results of repeat tests (second tests performed on a random sample of veterans) for 15 medical tests (Table 1). Electrocardiograms, medical history interviews, nerve conduction velocity examinations, thermal and vibratory sensation tests, and psychological tests were not repeated for any veteran. We analyzed interobserver variability, temporal trends, and test order for 13 psychological tests (Table 1). Only selected items from each medical and psychological examination were evaluated.

To examine potential sources of error, we performed two types of analyses. In the first type of analysis, we focused on the distribution of cohort status among certain subgroups of veterans. These subgroups were defined by observer for the medical examinations, by test order for the psychological examinations, and by magnification setting for the semen analysis. For example, we tested for a significant association between technician and cohort status to determine whether some technicians examined more Vietnam veterans than other technicians. If some technicians had examined proportionally more Vietnam veterans than other technicians, this disparity could bias the association between serving in Vietnam and the particular outcome of interest.

In the second type of analysis, we sought to determine whether certain subgroups of Vietnam veterans are at different risks for particular adverse health outcomes – that is, whether there was any interaction between the factor represented by the subgroups and cohort status. These subgroups were defined by observer, date of examination, test order (for psychological examinations), or magnification setting (for the semen analysis).

The observer may be an important source of measurement error since the measurements and evaluations they make may differ, despite their having been trained and certified to perform standardized examinations. For example, some physicians may be more capable than others of palpating a moderately enlarged spleen. If these physicians, compared with the physicians who were not capable of palpating moderately enlarged spleens, examined different proportions of veterans from the two groups, then bias could result. Therefore, we tested for any interaction between observer and cohort status in the distribution of the examination results. For 11 of 24 analyses of interobserver variability, some observers performed only a few examinations and were therefore grouped into one category, "Other." This resulted in grouping no more than 7.1% of the total number of subjects for any of these analyses. For all analyses of interobserver variability, subjects whose observers were unknown, because of missing information, were excluded (not more than 1.1% for any medical or psychological test).

Temporal trends were analyzed because the effects of time (such as minor modifications in test procedures or other factors that cannot be controlled) can bias estimates of the association between Vietnam service and health outcomes. As a crude method of screening for temporal trends, we defined four periods during which about the same number of veterans were examined. We defined the periods by the dates on which the laboratory tests were ordered:

1st Period: June 3 – September 23, 1985
2nd Period: September 24, 1985 – January 8, 1986
3rd Period: January 9 – April 15, 1986
4th Period: April 16 – September 25, 1986

We then tested for any interaction between time period and cohort status in the distribution of the examination results.

For the psychological examinations, we also investigated the effects of test order on performance. During examinations, veterans were randomly assigned to one of four test sequences. In the two morning periods, the group-administered Minnesota Multiphasic Personality Inventory alternated with the individually administered tests (Wechsler Adult Intelligence Scale-Revised, California Verbal Learning Test, Paced Auditory Serial Addition Test, Word List Generation, Wisconsin Card Sort, Grooved Pegboard, and Rey-Osterreith Complex Figure Drawing Test); in the afternoon, the individually administered Diagnostic Interview Schedule and Combat Exposure Questionnaire alternated with the group-administered tests (Army Qualification Test and Edinburgh Handedness Inventory). We tested for interactions between test order and cohort status. Veterans given tests in unusual orders (orders other than the above) (<1.1%) were excluded from the analysis.

We performed an additional analysis of the effect of the microscopic magnification settings on the results of semen analyses. Two magnifications (X 1.0 and X 1.5) of the ocular lens of the microscope were used to make video recordings of the semen specimens, and specimens were recorded at either one or the other setting. For details on this procedure, see Chapter 13, "Semen Analysis," of Volume III (*Medical Examination*). For analysis of sperm concentration and selected motility measures, we tested for an interaction between magnification and cohort status.

Repeat test measures were analyzed by using data obtained for those veterans who were examined twice. For selected medical examination components (except the laboratory component), veterans who make up about a 5% random sample were given one repeat test by a second observer who did not know the results of the first examination or the veteran's cohort status (Table 2). For most clinical laboratory determinations, a repeat sample, indistinguishable from other participant samples, was inserted into the same analytic run. Thus, the same technician, who was blind to the matching of repeat samples and the cohort status of the veteran providing the specimen, performed the original and the repeat determination. The percent of veterans selected for repeat tests is about the same between cohorts for every medical examination (Table 2). The percent of veterans examined twice ranges from 3.6 to 5.8 for all examinations except clinical laboratory determinations (14.9%) and the hypersensitivity skin test (14.7%). Using these repeat measures, we investigated any difference between cohorts with respect to the overall agreement of the paired observations for specific examination items for individual veterans.

A summary of statistical methods used for these analyses is presented by type of variable (categorical or continuous) in Table 3. In our analyses of interobserver variability, temporal trends, and test order, we used data obtained for all veterans. The same strategy was used for each of these analyses. We determined whether there was any interaction between the source of error and cohort status in the distribution of the examination result. If the outcome was categorical, we performed the Breslow-Day test for homogeneity of the odds ratio (Breslow and Day, 1980). Before performing the test, we dichotomized each polytomous categorical outcome into "normal" and "abnormal." If the outcome was continuous, we performed a two-way analysis of variance (ANOVA) in which "observers/time periods/test orders" and "cohort" were considered fixed effects and an F statistic was used to test for interaction (Kleinbaum *et. al.*, 1988).

For the repeat measures, we calculated a measure of agreement between the paired observations for each outcome for each cohort of veterans separately. The intraclass correlation coefficient (ICC) was calculated for continuous outcomes (Bartko, 1966). The ICC is based on the results of a two-way ANOVA in which "veterans" and "observers" are considered random effects. The ICC is the variance due to veterans over the sum of all variance components (*i.e.*, veteran, observer, residual error). A high ICC means that the variance of a single measurement on a subject is due largely to the subject, not the observer. The "percent agreement" and kappa statistic were calculated for categorical outcomes (Fleiss, 1981). Before calculating these statistics, however, we dichotomized each polytomous categorical outcome. The "percent agreement" is the number of veterans given the same diagnosis by both the first and second observers divided by the total number of veterans examined twice. The kappa statistic is a measure of interobserver concordance that corrects for agreement expected by chance alone. For example, "percent agreement," or proportion of agreement, may be high, but after it is corrected for the amount of agreement expected by chance alone, its value may be considerably reduced . In such cases, much of the overall agreement between the two observers can be explained by chance. When interpreting the kappa statistic, we used the following criteria: a kappa value greater than 0.75 represents excellent agreement beyond chance; values between 0.40 and 0.75 represent fair to good agreement; and values below 0.40 represent poor agreement (Landis and Koch, 1977). Fleiss has shown that the kappa statistic is equivalent to Bartko's intraclass correlation coefficient (Fleiss, 1975).

Although the kappa statistic is meant to be an improvement over the simpler measure "percent agreement," because it corrects for chance agreement, it is influenced by prevalence. Two observers who seem to have high agreement may nevertheless emerge with low values of kappa when the prevalence of the finding is low (Cicchetti, 1987). Since the prevalence of many of the outcomes measured in the VES was low, we developed guidelines for presenting the kappa statistic. We determined the guidelines by examining the stability of the kappa statistic for outcomes with different "prevalences." We estimated "prevalence" by calculating the average percent of veterans to whom the two observers apply a positive (or "abnormal") diagnosis. For each cohort, we present the "percent agreement" and "percent positive" for all categorical outcomes and the kappa statistic only for those with a percent positive greater than 5.

The results of our analyses are presented in two types of summary tables. The first type includes results of testing for homogeneity of the distribution of cohort status for the data quality factors: observer, date of examination, and test order. For instance, homogeneity of the distribution of cohort status for observer means that the proportion of Vietnam veterans (or non-Vietnam veterans) examined is the same for every observer. The second type includes results of tests for interaction between the data quality factor and cohort status, as well as measures of agreement for repeat test results. Each row of this summary table corresponds to a variable chosen for the assessment of data quality, and each column corresponds to one of the four analyses. For analyses of interobserver variability, temporal trends, and test order, the cell of the table indicates whether or not the interaction is statistically significant at the alpha = 0.01 level, assuming a two-sided test of significance. An alpha of 0.01 was chosen to be consistent with the approach used in the VES analyses for testing for significant interactions (see Chapter 2, Volume III). For repeat test analyses, we present the measure of agreement for each cohort separately.

3. RESULTS

3.1. DATA QUALITY FACTORS

Results of our analyses are presented for the three factors of data quality discussed above: observer, time period, and test order (for psychological outcomes only) We provide an overview of these factors to help in interpreting the main results presented in the next two sections. Tables 4-6 show the results of a chi-square (χ^2) test for homogeneity of the distribution of cohort status by observer for each medical and psychological examination. For each examination, the number of observers, the χ^2 statistic, and its p-value are given. In Tables 4 and 5, we present results for all medical examinations and, in Table 6, results for all psychological examinations. The results for each clinical laboratory determination are presented separately in Table 5 because different groups of technicians performed the laboratory tests. The distribution of cohort status does not vary significantly ($p<0.01$) among observers for any medical examination (Table 4), clinical laboratory determination (Table 5), or psychological examination (Table 6) except for sperm concentration, sperm motility, and hepatitis B surface antigen.

Table 7 shows the distribution of cohort status by time period. Results of the χ^2 test indicate that, though not statistically significant ($p<0.01$), the percentage of Vietnam veterans examined differs across time periods. The percentage of those examined in the first time period who were Vietnam veterans was higher than the percentage of those examined in later periods (59.2% in the first time period versus 53.6%-55.6% in the last three time periods). The distribution of cohort status does not vary by test order for any psychological test (Table 8).

In summary, the percentage of Vietnam veterans does not vary significantly among observers or test orders, indicating that assignment of veterans by observer and by test order was random, except for those veterans whose semen samples were analyzed for sperm concentration and motility and for hepatitis B surface antigen. The analysis of time period, however, indicates that slightly more Vietnam veterans than non-Vietnam veterans came to Lovelace Medical Foundation in the early part of the study.

3.2. MEDICAL HISTORY AND EXAMINATIONS

In Tables 9-27, we summarize the results of all data quality assessments by medical examination. For repeat tests only, results for categorical outcomes are presented separately from those of continuous outcomes. Tables 9-27 show that results for only 1 of the 399 tests for interobserver variability were statistically significant (reported frequent urination, $p=0.001$). The results for reported frequent urination in Table 28 show a significant interaction between the first two interviewers and cohort status, though the total percent of abnormalities found by either the first or second interviewer is less than 4.

Tables 9-27 show the "percent agreement," "percent positive," and kappa statistic for categorical outcomes and intraclass correlation coefficients for continuous outcomes measured for repeated examinations. The extent of agreement between the first and second examiners of veterans with repeat tests varies widely by examination, but it is roughly the same for Vietnam and non-Vietnam veterans. The agreement or reliability of our measurements on individual veterans was generally good for all medical examinations that were repeated, particularly for the audiometry examination, clinical laboratory determinations, and visual acuity examination. The reliability of several items in the dermatology examination,

general physical examination, hypersensitivity skin test, neurology examination, and peripheral vascular examination was low (Table 29).

Tables 9-27 show that results for only 2 of the 364 tests for temporal trends were statistically significant (palpable liver size, $p=0.002$, and pinprick sensation of the proximal ventral aspect of the right arm, $p=0.002$). Table 30 shows the mean and standard deviation of palpable liver size by time period and cohort status. Because of the small sample sizes in the last two periods, the standard deviation among all veterans measured is much larger in the last two time periods than in the first two. When measurements made during the last two periods are deleted before testing for temporal trends, the interaction is no longer significant. Table 31 shows the number and percent of veterans with abnormal pinprick sensation of the proximal ventral aspect of the right arm, by time period and cohort status. Although the difference in the percentage of veterans with abnormalities between cohorts changes across time periods, the total number of participants with abnormalities is small (13 or fewer) during each period.

Tables 32 and 33 show results of the additional analysis of ocular magnification settings for semen analysis assays. The percentage of Vietnam veterans whose specimens were video recorded at the two settings differs significantly ($p<0.01$) (Table 32). However, results of the test for interaction between magnification setting and cohort status were not significant for measures of either sperm concentration or motility (Table 33).

3.3. PSYCHOLOGICAL EXAMINATIONS

In Table 34, we summarize the results of all data quality assessments for psychological variables by psychological examination. In Table 34, none of the 32 statistical tests for interobserver variability resulted in p-values of less than 0.01. However, 4 out of the 32 p-values (12.5%) for the interobserver variability tests are between 0.01 and 0.05. The four tests that resulted in borderline p-values include the following variables: ever alcohol abuse or dependence, ever drug abuse only, and ≥ 3 childhood behavior problems in the Diagnostic Interview Schedule, and average correct F,A,S words in the Word List Generation examination. Controlling for interviewer made little difference in the magnitude of the effect of serving in Vietnam for any of these variables.

In Table 34, none of the 39 tests for temporal trends between cohorts resulted in p-values of less than 0.01. Two additional psychological outcomes measured for Vietnam veterans only were evaluated for temporal trends. These two outcome variables, Combat Exposure Index (Table 35) and Post-traumatic Stress Disorder (Table 36), are related to the extent of combat experience. For both variables, the tests for temporal trends were statistically significant ($p=0.002$ and $p=0.007$, respectively). The geometric mean for the Combat Exposure Index is higher in the first two time periods than in the last two time periods. The percentage of veterans with Post-traumatic Stress Disorder is greatest in the first time period. These findings are consistent with the trend toward greater Vietnam veteran participation during the first part of the study (Table 7).

In Table 34, none of the 30 tests for test order resulted in p-values of less than 0.01.

4. DISCUSSION

We found that more Vietnam veterans than non-Vietnam veterans came to Lovelace Medical Foundation in the early part of the study. Perhaps the Vietnam veterans were more motivated to participate and thus scheduled their visits earlier. This result suggests the importance of looking for time trends in the VES results. The Combat Exposure Index scores of Vietnam veterans varied over time, as did the proportion of veterans who met criteria for Post-traumatic Stress Disorder. These results indicate that not only more Vietnam veterans but also more Vietnam veterans with higher levels of self-reported combat experience were examined earlier in the study.

The analyses of interobserver variability, temporal trends, and test order taken together represent a large number of statistical tests applied to the same data. Thus, we expect a number of p-values to be less than 0.01 by chance alone. We examined the distribution of p-values to give us an overall view of the results of each set of tests and to help us interpret those test results that were statistically significant. For example, under the null hypothesis of no interaction between observer and cohort status, the significance probability p is uniformly distributed on the interval (0,1). Therefore, if N statistical tests are made on the same data, we will expect (N x p) of them to have p-values less than alpha. We can then compare this number with the number of significant test results we observed.

The observed number of significant tests is about the same or less than the expected number at nominal levels of alpha for each set of analyses we performed except for our analysis of interobserver variability for psychological examination measures (Table 37). We found that adjusting for interviewer in estimating the effect of Vietnam service did not change the overall odds ratio for any of the four psychological examination variables with results that were close to being statistically significant (Table 37, interobserver variability for psychological examinations, alpha = 0.05). The other sets of analyses resulted in few, if any, statistically significant findings. In view of the large number of analyses conducted, the few findings that were statistically significant are probably due to chance alone and, hence, are merely artifacts of multiple testing. We conclude, therefore, that, on the average, any differences reported between cohorts in the VES findings did not vary among observers or test orders, nor did they change over the length of the study.

The results of data quality assessments performed by using data obtained for participants in the semen analysis show that for sperm concentration and motility measures the proportion of Vietnam veterans who participated differs significantly among technicians and between the two magnification settings. We found, however, no significant interaction between either technician or magnification setting and cohort status for any of the sperm measures that we evaluated.

We should briefly mention the power of the statistical tests of interobserver variability, temporal trends, and test order. Our ability to detect significant interactions between any of these potential sources of error and cohort status is limited because power decreases as the prevalence of an "abnormal" outcome decreases. The prevalence of many of the measures in the VES was low (less than 5%). In addition, as the number of subgroups (*e.g.*, observers) increases, the power to detect a significant interaction with cohort status decreases. These factors may have lowered the number of statistically significant results.

The analysis of repeat tests involved estimating a measure of agreement for a large number of variables. To obtain an overall view of the reliability of the repeated examinations,

we made box-and-whisker plots (Tukey, 1977) that summarized the distribution of agreement measures for each examination by cohort (Figures 1 and 2). The "percent agreement" (used for categorical outcomes with "percent positive" less than or equal to 5.0) has been scaled to fit on the same plot with intraclass correlation coefficients and kappa values. The number of agreement measures summarized by each plot is given on the top of the graphs. In each plot, the box stretches from the first to the third quartile and contains a bar representing the median -- and the "whisker" extends from each end of the box to corresponding extreme measurements. The summary plots indicate that the reliability, or reproducibility, of our measurements for individual veterans was good, in general, for all repeated medical examinations, particularly for the audiometry examination, clinical laboratory determinations, and the visual acuity examination. Most of the measures for these three examination components were entirely or mostly automatic. However, for several items in the dermatologic, general physical, neurologic, and peripheral vascular examinations, and in the hypersensitivity skin tests, the reliability was low. The reliability was low for these clinical assessments and measurements because they involved subjective grading or personal interpretation by the observer (*e.g.*, estimating percussible liver size).

REFERENCES

Bartko JJ. The intraclass correlation coefficient as a measure of reliability. Psychol Rep 1966;19:3-11.

Breslow NE, Day NE. Statistical methods in cancer research, volume 1: The analysis of case-control studies. Lyon: International Agency for Research on Cancer, 1980:142.

Cicchetti DV. High examiner agreement and low kappa: probing users' misconceptions. Paper presented at Joint Statistical Association Meetings, San Francisco, California, August 1987.

Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley & Sons, 1981:212-36.

Fleiss JL. Measuring agreement between two judges on the presence or absence of a trait. Biometrics 1975;31:651-9.

Kleinbaum DG, Kupper LL, Muller KE. Applied regression analysis and other multivariable methods. 2nd ed. Boston: PWS-KENT Publishing Company, 1988:457-82.

Landis JR, Koch GC. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

Tukey JW. Exploratory data analysis. Reading, MA: Addison-Wesley Publishing Company, Inc., 1977: 39-41.

Table 1. Types of Analyses Performed for Medical and Psychological Examinations Included in Supplement B. Medical and Psychological Data Quality, by Examination

Examination	Type of Analysis Performed			
	Interobserver Variability	Repeat Tests	Temporal Trends	Test Order
Medical Examinations				
Audiometry	*	*	*	
Clinical Laboratory Determinations[a]	*	*	*	
Dermatology	*	*	*	
Electrocardiogram	*			
General Physical	*	*	*	
Hypersensitivity Skin Test	*	*	*	
Medical History	*		*	
Nerve Conduction Velocities	*		*	
Neurology	*	*	*	
Peripheral Vascular Test	*	*		
Pulmonary Function Test	*	*		
Radiology	*	*		
Thermal Test	*		*	
Vibratory Test	*		*	
Visual Acuity	*	*	*	
Psychological Examinations				
Army Qualification Test			*	*
California Verbal Learning Test	*		*	*
Combat Exposure Index			*	*
Diagnostic Interview Schedule	*		*	*
Edinburgh Handedness			*	*
Grooved Pegboard	*		*	*
Minnesota Multiphasic Personality Inventory			*	
Paced Auditory Serial Addition Test	*		*	*
Rey-Osterreith Complex Figure Drawing Test	*		*	*
Wechsler Adult Intelligence Scale - Revised	*		*	*
Wide Range Achievement Test Reading Subtest	*		*	
Wisconsin Card Sort	*		*	
Word List Generation	*		*	*

[a] Analysis of magnification setting was also performed for sperm concentration and selected motility measures.

Table 2. Number and Percent of Vietnam and Non-Vietnam Veterans Examined Twice, by Type of Repeated Examination

Repeated Examination	Vietnam		Non-Vietnam		Total	
	No.	%[a]	No.	%[a]	No.	%[a]
Audiometry	139	5.6	112	5.7	251	5.6
Clinical Laboratory Determinations	355	14.3	312	15.8	667	14.9
Dermatology	102	4.1	64	3.2	166	3.7
General Physical	128	5.1	117	5.9	245	5.5
Hypersensitivity Skin Test	353	14.2	304	15.4	657	14.7
Neurology	134	5.4	91	4.6	225	5.0
Peripheral Vascular Test	82	3.3	77	3.9	159	3.6
Pulmonary Function Test	124	5.0	109	5.5	233	5.2
Radiology	144	5.8	117	5.9	261	5.8
Visual Acuity	115	4.6	94	4.8	209	4.7

[a] For each examination, the numerator is the number of veterans examined twice and the denominator is 2490 (Vietnam cohort), 1972 (Non-Vietnam cohort), or 4462 (Total).

Table 3. Summary of Statistical Methods[a] Used for Data Quality Assessments, by Type of Statistical Analysis

Type of Statistical Analysis	Type of Examination Item	
	Categorical (i.e., "normal" and "abnormal")	Continuous
Interobserver variability, Temporal trends, Test order (Also magnification setting for semen analysis)	Hypothesis: Odds ratio is the same for all observers[b], time periods, or test orders Test: Breslow-Day test for homogeneity of the odds ratio	Hypothesis: Difference in cohort means[c] is the same for all observers[b], time periods, or test orders Test: F-test for interaction
Repeat tests	The "percent agreement," "percent positive," and kappa statistic were calculated for each cohort to measure agreement between two observers[b,d]	The intraclass correlation coefficient was calculated for each cohort to measure agreement between two observers[b,d]

[a] See text for detailed descriptions.
[b] Depending on the measurement, the observer is either a physician, technician, or interviewer.
[c] Some measures were log transformed before testing to satisfy the normality assumption of analysis of variance.
[d] For all clinical laboratory determinations, the first and second technician were the same person.

Table 4. Summary Statistics of Tests for Homogeneity of the Distribution of Cohort Status Among Observers, by Medical Examination

Medical Examination	Number of Observers	χ^2	P-Value
Audiometry	12	11.3	0.42
Dermatology	6	9.3	0.10
Electrocardiogram	20	25.4	0.15
General Physical	5	9.3	0.05
Hypersensitivity Skin Test	6	10.5	0.06
Medical History	3	2.1	0.35
Nerve Conduction Velocities	8	2.9	0.90
Neurology	8	5.1	0.65
Peripheral Vascular Tests	2	0.5	0.48
Pulmonary Function Tests	7	9.0	0.18
Radiology	5	1.7	0.79
Thermal Sensation Test	8	7.3	0.40
Vibratory Sensation Test	8	4.0	0.78
Visual Acuity	9	4.7	0.79

Table 5. Summary Statistics of Tests for Homogeneity of the Distribution of Cohort Status Among Technicians, by Clinical Laboratory Assay

Laboratory Assay	Number of Technicians	χ^2	P-Value
Hematology			
Hematocrit	13	17.4	0.13
Hemoglobin	13	17.3	0.14
Mean red blood cell volume	13	18.3	0.11
Mean corpuscular hemoglobin	13	17.4	0.13
Mean corpuscular hemoglobin concentration	13	18.4	0.10
Prothrombin time	12	13.3	0.28
Immunology			
Absolute B-lymphoctes	4	1.3	0.72
Absolute T-lymphocytes	3	1.2	0.55
Absolute T4-lymphocytes	3	1.5	0.46
Absolute T8-lymphocytes	3	1.0	0.60
Relative B-lymphoctes	3	1.1	0.58
Relative T-lymphocytes	4	1.1	0.77
Relative T4-lymphocytes	3	1.2	0.56
Relative T8-lymphocytes	3	0.5	0.78
T4/T8 ratio	3	1.1	0.57
Serum Chemistries			
Alanine aminotransferase	18	27.4	0.05
Albumin	17	27.9	0.03
Alkaline phosphatase	17	28.4	0.03
Aspartate aminotransferase	18	28.5	0.04
Blood urea nitrogen	17	28.4	0.03
Creatine phosphokinase	17	28.4	0.03
δ-aminolevulinic acid	16	28.3	0.02
Gamma glutamyl transferase	17	28.6	0.03
High-density lipoprotein (HDL) cholesterol	18	29.0	0.03
Lactic dehydrogenase	17	27.2	0.04
Serum creatinine	18	28.9	0.04
Serum immunoglobulin A	14	23.4	0.04
Serum immunoglobulin G	14	23.3	0.04
Serum immunoglobulin M	14	23.4	0.04
Total bilirubin	18	23.0	0.15
Total cholesterol	17	28.4	0.03
Total protein	17	27.9	0.03
Triglycerides	20	28.4	0.08
Unconjugated bilirubin	18	23.7	0.13
Semen Analysis			
Sperm concentration	3	13.3	0.004
Sperm morphology/morphometry	6	7.6	0.18
Sperm motility	3	13.3	0.004
Steroids/Hormones			
Dehydroepiandrosterone	15	23.9	0.05
Follicle-stimulating hormone	16	28.5	0.02
Luteinizing hormone	16	27.6	0.02
Testosterone	15	26.6	0.02
Urine Chemistries			
Coproporphyrin	6	2.6	0.77
D-Glucaric acid	14	26.2	0.02
Urine pH	16	15.9	0.39
Porphobilinogen	15	23.3	0.05
Uroporphyrin	5	2.5	0.64
Other Tests			
Antibody to HBcAg[a]	15	30.7	0.01
Antibody to HBsAg[a]	16	30.1	0.01
Hepatitis B surface antigen	15	32.5	0.003
Occult blood, feces	16	18.2	0.25
Serologic test syphilis (RPR)	13	15.1	0.24

[a] HBcAg = Hepatitis B core antigen; HBsAg = Hepatitis B surface antigen.

Table 6. Summary Statistics of Tests for Homogeneity of the Distribution of Cohort Status Among Interviewers, by Psychological Examination

Psychological Examination	Number of Interviewers	χ^2	P-Value
California Verbal Learning Test	24	24.7	0.36
Diagnostic Interview Schedule	17	14.1	0.59
Grooved Pegboard	24	24.9	0.35
Paced Auditory Serial Addition Test	24	24.9	0.36
Rey-Osterrieth Complex Figure Drawing Test	24	25.5	0.32
Wechsler Adult Intelligence Scale-Revised	24	25.2	0.34
Wide Range Achievement Test Reading Subtest	25	31.5	0.14
Wisconsin Card Sort	24	24.4	0.38
Word List Generation	24	24.4	0.38

Table 7. Number and Percent of Vietnam and Non-Vietnam Veterans Examined, by Time Period

	Vietnam		Non-Vietnam		Total	
Time Period	No.	%	No.	%	No.	%
Total	2490	55.8	1972	44.2	4462	100.0
1	664	59.2	458	40.8	1122	100.0
2	620	55.6	495	44.4	1115	100.0
3	598	53.6	518	46.4	1116	100.0
4	608	54.8	501	45.2	1109	100.0

($\chi^2 = 7.9$, df = 3, p = 0.05)

Table 8. Summary Statistics of Tests for Homogeneity of the Distribution of Cohort Status Among Test Orders, by Psychological Examination

Psychological Examination	Number of Test Orders	χ^2	P-Value
Army Qualification Test	3	3.7	0.16
California Verbal Learning Test	4	5.3	0.15
Diagnostic Interview Schedule	2	1.0	0.31
Edinburgh Handedness Inventory	3	3.3	0.19
Grooved Pegboard	4	5.5	0.14
Paced Auditory Serial Addition Test	2	1.4	0.24
Rey-Osterrieth Complex Figure Drawing Test	4	4.9	0.18
Wechsler Adult Intelligence Scale-Revised	4	5.2	0.15
Word List Generation	2	0.7	0.40

Table 9. Summary of Data Quality Assessments for Audiometry Examination Tone Frequencies

Audiometry Examination Tone Frequency and Ear	Interobserver Variability[a]	Repeat Tests						Temporal Trends[a]
		Vietnam			Non-Vietnam			
		Percent Agreement	Percent Positive[b]	Kappa[c]	Percent Agreement	Percent Positive[b]	Kappa[c]	
500 Hertz, Left	NO	99.3	1.1		100.0	1.8		NO
1000 Hertz, Right	NO	100.0	0.7		100.0	0.9		NO
2000 Hertz, Left	NO	99.3	2.5		100.0	1.8		NO
3000 Hertz, Right	NO	98.6	10.8	0.93	95.5	6.7	0.64	NO
4000 Hertz, Left	NO	96.4	20.5	0.89	95.5	20.1	0.86	NO
6000 Hertz, Right	NO	95.0	24.8	0.87	95.5	13.8	0.81	NO
8000 Hertz, Left	NO	95.7	20.9	0.87	95.5	14.7	0.82	NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.
[b] Percent positive defined in text.
[c] Kappa statistic has been presented only for outcomes with percent positive >5.0.

Note: Table 1 lists the types of analyses performed for each examination.

Table 10. Summary of Data Quality Assessments for Continuous Clinical Laboratory Assays

Laboratory Assay	Interobserver Variability[a]	Repeat Tests: Vietnam Intraclass Corr. Coef.	Repeat Tests: Non-Vietnam Intraclass Corr. Coef.	Temporal Trends[a]
Hematology				
Hematocrit	NO	0.92	0.94	NO
Hemoglobin	NO	0.96	0.97	NO
Mean red blood cell volume	NO	0.96	0.97	NO
Mean corpuscular hemoglobin	NO	0.93	0.97	NO
Mean corpuscular hemoglobin concentration	NO	0.69	0.79	NO
Prothrombin time	NO	0.97	0.97	NO
Immunology				
Absolute B-lymphoctes	NO	0.91	0.95	NO
Absolute T-lymphocytes	NO	0.97	0.98	NO
Absolute T4-lymphocytes	NO	0.97	0.95	NO
Absolute T8-lymphocytes	NO	0.96	0.96	NO
Relative B-lymphoctes	NO	0.84	0.89	NO
Relative T-lymphocytes	NO	0.84	0.86	NO
Relative T4-lymphocytes	NO	0.87	0.79	NO
Relative T8-lymphocytes	NO	0.93	0.90	NO
T4/T8 ratio	NO	0.92	0.93	NO
Serum Chemistries				
Alanine aminotransferase	NO	0.98	0.98	NO
Albumin	NO	0.90	0.87	NO
Alkaline phosphatase	NO	0.99	0.99	NO
Aspartate aminotransferase	NO	0.92	0.94	NO
Blood urea nitrogen	NO	0.99	0.99	NO
Creatine phosphokinase	NO	1.00	1.00	NO
δ-aminolevulinic acid	NO	0.57	0.60	NO
Gamma glutamyl transferase	NO	1.00	1.00	NO
High-density lipoprotein (HDL) cholesterol	NO	0.98	0.98	NO
Lactic dehydrogenase	NO	0.95	0.96	NO
Serum creatinine	NO	0.87	0.86	NO
Serum immunoglobulin A	NO	0.99	0.99	NO
Serum immunoglobulin G	NO	0.96	0.96	NO
Serum immunoglobulin M	NO	0.99	0.99	NO
Total bilirubin	NO	0.98	0.99	NO
Total cholesterol	NO	0.99	0.99	NO
Total protein	NO	0.90	0.89	NO
Triglycerides	NO	0.99	1.00	NO
Unconjugated bilirubin	NO	0.97	0.98	NO

Table 10. Summary of Data Quality Assessments for Continuous Clinical Laboratory Assays – Continued

Laboratory Assay	Interobserver Variability[a]	Repeat Tests: Vietnam Intraclass Corr. Coef.	Repeat Tests: Non-Vietnam Intraclass Corr. Coef.	Temporal Trends[a]
Semen Analysis				
Sperm concentration	NO	–	–	NO
Sperm Morphology/Morphometry				
Mean cell area	NO	–	–	NO
Mean cell perimeter	NO			NO
Mean cell length/width ratio	NO	–	–	NO
Mean major axis length	NO	–	–	NO
% normal class cells	NO	–	–	NO
Sperm Motility				
Mean linear velocity	NO	–	–	NO
% motile cells	NO	–	–	NO
Steroids/Hormones				
Dehydroepiandrosterone	NO	0.98	0.97	NO
Follicle-stimulating hormone	NO	0.93	0.88	NO
Luteinizing hormone	NO	0.81	0.81	NO
Testosterone	NO	0.96	0.93	NO
Urine Chemistries				
Coproporphyrin	NO	0.89	0.83	NO
D-Glucaric acid	NO	0.68	0.83	NO
Urine pH	NO	0.98	0.98	NO
Porphobilinogen	NO	0.91	0.92	NO
Uroporphyrin	NO	0.87	0.83	NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.

Note: Table 1 lists the types of analyses performed for each examination. A dash (–) denotes analysis was not performed.

Table 11. Summary of Data Quality Assessments for Categorical Clinical Laboratory Assays

Condition	Interobserver Variability[a]	Repeat Tests						Temporal Trends[a]
		Vietnam			Non-Vietnam			
		Percent Agreement	Percent Positive[b]	Kappa[c]	Percent Agreement	Percent Positive[b]	Kappa[c]	
Serologic test syphilis (RPR)	NO	100.0	0.8		100.0	1.0		NO
Occult blood, feces	NO	–	–	–	–	–	–	NO
Hepatitis B surface antigen	NO	100.0	0.0		100.0	0.9		NO
Antibody to HBsAg[d]	NO	100.0	11.3	1.00	100.0	6.4	1.00	NO
Antibody to HBcAg[d]	NO	99.1	14.3	0.96	99.1	12.4	0.96	NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.
[b] Percent positive defined in text.
[c] Kappa statistic has been presented only for outcomes with percent positive >5.0.
[d] HBsAg = Hepatitis B surface antigen; HBcAg = Hepatitis B core antigen.

Note: Table 1 lists the types of analyses performed for each examination.
A dash (–) denotes analysis was not performed.

Table 12. Summary of Data Quality Assessments for Dermatology Examination Items

Dermtology Examination Item	Interobserver Variability[a]	Repeat Tests						Temporal Trends[a]
		Vietnam			Non-Vietnam			
		Percent Agreement	Percent Positive[b]	Kappa[c]	Percent Agreement	Percent Positive[b]	Kappa[c]	
Abnormal Pigmentation	NO	80.2	22.8	0.44	75.0	23.4	0.30	NO
Hyperpigmentation	NO	94.1	4.0		93.8	4.7		NO
Hypopigmentation	NO	94.1	3.0		96.9	4.7		NO
Birthmarks	NO	86.1	16.8	0.51	82.8	14.8	0.33	NO
Hair Abnormalities	NO	74.3	39.6	0.48	75.0	39.1	0.49	NO
Alopecia, male pattern	NO	74.0	40.0	0.48	75.0	37.5	0.48	NO
Hirsutism	NO	100.0	0.0		98.4	0.8		NO
Infections	NO	61.4	61.9	0.26	56.3	59.4	0.21	NO
Acne, grade I	NO	81.2	11.4	0.12	79.7	13.3	0.15	NO
Acne, grade II	NO	91.1	7.4	0.35	89.1	7.0	0.18	NO
Acne, grade III	NO	97.0	1.5		98.4	2.3		NO
Acne, grade IV	NO	99.0	0.5		98.4	0.8		NO
Acne, atypical		100.0	0.0		100.0	0.0		
Comedones only	NO	95.0	2.5		85.9	8.6	0.11	NO
Folliculitis	NO	67.3	26.2	0.19	70.3	25.8	0.25	NO
Tinea of nails	NO	84.2	21.8	0.54	93.8	7.8	0.57	NO
Tinea versicolor	NO	98.0	1.0		100.0	1.6		NO
Other Tinea	NO	69.3	42.1	0.41	62.5	39.1	0.30	NO
Neoplastic	NO	70.3	57.4	0.39	75.0	53.1	0.50	NO
Acrochordon	NO	82.2	23.8	0.52	81.3	21.9	0.46	NO
Cancer of skin	NO	99.0	0.5		96.9	1.6		NO
Dermatofibromas	NO	92.0	11.0	0.53	93.8	6.4	0.64	NO
Epidermal inclusion cysts	NO	92.0	9.0	0.51	95.3	7.0	0.64	NO
Keratosis, actinic	NO	98.0	1.0		98.4	3.9		NO
Keratosis, seborrheic	NO	90.0	10.0	0.44	85.9	10.2	0.23	NO
Lipomas	NO	98.0	2.0		96.9	1.6		NO
Milia	NO	99.0	1.5		96.9	1.6		NO
Nevi atypical	NO	98.0	1.0		95.3	3.9		NO
Sebaceous hyperplasia	NO	96.0	6.0	0.65	93.8	4.7		NO
Warts, nongenital	NO	92.0	8.0	0.46	93.8	6.3	0.48	NO

Table 12. Summary of Data Quality Assessments for Dermatology Examination Items — Continued

Dermtology Examination Item	Interobserver Variability[a]	Repeat Tests						Temporal Trends[a]
		Vietnam			Non-Vietnam			
		Percent Agreement	Percent Positive[b]	Kappa[c]	Percent Agreement	Percent Positive[b]	Kappa[c]	
Vascular Conditions	NO	52.5	48.5	0.10	43.8	43.8	-0.07	NO
Hemangioma	NO	73.3	24.3	0.28	75.0	25.0	0.35	NO
Poikiloderma of Civatte	NO	97.0	2.5		96.9	3.1		NO
Spider angiomas	NO	92.1	5.0	0.17	95.3	2.3		NO
Varicosities	NO	97.0	8.4	0.81	95.3	7.0	0.64	NO
Sexually Transmitted Diseases	NO	99.0	2.5		100.0	0.0		NO
Herpetiform lesions	NO	99.0	2.5		100.0	0.0		NO
Condylomata	NO	100.0	0.0		100.0	0.0		NO
Trauma/Factitial	NO	77.2	81.7	0.24	82.8	85.2	0.32	NO
Drug tracks	NO	100.0	0.0		100.0	0.0		NO
Scars, postinflammatory	NO	79.0	17.5	0.28	85.9	19.5	0.55	NO
Tattoos	NO	96.0	13.1	0.82	98.4	8.6	0.90	NO
Inflammatory Conditions	NO	71.0	22.5	0.17	75.0	23.4	0.32	NO
Bullae		99.0	0.5		100.0	0.0		
Vesicles		100.0	0.0		100.0	0.0		
Eczematous dermatitis	NO	92.0	?	4.0	93.8	3.1		NO
Dyshidrosis	NO	99.0	0.5		96.9	1.6		NO
Lichen simplex chronicus	NO	97.0	1.5		100.0	0.0		NO
Psoriasis	NO	100.0	2.0		98.4	0.8		NO
Seborrheic dermatitis	NO	78.0	13.0	0.04	82.8	16.4	0.37	NO
Miscellaneous Causes[d]								
Keratosis pilaris	NO	86.0	7.0	-0.07	84.4	10.9	0.21	NO
Photodermatitis		100.0	0.0		100.0	0.0		
Sunburn	NO	99.0	1.5		96.9	1.6		NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.
[b] Percent positive defined in text.
[c] Kappa statistic has been presented only for outcomes with percent positive >5.0.
[d] "Miscellaneous Causes" is a subheading.

Note: Table 1 lists the types of analyses performed for each examination.
A blank ( ) denotes tests not performed because of <10 abnormalities.

Table 13. Summary of Data Quality Assessments for Electrocardiogram Examination Items

Electrocardiogram Examination Item	Interobserver Variability[a]
Ventricular rate	NO
PR interval	NO
QRS duration	NO
QT interval	NO
QTC interval	NO
P-Axis	NO
R-Axis	NO
T-Axis	NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.

Note: Table 1 lists the types of analyses performed for each examination.

Table 14. Summary of Data Quality Assessments for Continuous General Physical Examination Items

General Physical Examination Item	Interobserver Variability[a]	Repeat Tests: Vietnam Intraclass Corr.Coef.	Repeat Tests: Non-Vietnam Intraclass Corr.Coef.	Temporal Trends[a]
Height	NO	0.99	0.94	NO
Weight	NO	0.76	0.98	NO
Pulse rate	NO	0.72	0.77	NO
Respirations	NO	0.28	0.36	NO
Systolic blood pressure—right arm-1st meas.	NO	0.60	0.73	NO
Diastolic blood pressure—right arm-1st meas.	NO	0.61	0.70	NO
Systolic blood pressure—left arm-1st meas.	NO	0.52	0.72	—
Diastolic blood pressure—left arm-1st meas.	NO	0.61	0.55	—
Systolic blood pressure—right arm-2nd meas.	NO	0.51	0.72	—
Diastolic blood pressure—right arm-2nd meas.	NO	0.58	0.60	—
Systolic blood pressure—left arm-2nd meas.	NO	0.56	0.54	—
Diastolic blood pressure—left arm-2nd meas.	NO	0.65	0.49	—
Palpable liver size[b]	NO	—	—	YES
Percussible liver size	NO	0.20	0.33	NO
Right testis size[c]	NO	0.40	0.18	NO
Right testis size[d]	NO	0.53	0.42	NO
Left testis size[c]	NO	0.24	0.42	NO
Left testis size[d]	NO	0.31	0.24	NO
Body mass index	NO	0.99	0.99	NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.
[b] Too few values to calculate intraclass correlation coefficient.
[c] Measured with calipers.
[d] Measured with orchidometers.

Note: Table 1 lists the types of analyses performed for each examination.
A dash (—) denotes analysis was not performed.

Table 15. Summary of Data Quality Assessments for Categorical General Physical Examination Items

General Physical Examination Item	Interobserver Variability[a]	Repeat Tests						Temporal Trends[a]
		Vietnam			Non-Vietnam			
		Percent Agreement	Percent Positive[b]	Kappa[c]	Percent Agreement	Percent Positive[b]	Kappa[c]	
Skull condition	NO	98.9	0.6		100.0	0.0		NO
Eyes-Conjunctival discharge		97.8	1.1		100.0	0.0		
Eyes-Corneal Abnormality	NO	97.8	1.1		98.8	0.6		NO
Scarring	NO	98.9	0.6		100.0	0.0		NO
Cataract	NO	100.0	0.0		100.0	0.0		NO
Eyes-Retinal Abnormality	NO	94.4	2.8		96.3	1.9		NO
Arteriovenous nicking	NO	96.7	1.7		97.5	1.2		NO
Arteriolar spasm	NO	97.8	1.1		98.8	0.6		NO
Ear-Canal abnormalities	NO	89.0	11.0	0.44	89.2	11.4	0.47	NO
Ear-Impacted cerumen	NO	89.0	11.0	0.44	89.2	11.4	0.47	NO
Ear-Middle ear abnormalities	NO	87.8	6.1	-0.06	94.7	4.0		NO
Ear-Drum scarred	NO	91.5	4.3		96.0	3.3		–
Nose	NO	97.8	2.2		100.0	0.0		NO
Throat	NO	96.7	1.6		98.8	1.8		NO
Mouth-Dental status	NO	60.4	44.0	0.20	56.6	37.3	0.07	NO
Sinuses	NO	94.5	2.8		97.6	2.4		NO
Salivary glands	NO	100.0	0.0		100.0	0.0		NO
Thyroid abnormalities	NO	98.9	0.6		98.8	0.6		NO
Thyroid-Size	NO	98.9	0.5		98.8	0.6		NO
Carotid pulses		98.9	0.6		100.0	0.0		
Neck masses	NO	100.0	0.0		100.0	0.0		NO
Diminished breath sounds	NO	96.7	1.6		91.6	4.2		NO
Adventitial lung sounds	NO	97.8	2.2		97.6	1.2		NO
Lung crackles	NO	100.0	0.0		100.0	0.0		NO
Lung wheezes	NO	97.8	2.2		97.6	1.2		NO
Heart-Abnormal Sounds	NO	75.8	14.3	0.03	85.5	13.3	0.37	NO
Heart murmurs	NO	84.6	8.8	0.04	90.2	9.8	0.45	NO
Systolic clicks	NO	93.4	3.3		92.6	3.7		NO
Gallop sounds	NO	91.2	4.4		98.8	0.6		NO
Gynecomastia	NO	91.2	5.5	0.17	92.8	3.6		NO
Abdomen-Visible abnormality	NO	100.0	0.0		100.0	0.0		NO
Abdomen-Palpable mass	NO	–	–	–	–	–	–	NO
Abdomen-Tenderness	NO	97.8	1.1		98.8	0.6		NO

Table 15. Summary of Data Quality Assessments for Categorical General Physical Examination Items — Continued

General Physical Examination Item	Interobserver Variability[a]	Repeat Tests						Temporal Trends[a]
		Vietnam			Non-Vietnam			
		Percent Agreement	Percent Positive[b]	Kappa[c]	Percent Agreement	Percent Positive[b]	Kappa[c]	
Palpable liver	NO	91.2	4.4		94.0	3.0		NO
Spleen palpable		98.9	0.6		100.0	0.0		
CVA tenderness	NO	98.9	0.6		98.8	0.6		NO
Bruits	NO	100.0	0.0		98.8	0.6		NO
Femoral bruit-RT		100.0	0.0		100.0	0.0		
Femoral bruit-LT		100.0	0.0		100.0	0.0		
Hernias	NO	95.6	2.2		94.0	3.0		NO
Abnormal penis	NO	100.0	0.0		100.0	1.2		NO
Epididymis thickened/tender	NO	—	—	—	—	—	—	NO
Varicocele	NO	90.1	8.2	0.35	86.7	9.0	0.21	NO
Scrotal mass	NO	—	—	—	—	—	—	NO
Prostate Abnormality	NO	94.5	2.8		89.0	5.5	-0.05	NO
Enlarged prostate	NO	94.5	2.8		93.9	3.0		NO
Tender prostate	NO	100.0	0.0		100.0	0.0		NO
Rectal abnormalities	NO	60.4	23.1	-0.06	79.3	22.6	0.41	NO
Anal sphincter tone	NO	—	—	—	—	—	—	NO
Stool sampled	NO	—	—	—	—	—	—	NO
Absent extremities	NO	100.0	0.0		100.0	0.0		NO
Clubbing of fingers	NO	97.8	2.2		98.8	0.6		NO
Edema	NO	94.5	2.8		95.2	2.4		NO
Acrocyanosis	NO	96.7	1.6		100.0	0.0		NO
Varicose leg veins	NO	96.7	4.9		96.4	4.2		NO
Leg veins inflamed		100.0	0.0		100.0	0.0		
Soft tissue mass-extremity	NO	95.6	2.2		98.8	0.6		NO
Range of motion-dec-extremity	NO	81.3	12.6	0.16	84.3	11.4	0.23	NO
Straight leg raising	NO	63.7	40.1	0.26	62.7	41.6	0.23	—
Joint swelling	NO	95.6	2.2		96.8	3.0		NO
Lymph nodes	NO	92.3	3.8		91.6	4.2		NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.
[b] Percent positive defined in text.
[c] Kappa statistic has been presented only for outcomes with percent positive >5.0.

Note: Table 1 lists the types of analyses performed for each examination.
A dash (—) denotes analysis was not performed.
A blank ( ) denotes tests not performed because of <10 abnormalities.

Table 16. Summary of Data Quality Assessments for Continuous Hypersensitivity Skin Test Items

Hypersensitivity Skin Test Item	Interobserver Variability[a]	Repeat Tests: Vietnam Intraclass Corr. Coef.	Repeat Tests: Non-Vietnam Intraclass Corr. Coef.	Temporal Trends[a]
Proteus	NO	0.26	0.21	NO
Trichophyton	NO	0.59	0.67	NO
Candida	NO	0.50	0.47	NO
Tetanus	NO	0.47	0.49	NO
Diphtheria	NO	0.60	0.65	NO
Streptoccus	NO	0.56	0.51	NO
Tuberculin	NO	0.66	0.61	NO
Total induration[b]	NO	0.34	0.32	NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.

[b] Sum of values for the seven antigens.

Note: Table 1 lists the types of analyses performed for each examination.

Table 17. Summary of Data Quality Assessments for Categorical Hypersensitivity Skin Test Items

Hypersensitivity Skin Test Item	Interobserver Variability[a]	Repeat Tests						Temporal Trends[a]
		Vietnam			Non-Vietnam			
		Percent Agreement	Percent Positive[b]	Kappa[c]	Percent Agreement	Percent Positive[b]	Kappa[c]	
Anergy[d]	NO	96.3	1.8		97.7	3.8		NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.
[b] Percent positive defined in text.
[c] Kappa statistic has been presented only for outcomes with percent positive >5.0.
[d] Abnormal is defined as <2 mm response to all seven recall antigens in cell-mediated immunity (CMI) test (see Chapter 8 in Volume III).

Note: Table 1 lists the types of analyses performed for each examination.

Table 18. Summary of Data Quality Assessments for Medical History Items

Medical History Item	Interobserver Variability[a]	Temporal Trends[a]
Alcohol-No. days drinking per month	NO	NO
Alcohol-No. drinks per day	NO	NO
Alcohol-No. times ≥5 drinks per day	NO	NO
Alcohol-No. times drink and drive	NO	NO
Current medications	NO	NO
Special diet	NO	NO
Food midnight-am	NO	NO
Hospitalized since discharge from service	NO	NO
Broken bones	NO	NO
Motor vehicle injury	NO	NO
Head injury	NO	NO
Arthritis	NO	NO
Gout	NO	NO
Diabetes	NO	NO
Overactive thyroid	NO	NO
Underactive thyroid	NO	NO
Eczema	NO	NO
Psoriasis	NO	NO
Chloracne	NO	NO
Asthma	NO	NO
Chronic bronchitis	NO	NO
Pneumonia	NO	NO
Hypertension	NO	NO
Heart murmur	NO	NO
Angina	NO	NO
Heart attack	NO	NO
Peripheral vascular disease	NO	NO
Stomach or duodenal ulcer	NO	NO
Gastritis	NO	NO
Irritable bowel syndrome	NO	NO
Hemorrhoids	NO	NO
Liver damage/alcohol-induced	NO	NO
Hepatitis	NO	NO
Cirrhosis	NO	NO
Anemia	NO	NO
Migraine headaches	NO	NO
Peripheral neuropathy	NO	NO
Post-traumatic stress disorder	NO	NO
Kidney/bladder stones	NO	NO
Urinary tract infection	NO	NO
Chronic kidney disease		
Prostatitis	NO	NO
Epididymitis	NO	NO
Varicocele	NO	NO
Gonorrhea	NO	NO
Syphilis	NO	NO
Genital herpes	NO	NO
Infectious mononucleosis	NO	NO
Benign tumor	NO	NO
Any cancer	NO	NO
Allergies/ever had	NO	NO
Cold present/now	NO	NO
Skin boils or abscesses	NO	NO
Skin darkening	NO	NO
Abnormal hair growth		
Vision loss	NO	NO
Double vision	NO	NO
Bright light pain	NO	NO
Ringing in ear	NO	NO
Spinning sensation	NO	NO
Nose bleed	NO	NO
Shortness of breath	NO	NO

Table 18. Summary of Data Quality Assessments for Medical History Items – Continued

Medical History Item	Interobserver Variability[a]	Temporal Trends[a]
Persistent cough	NO	NO
Wheezing	NO	NO
Cough blood	NO	NO
Chest pain	NO	NO
Rapid heart beating	NO	NO
Calf pain with exercise	NO	NO
Appetite loss	NO	NO
Weight loss	NO	NO
Abdominal pain/recurrent	NO	NO
Vomiting up blood	NO	NO
Black stools	NO	NO
Loose stools	NO	NO
Bleed or bruise easily	NO	NO
Frequent urination	YES	NO
Bladder control loss	NO	NO
Night urination frequently	NO	NO
Inability to urinate	NO	NO
Urine dribble	NO	NO
Blood in urine	NO	NO
Penile discharge	NO	NO
Sores on penis	NO	NO
Swollen testicles	NO	NO
Impotence-Erection	NO	NO
Impotence-Ejaculation	NO	NO
Headaches	NO	NO
Seizure/Convulsions	NO	NO
Memory loss	NO	NO
Numbness of limbs	NO	NO
Tingling of limbs	NO	NO
Burning of limbs	NO	NO
Weakness-Leg	NO	NO
Weakness-Hands	NO	NO
Rheumatism-Low back	NO	NO
Rheumatism-Other areas	NO	NO
Currently employed	NO	NO
Work exposure to chemicals	NO	NO
Cigarette smoker-current	NO	NO
Smoke-Marijuana or hashish	NO	NO
Use cocaine	NO	NO
Use heroin	NO	NO
Counseling for alcohol or drug use	NO	NO
Treatment for alcohol or drug use	NO	NO
General health	NO	NO
Infectious diseases	NO	NO
Neoplasms	NO	NO
Endocrine diseases	NO	NO
Diseases of blood		
Mental disorders	NO	NO
Diseases of nervous system	NO	NO
Circulatory diseases	NO	NO
Respiratory diseases	NO	NO
Digestive system diseases	NO	NO
Diseases of genitourinary system	NO	NO
Diseases of skin	NO	NO
Musculoskeletal diseases	NO	NO
Congenital anomalies		
Symptoms, signs, and ill-defined cond.	NO	NO
Injuries and poisonings	NO	NO
Any condition	NO	NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.0$). See Methods Section for explanation.

Note: Table 1 lists the types of analyses performed for each examination.
A blank ( ) denotes tests not performed because of <10 abnormalities.

Table 19. Summary of Data Quality Assessments for Nerve Conduction Velocities Examination Items

Nerve Conduction Velocities Examination Item	Interobserver Variability[a]	Temporal Trends[a]
Median motor distal onset latency	NO	NO
Median motor distal amplitude	NO	NO
Median sensory distal onset latency	NO	NO
Median sensory distal amplitude	NO	NO
Median sensory proximal onset latency	NO	NO
Median sensory proximal distance	NO	NO
Median sensory distal distance	NO	NO
Sural sensory distal onset latency	NO	NO
Sural sensory distal amplitude	NO	NO
Sural sensory distance	NO	NO
Temperature of foot	NO	NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.

Note: Table 1 lists the types of analyses performed for each examination.

Table 20. Summary of Data Quality Assessments for Neurology Examination Items

Neurology Examination Item	Interobserver Variability[a]	Repeat Tests						Temporal Trends[a]
		Vietnam			Non-Vietnam			
		Percent Agreement	Percent Positive[b]	Kappa[c]	Percent Agreement	Percent Positive[b]	Kappa[c]	
Optic disc-RT	NO	96.1	2.0		96.6	1.7		NO
Pupil size-RT	NO	93.8	3.1		95.4	2.3		NO
Nystagmus-RT	NO	99.2	0.4		98.9	0.6		NO
Facial muscles-RT	NO	100.0	0.8		100.0	0.0		NO
Gait	NO	100.0	0.0		100.0	0.0		NO
Tandem gait	NO	98.3	0.8		97.5	1.2		NO
Station/eyes closed	NO	100.0	0.0		100.0	0.0		NO
Strength-RT knee ext	NO	100.0	0.8		100.0	0.0		NO
Tremors-LT arm	NO	97.7	2.0		97.7	2.3		NO
Finger-Nose ataxia	NO	99.2	0.4		100.0	1.2		NO
Heel-Shin ataxia	NO	98.4	1.6		98.9	2.9		NO
Arm drift	NO	100.0	0.8		100.0	0.0		NO
Reflex-RT knee-1[d]	NO	83.5	15.6	0.37	78.9	18.4	0.30	NO
Reflex-RT knee-2[e]	NO	83.0	8.5	-0.08	84.6	9.2	0.08	NO
Reflex-LT knee-1[d]	NO	83.5	15.6	0.37	81.6	17.1	0.35	NO
Reflex-LT knee-2[e]	NO	84.0	9.0	0.04	85.1	9.0	0.09	NO
Reflex-RT plantar-1[f]	NO	98.9	1.7		100.0	0.0		NO
Reflex-RT plantar-2[g]	NO	80.2	13.5	0.16	75.9	18.4	0.20	NO
Reflex-LT plantar-1[f]	NO	98.9	1.6		100.0	0.0		NO
Reflex-LT plantar-2[g]	NO	81.4	11.9	0.12	75.0	18.8	0.16	NO
Pinprick-RT arm-PD	NO	100.0	0.0		98.8	0.6		NO
Pinprick-LT arm-PD	NO	96.8	1.6		98.8	0.6		NO
Pinprick-RT arm-PV	NO	99.2	0.4		97.7	1.2		YES
Pinprick-LT arm-PV	NO	98.4	0.8		97.7	1.2		NO
Vibratory-RT-Lateral Malleolus	NO	98.4	1.6		95.3	2.3		NO
Vibratory-LT-Lateral Malleolus	NO	99.2	1.2		94.2	2.9		NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.
[b] Percent positive defined in text.
[c] Kappa statistic has been presented only for outcomes with percent positive >5.0.
[d] Absent or hypo reflex defined as abnormal.
[e] Hyper reflex, or unsustained or sustained clonus defined as abnormal .
[f] Reversed reflex defined as abnormal.
[g] Absent reflex defined as abnormal.

Note: Table 1 lists the types of analyses performed for each examination.

Table 21. Summary of Data Quality Assessments for Continuous Peripheral Vascular Examination Items

Peripheral Vascular Examination Item	Interobserver Variability[a]	Repeat Tests	
		Vietnam Intraclass Corr. Coef.	Non-Vietnam Intraclass Corr. Coef.
Resting ankle blood pressure-RT	NO	0.76	0.74
Resting ankle blood pressure-LT	NO	0.73	0.77
Maximum brachial blood pressure-RT	NO	0.60	0.61
Maximum brachial blood pressure-LT	NO	0.78	0.74
Resting brachial/ankle blood pressure index-RT	NO	0.17	0.42
Resting brachial/ankle blood pressure index-LT	NO	0.13	0.35

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.

Note: Table 1 lists the types of analyses performed for each examination.

Table 22. Summary of Data Quality Assessments for Categorical Peripheral Vascular Examination Items

Peripheral Vascular Examination Item	Interobserver Variability[a]	Repeat Tests					
		Vietnam			Non-Vietnam		
		Percent Agreement	Percent Positive[b]	Kappa[c]	Percent Agreement	Percent Positive[b]	Kappa[c]
Post tibial waveform morphology-RT		100.0	0.0		100.0	0.0	
Post tibial waveform morphology-LT		100.0	0.0		100.0	0.0	

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.
[b] Percent positive defined in text.
[c] Kappa statistic has been presented only for outcomes with percent positive >5.0.

Note: Table 1 lists the types of analyses performed for each examination.
A blank ( ) denotes tests not performed because of <10 abnormalities.

Table 23. Summary of Data Quality Assessments for Pulmonary Function Examination Items

Pulmonary Function Examination Item	Interobserver Variability[a]	Repeat Tests: Vietnam Intraclass Corr. Coef.	Repeat Tests: Non-Vietnam Intraclass Corr. Coef.
Slow vital capacity	NO	0.79	0.94
Forced vital capacity (FVC)	NO	0.95	0.94
Forced expiratory volume in one second (FEV1)	NO	0.94	0.94
FEV1/FVC	NO	0.90	0.72
Peak expiratory flow	NO	0.59	0.69
Mean maximal expiratory flow	NO	0.83	0.82
Mean maximal inspiratory flow	NO	0.46	0.56

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.

Note: Table 1 lists the types of analyses performed for each examination.

Table 24. Summary of Data Quality Assessments for Radiology Examination Items

Radiology Examination Item	Interobserver Variability[a]	Repeat Tests					
		Vietnam			Non-Vietnam		
		Percent Agreement	Percent Positive[b]	Kappa[c]	Percent Agreement	Percent Positive[b]	Kappa[c]
Chest X-ray abnormality	NO	81.2	35.8	0.59	78.6	38.0	0.55

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.
[b] Percent positive defined in text.
[c] Kappa statistic has been presented only for outcomes with percent positive >5.0.

Note: Table 1 lists the types of analyses performed for each examination.

Table 25. Summary of Data Quality Assessments for Thermal Sensation Test Items

Thermal Sensation Test Item	Interobserver Variability[a]	Temporal Trends[a]
Temperature of index finger	NO	NO
Thermal threshold of index finger	NO	NO
Temperature of great toe	NO	NO
Thermal threshold of great toe	NO	NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.

Note: Table 1 lists the types of analyses performed for each examination.

Table 26. Summary of Data Quality Assessments for Vibratory Sensation Test Items

Vibratory Sensation Test Item	Interobserver Variability[a]	Temporal Trends[a]
Vibration threshold of index finger	NO	NO
Vibration threshold of great toe	NO	NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.

Note: Table 1 lists the types of analyses performed for each examination.

Table 27. Summary of Data Quality Assessments for Visual Acuity Examination Items

Visual Acuity Examination Item	Interobserver Variability[a]	Repeat Tests						Temporal Trends[a]
		Vietnam			Non-Vietnam			
		Percent Agreement	Percent Positive[b]	Kappa[c]	Percent Agreement	Percent Positive[b]	Kappa[c]	
Distance-Uncorrected-LT	NO	98.3	24.3	0.95	97.9	18.1	0.93	NO
Distance-Uncorrected-RT	NO	92.2	25.7	0.80	96.8	21.8	0.91	NO
Distance-Uncorrected-both	NO	93.7	22.8	0.82	97.8	16.5	0.92	NO
Near-Uncorrected-LT	NO	98.3	14.8	0.93	96.8	4.8		NO
Near-Uncorrected-RT	NO	94.8	15.6	0.80	98.9	10.1	0.94	NO
Near-Uncorrected-both	NO	94.9	12.8	0.77	97.7	4.7		NO
Right-Eye temporal 85 degrees	NO	98.3	0.9		98.9	1.6		NO
Left-Eye nasal 35 degrees	NO	99.1	0.4		100.0	0.0		NO
Left-Eye temporal 85 degrees	NO	97.4	2.2		100.0	1.1		NO
Right-Eye nasal 35 degrees	NO	100.0	0.0		100.0	0.0		NO

[a] Yes/no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.
[b] Percent positive defined in text.
[c] Kappa statistic has been presented only for outcomes with percent positive >5.0.

Note: Table 1 lists the types of analyses performed for each examination.

Table 28. Number and Percent of Vietnam and Non-Vietnam Veterans Who Reported the Symptom of Frequent Urination During Medical Interview, by Interviewer

	Vietnam		Non-Vietnam		Total	
Interviewer	No.	%	No.	%	No.	%
A	58	4.7	12	1.2	70	3.1
B	26	2.3	22	2.6	48	2.4
Other	4	3.4	1	0.9	5	2.2

(χ^2 for interviewer by cohort interaction = 13.1, df=2, p=0.001)

Table 29. Items of Repeated Medical Examinations With Low Reliability, by Repeated Medical Examination

Repeated Medical Examination	Item[a]
Dermatology	Infections
	Acne, grade I
	Folliculitis
	Vascular Conditions
	Hemangioma
	Trauma/Factitial
	Inflammatory Conditions
General Physical	Left Testis Size
	Percussible Liver Size
	Respirations
	Right Testis Size
	Mouth-Dental Status
	Rectal Abnormalities
	Range of Motion-Decreased-Extremity
	Straight Leg Raising
Hypersensitivity Skin Test	Proteus
	Total Induration
Neurology	Reflex–Right Knee
	Reflex–Left Knee
	Reflex–Right Plantar
	Reflex–Left Plantar
Peripheral Vascular Test	Resting Brachial/Ankle Blood Pressure Index–Right
	Resting Brachial/Ankle Blood Pressure Index–Left

[a] Refer to Tables 9-27 for specific level of agreement.

Table 30. Mean and Standard Deviation of Palpable Liver Size (cm) for Vietnam and Non-Vietnam Veterans, by Time Period

	Vietnam			Non-Vietnam			Total		
Time Period	No.	Mean	S.D.	No.	Mean	S.D.	No.	Mean	S.D.
1	90	1.91	1.12	52	1.96	1.43	142	1.93	1.24
2	19	2.53	1.17	13	3.15	1.46	32	2.78	1.31
3	6	3.67	3.61	4	2.00	0.82	10	3.00	2.87
4	4	1.50	0.58	4	4.75	2.63	8	3.13	2.47

(F statistic for time period by cohort interaction = 5.0, df=3, p=0.002)

Table 31. Number and Percent of Vietnam and Non-Vietnam Veterans With Abnormal Pinprick Sensation of the Proximal Ventral Aspect of the Right Arm, by Time Period

	Vietnam		Non-Vietnam		Total	
Time Period	No.	%	No.	%	No.	%
1	8	1.4	1	0.2	9	0.9
2	10	1.6	3	0.6	13	1.2
3	4	0.7	6	1.2	10	0.9
4	1	0.2	9	1.8	10	0.9

(χ^2 for time period by cohort interaction = 14.8, df=3, p=0.002)

Table 32. Summary Statistics of Tests for Homogeneity of the Distribution of Cohort Status Among Semen Analysis Participants With Specimens Video Recorded at Two Ocular Magnification Settings[a], by Sperm Measure

Measure	χ^2	P-Value
Sperm concentration and motility	8.5	0.004

[a] For each semen analysis participant, the semen specimen was video recorded at one of two different microscope ocular lens settings (X 1.0 versus X 1.5). This analysis tested for homogeneity of the distribution of Vietnam and non-Vietnam veterans whose specimens were recorded at the two settings.

Table 33. Summary of Data Quality Assessments for Selected Semen Analysis Measures

Measure	Ocular Magnification Setting[a]
Sperm Concentration	NO
Sperm Motility	
Linear velocity	NO
% motile cells	NO

[a] Yes/no indicates whether interaction between magnification setting and cohort status is statistically significant (p<0.01).

NOTE: Table 1 lists the types of analyses performed for each examination.

Table 34. Summary of Data Quality Assessments for Psychological Examination Items, by Psychological Examination

Psychological Examination and Item	Interobserver Variability[a]	Temporal Trends[a]	Test Order[a]
Army Qualification Test			
General technical score, adjusted	–	NO	NO
California Verbal Learning Test (CVLT)			
CVLT delayed free recall	NO	NO	NO
CVLT immediate free recall	NO	NO	NO
CVLT middle recall	NO	NO	NO
CVLT primary recall	NO	NO	NO
CVLT recency recall	NO	NO	NO
Total recall 5 trials	NO	NO	NO
Combat Exposure Index (Vietnam Veterans only)[b]	YES	NO	
Diagnostic Interview Schedule			
Post-traumatic stress disorder (Vietnam Veterans only)[c]	NO	YES	NO
Ever depressed	NO	NO	NO
Ever alcohol abuse or dependence	NO	NO	NO
Ever alcohol abuse only	NO	NO	NO
Ever alcohol dependence only	NO	NO	NO
Ever drug abuse or dependence	NO	NO	NO
Ever drug abuse only	NO	NO	NO
Ever drug dependence only	NO	NO	NO
Ever generalized anxiety	NO	NO	NO
Antisocial personality	NO	NO	NO
≥4 adult behavior problems	NO	NO	NO
≥3 childhood behavior problems	NO	NO	NO
Duration of interview	NO	NO	NO
Edinburgh Handedness Inventory	–	NO	NO
Grooved Pegboard			
Dominant hand-seconds completed	NO	NO	NO
Other hand-seconds completed	NO	NO	NO
Minnesota Multiphasic Personality Inventory			
F scale – T score	–	NO	–
K scale – T score	–	NO	–
Scale 1 (HS) – T score	–	NO	–
Scale 2 (D) – T score	–	NO	–
Scale 8 (SC) – T score	–	NO	–
Paced Auditory Serial Addition Test			
% with correct response	NO	NO	NO
Sum of correct responses – 4 trials	NO	NO	NO
Rey-Osterrieth Complex Figure Drawing Test			
Total immediate memory	NO	NO	NO
Total delayed memory	NO	NO	NO
Wechsler Adult Intelligence Scale – Revised (WAIS-R)			
WAIS-R information, adjusted	NO	NO	NO
WAIS-R block design, adjusted	NO	NO	NO
Wide Range Achievement Test Reading Subtest			
Raw reading score	NO	NO	–
Wisconsin Card Sort			
Average trials per sort	NO	NO	–
Number of loss set ratios	NO	NO	–
Perseverations/countables ratio	NO	NO	–
Word List Generation			
Average correct F,A,S words	NO	NO	NO
Number correct animals	NO	NO	NO

[a] Yes /no indicates whether interaction with cohort status is statistically significant ($p<0.01$). See Methods Section for explanation.

[b] Index derived from responses to a 12-item questionnaire administered after the Diagnostic Interview Schedule. Tests performed only for Vietnam veteran cohort.

[c] Tests performed only for Vietnam veteran cohort.

Note: Table 1 lists the types of analyses performed for each examination.
A dash (–) denotes analysis was not performed.

Table 35. Geometric Mean and Geometric Standard Deviation of Combat Exposure Index for Vietnam Veterans, by Time Period

Time Period	No.	Geometric Mean	Geometric Standard Deviation
1	661	18.66	2.67
2	618	19.65	2.47
3	596	16.89	2.58
4	607	16.42	2.64
(F statistic for time period = 4.8, df=3, p=0.002)			

Table 36. Number and Percent of Vietnam Veterans Who Ever Had Post-traumatic Stress Disorder, by Time Period

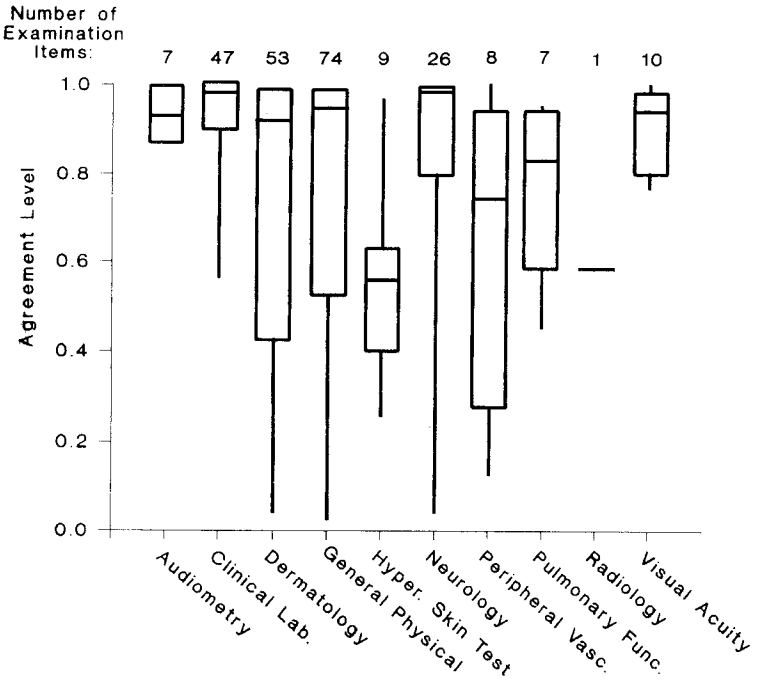
Time Period	No.	%	Total
1	128	19.3	663
2	94	15.2	620
3	89	14.9	597
4	75	12.3	608
(χ^2 for time period = 12.2, df=3, p=0.007)			

Table 37. Expected and Observed Number of Significant Tests for Data Quality Assessments, by Type of Examination and Analysis

Type of Examination and Analysis	Number of Tests (N)	Significance Level of Test (p)	Expected Number of Significant Tests[a] (N x p)	Observed Number of Significant Tests (Number < p)
Medical Examinations				
Interobserver Variability	387	0.01	3.9	1
		0.05	19.4	14
		0.10	38.7	38
		0.20	77.4	70
Temporal Trends	357	0.01	3.6	2
		0.05	17.9	14
		0.10	35.7	34
		0.20	71.4	78
Psychological Examinations				
Interobserver Variability	32	0.01	0.3	0
		0.05	1.6	4
		0.10	3.2	6
		0.20	6.4	9
Temporal Trends	39	0.01	0.4	0
		0.05	2.0	1
		0.10	3.9	4
		0.20	7.8	7
Test Order	30	0.01	0.3	0
		0.05	1.5	0
		0.10	3.0	1
		0.20	6.0	5

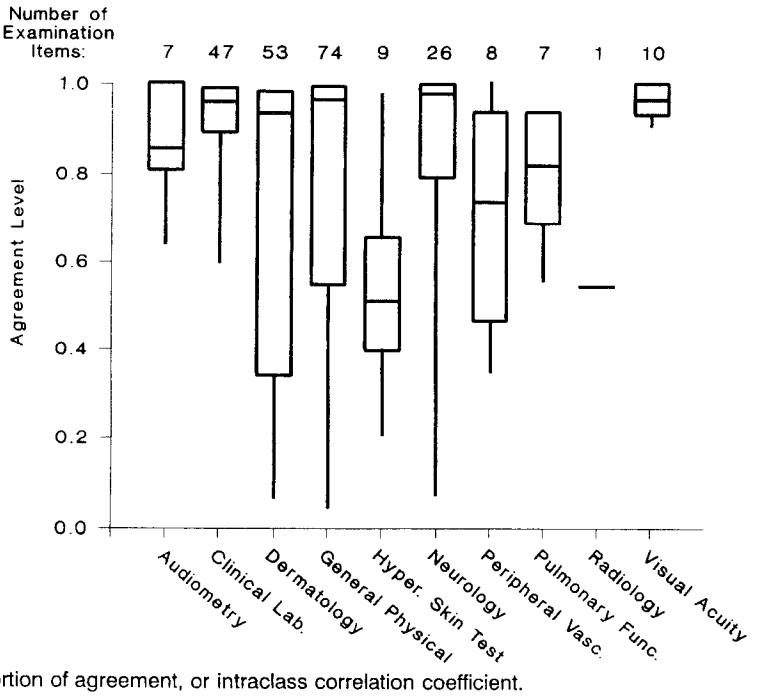
[a] Under the null hypothesis of no interaction, the significance level will be distributed uniformly between 0 and 1.

Figure 1. Box-and-Whisker Plots of Agreement Measures[a], by Medical Examination Component for Vietnam Veterans



[a] Kappa, proportion of agreement, or intraclass correlation coefficient.

Figure 2. Box-and-Whisker Plots of Agreement Measures[a], by Medical Examination Component for Non-Vietnam Veterans



[a] Kappa, proportion of agreement, or intraclass correlation coefficient.