PHYSICAL EXAMINATION QUALITY CONTROL

QC was emphasized in the physical examination, as this data source provided most of the medical information for clinical and epidemiologic analyses.

Initial concern for a high-quality physical examination was addressed by a stringent SCRF selection process for all personnel who were to directly interact with the participants. Each staff member was hand-selected for the AFHS on the basis of expertise, experience, and a commitment to remain with the study throughout the examination cycle. Further, the Air Force reviewed the credentials of all key staff members and approved their participation in the study.

A complete pretest physical examination, interview, psychological test, and laboratory workup was done for 11 volunteers several weeks before the scheduled start of the study. Refresher training was given to the dermatologists to enhance their skill in diagnosing chloracne, techniques for detecting specific heart sounds were reviewed with the internists, and diagnosticians were reminded of the need to review Baseline and 1985 examination data as they formulated all diagnoses. Additionally, automatic monitors to measure blood pressure were instituted for more accurate readings. Further, all aspects of patient contact were reviewed: the initial inbriefing of the participants, the logistics of transportation and patient flow within the clinic, and the final outbriefing by the diagnostician.

During the examinations, refinements continued whenever operational problems were detected by the SCRF staff and the Air Force onsite monitor, or when participants identified areas requiring improvement. Both of these types of information were addressed during the weekly clinical QA meeting of key SCRF staff, chaired by the SCRF Medical Project Director and attended by an Air Force representative. In addition, written critique forms submitted by all participants were reviewed in detail at the SCRF weekly meetings, providing additional insight to both temporary shortcomings of the entire logistic process as well as the numerous strong points of the programs.

Following examination of each participant group, all physical examination forms were reviewed by the SCRF staff for omissions, incomplete examinations, and inconsistencies. The examiners or technicians were quickly contacted to correct the data. Special effort was made to complete this review while the participants were at the examination site. In all cases of data correction, a complete audit trail was maintained. Finally, all mark-sense physical examination forms were read by an optical scanner. (This subject is discussed in more detail in the Data Management Quality Control section of this chapter.)

Compliance with all aspects of the physical examination was monitored daily by the Air Force onsite monitor and the SCRF Medical Project Director. Additional periodic inspections were conducted by the SCRF Chief of Medicine and the Science Applications International Corporation (SAIC) Principal Investigator. All such clinical reviews were done unobtrusively, and with the full consent of the participant; suggestions or corrections to the examination procedure were always discussed privately with the attending physician. These inspections emphasized aspects of clinical techniques,

sequencing and completeness of the clinical data with respect to the examination forms, and the total blindness of the examinations. Of particular note were the detailed daily log entries of the six Air Force monitors. These entries ensured continuity of knowledge (the monitors rotated approximately every 2 weeks) by documenting examination procedural changes and recording events requiring followup by either the Air Force or the prime contractor.

Establishment of rapport with each study participant was a primary goal of all organizations involved in this study. Although "rapport building" may not be a traditional QA parameter in most research studies, it is paramount in the AFHS because maintaining the satisfaction of participants encourages them to continue in the study, and thus a significant reduction in future statistical power or bias, or both, is avoided. Therefore, every staff member, from the initial telephone recruiter to the nurse coordinator and the Project Manager, emphasized courtesy, empathy, assistance, and personalized treatment of each participant. Based on the evaluation forms, 67 percent of the participants evaluated their experience in the 1987 followup as excellent and 27 percent classified it as good. Five percent of the participants rated the experience as satisfactory and only 1 percent felt that it was unsatisfactory.

LABORATORY QUALITY CONTROL

Before the study was begun, specific QC laboratory procedures were designed, developed, and implemented to rapidly detect problems related to test/assay performance, validity of reagents, analysis of data, and reporting of results. All laboratory assays for the study were done with state-of-the-art laboratory equipment and techniques. Laboratory facilities all had the equivalent of National Institutes of Health Biosafety Level 2 approval ratings and were certified by the College of American Pathology.

Quality Control Procedures for the Clinical Laboratory

Hematology assays were performed on Coulter 5-Plus® equipment; sedimentation rate determinations were performed using the large-tube Westergren method. The Dupont Automated Chemical Analyzer® was used to perform the biochemical assays; radioimmunoassays were done with standard test kits. Electrophoresis and occult blood tests were performed manually. Hepatitis B tests were performed using Abbott Diagnostic kits. Monospecific antibodies were used for immunologlobulin assays using the Beckman Array Protein System®. Blood-cell counts were performed with standard microscopy, and Clinitek®, a reflectance spectometry urinalysis, was used for all urinalyses. All other assays were done using industry-approved equipment and techniques.

All laboratory operations were controlled with the use of an integrated medical laboratory management information system that incorporated direct device to data base interfaces for automated testing equipment, and data entry for manual tests was performed by the laboratory technologists. An automated audit trail and a set of comments for technologist remarks were kept for each test so that any QC results could be retraced.

Procedural QC included using instrumentation and reagents from the same lot numbers throughout the study. Strict standards of calibration for all automated laboratory equipment were maintained at all times.

Trilevel or bilevel controls were used as the primary means for monitoring the quality of all tests. On every group of participant samples, one control (low, medium, or high) was run at the start, after every ninth sample, and at the end of each test run. Each trilevel control was used before repeating it in the run, when more than 18 experimental samples were analyzed. In addition, split aliquots were made from every tenth patient sample and were analyzed separately to measure test reproducibility.

All QC data were analyzed and summarized in formal QC reports generated weekly. QC data were subjected to independent statistical analysis to produce and analyze time-dependent trends. For all equipment malfunctions or other exceptions, a formal QC exception report was prepared by the responsible individual and forwarded to the QA officer and the project management team.

An additional measure of quality control introduced during the study was the cumulative sum (CUSUM) tests run with trilevel controls.[1] In particular, the fast initial response (FIR) CUSUM QC technique was used. It has an advantage in detecting long-term subtle drift that could have substantial adverse analytical consequences.[2] FIR is a special case of the CUSUM QC scheme that increases the overall effectiveness of the QC procedure. Unlike QC procedures using standard control charts, which compare each observation to designated limits, these tests utilize the cumulative sum of deviations from a target value.

CUSUM statistics were accumulated for each of the trilevels to quickly detect instrument calibration problems as identified by excessive drift. If an out-of-control situation was indicated, the graph showed when the change first occurred. When CUSUM indicated an out-of-control situation, all adjacent patient samples were reanalyzed after the equipment was thoroughly checked and fresh controls were run. Coefficient of variation (CV) requirements were established before the study for each test.

FIR CUSUM generally has been applied to QC in industry, particularly in high-volume, high-precision applications. It is believed that FIR CUSUM has not generally been applied in a biomedical setting. This procedure has proven to be effective and is now being used regularly in the SCRF clinical laboratory.

As the examination portion of this study ended, laboratory outliers were analyzed for logical validity by an independent clinician. All out-of-range test results were examined and scored as clinically explainable, clinically possible, or clinically unexplained. No clinical laboratory data were excluded because all out-of-range results were found to be clinically explainable or clinically possible.

Quality Control Procedures for the Immunology Laboratory

The QC procedures for the Cellular Immunology section of the AFHS were structured to rapidly detect any problems in four major test parameters:

(1) assay performance, (2) reagent validity, (3) data analysis, and (4) results reporting. The QC measures were detailed in the Quality Procedures Plan and documented before testing started. Compliance was monitored daily by the Cellular Immunology laboratory supervisor. Key aspects of the program included instrument and equipment calibration and maintenance, assay controls, accuracy and precision determination, and system failure checks.

QC measures followed in all Cellular Immunology assays included:

- Testing of a blood sample from a normal, healthy control individual with each group of AFHS patient samples
- Duplicate testing of one random patient sample in each assay
- Quadruplicate testing of each patient sample for each variable in each of the functional assays (e.g., phytohemagglutinin [PHA] stimulation, natural killer cell, and mixed lymphocyte culture)
- Parallel testing and monitoring reactivity of various lots of reagents when appropriate
- Verification of patient and specimen identification by at least two individuals before final reporting to the data base
- Note codes attached to any data point with a detected deviation due to procedural setup error, assay malfunction, equipment malfunction, or assay technical error
- Note codes attached to any data point outside the range of expected values as identified by the Cellular Immunology laboratory supervisor
- Review of all final assay reports by the Cellular Immunology laboratory supervisor prior to entry into the data base.

QC for each functional assay including PHA, mixed lymphocyte culture, and natural killer cell consisted of monitoring assay controls, duplicate sample reproducibility, and trends in reagent reactivity. Assay precision was determined by calculating the CV of the quadruplicates for each variable tested. Also, a mean value of the CV for each assay was calculated. Individual CV's of 15 percent or less were the target values for the stimulated samples in the mitogen and natural killer cell assays. The Student's t-test was applied to duplicates to determine if there was a significant difference in sampling for the functional assays. Critical t-values at the 0.05 significance level were used to determine if duplicate sample results varied significantly. Positive and negative values were assigned, arbitrarily subtracting the second duplicate value from the first, to determine if there was a systematic bias in one direction. Grubbs' statistical test[3] was used to identify any statistically significant outlier. This test was applied only to samples whose CV's were greater than 20 percent at a p-value of 0.01. The mitogen stimulation (PHA) effect was followed by daily evaluation of the radioactive counts in counts per minute. When counts fell below expected values, suggesting that reagent deterioration had occurred, new aliquots were used.

QC measures for the cell surface marker assays included: calculation of (CD4 + CD8)/CD2 (formerly $[T_4 + T_8]/T_{11}$) cell ratios, evaluation of flow cytometer computer outputs (cytograms and histograms), and duplicate sample testing. The cellular ratios should approximate the value 1.0 for a normal population. Validity of cytogram and histogram distributions generated by the flow cytometer was confirmed by the Cellular Immunology laboratory supervisor for each sample analyzed. The proportional difference between duplicate samples was calculated and monitored for significant differences.

On completion of this followup effort, the entire cellular immunology data base was reviewed by the Air Force team, laboratory staff, and an immunology consultant. Comments attached to the data points were also reviewed. Any data point that appeared to be a significant outlier was reviewed and coded as an unexplained outlier. Unexplained outliers were deleted from the data base as errors of an unknown nature. This review was conducted without knowledge of exposure status. The results of this review are presented in Chapter 19.

DATA MANAGEMENT QUALITY CONTROL

Overview of Quality Control Procedures

The QC program for the data management activity consisted of multiple checks at all steps of the examination, data collection, and data processing cycle. Data QC procedures for data collection, conversion, and integration were developed before the clinical examinations began. Pretesting of all forms, procedures, and logistic arrangements was conducted 3 weeks before the examinations actually began. Additionally, during the first 2 months of the clinical examinations, all data collection activities were intensely scrutinized to detect and correct procedural deficiencies.

QC activities also included automated QC techniques applied to laboratory data; clinical evaluations of all laboratory outliers; review of all physical examination findings by one of two diagnosticians who was not involved in the conduct of the physical examinations; and automated and manual data quality checking of hard copy against transcribed computer files for all questionnaire, physical examination, and medical coding data streams.

Five interwoven layers of QC were instituted to ensure data integrity. Efforts focused on (1) data processing system design, (2) design and administration of all exams or questionnaires, (3) data completeness checks, (4) data validation techniques, and (5) quality control of medical records coding. In some cases, the QC procedures described in this section were implemented throughout the data management task rather than assigned to a particular activity. These comprehensive QC procedures will be mentioned where appropriate throughout the remainder of this section.

Data Processing System Design

For each data stream, standards were set to establish data element format (character or numeric), data element naming conventions, data element

text labels, numeric codes for qualitative responses and results, QC range checks for continuous data elements, and QC validity checks for categorical data. A data dictionary provided detailed information on each data element.

A systems integration approach was applied to the design and implementation of data collection procedures and techniques so that data emanating from the various study sources (physical examination, questionnaire, laboratory) were consistent in file format and structure. This was necessary to ensure that all data could be integrated into a single data base management system for analysis. Figure 6-1 provides an overview of the QC activities used in the data management process.

Forms and questionnaires were carefully designed to ensure that all required data elements would be collected in accordance with the Study Protocol and in a standardized format. The design of these instruments was such that they reflected the order in which the examination itself would be administered and provided for the sequential recoding of information to streamline remaining data management activities.

Completed medical records and questionnaires were converted from hard copy to machine-readable images using customized data-entry systems or state-of-the-art optical mark reading equipment. Verification procedures were performed to ensure that a uniquely identified participant record existed within each data file, and that the appropriate number of responses for each applicable field was provided. Data files were then verified against original data sheets and corrected as necessary.

Data files were then subjected to validity checks. Any potentially conflicting results as well as any data values falling at the extremes of expected ranges were manually reviewed. Extreme values were reverified against the original raw data copies and either corrected or documented as valid results. Potentially conflicting results were returned to the examiners for review. These results were then documented as correctly recorded, corrected, or flagged for exclusion from analysis because of unresolvable examiner errors or omissions. This process was continued until all results were properly documented.

Once the edits were completed and the data reverified, the "cleaned" files or tapes were transferred to the data analysis center for final inspection and integration into the study data base. For this QC measure, each data file was loaded into a SAS® data set, and descriptive analyses were run. The validation, correction, transmission, and analysis QC procedures were repeated as necessary to ensure that all extreme or suspicious values had been validated.

Design and Administration of Physical and Psychological Examination Forms

As mentioned, the examination forms were designed to solicit all required data such that recording time was minimized, comprehension was enhanced, and data input could occur with a minimum of transcription errors. Optical Mark Recognition (OMR) technologies were selected to eliminate the risk of transcription errors and were applied to all psychological tests. Customized mark-sense forms were also developed and OMR technology was used

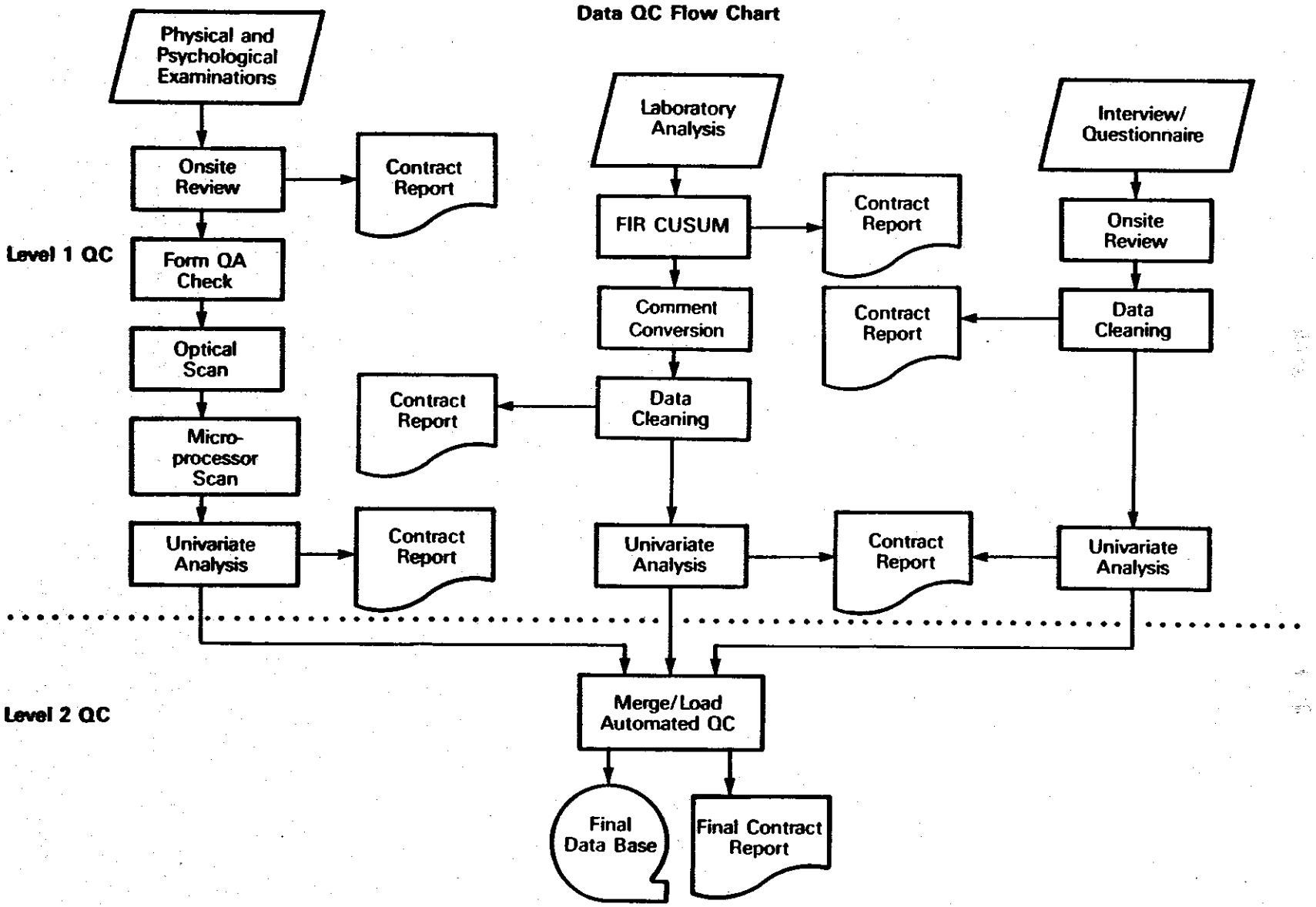

Figure 6-1.
Two Levels of Quality Control Applied to All Collected Data Prior to Statistical Analysis

to achieve these same objectives for segments of the physical examination and the self-administered questionnaires. The use of mark-sense forms allowed the creation of computerized data files directly from the raw data recorded on these forms.

QC procedures for all data collection instruments began with a review of all forms as they were completed. Any forms containing missing examination results were returned to the examining physician for completion before the participants left the site. Any questionable results or "hard-to-diagnose" conditions (such as heart sounds or peripheral pulses) were verified by the diagnostician at the outbriefing. All examination forms were signed by the examining physician, and the examiner identification number was coded in the data base. Detailed QC records were maintained, which indicated the examining physician and the type of deficiency detected. Deficiency reports were reviewed by the study coordinator to detect any patterns of physician data entry error. A final level of QC audit was accomplished by Air Force statisticians, who conducted a detailed screening of the data and checked for errors.

Data Completeness Checks

Customized programming of the OMR allowed for the identification of those forms (and their corresponding data records) with missing responses, as well as those with multiple responses to questions that required a single response. The OMR scanner was programmed to reject forms that failed completeness and multiple response checks and to output a control code for each rejected form. The control code identified the location of the first three verification checks failed for a given form.

When a raw data form was rejected, the reason for the rejection was determined and the exact data element was corrected by comparing the rejected raw data form to the values recorded in the data record created by the scanner. A customized set of rejection and resolution codes was developed for the study to describe all the reasons for a form's rejection and any subsequent reasons for changing a data value. Various codes identified values recovered from light marks, missing marks explained by examiner comments, and missing comment flags resolved by the presence or absence of text in the comment areas. These codes ensured data completeness by accounting for all questionable or missing responses.

Some of the rejected forms did not contain actual data errors but rather anomalies created in using mark-sense cards for data collection. For instance, incompletely erased responses and responses marked with too little carbon or graphite were incorrectly counted or missed, respectively, by the scanner. Examiners also tended to clearly mark responses for abnormal findings while bypassing or lightly marking responses for expected or desired findings. Failure of the form to provide the correct number of expected responses always resulted in rejection. These technology-based errors were resolved, as were the anticipated, more traditional errors.

The rejection code, data location code, resolution code, data inspector's initials, and correct data value were directly posted to a participant's data record. This innovative technique not only effectively

maintained a comprehensive audit trail of all record manipulations, it also provided a mechanism for measuring the frequency of specific errors.

Statistics were compiled on out-of-range results and data omissions that had been accepted in the previous QC audits. The results were monitored to detect trends, possible bias situations, and other data quality problems. This information was reviewed and relayed to examiners and internal auditors to assist in preventing or correcting chronic, but avoidable, problems. Refresher training was provided to examining physicians to avoid data omissions. Physicians were consulted to recover missing data, and out-of-range results were reviewed for logical validity by an independent clinician.

Data Validation Techniques

QC activities also included data validation techniques. As mentioned earlier, data files were examined in a series of verification and validation procedures developed to check the results within each participant's record for logical consistency and abnormal findings. Any records noted to have ambiguous findings, incongruent observations, extreme results, or errors or omissions were listed and submitted for review to a physician.

Again, clinical judgments were made by the auditing physician in assigning a validation code for each extreme or questionable data result. The validation codes allowed for indicating that data were deciphered from examiner comments or from related findings from another specialty area, or were accurately recorded and logically consistent with other findings for the participant. Data points that could not be definitively validated or recovered through clinical judgment and consultation with the original examiner were assigned codes noting missing or invalid data values. Some reasons for data not being available for analysis included participant refusal; incomplete, confusing, ambiguous, or unclassifiable information; contaminated samples; unscorable psychological examinations; use of data from previous Air Force studies at which the 1987 followup participant was not present; and an exemption from testing (e.g., exemption from delayed skin testing to prevent confounding of immunology panel results). These unrecoverable data points were excluded from subsequent analysis. The number of values that were not available for analyses is presented in Chapters 9 through 20 by variable and group.

Medical Records Coding Quality Control

After inventory, SAIC forwarded completed questionnaires and physical examination records to the Air Force at Brooks AFB, Texas, for diagnostic coding and verification of all subjectively reported conditions. The Air Force used the International Classification of Diseases, 9th Revision, Clinical Modification for morbidity coding; the International Classification of Diseases, 9th Revision, for mortality coding; the Systematized Nomenclature of Medicine for anatomic site coding; and the American Hospital Formulary Service for medication coding. Two coders independently processed each questionnaire and physical examination. Both codings were then subjected to a 100-percent QA and QC review, during which every posted code was checked against medical records. A third party adjudicated any discordances.

After QA and QC review and/or adjudication, information from the coding sheets was placed into the AFHS data base using a 100-percent double blind data entry and verification scheme. Any discordances were reviewed, corrected, and again subjected to double blind entry and verification. After coding and data entry, the Air Force batched the questionnaires and forwarded them to NORC in Chicago, Illinois, for data processing. The Air Force then obtained the NORC questionnaire data tape, matched this information to the Air Force data file, and resolved any differences. A single, final combined data base was produced by the contractor, and a copy was sent to the Air Force.

STATISTICAL ANALYSIS QUALITY CONTROL

Specific QC measures were developed for activities falling within the statistical analysis task: construction of data bases for the statistical analysis of each clinical chapter, the statistical analysis itself, and the preparation of the clinical chapters.

Each specialized statistical data base was constructed by defining and locating each variable within the many subparts of the composite followup data base. Although the data had been subjected to QC procedures during collection, statistical checks for outliers and other improbable values were conducted; anomalies identified by the statisticians were discussed with those responsible for the data collection, i.e., either NORC or SCRF.

The data base was frozen prior to starting the statistical analysis. However, during the data analysis, some discrepancies or data problems were identified. Each issue was investigated to determine the nature and impact on the outcome of the analyses and documented. For all but two issues, described below, the analyses were reaccomplished using revised data.

1) One Black Ranch Hand was inadvertently coded as a nonblack in the data base. Since all of the 1987 followup analyses had been completed before the error was identified, selected variables were reanalyzed to determine the impact of having one Ranch Hand misclassified on race. (Only the analyses that utilized race could be affected by this error.) Race was used in the adjusted analyses (group contrast and one stratum of the exposure index), interaction analyses, dependent variable associations, and unadjusted skin cancer analyses since Blacks were excluded. Variables were selected where (1) the result of the adjusted group contrast was significant, (2) the misclassified participant was abnormal, and/or (3) Blacks were excluded.

   For group contrasts, race was used indirectly (i.e., exclusion or covariate). For most analyses, the effect was in the third decimal place of the p-value. Changes of this order of magnitude in the significance level could result from using two different statistical methods or different software manufacturers of the same analysis method. The change in the p-value was larger for stratified analyses and nonsignificant results but would not change the overall statistical conclusion. The change in the p-values for covariate associations was slightly larger (second decimal place). However, the dependent variable-race associations are strictly summary statistics and auxiliary information with no relevance to the statistical conclusion on group differences.

The misclassified Ranch Hand was an enlisted flyer. Since the sample size for the enlisted flyer cohort is smaller (171) than for group contrasts (2,294), the change in the p-value was also slightly larger, and the change followed the same pattern as group contrasts. However, minimal emphasis is placed on the results of the exposure analysis, and the change in results would not impact the overall statistical conclusions of a clinical area.

Thus, the effect of having one participant misclassified on race does not have a substantial effect on the analysis results and did not warrant reanalysis of the data.

2) In reviewing the medical records for diabetes, it was determined that 13 participants had been misclassified (11 participants were coded in error as having a verified history of diabetes, and 2 participants coded as normal actually have a history of diabetes as verified by medical record). Verified history of diabetes was used as a dependent variable in the endocrine assessment, a candidate covariate for neurological and renal analyses, an exclusion for 2-hour postprandial glucose in the endocrine assessment, and an exclusion in the cardiovascular assessment.

   In the dependent variable analysis of verified history of diabetes, the classification of the 13 participants was corrected, and the analysis was reaccomplished. When verified history of diabetes was used as a covariate or exclusion, the misclassification of the 13 participants was judged to be negligible, and reanalysis using revised data showed little difference or was not deemed necessary.

QA largely depended on regular communication and general agreement among statisticians. Several meetings and consultations among the Air Force team, the SAIC Principal Investigator, the SAIC statisticians, and the University of Chicago staff members were held in conjunction with the development of the data analysis plan. During the course of the analysis there were frequent telephone conversations. Any problems arising in the statistical analysis were resolved by team discussion. The software was checked by comparing results from analyses on the same variable by different programs (for example, BMDP®-LR [logistic regression] and BMDP®-4F [log-linear model] will give the same results for dichotomous variables when the program options are appropriately chosen). The statisticians frequently checked that the number of observations used in an analysis was correct, and peer review ensured that the program code was appropriate for the chosen procedure. The analyses were conducted in accordance with the data analysis plan, which was reviewed extensively. Throughout the study, duplicate data bases were maintained by the Air Force and SAIC. Upon completion of the analyses, SAIC delivered all analysis software and SAS® data sets for each clinical area to the Air Force for final review and archiving.

All tables and statistical results were checked against the computer output from which they were derived, and all statistical statements in the text were checked for consistency with the results given in the tables. Additionally, drafts of chapters in the report were reviewed by the Air Force and SAIC investigators, and the QRC.

CHAPTER 6

REFERENCES

1. Bissell, A.F. 1969. CUSUM techniques for quality control. Appl. Stat. Vol. 18.

2. Lucas, J.M., and R. Crosier. 1982. Fast initial response for CUSUM quality control schemes: Give your CUSUM a headstart. Technometrics Vol. 24.

3. Grubbs, F.E. 1969. Procedures for detecting outlying observations in samples. Technometrics XI:1-21.

CHAPTER 7

STATISTICAL METHODS

This chapter summarizes the statistical approach used in the data analysis of the 1987 followup of the Air Force Health Study (AFHS). The statistical analysis emphasizes the evaluation of possible differences in health status between the Ranch Hand and Comparison group members. After preliminary analysis to check for data anomalies and to obtain a general overview of the data, the analysis comprised both simple contrasts between the two groups and more complex methods employing adjustment for important covariates. To augment these analyses, the possibility of a greater frequency of medical problems with increasing herbicide dose was assessed in the Ranch Hand group. The exposure index was used to approximate the potential herbicide exposure of each individual. The exposure index analyses paralleled the analyses of group contrasts and used the same candidate covariates. Further, longitudinal analyses were conducted for selected variables to examine group differences in the changes in these variables over time. A summary of the statistical techniques used is provided in Table 7-1. This basic approach was employed in the analyses for each clinical category.

The computer software used throughout for the more complex adjusted analyses included BMDP[1]®-LR and BMDP®-4F for discrete dependent variables, and SAS[2]® GLM for continuous dependent variables. During the analyses, assumptions underlying the statistical methods were checked and, if necessary, appropriate corrective steps were taken. For example, asymmetrically distributed data were transformed to enhance normality in continuous analyses and sparse cells were occasionally collapsed in discrete analyses.

PRELIMINARY ANALYSIS

The preliminary analysis included the calculation of basic descriptive measures for the dependent and independent variables (covariates) for each group (Ranch Hand and Comparison). The descriptive measures included frequency distributions, histograms, mean, median, standard deviation, and range. These analyses provided an overview of each variable and the relationship of the Ranch Hand group to the Comparison group. In addition, the preliminary analysis provided insight regarding the specification of normal/abnormal limits and cutpoints, and the choice of possible transformations of asymmetrically distributed data for continuous dependent variables.

Another purpose of the preliminary analysis was to examine the relationship between the covariates and the dependent variables and the relationships among the covariates. To accomplish this, cross tabulations of discrete variables were constructed and analyzed by the chi-square test or Fisher's exact test. For continuous variables, simple t-tests and analyses of variance were performed, and product-moment correlation coefficients were computed as appropriate. The preliminary analyses were accomplished with the use of SAS®. Covariate tables are presented for the dependent variable and relevant covariates and contain both descriptive statistics and corresponding p-values showing the strength and statistical significance of the associations. Associations with a p-value less than or equal to 0.05 are described as significant, and associations with a p-value greater than 0.05 but less than or equal to 0.10 are termed marginally significant or borderline significant.

TABLE 7-1.

Summary of Statistical Procedures

Chi-square Contingency Table Test

The chi-square test of independence[3] is calculated for a contingency table by the following formula:

$$X^2 = \Sigma(f_o - f_e)^2/f_e$$

where the sum is taken over all cells of the contingency table and

f_o = observed frequency in a cell

f_e = expected frequency under the hypothesis of independence.

Large values indicate deviations from the null hypothesis and are tested for significance by comparing the calculated X^2 to the tables of the chi-square distribution.

Correlation Coefficient (Pearson's Product-Moment)

The population correlation coefficient,[4] ρ, measures the strength of the linear relationship between two random variables X and Y. A commonly used sample-based estimate of this correlation coefficient is

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{[\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2]^{1/2}}$$

where the sum is taken over all (x,y) pairs in the sample. A Student's t-test based on this estimator is used to test for a significant correlation between the two random variables of interest. For the sample size of 2,294 in this study, a sample correlation coefficient of ±0.041 is sufficient to attain a statistically significant correlation at a 5-percent level for a two-sided hypothesis test, assuming normality of X and Y.

Fisher's Exact Test

Fisher's exact test[3] is a randomization test of the hypothesis of independence for a 2x2 contingency table. This technique is particularly useful for small samples and sparse cells. This is a permutation test based on the exact probability of observing the particular set of frequencies, or of sets more extreme, under the null hypothesis. The p-value presented for this hypothesis test is twice the one-tail p-value[5] with an upper bound of 1. In most cases, this p-value is quite close to the p-value associated with the continuity-corrected chi-square test statistic.[5]

TABLE 7-1. (continued)

Summary of Statistical Procedures

General Linear Models Analysis

The form of the general linear model[6] for two independent variables is:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \varepsilon$$

where

Y = dependent variable (continuous)

α = level of Y at $X_1 = 0$ and $X_2 = 0$, i.e., the intercept

X_1, X_2 = measured value of the first and second independent variables, respectively, which may be continuous or discrete

β_1, β_2 = coefficient indicating linear association between Y and X_1, Y and X_2, respectively; each coefficient reflects the effect on the model of the corresponding independent variable adjusted for the effect of the other independent variable

β_{12} = coefficient reflecting the linear interaction of X_1 and X_2, adjusted for linear main effects

ε = error term.

This model assumes that the error terms are independent and normally distributed with a mean of 0 and a constant variance. Extension to more than two independent variables and interaction terms is immediate.

Linear regression, multiple regression, analysis of variance, analysis of covariance, and repeated measures analysis of variance[7] are all examples of general linear models analyses.

Logistic Regression Analysis

The logistic regression model[3,8] enables a dichotomous dependent variable to be modeled in a regression framework with continuous and/or discrete independent variables. For two risk factors, such as group and age, the logistic regression model would be:

$$\text{logit } P = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \varepsilon$$

where

P = probability of disease for an individual with risk factors X_1 and X_2

TABLE 7-1. (continued)

Summary of Statistical Procedures

logit P = ln (P/1-P), i.e., the log odds for disease

X_1 = first risk factor, e.g., group

X_2 = second risk factor, e.g., age.

The parameters are interpreted as follows:

α = log odds for the disease when $X_1 = 0$ and $X_2 = 0$

β_1 = coefficient indicating the group effect adjusted for age

β_2 = coefficient indicating the age effect adjusted for group

β_{12} = coefficient indicating the interaction between group and age, adjusted for linear main effects

ϵ = error term.

In the absence of an interaction ($\beta_{12} = 0$), $\exp(\beta_1)$ reflects the adjusted odds ratio for individuals in Group 1 ($X_1 = 1$) relative to Group 0 ($X_1 = 0$). If the probability of disease is small, the odds ratio will be approximately equal to the relative risk.

Throughout this report, the adjusted odds ratios will be referred to as adjusted relative risks. Correspondingly, in the absence of covariates (i.e., unadjusted analysis), the odds ratios will be referred to as estimated relative risks.

Log-linear Analysis

Log-linear analysis[3] is a statistical technique for analyzing cross-classified data or contingency tables. A saturated log-linear model for a three-way table is:

$$\ln (Z_{ijk}) = U_o + U_{1(i)} + U_{2(j)} + U_{3(k)} + U_{12(ij)} + U_{23(jk)} + U_{13(ik)} + U_{123(ijk)}$$

where

Z_{ijk} = expected cell count

$U_{1(i)}$ = specific one-factor effect

$U_{12(ij)}$ = specific two-factor effect or interaction

$U_{123(ijk)}$ = three-factor effect or interaction.

TABLE 7-1. (continued)

Summary of Statistical Procedures

The simplest models are obtained by including only the significant U-terms. Adjusted relative risks are derived from the estimated U-terms from an adequately fitting model.

Proportional Odds Model Analysis

The proportional odds model[9] allows for the analysis of an ordered categorized dependent variable. The model assumes that the odds of falling below a certain level rather than above it for individuals at different levels of an independent variable $\underline{X}$ are in constant ratio. For example, if the response takes one of the four values "excellent," "good," "fair," or "poor," and $\underline{X}$ is a simple indicator variable designating group (Ranch Hand versus Comparison), then the proportional odds model states that the odds for responding "poor" versus "fair," "good," or "excellent" in the Ranch Hand group are a multiple, $\exp(\beta)$, of the corresponding odds in the Comparison group. Likewise, the odds for responding "poor" or "fair" versus "good" or "excellent" in the Ranch Hand group are the same multiple, $\exp(\beta)$, of the corresponding odds in the Comparison group, as are the odds for responding "poor," "fair," or "good" versus "excellent" in the two groups. Thus, the model is appropriate whenever one frequency distribution is "shifted left" relative to another distribution. Incorporation of other variables into $\underline{X}$ allows the estimation of proportional odds ratios adjusted for covariates.

Let the ordered response Y take values in the range 1 to K, and let $\pi_i(\underline{X})$, i=1,...,K, denote the probability of responding at level i for an individual with covariate vector $\underline{X}$. Let $\kappa_j(\underline{X})$ be the odds that $Y \leq j$ given $\underline{X}$, i.e.,

$$\kappa_j(\underline{X}) = \frac{\pi_1(\underline{X}) + \pi_2(\underline{X}) + \ldots + \pi_j(\underline{X})}{\pi_{j+1}(\underline{X}) + \pi_{j+2}(\underline{X}) + \ldots + \pi_k(\underline{X})}, \quad j=1,\ldots,K-1$$

The proportional odds model specifies that

$\kappa_j(\underline{X}) = \kappa_j \exp(\beta'\underline{X})$, for constant κ_j.

Thus, the ratio of odds for individuals at covariate levels $\underline{X}_1$ and $\underline{X}_2$ is

$$\frac{\kappa_j(\underline{X}_1)}{\kappa_j(\underline{X}_2)} = \exp\{\beta'(\underline{X}_1 - \underline{X}_2)\}$$

and depends only on $\underline{X}_1 - \underline{X}_2$ and not on j.

TABLE 7-1. (continued)

Summary of Statistical Procedures

Two Sample t-Test
A statistical test for determining whether or not it is reasonable to conclude that two population means are unequal utilizes the t-distribution.[10] Tests can be performed when population variances are equal or unequal; different t-distributions are used, however. This test can be used when the two populations are independent (e.g., Ranch Hand and Comparison) or dependent (e.g., 1982 and 1987 measurements on the same participant for a longitudinal analysis).

GROUP CONTRASTS

Contrasts of the Ranch Hand and Comparison groups, termed core analyses, consisted of a series of steps taken to ascertain whether or not a statistically significant difference existed between these groups for every dependent variable examined.

Both unadjusted and adjusted analyses were performed and are presented for each clinical chapter. Unadjusted analyses consisted of contrasts between the Ranch Hand and Comparison groups of the mean values, or proportion with abnormal values, of each dependent variable by t-tests, Fisher's exact test, or chi-square tests, as appropriate. Adjusted analyses have taken into account significant covariates in the assessment of possible group differences using general linear, logistic regression, proportional odds, or log-linear models. Covariates measured in 1985 but not in 1987 were used where necessary. The terms significant, marginally significant, and borderline significant, as defined previously, are also used for the descriptions of the group contrast results and the adjusting models.

Continuous Dependent Variables

When the dependent variable was continuous, the general linear models procedure of SAS® was used to fit a model of the dependent variable in terms of group (Ranch Hand or Comparison), appropriate covariates, group-by-covariate interactions, and interactions between covariates. The covariates were either continuous or discrete. If necessary, the dependent variable was transformed prior to analysis to enhance the normality of its distribution.[11] When a "best" model was fitted, according to the strategy outlined below, the test for significance of the group difference was then performed on the adjusted group means,[12] provided there were no significant interactions between group and any of the covariates. Group differences in the presence of interactions were assessed using stratification by different levels of the covariate(s) involved in the interaction.

Discrete Dependent Variables

Discrete dependent variables were analyzed by methods parallel to those used for continuous variables. For dichotomous variables, logistic regression was carried out by the BMDP®-LR program; for this analysis, the covariates could be either continuous or discrete. For polychotomous dependent variables, where the number of categories is three or more, log-linear modeling was performed by the use of the BMDP®-4F program. For this type of analysis, all covariates must be categorized. The logistic and log-linear models were fitted by the method of maximum likelihood.

To make the results parallel to those obtained by logistic regression, i.e., to maintain the distinction between dependent and independent variables, the marginals were fixed in the model[13] by incorporating the full k-factor interaction term involving the k covariates used in the model, effectively converting the log-linear model into a logit model. The significance of the relative risk for group was determined by examination of the appropriate model, as determined by the model that included all statistically significant effects and group, or by examination of the significant interactions. Adjusted relative risks were derived from the coefficients of the appropriate model.

Modeling Strategy

In each clinical category, many covariates were considered for inclusion in the statistical models for adjusted group contrasts. The large number of such covariates and consequent interaction terms and the resulting difficulties of interpretation obligated the adoption of a strategy for identifying a moderately simple model involving only significant effects. Interpretation of possible group differences was then made in the context of this simple model. A schematic representation of the generalized modeling strategy is provided in Figure D-1 of Appendix D.

An initial model including all two-factor interactions was examined. Global tests at the 0.05 level, or individual tests at the 0.15 level, were used to screen out unnecessary two-factor interactions. Thereafter, a hierarchical stepwise deletion strategy was used, eliminating effects with a p-value greater than 0.05 (except the main group effect) and retaining lower order effects if involved in higher order interactions, to result in the simplest model. Interactions between covariates, if significant, were retained as effects.

Occasionally, because of numerous covariates and the resulting sparse cell sizes, preliminary investigations of unadjusted and adjusted dependent variable-covariate associations were conducted to identify initial models using a subset of the original candidate covariates. These methods are specific to the dependent variables and the relevant covariates for a clinical area and are discussed in the individual chapters.

In the analysis for a particular dependent variable, when no group-by-covariate interactions were significant at the 0.05 significance level, adjusted group means or relative risks are presented. If any group-by-covariate interaction was significant at the 0.05 significance level, then

the behavior of the group difference was explored for different levels (categories) of the covariate to identify the subpopulation(s) for which a group difference existed. Further, if any group-by-covariate interaction was significant at a level between 0.01 and 0.05, the adjusted group means or relative risks are also presented, after dropping the interaction term from the model.

Power

Conducting a statistical test using a Type I error, also called alpha level, of 0.05 (α=0.05) means that on the average, in 5 cases out of 100, a false conclusion would be made that an association (herbicide effect) exists when, in reality, there is no association. The other possible inference error (called a Type II error) is that of failing to detect an association when it actually exists. The probability of a Type II error (β) for a statistical test is 1 minus the power of the test. The power of the test is the probability that the test will reject the hypothesis of no herbicide effect when an effect does in fact exist. The power of a test depends on the group sample sizes, the disease prevalence rate, and the true group difference measured in terms of relative risk.

Table 7-2 contains the approximate sample size required to detect specific relative risks with an approximate power of 0.8 (β=0.2) using an alpha level of 0.05 for a two-sided test and assuming equal Ranch Hand and Comparison group sizes and unpaired analyses. Relative risk is the ratio of the disease prevalence rate of the Ranch Hand and Comparison groups. Conditions or diseases with Comparison population prevalence rates and exposed group relative risks corresponding to those below the heavy black line on the table can be detected with a probability of at least 0.8 with the sample sizes used in this study. That is, the sample sizes used for this study are greater than the sample size requirements appearing in this table below the heavy black line, implying a power of at least 0.8 in these situations. These tables imply that this study has adequate power to detect relative risks of 2.0 or more for major aggregates of disease such as heart disease and total cancer.

Table 7-3 provides the same information for continuous variables in terms of percentage mean shift and variability, assuming unpaired testing of a normally distributed variable and equal sample sizes.

In the 1987 followup of the AFHS, 995 Ranch Hands participated in the physical examination. In this size group, the chance of identifying zero cases of a disease with a prevalence of 1/500 or less is greater than 10 percent. Table 7-4 contains the probability of encountering no cases of disease states for cumulative prevalence rates of 1/200, 1/500, 1/1,000, 1/2,000, 1/5,000, and 1/10,000.

EXPOSURE INDEX ANALYSES

The exposure index was constructed to approximate the level of dose of the herbicide received by each member of the Ranch Hand group. Exposure index analyses were conducted to determine if differences existed in the levels of the dependent variable corresponding to the levels of the exposure index.

TABLE 7-2.

Required Sample Sizes to Detect Group Differences
in Two-Sample Testing Assuming Equal Sample Sizes*
(Relative Risk Calculations)

Prevalence Rate of Disease in Comparison Population	Relative Risk (Multiplicative Factor of Prevalence Rate for Ranch Hand Group)										
	1.25	1.50	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00
$\frac{1}{10,000}$	2,822,082	783,901	235,164	78,384	43,544	29,391	21,944	17,415	14,393	12,243	10,640
$\frac{1}{5,000}$	1,410,882	391,901	117,564	39,184	21,766	14,691	10,968	8,703	7,193	6,118	5,317
$\frac{1}{1,000}$	281,922	78,301	23,484	7,824	4,344	2,931	2,187	1,735	1,433	1,218	1,058
$\frac{1}{500}$	140,802	39,101	11,724	3,904	2,166	1,461	1,089	863	713	606	526
$\frac{1}{100}$	27,906	7,741	2,316	768	424	285	211	167	137	116	100
$\frac{1}{50}$	13,794	3,821	1,140	376	206	137	101	79	65	54	47
$\frac{1}{10}$	2,504	685	199	62	32	20	14	10	7	5	4

*This study has unequal sample sizes; therefore, the tabled values are understated.

TABLE 7-3.

Required Sample Sizes to Detect Group Differences in Two-Sample Testing Assuming Equal Sample Sizes* (Mean Shift Calculations)

	Variability (σ/μ)				
Mean Shift	0.05	0.10	0.25	0.50	0.75
0.5%	1,568	6,272	39,200	156,800	352,800
1.0%	392	1,568	9,800	39,200	88,200
1.5%	175	697	4,356	17,423	39,200
2.0%	98	392	2,450	9,800	22,050
2.5%	63	251	1,568	6,272	14,112
5.0%	16	63	392	1,568	3,528
7.5%	7	28	175	697	1,568
10.0%	4	16	98	392	882

*This study has unequal sample sizes; therefore, the tabled values are understated.

TABLE 7-4.

Probability of Zero Cases as a Function of Prevalence

Disease Prevalence	Probability of Finding Zero Cases in a Group of 995 Participants
1/10,000	0.905
1/5,000	0.820
1/2,000	0.608
1/1,000	0.370
1/500	0.136
1/200	0.007

The exposure index was trichotomized as high, medium, and low, separately, for each of the three occupational groups (officer, enlisted flyer, enlisted groundcrew). Separate analyses were conducted for each occupational cohort, since relative differences in exposure between the groups could not be determined from historical records. Discrete dependent variables were evaluated using log-linear and logistic regression models, treating exposure level as a discrete variable (by means of two indicator variables) and adjusting for covariates. For continuous dependent variables, a general linear model was fit, adjusting for covariates and using two indicator variables to designate exposure level. Contrasts between medium and low, and between high and low exposure levels, were also performed.

The modeling strategy used for the exposure index analysis follows: First, the initial model did not include covariate-by-covariate interactions, and secondly, all the covariates were included as main effects in the final model. Further, in the presence of small frequencies of abnormalities, exposure index analyses were occasionally carried out using only the main effects model (i.e., using exposure index and all the covariates but not including interaction terms).

The terms significant, marginally significant, and borderline significant, as defined for the dependent variable-covariate associations, are used for the descriptions of the exposure index results.

LONGITUDINAL ANALYSES

General

Another objective of the AFHS is to observe and contrast the change in various laboratory parameters or the presence of abnormalities and disease between the Ranch Hand and the Comparison groups. This followup objective is

not without scientific, logistic, and interpretive challenge, considering mobile populations, problems of loss to study, changing laboratory methods and diagnostic criteria, and the diversity of many changing factors over a period encompassing numerous followup examinations. The following sections describe the statistical procedures used for both continuous and discrete longitudinal data. In general, the analyses used data from two timepoints: Baseline and the 1987 followup. Tabulations include 1985 summary statistics in addition to those from Baseline and the 1987 followup for reference purposes. The summary statistics for the 1985 followup are limited to those participants included in the Baseline to 1987 longitudinal analysis who also participated in the 1985 followup examination.

Continuous Data

A repeated measures analysis of variance procedure[7] was used to analyze the variables measured on a continuous scale. The model describing the effects on the dependent variable (Y) for the kth participant (π_k) in the ith group (α_i) at the jth time (β_j) is as follows:

$$Y_{ijk} = \mu + \alpha_i + \pi_{k(i)} + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

The sources of variation and associated degrees of freedom are given below:

Source	Degrees of Freedom*
Group (Ranch Hand vs. Comparison)	1
Subject/Group	n_1+n_2-2
Time	1
Group-by-Time	1
(Subject-by-Time)/Group	n_1+n_2-2

*Based on n_1=944 Ranch Hands and n_2=1,113 Comparisons when no data are missing at either time endpoint for any participants.

The primary source of interest is the group-by-time interaction ($\alpha\beta_{ij}$). Using measurements on each participant at two times (Baseline and 1987 followup), a test on this interaction is equivalent to a test on the equality of mean differences (over time) between the Ranch Hand and Comparison groups.

Care must be taken in the interpretation of the main effect, time (β_j) (i.e., the difference in the means between the two timepoints). This effect is confounded by laboratory differences.

The source of variation due to group (α_i) reflects a difference between the overall Ranch Hand and Comparison means (averaged over both times). This source should complement the group difference findings at Baseline and at the 1987 followup, provided the group changes are consistent (no significant group-by-time interaction). All available participants were used in the group contrast analyses at each timepoint, whereas only the participants with both measures were included in the repeated measures analysis.

Discrete Data

Frequently, data were collected as normal-abnormal, or continuous measurements were discretized into this binomial response. For the Ranch Hand and Comparison groups, a Baseline versus 1987 followup 2x2 (normal-abnormal) table of frequencies was prepared (paired data):

		Followup: Ranch Hand		Followup: Comparison	
		Abnormal	Normal	Abnormal	Normal
Baseline: Abnormal			b		d
Baseline: Normal		a		c	

As with the McNemar test,[8] only the Normal to Abnormal and Abnormal to Normal off-diagonal data were used in further contrasts. A conventional chi-square test was used to test the null hypothesis of a comparable pattern of change for the two groups (unpaired data).[14]

		Pattern of Change: Normal→ Abnormal	Pattern of Change: Abnormal→ Normal
Group: Ranch Hand		a	b
Group: Comparison		c	d

This test is equivalent to testing no group-by-time-by-endpoint interaction in a matched pair analysis.[15]

SUMMARY

The statistical methods and modeling strategies employed in this study are commonly applied in large cohort studies. The use of stepwise procedures and the descriptions of group-by-dependent variable-by-covariate interactions are also common to all large studies. The many analyses and corresponding tabulations have been prescribed in an analytical plan[16] and are intended to address many different approaches to data analysis and to allow the reader to check the results.

CHAPTER 7

REFERENCES

1. BMDP statistical software manual. 1985. Berkeley, California: University of California Press.

2. SAS/STAT guide for personal computers, version 6 ed. 1985. Cary, North Carolina: SAS Institute Inc.

3. Bishop, Y.M.M., S.E. Feinberg, and P.W. Holland. 1975. Discrete multivariate analysis: Theory and practice. Cambridge: MIT Press.

4. Kleinbaum, D.G., and L.L. Kupper. 1978. Applied regression analysis and other multivariable methods. North Scituate, Massachusetts: Duxbury Press.

5. Armitage, P. Statistical methods in medical research. Blockwell Scientific Publications.

6. Neter, J., and W. Wasserman. 1974. Applied linear statistical models. Homewood, Illinois: Richard D. Irwin, Inc.

7. Winer, B.J. 1971. Statistical principles in experimental design. 2d ed. New York: McGraw Hill.

8. Fleiss, J.L. 1981. Statistical methods for rates and proportions. 2d ed. New York: John Wiley & Sons.

9. McCullagh, P. 1980. Regression models for ordinal data. J. Royal Stat. Soc. B42(2):109-142.

10. Daniel, W.W. 1978. Biostatistics: A foundation for analysis in the health sciences. New York: John Wiley & Sons.

11. Box, G.E.P., and D.R. Cox. 1964. An analysis of transformations. J. Royal Stat. Soc. B26:211.

12. Searle, S.R., F.M. Speed, and G.A. Milliken. 1980. Population marginal means in the linear model: An alternative to least squares means. The American Statistician 34:216-221.

13. Feinberg, S.E. 1981. The analysis of cross-classified data. Cambridge: MIT Press.

14. Breslow, N.E., and N.E. Day. 1980. Statistical methods in cancer research. Volume 1. The Analysis of Case-Control Studies, Lyon, International Agency for Research on Cancer (IARC Scientific Publications No. 32).

15. Breslow, N.E. 1982. Covariance adjustment of relative risk in matched studies. Biometrics 38:661-672.

16. Air Force Health Study Analytical Plan for the 1987 Followup. 14 October 1988. McLean, Virginia: Science Applications International Corporation.

CHAPTER 8

EXPOSURE INDEX

An increased incidence of adverse health effects at higher levels of exposure represents a classic increasing dose-response relationship. The potential relationship of clinical endpoints with herbicide exposure can be tested using an estimate of exposure, hereinafter called an exposure index, for each member of the Air Force Health Study Ranch Hand cohort.

An index of potential exposure to any of four 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD)-containing herbicides from fixed-wing spray missions was constructed for each Ranch Hand from the available historical data. The index serves as an estimate only, since the actual concentration of TCDD in the herbicides varied from lot to lot and individual assessments of actual body burden during or just after exposure in Vietnam were not feasible. The four TCDD-containing herbicides used in the development of the index are Herbicide Orange, Herbicide Purple, Herbicide Pink, and Herbicide Green. The exposure index was designed to correlate as closely as possible with exposure and is not an exact measure of actual individual exposures. Although the index contains errors when used to assess the exposure of a specific individual, it was thought to provide some degree of useful inference for groups of similarly exposed individuals.

The exposure index for each subject is defined as the product of the TCDD weighting factor, the gallons of TCDD-containing herbicide sprayed in the Republic of Vietnam (RVN) theater during the tour of the subject, and the inverse of the number of men sharing the subject's duties during the tour of the subject. Each of these factors is described below.

The TCDD weighting factor reflects the estimated relative concentration of TCDD in the herbicides sprayed. The estimated mean concentrations of TCDD in Herbicide Orange, Herbicide Purple, Herbicide Pink, and Herbicide Green are 2 parts per million (ppm), 33 ppm, 66 ppm, and 66 ppm, respectively. Archived samples of Herbicide Purple indicate a mean concentration of approximately 33 ppm, and samples of Herbicide Orange had a mean concentration of about 2 ppm. Since Herbicide Pink and Herbicide Green contained twice as much 2,4,5-T as Herbicide Purple, the estimated mean concentration of TCDD in these two herbicides was approximately 66 ppm. Based on procurement records and dissemination information, a combination of Herbicide Green, Herbicide Pink, and Herbicide Purple was sprayed between January 1962 and 1965. Using available data on the number of gallons procured and sprayed,[1] the estimated mean concentration of TCDD for this time period was 48.0 ppm.

The Herbs Tape and other data sources[1] indicate that only Herbicide Orange was disseminated after 1 July 1965. Normalizing to Herbicide Orange, the weighting factor becomes 24.0 before 1 July 1965 and 1.0 after 1 July 1965.

Using the Herbs Tape, Contemporary Historical Evaluation and Combat Operations Reports, and quarterly operations reports, a table of gallons of TCDD-containing herbicide sprayed for each month of the operation was constructed. Gallons of Herbicides Purple, Pink, and Green were converted to Herbicide Orange equivalent gallons based on the TCDD weighting factor of 24.0. This information is provided in Table E-1 of Appendix E.

The dates and occupational category of each Ranch Hand's tour(s) in the RVN were obtained by a manual review of military records. The study design specified five occupational categories: (1) officer-pilot, (2) officer-navigator, (3) officer-nonflying, (4) enlisted flyer, and (5) enlisted groundcrew. Based on the review of the records, the Ranch Hand manning for each occupational category by month was compiled.

A numeric exposure index reflecting the effective number of gallons of Herbicide Orange to which each individual was potentially exposed was computed. For analysis purposes, the values were categorized as high, medium, or low for each occupational category. Only three occupational categories were used. The three officer categories were combined into one since pilots and navigators were exposed in the same manner and the officer-nonflying category, which included a relatively small number of participants, consisted of administrators whose exposure was considered to be essentially zero. The overall group of "nonexposed" Ranch Hands, estimated at approximately 2 percent of the Ranch Hand group, was analyzed in the low exposure category (see Table 8-1), conceivably leading to dilution of the exposure analyses and group contrasts. The exposure index categorizations developed for the Baseline study and used in this report are provided in Table 8-1, along with the frequencies of Ranch Hand participants by occupation and exposure level. The cutpoints for the categories of the exposure index were the 33rd and 66th percentiles of the exposure index distributions within each of the three occupational strata (officer, enlisted flyer, and enlisted groundcrew). Ranch Hands with administrative duties were assigned an index of zero.

TABLE 8-1.

Exposure Index Categorization of 995 Compliant Ranch Hands

Occupational Group	Exposure Category	Effective Herbicide Orange Gallons Corresponding to Exposure Category	Number of Ranch Hand Participants in Exposure Category
Officer	Low	<35,000	130
	Medium	35,000-70,000	124
	High	>70,000	125
Enlisted Flyer	Low	<50,000	55
	Medium	50,000-85,000	63
	High	>85,000	53
Enlisted Groundcrew	Low	<20,000	147
	Medium	20,000-27,000	158
	High	>27,000	140
Total			995

The calculated exposure index is not specific to individual and, therefore, may underestimate exposure for those individuals whose jobs required routine handling of herbicide. For example, maintenance schedules for the aircraft herbicide spray tank required that an emergency dump valve be periodically greased, requiring entry into the tank. The current exposure index cannot distinguish between men who received such exposure and men who did not. The extent to which individuals are misclassified by the current exposure index is not known, precluding bias calculations at this time.

Every laboratory and physical examination endpoint in this study was assessed for dose-response effects versus the calculated exposure index. Current TCDD assay results did not correlate with the exposure index, with or without adjustment for time since exposure. These exposure index analyses are presented because some members of the Advisory Committee of the Science Panel of the Agent Orange Working Group advised that they be included in this report.

Because of the acknowledged imprecision of the exposure index, Air Force efforts are under way to measure TCDD levels in serum collected from participants in the 1987 followup. Serum was obtained for 1,999 of the 2,294 participants and is currently being analyzed by the Centers for Disease Control. As of September 1989, results of 1,366 serum specimens (888 Ranch Hands and 468 Comparisons) have been reported. These results are summarized in Table 8-2.

TABLE 8-2.

Serum TCDD Results

	Ranch Hand			Comparison		
Stratum	Sample Size	Median*	Range*	Sample Size	Median*	Range*
Officer--Pilot	247	7.3	0.0-42.6	118	4.7	0.0-13.1
Officer--Navigator	63	9.3	1.1-36.0	27	4.9	2.4-7.9
Officer--Nonflying	19	6.7	3.0-24.9	4	4.0	0.0-4.6
Enlisted Flyer	152	17.2	0.0-195.5	76	4.3	0.0-12.8
Enlisted Groundcrew	407	23.6	0.0-617.8	243	4.2	0.0-54.8
All Personnel	888	12.4	0.0-617.8	468	4.4	0.0-54.8

*In parts per trillion.

These results indicate that (1) Comparisons have background levels; (2) Ranch Hands have higher current TCDD levels than Comparisons; and (3) among Ranch Hands, nonflying enlisted personnel have the highest and officers have the lowest TCDD levels.

The relationship between current TCDD body burden and the constructed exposure index will be described in a future report. This report is expected in early 1991.

CHAPTER 8

REFERENCES

1. Young, A.L., J.A. Calcagni, C.E. Thalken, and J.W. Tremblay. 1978. The toxicology, environmental fate, and human risk of herbicide orange and its associated dioxin. Technical report OEHL-TR-78-92, USAF Occupational and Environmental Health Laboratory, Brooks AFB, Texas. 247 pp.

CHAPTER 9

GENERAL HEALTH

INTRODUCTION

Background

The effects of heavy, acute exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD or dioxin) have been demonstrated in a number of different organ systems. It is plausible, therefore, that chronic low-dose exposure to TCDD might induce subtle, interrelated effects that are not organ-system specific, but are manifest only in general terms, or affect the state of "well-being." Numerous animal studies and studies of exposed populations have shown that many enzyme induction systems throughout the body are affected by TCDD, which may have wide-ranging results.[1-4] However, it is difficult to measure overall health objectively. For this reason, general health outcomes, as defined by this study, should be judged in context with other more specific clinical endpoints.

Baseline Summary Results

Five general health variables were included in the 1982 Baseline examination: self-perception of health, appearance of illness or distress, relative age, sedimentation rate, and percent body fat. In the analysis of the Baseline examination data, a statistically significant difference in self-perception of health was found between the Ranch Hand and Comparison groups, with a greater percentage of Ranch Hands reporting their health as fair or poor than Comparisons (20.6% vs. 14.2%). This was true in both the younger and older age groups (Est. RR: 1.82, p=0.017 for individuals 40 or less and Est. RR: 1.35, p=0.025 for individuals older than 40). Since only 9 of 1,811 individuals were reported by the examining physician as appearing ill or distressed, this designation was apparently reserved for only very ill or distressed individuals. Nevertheless, eight of the nine individuals were Ranch Hands, the difference being of borderline significance (p=0.056). Conversely, more Ranch Hands than Comparisons were reported by the examiners as appearing younger than their actual ages (4.9% vs. 2.5%, p=0.029). No overall differences in percent body fat or sedimentation rate were found, although a significant interaction between group and age for sedimentation rate was noted; younger Ranch Hands had fewer sedimentation rate abnormalities than did their Comparisons, whereas no difference was found in participants older than 40. In the exposure index analyses conducted in the Ranch Hand group, no statistically significant dose-response relationships were detected.

1985 Followup Study Summary Results

General physical health was evaluated by the same five measures used in the Baseline examination (self-perception of health, appearance of illness or distress, relative age, percent body fat, and sedimentation rate).

The Ranch Hands again rated their health as fair or poor more often than the Comparisons (9.1% vs. 7.3%, respectively), although this difference was not statistically significant. However, further analysis revealed a significant group-by-occupation interaction; differences were largely confined to the enlisted groundcrew category where the adjusted relative risk was 1.90 ($p=0.003$).

Ten individuals were reported as appearing acutely ill or distressed at the followup examination. In contrast to the Baseline examination, four were Ranch Hands and six were Comparisons; thus, no group difference was suggested.

Relative age, as determined by the examining physician, was not significantly different in the two groups. There was a significant group-by-occupation interaction, but none of the estimated relative risks for the occupational categories was significant.

The (geometric) mean sedimentation rates did not differ significantly, either unadjusted or after adjustment for age, race, occupation, personality score, and an age-by-personality score interaction. However, in the discrete analysis, 5.8 percent of the Ranch Hands had sedimentation rate abnormalities (>20 mm/hr), contrasted to 3.6 percent in the Comparison group. This difference was significant both unadjusted ($p=0.013$) and adjusted for age and personality score ($p=0.011$).

The mean percent body fat of the Ranch Hands was significantly lower than the Comparisons (21.10 vs. 21.54, respectively; $p=0.037$), and the difference was of nearly the same magnitude after adjustment for age, race, and occupation. However, both unadjusted and adjusted tests of the discretized data did not reveal significant group differences, although the percent obese (>25% body fat) was lower in the Ranch Hands than in the Comparisons.

Detailed exposure analyses were done on four general health variables (appearance of illness or distress was too sparse for testing). Only one analysis detected a significant effect, namely, a positive association between sedimentation rate abnormalities and increasing exposure in the enlisted flyer cohort. Overall, no consistent pattern of exposure effects was discernible, and the exposure findings at the 1985 followup were similar to the findings at Baseline.

Longitudinal differences between the 1982 Baseline and the 1985 followup examination were assessed by analyses of two discrete variables: self-perception of health and sedimentation rate. Analysis of self-perception of health showed no significant group differences in the change over time, with the Ranch Hand and Comparison groups reporting symmetrical improvements in their perceptions over the 3-year period. The sedimentation rate analysis, however, revealed a highly significant group difference ($p=0.002$), due to a reversal of findings between examinations; i.e., a significant detriment in the (younger) Comparisons at the Baseline examination versus a significant detriment in the Ranch Hands at the followup examination.

Parameters of the 1987 General Health Assessment

Dependent Variables

The 1987 general health assessment was based on questionnaire, physical examination, and laboratory examination data. The variables analyzed were identical to those in the 1982 Baseline and 1985 followup examinations.

Questionnaire Data

During the questionnaire health interview, each study participant was asked, "Compared to other people your age, would you say your health is excellent, good, fair, or poor?" This self-reported perception was analyzed as a measure of the general health status of each participant, though susceptible to varying degrees of conscious and subconscious bias.

No participants were excluded for medical reasons from the analysis of this variable.

Physical Examination Data

Three variables derived from the physical examination were analyzed in the assessment of general health. The physician at the examination recorded the appearance of illness or distress (yes/no) of the study participant. The physician also noted the appearance of the subject as younger than, older than, or the same as his stated age. To the degree that the examining physicians were kept blind to the participant's group membership, these assessments were less subject to bias than the self-perception of health.

Percent body fat, a measure of the relative body mass of an individual and calculated from height and weight recorded at the physical examination, was also analyzed. Percent body fat was calculated from a metric body mass index,[3] and the formula was

$$\text{Percent Body Fat} = \frac{\text{Weight (kg)}}{[\text{Height (m)}]^2} \times 1.264 - 13.305.$$

This variable was analyzed in both the discrete and continuous forms. For purposes of discrete analyses, percent body fat was dichotomized as lean/normal ($\leq$25%) and obese (>25%).

No participants were excluded for medical reasons from the analyses of these three variables.

Laboratory Examination Data

The erythrocyte sedimentation rate (mm/hr), measured at the laboratory examination, was analyzed. Although nonspecific, a high sedimentation rate is

a generally accepted indicator of an ongoing disease process. This variable was analyzed in both the discrete and continuous forms. The logarithmic transformation was used to enhance statistical normality for continuous analyses.

No participants were excluded for medical reasons from the analysis of this variable.

Covariates

The effects of the covariates age, race, occupation, and personality type were examined in the assessment of general health, both in pairwise associations with the dependent variables and in adjusted statistical analyses. Age, race, and occupation were matching variables and were used for analyses with all dependent variables. Age was used in its continuous form for all adjusted analyses. Personality type was used in the analysis of self-perception of health and sedimentation rate only. Personality type was determined from the Jenkins Activity Survey administered during the 1985 followup examination. This variable was derived from a discriminant-function equation based on questions that best discriminate men judged to be Type A from those judged as Type B.[6] Positive scores reflect the Type A direction and negative scores the Type B direction. The personality-type score was used in its continuous form for all adjusted analyses. Participants at the 1987 followup examination who had not attended the 1985 followup examination had missing information for personality type, as did a few participants who could not be classified in 1985, because the Jenkins Activity Survey was not administered at the 1987 followup examination.

Relation to Baseline and 1985 Followup Studies

As noted above, the same variables were analyzed for the 1987 followup study as for the Baseline and 1985 followup studies.

For longitudinal analyses, sedimentation rate was analyzed as a discrete variable. The normal range for sedimentation rate for the Baseline examination was less than or equal to 12 mm/hr; the Scripps Clinic and Research Foundation (SCRF) normal range for sedimentation rate for the 1987 followup was less than or equal to 20 mm/hr. Self-perception of health was also analyzed in the longitudinal analyses.

Statistical Methods

The basic statistical analysis methods used in this chapter are described in Chapter 7. In addition, proportional odds model analysis, also described in Chapter 7, was used.

Table 9-1 summarizes the statistical analyses performed for the 1987 general health assessment. The first part of this table describes the dependent variables (including units for laboratory measurements), the source of the data used for the analysis, the form(s) of the data (discrete and/or continuous), and cutpoints. This table also presents candidate covariates examined in adjusted Ranch Hand versus Comparison contrasts (also referred to

TABLE 9-1.

Statistical Analysis for the General Health Assessment

Dependent Variables

Variable (Units)	Data Source	Data Form	Cutpoints	Candidate Covariates	Statistical Analyses
Self-Perception of Health	Q-SR	D	Excellent Good Fair Poor	AGE RACE OCC PERS	UC:CS,PO AC:LR,PO CA:CS UE:CS,FT AE:LR L:OR
Appearance of Illness or Distress	PE	D	Yes No	AGE RACE OCC	UC:CS,FT AC:LR CA:CS,FT UE:CS,FT AE:LR
Relative Age	PE	D	Younger Same Older	AGE RACE OCC	UC:CS,PO AC:LR,PO CA:CS UE:CS,FT AE:LR
Percent Body Fat	PE	D/C	Lean/Normal: $\leq$ 25% Obese: >25%	AGE RACE OCC	UC:CS,FT,TT AC:LR,GLM CA:CC,TT, GLM,CS,FT UE:CS,FT, GLM AE:LR,GLM
Sedimentation Rate (mm/hr)	LAB	D/C	Normal: $\leq$ 20 Abnormal: >20	AGE RACE OCC PERS	UC:CS,FT,T AC:LR,GLM CA:CC,TT, GLM,CS,FT UE:CS,FT, GLM,TT AE:LR,GLM L:OR

TABLE 9-1. (continued)

Statistical Analysis for the General Health Assessment

Covariates

Variable (Abbreviations)	Data Source	Data Form	Cutpoints
Age (AGE)	MIL	D/C	Born ≥1942 Born 1923-1941 Born ≤1922
Race (RACE)	MIL	D	Nonblack Black
Occupation (OCC)	MIL	D	Officer Enlisted Flyer Enlisted Groundcrew
Personality Type (PERS)	PE (1985)	D/C	A Direction B Direction

Abbreviations:

Data Source:
LAB--1987 SCRF laboratory results
MIL--Air Force military records
PE (1985)--1985 SCRF physical examination
PE--1987 SCRF physical examination
Q-SR--1987 NORC questionnaire (self-reported)

Data Form:
D--Discrete analysis only
D/C--Discrete and continuous analyses for dependent variables; appropriate form for analysis (either discrete or continuous) for covariates

Statistical Analyses:
UC--Unadjusted core analyses
AC--Adjusted core analyses
CA--Dependent variable-covariate associations
UE--Unadjusted exposure index analyses
AE--Adjusted exposure index analyses
L--Longitudinal analyses

Statistical Methods:
CC--Pearson's product moment correlation coefficient
CS--Chi-square contingency table test
FT--Fisher's exact test
GLM--General linear models analysis
LR--Logistic regression analysis
OR--Chi-square test on the odds ratio
PO--Proportional odds model analysis
TT--Two-sample t-test

TABLE 9-2.

Number of Participants With Missing Data for the General Health Assessment

Variable	Group: Analysis Use	Ranch Hand	Comparison	Total
Self-Perception of Health	DEP	0	1	1
Appearance of Illness or Distress	DEP	0	1	1
Sedimentation Rate	DEP	1	3	4
Personality Type (1985 data)	COV	39	78	117

Abbreviations: DEP--Dependent variable (missing data)
COV--Covariate (missing data)

as core analyses), exposure index analyses, and dependent variable-covariate associations. To conserve space, abbreviations are used extensively in the body of the table and are defined in footnotes.

The second part of this table provides a further description of candidate covariates. Standard abbreviations for these variables, which will be used subsequently in this chapter, are presented, as well as data source, data form, and cutpoints.

Table 9-2 provides a list of the number of participants with missing data for the dependent variables and covariates described in Table 9-1.

RESULTS

Ranch Hand and Comparison Group Contrast

Questionnaire Variable

Self-Perception of Health

Table 9-3 gives the frequency distribution of self-perception of health for the Ranch Hand and Comparison groups, as well as the estimated relative risk of reporting one's health as fair or poor. The two distributions were

TABLE 9-3.

Unadjusted Analysis for General Health Variables by Group

Variable	Statistic	Group: Ranch Hand		Group: Comparison		Contrast	Est. Relative Risk (95% C.I.)	p-Value
Self-Perception of Health	n	995		1,298		Overall		0.250
	Number/%							
	Excellent	474	47.6%	651	50.2%	Fair/Poor	1.01 (0.72,1.40)	0.975
	Good	454	45.6%	560	43.1%	vs.		
	Fair	51	5.1%	75	5.8%	Exc./Good		
	Poor	16	1.6%	12	0.9%			
Appearance of Illness or Distress	n	995		1,298				
	Number/%							
	Yes	9	0.9%	7	0.5%	Yes vs. No	1.68 (0.62,4.54)	0.300
	No	986	99.1%	1,291	99.5%			
Relative Age	n	995		1,299		Overall		0.671
	Number/%							
	Younger	11	1.1%	10	0.8%	Older	0.94 (0.66,1.35)	0.741
	Same	929	93.4%	1,213	93.4%	vs.		
	Older	55	5.5%	76	5.8%	Younger/Same		
Percent Body Fat	n	995		1,299				
	Mean	21.46		21.67			—	0.335
	95% C.I.	(21.14,21.79)		(21.39,21.95)				
	Number/%							
	Lean/Normal	803	80.7%	1,013	78.0%	Obese	0.85 (0.69,1.04)	0.111
	Obese	192	19.3%	286	22.0%	vs.		
						Lean/Normal		
Sedimentation Rate	n	994		1,296				
	Mean[a]	5.30		5.09			—	0.255
	95% C.I.[a]	(5.02,5.60)		(4.87,5.32)				
	Number/%							
	Abnormal	70	7.0%	54	4.2%	Abnormal vs.	1.74 (1.21,2.51)	0.003
	Normal	924	93.0%	1,242	95.8%	Normal		

— Estimated relative risk not applicable for continuous analysis of a variable.
[a] Transformed from natural logarithm scale.

similar, with 6.7 percent of the members from each group reporting their health as fair or poor. Slightly fewer Ranch Hands than Comparisons reported their health as excellent, but neither the overall comparison of the frequency distributions nor a proportional odds model fit to the ordinal data revealed a significant group difference (p=0.250 and p=0.267, respectively). The downward trend in the percentage of individuals reporting their health as fair or poor noted in the 1985 followup report continued: 20.4 percent at Baseline, 9.1 percent at the 1985 followup examination, and 6.7 percent at the 1987 followup examination in the Ranch Hand group; 15.9 percent, 7.3 percent, and 6.7 percent, respectively, in the Comparisons.

Tests of association between self-perception of health and each of the covariates (age, race, occupation, and personality type) appear in Appendix F, Table F-1. These tests indicated an association of borderline significance with age (p=0.062), with slightly fewer individuals born in or after 1942 perceiving their health as fair or poor compared to those born between 1923 and 1941 or those born in or before 1922 (5.6% vs. 7.5% and 7.2%, respectively).

There was a highly significant association (p<0.001) between self-perception of health and occupation: 4.1 percent of the officers reported their health as fair or poor compared to 8.6 percent of the enlisted flyers and 8.2 percent of the enlisted groundcrew. There was also a highly significant (p<0.001) association with personality type. Equal percentages of Type A's and Type B's reported their health as fair or poor (6.6%), but 54.5 percent of the Type A's reported their health as excellent compared to 45.6 percent of the Type B's.

The results of adjusted analyses of self-perception of health are presented in Table 9-4. A logistic regression model with the outcome dichotomized as fair/poor or excellent/good was used to analyze this variable (age and personality type were incorporated as continuous independent variables).

There was a significant age effect (p=0.005) as well as a significant occupation-by-personality type interaction (p=0.012). In contrast to the 1985 examination, however, there was no significant interaction between group and occupation (p=0.632). A proportional odds model adjusting for age, race, occupation, and personality type also did not reveal any statistically significant group difference (adjusted proportional odds: 1.09, 95% C.I.: [0.92, 1.29], p=0.305).

Physical Examination Variables

Appearance of Illness or Distress

A total of 16 individuals were reported by the examining physicians as appearing ill or distressed (see Table 9-3). Nine were from the Ranch Hand group and seven from the Comparisons. Upon examination of the dependent variable-by-covariate associations, a significant association between the appearance of illness or distress and age was detected (p=0.016). All but 1 of the 16 ill or distressed individuals were born in or before 1941 (Appendix F, Table F-1).

TABLE 9-4.

Adjusted Analysis for General Health Variables by Group

Variable	Statistic	Group: Ranch Hand	Group: Comparison	Contrast	Adj. Relative Risk (95% C.I.)	p-Value	Covariate Remarks
Self-Perception of Health	n	956	1,220	Fair/Poor vs. Exc./Good	1.01 (0.72,1.42)	0.999	AGE (p=0.005) OCC*PERS (p=0.012)
Appearance of Illness or Distress	n	995	1,298	Yes vs. No	1.67 (0.62,4.52)	0.308	AGE (p=0.004)
Relative Age	n	995	1,299	Older vs. Younger/Same	0.94 (0.66,1.34)	0.726	OCC (p<0.001)
Percent Body Fat	n Adj. Mean 95% C.I.	995 21.58 (21.02,22.13)	1,299 21.80 (21.26,22.33)		—	0.314	AGE*RACE (p=0.032) AGE*OCC (p=0.002)
	n	995	1,299	Obese vs. Lean/Normal	0.84 (0.69,1.04)	0.104	AGE (p<0.001) OCC (p<0.001)
Sedimentation Rate	n Adj. Mean[a] 95% C.I.[a]	955 5.32 (5.04,5.61)	1,218 5.16 (4.92,5.42)		—	0.413	OCC (p<0.001) AGE*PERS (p=0.006)
	n	955	1,218	Abnormal vs. Normal	1.70 (1.17,2.48)	0.005	AGE (p<0.001) OCC (p=0.002) PERS (p=0.042)

— Adjusted relative risk not applicable for continuous analysis of a variable.

[a]Transformed from natural logarithm scale.

Due to the sparseness of the data, an analysis was performed adjusting only for age (in continuous form); the results are shown in Table 9-4. Age was again highly significant (p=0.004), but the adjusted relative risk was essentially unchanged from the unadjusted relative risk.

Relative Age

Table 9-3 shows very little difference between the Ranch Hand and Comparison groups in relative age. Five-and-one-half percent of the Ranch Hands appeared older than their stated age and 94.5 percent appeared younger than or the same as their stated age. In the Comparisons, 5.8 percent appeared older and 94.2 percent appeared younger than or the same as their stated age, giving an estimated relative risk slightly less than 1 for this dichotomization of the outcomes. A proportional odds model fit to the ordinal responses also did not reveal any significant group difference (estimated proportional odds: 0.90, 95% C.I.: [0.65, 1.26], p=0.544).

Examination of the covariate effects (Table F-1 of Appendix F) revealed a significant association between relative age and age itself (p<0.001) (a higher percentage of older individuals than younger individuals were reported as appearing younger than their stated age), race (p=0.039) (Blacks more often appeared younger than their stated ages than nonblacks), and occupation (p<0.001) (relatively more officers appeared younger than their stated ages and fewer appeared older than their stated ages as compared to enlisted personnel).

Logistic regression analyses detected only a significant main effect of occupation (p<0.001) (Table 9-4). The adjusted relative risk was nearly identical to the unadjusted value. A proportional odds model fit to the ordinal responses revealed significant age and occupation effects (p=0.032 and p<0.001, respectively), but no group difference was evident (adjusted proportional odds: 0.90, 95% C.I.: [0.64, 1.25], p=0.520).

Percent Body Fat

Percent body fat was analyzed both as a continuous variable and trichotomized into lean (<10%), normal (10-25%), and obese (>25%) categories. Few individuals were lean (four Ranch Hands and five Comparisons) and thus relative risk estimates and logistic regression analyses were based upon a dichotomization into obese versus lean/normal categories. Mean percent body fat was not significantly different in the two groups (21.46% in the Ranch Hands vs. 21.67% in the Comparisons). The percent obese in the Ranch Hand group was less than that in the Comparisons, but not significantly so.

Examination of dependent variable-by-covariate associations (Table F-1) found significant age and occupation effects. Percent body fat was significantly correlated with age (p=0.032), and the percent obese was highest in those born between 1923 and 1941 (p=0.008). There was no statistically significant difference in mean percent body fat across the three occupational groups, but the percent obese was higher in the enlisted flyers than in the officers and higher still in the enlisted groundcrew (p=0.007).

Adjusted analyses of the percent body fat as a continuous variable detected significant age-by-race ($p=0.032$) and age-by-occupation ($p=0.002$) interactions (Table 9-4). The adjusted means in the Ranch Hand and Comparison groups, however, were not significantly different. Discrete analyses of the percent obese detected significant age and occupation effects ($p<0.001$ for both), but the adjusted relative risk was not significantly different from 1.

Laboratory Examination Variable

Erythrocyte Sedimentation Rate

The erythrocyte sedimentation rate was also analyzed in both continuous and discrete forms. Histograms generated for each group were skewed markedly to the right and thus the data were analyzed after transformation to a (natural) logarithm scale, which led to more symmetrical distributions. For the discrete analysis, the values were dichotomized into abnormal (>20 mm/hr) or normal ($\leq$20 mm/hr) categories.

The group means were not significantly different, but the percent abnormal was significantly greater in the Ranch Hand group than in the Comparison group (Est. RR: 1.74, 95% C.I.: [1.21, 2.51], $p=0.003$). A similar finding was noted in the 1985 followup report.

Age, occupation, and personality type were all significantly associated with the sedimentation rate (Appendix F, Table F-1). Older individuals had significantly higher sedimentation rates ($p<0.001$), although the correlation was only 0.230. The percent abnormal increased steadily with age. Enlisted flyers exhibited the highest mean sedimentation rates and the highest percent abnormal; officers had the lowest mean and lowest percent abnormal. P-values for the association with occupation were 0.006 and 0.034 for the continuous and discrete forms of sedimentation rate, respectively. Personality type was negatively associated with sedimentation rate; 6.6 percent of Type B individuals were abnormal compared to 4.2 percent of Type A's ($p=0.017$).

Adjusted analyses led to essentially the same conclusions as the unadjusted analyses (Table 9-4). There was a significant occupation effect ($p<0.001$) and an age-by-personality type interaction ($p=0.006$) in the continuous analysis, but the adjusted group means were not significantly different. Logistic regression analysis revealed significant effects of age ($p<0.001$), occupation ($p=0.002$), and personality type ($p=0.042$), and a significant adjusted relative risk of 1.70 (95% C.I.: [1.17, 2.48], $p=0.005$).

Exposure Index Analysis

The exposure index, expressed in equivalent gallons of dioxin-containing herbicide potentially encountered by each Ranch Hand during his tour of duty in Vietnam, was categorized as low, medium, or high. Separate analyses were performed within each occupational cohort. (A detailed description of the exposure index can be found in Chapter 8.) The frequency distributions for each variable and associated tests and comparisons within each occupational cohort are shown in Table 9-5. "M vs. L" and "H vs. L" are the estimated

TABLE 9-5.

Unadjusted Exposure Index for General Health Variables by Occupation

Variable	Occupation	Statistic	Exposure Index Low	Exposure Index Medium	Exposure Index High	Exposure Index Contrast	Est. Relative Risk (95% C.I.)	p-Value
Self-Perception of Health	Officer	n	130	124	125	Overall[a]		0.300
		Number/%						
		Excellent	81 62.3%	89 71.8%	71 56.8%	M vs. L[a]	0.83 (0.22,3.18)	0.787
		Good	44 33.8%	31 25.0%	45 36.0%	H vs. L[a]	1.94 (0.63,5.96)	0.246
		Fair	4 3.1%	3 2.4%	6 4.8%			
		Poor	1 0.8%	1 0.8%	3 2.4%			
	Enlisted Flyer	n	55	63	53	Overall[a]		0.416
		Number/%						
		Excellent	21 38.2%	20 31.8%	23 43.4%	M vs. L[a]	0.64 (0.14,2.98)	0.569
		Good	30 54.6%	40 63.5%	24 45.3%	H vs. L[a]	1.63 (0.43,6.13)	0.472
		Fair	2 3.6%	2 3.2%	4 7.6%			
		Poor	2 3.6%	1 1.6%	2 3.8%			
	Enlisted Groundcrew	n	147	158	140	Overall[a]		0.107
		Number/%						
		Excellent	59 40.1%	57 36.1%	53 37.9%	M vs. L[a]	0.92 (0.43,1.96)	0.834
		Good	73 49.7%	86 54.4%	81 57.9%	H vs. L[a]	0.39 (0.15,1.05)	0.061
		Fair	13 8.8%	13 8.2%	4 2.9%			
		Poor	2 1.4%	2 1.3%	2 1.4%			

TABLE 9-5. (continued)

Unadjusted Exposure Index for General Health Variables by Occupation

Variable	Occupation	Statistic	Exposure Index: Low	Exposure Index: Medium	Exposure Index: High	Exposure Index Contrast	Est. Relative Risk (95% C.I.)	p-Value
Appearance of Illness or Distress	Officer	n	130	124	125	Overall		0.362
		Number/%						
		Yes	1 0.8%	0 0.0%	2 1.6%	M vs. L	—	0.999
		No	129 99.2%	124 100.0%	123 98.4%	H vs. L	2.10 (0.19,23.43)	0.970
	Enlisted Flyer	n	55	63	53	Overall		0.118
		Number/%						
		Yes	2 3.6%	0 0.0%	0 0.0%	M vs. L	—	0.430
		No	53 96.4%	63 100.0%	53 100.0%	H vs. L	—	0.430
	Enlisted Groundcrew	n	147	158	140	Overall		0.139
		Number/%						
		Yes	1 0.7%	0 0.0%	3 2.1%	M vs. L	—	0.964
		No	146 99.3%	158 100.0%	137 97.9%	H vs. L	3.20 (0.33,31.11)	0.586

TABLE 9-5. (continued)

Unadjusted Exposure Index for General Health Variables by Occupation

Variable	Occupation	Statistic	Exposure Index: Low	Exposure Index: Medium	Exposure Index: High	Exposure Index Contrast	Est. Relative Risk (95% C.I.)	p-Value
Relative Age	Officer	n	130	124	125	Overall[b]		0.368
		Number/%						
		Younger	1 0.8%	3 2.4%	3 2.4%	M vs. L[b]	0.69 (0.19,2.50)	0.569
		Same	123 94.6%	117 94.4%	120 96.0%	H vs. L[b]	0.34 (0.07,1.70)	0.187
		Older	6 4.6%	4 3.2%	2 1.6%			
	Enlisted Flyer	n	55	63	53	Overall[b]		0.289
		Number/%						
		Younger	0 0.0%	0 0.0%	0 0.0%	M vs. L[b]	0.34 (0.08,1.40)	0.136
		Same	48 87.3%	60 95.2%	49 92.4%	H vs. L[b]	0.56 (0.15,2.04)	0.379
		Older	7 12.7%	3 4.8%	4 7.6%			
	Enlisted Groundcrew	n	147	158	140	Overall[b]		0.806
		Number/%						
		Younger	3 2.0%	1 0.6%	0 0.0%	M vs. L[b]	1.30 (0.51,3.33)	0.582
		Same	136 92.5%	146 92.4%	130 92.9%	H vs. L[b]	1.34 (0.51,3.49)	0.555
		Older	8 5.4%	11 7.0%	10 7.1%			

TABLE 9-5. (continued)

Unadjusted Exposure Index for General Health Variables by Occupation

Variable	Occupation	Statistic	Exposure Index Low	Exposure Index Medium	Exposure Index High	Exposure Index Contrast	Est. Relative Risk (95% C.I.)	p-Value
Percent Body Fat	Officer	n	130	124	125	Overall		0.997
		Mean	21.42	21.45	21.40	M vs. L	—	0.964
		95% C.I.	(20.68,22.16)	(20.51,22.38)	(20.71,22.08)	H vs. L	—	0.969
		Number/%						
		Lean/Normal	106 81.5%	103 83.1%	107 85.6%	Overall		0.677
		Obese	24 18.5%	21 16.9%	18 14.4%	M vs. L	0.90 (0.47,1.72)	0.749
						H vs. L	0.74 (0.38,1.45)	0.384
	Enlisted Flyer	n	55	63	53	Overall		0.163
		Mean	20.17	21.72	22.20	M vs. L	—	0.148
		95% C.I.	(18.94,21.39)	(20.11,23.35)	(20.59,23.80)	H vs. L	—	0.071
		Number/%						
		Lean/Normal	48 87.3%	50 79.4%	40 75.5%	Overall		0.268
		Obese	7 12.7%	13 20.6%	13 24.5%	M vs. L	1.78 (0.66,4.85)	0.258
						H vs. L	2.23 (0.81,6.12)	0.119
	Enlisted Groundcrew	n	147	158	140	Overall		0.896
		Mean	21.67	21.57	21.37	M vs. L	—	0.876
		95% C.I.	(20.80,22.53)	(20.71,22.43)	(20.47,22.27)	H vs. L	—	0.645
		Number/%						
		Lean/Normal	114 77.6%	127 80.4%	108 77.1%	Overall		0.754
		Obese	33 22.4%	31 19.6%	32 22.9%	M vs. L	0.84 (0.49,1.46)	0.542
						H vs. L	1.02 (0.59,1.78)	0.936

TABLE 9-5. (continued)

Unadjusted Exposure Index for General Health Variables by Occupation

Variable	Occupation	Statistic	Exposure Index Low	Exposure Index Medium	Exposure Index High	Exposure Index Contrast	Est. Relative Risk (95% C.I.)	p-Value
Sedimentation Rate	Officer	n	130	124	124	Overall		0.869
		Mean[c]	4.91	5.18	4.93	M vs. L	—	0.630
		95% C.I.[c]	(4.26,5.66)	(4.42,6.06)	(4.22,5.77)	H vs. L	—	0.965
		Number/%						
		Abnormal	7 5.4%	8 6.4%	4 3.2%	Overall		0.477
		Normal	123 94.6%	116 93.6%	120 96.8%	M vs. L	1.21 (0.43,3.45)	0.719
						H vs. L	0.59 (0.17,2.15)	0.401
	Enlisted Flyer	n	55	63	53	Overall		0.849
		Mean[c]	6.25	6.28	5.79	M vs. L	—	0.980
		95% C.I.[c]	(5.08,7.70)	(5.15,7.65)	(4.49,7.47)	H vs. L	—	0.634
		Number/%						
		Abnormal	5 9.1%	5 7.9%	7 13.2%	Overall		0.629
		Normal	50 90.9%	58 92.1%	46 86.8%	M vs. L	0.86 (0.24,3.15)	0.826
						H vs. L	1.52 (0.45,5.13)	0.497
	Enlisted Groundcrew	n	147	158	140	Overall		0.720
		Mean[c]	5.14	5.15	5.54	M vs. L	—	0.988
		95% C.I.[c]	(4.45,5.95)	(4.46,5.94)	(4.81,6.39)	H vs. L	—	0.479
		Number/%						
		Abnormal	12 8.2%	12 7.6%	10 7.1%	Overall		0.948
		Normal	135 91.8%	146 92.4%	130 92.9%	M vs. L	0.92 (0.40,2.13)	0.857
						H vs. L	0.86 (0.36,2.07)	0.749

[a]Outcome categories: Fair/Poor vs. Excellent/Good.
[b]Outcome categories: Older vs. Younger/Same.
[c]Transformed from natural logarithm scale.
—Estimated relative risk/confidence interval not given due to cells with zero frequency; estimated relative risk not applicable for continuous analysis of a variable.

relative risks for medium versus low exposure and high versus low exposure, respectively. The results of adjusted exposure index analyses are presented in Table 9-6. Covariates examined included age, race, and personality type; on certain occasions when data were sparse, fewer terms were retained in the final model. The final interpretation of these exposure data must await the reanalysis of the clinical data using the results of the serum dioxin assay. This report is expected in 1991.

Questionnaire Variable

Self-Perception of Health

No statistically significant differences overall, nor any significant contrasts for any of the occupational cohorts, were found.

There were also no statistically significant findings from the adjusted analyses. There was a borderline overall effect in the enlisted groundcrew category ($p=0.074$), but this was due to a relative risk for the high vs. low contrast that was less than 1, and not indicative of an increasing dose-response relationship.

Physical Examination Variables

Results from the exposure index analyses of the appearance of illness or distress, relative age appearance, and percent body fat are also given in Tables 9-5 and 9-6.

Appearance of Illness or Distress

The number of abnormalities was quite sparse for the appearance of illness or distress; none of the overall tests was statistically significant. Adjusted analyses were not carried out for this variable.

Relative Age

There were no significant dose-response relationships for relative age in either the unadjusted or adjusted analyses.

Percent Body Fat

Percent body fat was analyzed in both the continuous and discrete forms. For the unadjusted analyses, there were no significant differences among the mean percent body fat levels across the three exposure level categories in any of the three occupational cohorts, nor were significant differences obtained in any of the discrete analyses. Adjusted analyses also did not reveal any significant exposure level effects in the officers or enlisted groundcrew. When analyzed in the discrete form, there was a highly significant ($p=0.005$) exposure index-by-age interaction in the enlisted flyer cohort, however. This

TABLE 9-6.

Adjusted Exposure Index for General Health Variables by Occupation

Variable	Occupation	Statistic	Exposure Index			Exposure Index Contrast	Adj. Relative Risk (95% C.I.)	p-Value
			Low	Medium	High			
Self-Perception of Health	Officer	n	122	121	118	Overall[a]		0.516
						M vs. L[a]	0.73 (0.18,2.87)	0.646
						H vs. L[a]	1.50 (0.46,4.95)	0.503
	Enlisted Flyer	n	53	63	51	Overall[a]		0.398
						M vs. L[a]	1.13 (0.21,6.12)	0.887
						H vs. L[a]	2.45 (0.56,10.63)	0.234
	Enlisted Groundcrew	n	144	151	133	Overall[a]		0.074
						M vs. L[a]	1.17 (0.54,2.56)	0.689
						H vs. L[a]	0.41 (0.15,1.11)	0.078
Relative Age	Officer	n	130	124	125	Overall[b]		0.332
						M vs. L[b]	0.62 (0.16,2.34)	0.478
						H vs. L[b]	0.31 (0.06,1.62)	0.165
	Enlisted Flyer	n	55	63	53	Overall[b]		0.325
						M vs. L[b]	0.36 (0.09,1.48)	0.159
						H vs. L[b]	0.55 (0.15,2.02)	0.368
	Enlisted Groundcrew	n	147	158	140	Overall[b]		0.816
						M vs. L[b]	1.27 (0.49,3.28)	0.617
						H vs. L[b]	1.34 (0.51,3.50)	0.555