### TABLE 5-28.

**Medication Use of Partially Compliant Versus Fully Compliant Nonblack Participants**

<table>
<thead>
<tr>
<th>Compliance Status</th>
<th>Medication Use</th>
<th>Number</th>
<th>Percent</th>
<th>Ranch Hand</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>123</td>
<td>42</td>
<td>167</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>832</td>
<td>44</td>
<td>1,043</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>955</td>
<td></td>
<td>1,210</td>
<td>2,165</td>
</tr>
<tr>
<td>Partial</td>
<td>Yes</td>
<td>1</td>
<td>25</td>
<td>3</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>7</td>
<td>21</td>
<td>27</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>8</td>
<td></td>
<td>30</td>
<td>38</td>
</tr>
</tbody>
</table>

### TABLE 5-29.

**Work Loss of Partially Compliant Versus Fully Compliant Nonblack Participants**

<table>
<thead>
<tr>
<th>Compliance Status</th>
<th>Work Loss</th>
<th>Number</th>
<th>Percent</th>
<th>Ranch Hand</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>796</td>
<td>44</td>
<td>1,010</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>155</td>
<td>44</td>
<td>200</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>951</td>
<td></td>
<td>1,210</td>
<td>2,161</td>
</tr>
<tr>
<td>Partial</td>
<td>Yes</td>
<td>8</td>
<td>22</td>
<td>28</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>8</td>
<td></td>
<td>30</td>
<td>38</td>
</tr>
</tbody>
</table>
CHAPTER 6
QUALITY CONTROL

During the first AFHS followup, stringent adherence to quality assurance (QA) was planned for and upheld throughout the study, from project initiation to final product delivery and acceptance by the Air Force. A quality program plan was developed for this study cycle, outlining all contract activities requiring periodic and/or systematic QA and quality control (QC) monitoring.

The purpose of this chapter is to provide an overview of the specific QA measures developed and used by the project team, specifically in the areas of administrative QC; questionnaire, physical, and psychological examination QC; laboratory QC measures; data base management QA; and statistical QC.

ADMINISTRATIVE QUALITY ASSURANCE

In recognition of the magnitude, complexity, and importance of the AFHS, a Quality Review Committee (QRC) was established at the initiation of the third-year followup for the purpose of providing general oversight to the AFHS QA Program and advice on the appropriateness of program management and QC actions. The QRC was composed of senior corporate personnel from the prime contractor. These independent reviewers remained separate from the project management staff. The QRC met formally each quarter to review recent study progress and any issues that either had an impact on study quality or were perceived as a potential problem.

Assisting the QRC in day-to-day oversight responsibilities was a QA officer responsible for reviewing procedures, performance, and work products from all task managers and key project staff. As part of the monitoring function, the QA officer received exception reports from project task managers whenever an incident occurred that appeared to affect study quality. Monthly reports were also prepared for the Air Force, documenting project compliance with project QA criteria and noting any instances of non-compliance.

An additional measure of corporate QC was implemented through independent QA audits of individual project tasks. Members of the QRC determined first-hand whether QA procedures for a particular task were being conducted, whether procedures were appropriate for the task, and whether QA was complete for all aspects of each task.

The remainder of this chapter comprises specific QA procedures followed for the individual tasks.

QUESTIONNAIRE QUALITY CONTROL

NORC used both onsite and home-office QA procedures to produce a comprehensive data set. All AFHS questionnaires were pretested to evaluate
their completion time and participant acceptability before they were used at the SCRF. Onsite QC procedures included weekly observation and rating of each interviewer, editing of every questionnaire at the completion of the interview, and monitoring of participant evaluations. The Air Force also continuously conducted QA observations of all onsite activities. QC of data processing included manually editing each questionnaire, including a 100-percent verification of critical items for each questionnaire, computerized cleaning (with both single item and interitem review for range and consistency), identifying outliers, and reviewing the actual questionnaire copy to reconcile or correct detected errors.

All telephone surveys were monitored for quality and accuracy of interviewer performance by NORC supervisors. The telephone survey supervisor monitored 3 percent of each interviewer's calls to assure an appropriate presentation and an accurate transcription of responses. An additional 5 percent of the participants were recontacted after the interview to evaluate interviewer performance and validate that the correct respondent had been contacted.

NORC recruited and trained interviewers according to the detailed procedures described in Chapter 3. A minimum number of interviewers was selected to reduce interviewer variability. Additionally, these individuals were blinded to the participants' exposure status to avoid any bias. Interviewers were required to ask questions exactly as recorded, and in the order in which they appeared. No personal interpretation was allowed.

An onsite field manager closely supervised each interviewer's work regularly, observing individual interviews weekly during the examination schedule. The field manager reported directly to the NORC Project Director weekly, and was reviewed by the Project Director during quarterly site visits, to ensure direct accountability by the home office and the field manager for promptly resolving any issues.

Specifically, interviewers were checked for accuracy in questionnaire skip patterns, probing, circling of the correct code, control of the interview, voice quality, reading, and use of associated documents. When called for, the onsite manager gave immediate retraining after each observation and documented the content of this training. At weekly meetings, held with all interviewers, the field manager used generalizations from individual interviewer performance observations to train an entire group of interviewers.

The NORC field manager also monitored participant evaluations of the study closely and used the information gathered to plan and implement retraining. The manager and staff edited each completed questionnaire before it was shipped to Chicago, attempting to retrieve missing data while the study participant was at the physical examination site. Missing or ambiguous data were also retrieved by telephone when necessary.

Spouse fertility data were obtained independently of the participant interview by sending the mail questionnaire while the study participant was at the examination site, and by having a group meeting for wives who accompanied their spouses to the clinic site, where they could complete their questionnaires in private. The Assistant Survey Director in Chicago supervised and edited all interviews conducted at home with participants and spouses.
Once the participant and spouse questionnaires were received in Chicago, they were edited for completeness by a coding supervisor and staff dedicated to the AFHS for the entire project. Resolution of inconsistencies was accomplished by staff members, who standardized all responses prior to keypunching. Questionnaires were then coded, and a 10-percent recode was done on open-ended items. When a batch failed the 10-percent recode, the entire batch was recoded and the coding staff was retrained. One hundred percent quality control was accomplished by the Air Force.

During data entry, range validity checks were performed and 10 percent of the most important items in each questionnaire was verified. Data were then passed through a computer program that checked for inter- and intra-column errors. When errors were detected, the questionnaires were reviewed and the errors corrected. The process continued until no errors were detected by the cleaning program. Then, frequencies were reviewed and any anomalies or errors previously undetected were corrected by reviewing the questionnaires on a case-by-case basis. All corrections were entered into the data tape, but no changes were made to the data recorded in the questionnaires. QA reports were generated monthly, detailing the summary statistics on the number of questionnaires reviewed, the number and types of transcriptions failing QC checks, and the average number of coding errors per batch processed.

PHYSICAL EXAMINATION QUALITY CONTROL

QC was emphasized in the physical examination, as this data source provided most of the medical information for clinical and epidemiological 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 Technical Team 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 10 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 examination data as they formulated all diagnoses. 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 diagnosticians.

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 to ensure total continuity and sensibility of the final examination contents. (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 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 five 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. Every staff member, therefore, from the initial telephone recruiter to the nurse coordinator and the Project Manager, emphasized courtesy, empathy, assistance, and personalized treatment of each participant.

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 (BSL-2) approval ratings and were certified by the College of American Pathology (CAP).

Hematology assays were performed on Coulter S Plus® equipment; sedimentation rate determinations were performed using the large-tube Westergren method. The DuPont Automated Chemical Analyzer® (ACA) was used to perform the biochemical assays; radioimmunoassays (RIA) were done with standard test kits; and porphyrin was assayed by high-performance liquid
chromatography at the Mayo Clinic in Rochester, Minnesota. Hepatitis B tests
were performed using Abbott kits, and manually performed electrophoresis and
monospecific antibodies were used for immunoglobulin assays. Blood-cell
counts were performed with standard microscopy, and Clinitek, a reflectance
spectrometry 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 entries were
kept for each test so that any QC results could be retraced.

Procedural QC included using instrumentation and reagents from one lot
number 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 respon­
sible individual and forwarded to the QA officer and the project management

team.

An additional measure of quality control introduced during the study was
the CUSUM tests run with trilevel controls. In particular, the fast initial
response cumulative sum (FIR CUSUM) QC technique was used. It has an
advantage in detecting long-term subtle drift that could have substantial
adverse analytical consequences. 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. Coefficient of variation (CV) standards were
established before the study for each test. All adjacent patient samples
were reanalyzed after the equipment was thoroughly checked and fresh controls
were run.

FIR CUSUM generally has been applied to QC in industry, particularly in
high-volume, high-precision applications. To our knowledge, FIR CUSUM has
not generally been applied in a biomedical setting. According to SCRF
laboratory personnel, this procedure proved so successful in the AFHS that
most of the SCRF clinical laboratory will begin using it in the near future.
As the examination portion of this study ended, all 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.

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:

- 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., PHA stimulation, natural killer cell effector/target ratios)
- 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 from normal due to procedural setup error, assay malfunction, equipment malfunction, or assay technical error
- 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 phytohemagglutinin (PHA), pokeweeds, mixed lymphocyte culture (MLC), and natural killer cell consisted of monitoring assay controls, duplicate sample reproducibility, and any 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. Grubbs' statistical test 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, pokeweeds) effect was
followed by daily evaluation of the radioactive counts in counts per minute (cpm) for each mitogen. When counts fell below expected values, suggesting that reagent deterioration had occurred, new aliquots were used.

QC measures for the cell surface marker assays were calculation of T<sup>+</sup> + T<sub>8</sub> / T<sub>11</sub> cell ratios, evaluation of flow cytometer computer outputs (cytograms and histograms), and duplicate sample testing. T<sup>+</sup> + T<sub>8</sub> / T<sub>11</sub> 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 percent positive cells for each surface marker was determined in the duplicates and viewed graphically using a microcomputer program. Any significant differences between duplicates were noted and followed for abnormal trends.

On completion of this followup effort, the entire cellular immunology data base was reviewed by the Air Force team, laboratory staff, and consultants. Comments attached to the data points were also reviewed. Any data point that appeared unusual was reviewed and identified 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.

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 an independent diagnostician, 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 about to be described 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.

6-7
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 base management process.

Forms and questionnaires were carefully designed to ensure that all required data elements would be collected according to the Study Protocol. 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.

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 Statistical Analysis System (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.
Two Levels of Quality Control Applied to All Collected Data Prior to Statistical Analysis
Customized mark-sense forms were also developed and OMR technology was used 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 immediately 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 so that interexaminer variation could be analyzed. 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. (See examples of mark-sense forms in Figures 4-3 and 4-4.)

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.

Careful monitoring identified trends where individual data values were missed as well as the frequency with which individual examiners incorrectly marked their examination forms. 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.

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 nonobvious 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. These unrecoverable data points were excluded from subsequent analysis.

Medical Records Coding Quality Control

Upon completion of the NORC data processing, all AFHS questionnaires were forwarded to SAIC for the medical coding of reported conditions. The International Classification of Diseases, 9th Revision, Clinical Modification (morbidity); International Classification of Diseases, 9th Revision (mortality); Systematized Nomenclature of Medicine (anatomic site); and American Hospital Formulary Service (medications) coding schemes were used, suitably modified. Each questionnaire was coded by two coders working independently. The results of the two coders were forwarded to the USAF for 100-percent QA/QC and final adjudication. The information from the physical examination was coded similarly.

After the coding data were adjudicated, they were returned to SAIC for data entry. The coding sheets were batched, key entered, verified, and corrected. The corrections were also verified. The key entry and verification functions were performed by various operators. Five percent, or 100 records of each batch (whichever was larger), was randomly selected and subjected to manual reverification. An error rate of greater than 1 percent of this sample mandated reverification of the entire batch. In this final QA/QC check, the automated files were reviewed and compared to the hard copy by trained medical record coders, all of whom satisfied the minimum requirement of Accredited Record Technician or Registered Record Administrator eligibility.
A manual tracking system was used to retrieve medical records. A chronological log was maintained to track participant requests for authorization to obtain medical record(s), receipt of the authorizations, requests for records from the provider, and receipt of the records from the provider. Identifying information in these logs included participant name, case number, date of action, condition(s) to be verified, dependent name (if appropriate), and type of medical provider (Federal/non-Federal).

Due to the intricacies of obtaining medical records from Federal facilities, this task ultimately became the responsibility of 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 production of statistical reports to serve as the basis for 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. Lists of variables and their data sources were submitted to the Air Force for approval. 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.

QA largely depended on regular communication and general agreement among statisticians. Several meetings and consultations among the Air Force team, the 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 chosen properly). 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 USAF and SAIC. Upon completion of the analyses, SAIC delivered all analysis software and SAS data sets for each clinical area to the USAF 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 USAF and SAIC investigators, and the QRC.
Chapter 6

References


CHAPTER 7
STATISTICAL METHODS

This chapter summarizes the key statistical elements of the study design, the statistical analysis issues, and the specific statistical methods used in the analysis. Additional details may be found in the USAF Study Protocol.

The primary focus of the statistical analysis was a contrast of health status of the Ranch Hand and Comparison groups. Assessments were made of the proportions of participants with abnormal findings and of mean levels of key laboratory measurements. The analyses encompassed both simple contrasts between the two groups and more complex methods, in which adjustment was made for important covariates.

In addition to these analyses, the possibility of an increasing response of medical problems with herbicide dose was explored, since if indeed there were an effect, more problems would be expected among the more heavily exposed. Although exact dosage information is not available, an exposure index was developed for the exposed population (the Ranch Hands) that approximates the potential herbicide exposure of each individual, incorporating information such as the occupation of the individual, his period of duty in the spraying operation, and the numbers of barrels per day of herbicide used during that period. Details on the exposure index are given in Chapter 8. Dose-response analyses were conducted for the Ranch Hands only, using this exposure index as a surrogate measure of dose.

Interpretation of the results of the exposure index analyses, however, depends critically on the accuracy of the exposure index, which presently can be regarded as only fair. (Improved dosage information will be obtained for future studies from recently developed serum dioxin assay techniques.) Thus, the analyses of overall group differences between the Ranch Hands and the Comparisons are given primary emphasis, and the exposure index analyses merely supplement them.

STATISTICAL STUDY DESIGN

An overt herbicide effect would be characterized by more symptoms, signs, abnormal laboratory tests, syndromes, or diseases in the Ranch Hand group than in the Comparison group. If the disease(s) were fatal, increased mortality might also be observed. A subclinical herbicide effect would be detected as an increase in abnormal findings on the physical examination (particularly laboratory tests) that may or may not also be associated with symptom reporting or increased mortality. Thus, the basic objective of the statistical analysis is to test for differences between the Ranch Hand (exposed) group and the Comparison (nonexposed) group.
In general, two types of data are used in the analysis. First, there are subjective data on symptoms reported by the participant in the questionnaire and in the review-of-systems section of the physical examination. Second, there are objective data, which include medical findings or signs identified during the physical examination, or by reviews of laboratory results, medical records, and death certificates.

Symptoms reported by the study participants are subjective by definition, and are subject to influences that could result in erroneous conclusions. An association found between reported symptoms and herbicide exposure must be subjected to further confirmation, as the observations may result from over- or under-reporting bias and may not be indicative of a true herbicide effect. On the other hand, the medical findings data do not suffer from the same degree of participant influence.

The medical findings and medical records review were conducted by highly trained individuals employed for the duration of the data collection and assessment phases of the study. They were held to stringent QC standards, as described in Chapter 6, to ensure that these data were as objective and accurate as possible.

Incorporated in the study design is a feature that attempts to check for and correct symptom-reporting errors. A key component is a reported symptom verification process conducted by reviewing participant medical records and findings from the physical examination. In the retrospective morbidity portion of the study, the participant is questioned on past illnesses and medical conditions. With the participant's consent, an effort is made to obtain the medical records to verify the reported condition and, thus, to substantiate any unverified conditions. In addition, the study design includes verification of negative responses to determine unreported conditions. The medical record review process is time intensive and only a portion of the data was available for analysis in this study. Over-reporting was assessed by comparing the reported illness rates with the results of the physical examination and medical record review. Similarly, the assessment and correction of under-reporting requires the review of medical records to identify unreported illnesses. Obviously, this under-reporting assessment is restricted to conditions for which medical care was obtained or that were identifiable at the physical examination.

**STATISTICAL ISSUES**

In conducting the statistical analysis of the data in this study, there are a number of underlying issues. Except for bias, which is the topic of Chapter 5, these issues are discussed in this section. However, based upon the results of the bias analysis presented in Chapter 5, all statistical analyses in the clinical chapters use the contrast of Ranch Hands versus the total Comparison group. For the purposes of completeness and cross-reference to the Baseline report, identical analyses using the contrast of the Ranch Hands versus the Original Comparisons have been conducted, and these results are presented in the form of summary tables in each chapter appendix.
Intervening Variables

When comparing any two groups of individuals, the exact proportion of diseased individuals in each group is usually found to differ. The purpose of classical statistical hypothesis testing is to determine whether the observed difference in disease rates could be due to chance alone. If the observed difference is not attributable to chance, the two groups are considered representative of two truly different populations.

If a statistically significant difference is found between the Ranch Hand group and the Comparison group, results from more rigorous statistical procedures must be examined and the medical context considered before the possibility of a causal relationship between disease and group (exposure) can be entertained. Alternatively, the absence of a statistically significant difference between groups does not exclude the possibility of a true causal relationship between exposure and disease. Thus, group associations, whether significant or not, should be examined with adjustment for other variables called intervening variables (explanatory variables, risk factors, or covariates) that may account for, or mask, a true effect. For example, the two groups might differ with respect to age or racial composition, each of which may affect the outcome of the study. To protect against this, the technique of matching was used: The Ranch Hands and Comparisons were matched on age, race, and military occupation.

Since it is not feasible to perfectly match a Comparison to an exposed individual with respect to all important explanatory variables, statistical procedures may be used to adjust for such explanatory variables so that valid interpretations can be made of apparent group differences. Thus, it was necessary to identify and collect data on suspected explanatory variables. Unfortunately, there is no way to ensure that all important intervening variables are taken into account. The best method that can be achieved is to incorporate all known covariates in the data collection and analysis.

In most studies, covariates are variables measured prior to exposure. However, in the AFHS, except for the matching variables and historical data related to events prior to service in Southeast Asia, most covariate values were obtained at the Baseline or first followup interview and physical examination, which occurred 10 to 20 years following exposure. These covariates can generally be referred to as time-dependent covariates. They can elucidate the causal path between exposure and a particular disease; however, they are in a sense both dependent and independent variables, and therefore, analyses involving such covariates require careful interpretation.

Besides covariates, both confounding variables and interactions must also be considered. A confounding variable is an intervening variable associated with both disease and exposure. (This is in contrast with a covariate that is associated only with disease.) Adjustments must be made for confounding variables to avoid a biased estimate of the group-disease relationship. An interaction exists when the effect of one variable varies across the levels of another variable. For example, the group difference might be large in one occupation group and negligible in another. Incorporating interactions in the analysis allows for the identification of subpopulations at increased or decreased risk.
Conducting a statistical test using a Type I error, also called alpha level, of 0.05 ($\alpha = 0.05$) means that, on the average, in 5 cases out of 100, a false conclusion that an association (herbicide effect) exists would be made 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 ($\beta$) 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 7-1 contains the approximate sample size required to detect specific relative risks with an approximate power of 0.8 ($\beta = 0.2$) 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 an approximate 0.8 probability with the sample sizes used in this study.

Table 7-2 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 first follow-up of the AFHS, 1,016 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 7-3 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.

Multiple Endpoints and Comparisons

In developing the Protocol for the AFHS, previous animal and epidemiologic studies, case reports, and veterans' concerns were reviewed to delineate the possible effects of exposure. The conclusion was reached that a comprehensive evaluation was needed due to the lack of an easily identifiable symptom complex in individual patients. Consequently, the morbidity study is very broad in scope, involving the collection and analysis of data related to general health indices as well as specific organ systems and clinical disease categories.

The large number of endpoints under consideration presents a difficult problem in the assessment of Type I error rates. More than 150 dependent variables were tested, not to mention tests for interaction and multiple contrasts among the low, medium, and high exposure-level categories in the exposure index analyses. Furthermore, the dependent variables were correlated to varying degrees, and this makes it even more difficult to assess the attained significance levels. To allow for multiple endpoints, Bonferroni's inequality, which requires significance at the $\alpha / K$ level where $K$ is the number of endpoints considered, may be used, but this procedure
<table>
<thead>
<tr>
<th>Occurrence Rate of Disease in Control Population</th>
<th>Relative Risk (Multiplicative Factor of Occurrence Rate for Exposed Group)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.25</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10,000</td>
<td>2,822,082</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5,000</td>
<td>1,410,882</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1,000</td>
<td>281,922</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>140,802</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>27,906</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>13,794</td>
</tr>
</tbody>
</table>

*This study has unequal sample sizes; therefore, the tabled values are understated. The similar table in the Baseline Morbidity Report, 24 February 1984, is in error because tabulated sample sizes were only one-half of their correct values.
<table>
<thead>
<tr>
<th>Mean shift</th>
<th>Variability (σ/μ)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.05</td>
<td>0.10</td>
<td>0.25</td>
<td>0.50</td>
<td>0.75</td>
</tr>
<tr>
<td>0.5%</td>
<td>1,568</td>
<td>6,272</td>
<td>39,200</td>
<td>156,800</td>
<td>352,800</td>
</tr>
<tr>
<td>1.0%</td>
<td>392</td>
<td>1,568</td>
<td>9,800</td>
<td>39,200</td>
<td>88,200</td>
</tr>
<tr>
<td>1.5%</td>
<td>175</td>
<td>697</td>
<td>4,356</td>
<td>17,423</td>
<td>39,200</td>
</tr>
<tr>
<td>2.0%</td>
<td>98</td>
<td>392</td>
<td>2,450</td>
<td>9,800</td>
<td>22,050</td>
</tr>
<tr>
<td>2.5%</td>
<td>63</td>
<td>251</td>
<td>1,568</td>
<td>6,272</td>
<td>14,112</td>
</tr>
<tr>
<td>5.0%</td>
<td>16</td>
<td>63</td>
<td>392</td>
<td>1,568</td>
<td>3,528</td>
</tr>
<tr>
<td>7.5%</td>
<td>7</td>
<td>28</td>
<td>175</td>
<td>697</td>
<td>1,568</td>
</tr>
<tr>
<td>10.0%</td>
<td>4</td>
<td>16</td>
<td>98</td>
<td>392</td>
<td>882</td>
</tr>
</tbody>
</table>

*This study has unequal sample sizes; therefore, the tabled values are understated. The similar table in the Baseline Morbidity Report, 24 February 1984, is in error because tabulated sample sizes were only one-half of their correct values.
becomes increasingly more conservative as the correlation among the endpoints increases. For the analysis results in this report, an alpha level of 0.05 was used for each dependent variable. In addition, group contrasts in strata defined by levels of a covariate involving in a group-by-covariate interaction were assessed by an alpha level of 0.05. The same was true for exposure level strata.

In light of the multiple-endpoints problem, extreme caution in the interpretation of statistical results was required. A first consideration was the strength of the association in terms of the significance of the relative risk or difference in group means. All associations with p-values of 0.10 or less were examined and are described in this report. Then, careful consideration was given to the pattern of statistically significant results. Were only a few sporadic endpoints statistically significant, or was significance achieved on a number of endpoints indicating the same organ system failure? Were the significant results all in the same direction, and did they make biological and clinical sense? Did they confirm previous studies, or were they new findings?

**Paired Versus Unpaired Analyses**

Matching subjects in a study design on selected variables improves the comparability of the groups to be compared and, depending on the relationship of the matching variables to the study objective, the matching can be used explicitly in the analysis. In this study, the Comparison group was matched to the exposed group on age (to the nearest month of birth), race (Black, nonblack), and occupational category (officer-pilot, officer-navigator, officer-nonflyer, enlisted flyer, enlisted groundcrew). The matching was exact for occupational category, nearly exact for race (three mismatches occurred because of recording errors), and very close with respect to age (69% of the mortality population was matched to the nearest month of birth and more than 95% to the nearest year of birth).

The general approach in this report, however, was to conduct unpaired analyses using all available data, based on stratification and/or covariate adjustment. In an unpaired analysis, the matching still serves to improve...
the comparability of the two groups, and precision is usually gained from the stratification and covariate adjustment.

Mortality and Morbidity Data

The AFHS incorporated both mortality and morbidity endpoints. The mortality data have been, and will continue to be, subjected to separate analysis. Interpretation of the morbidity analyses must be made in the light of the mortality results, particularly as the study continues and the number of deaths increases. Differential mortality in the two groups could obviously have an important impact on contrasts of physical examination findings in the surviving cohorts. This issue was examined in the analysis of selected diseases, for example, cancer.

Cutpoints

The variables in this study were discrete, categorical, or continuous. Many served primarily as dependent variables, and when in the continuous form, powerful analyses were possible. In other settings, particularly when log-linear or logistic regression models were fitted, it is often necessary to dichotomize or discretize the continuous variables. Discretization, by establishing suitable nonoverlapping intervals or cutpoints, was often the result of a judgment requiring both statistical and clinical input.

In general, cutpoint decisions considered the form of the variable, distribution of the variable, established values (e.g., cholesterol, normal-abnormal, as specified by a given technique in a given laboratory), scientific values set by precedence (e.g., systolic and diastolic normal threshold 140/90), and error induction by another variable (e.g., use of the blood pressure threshold in obese-armed individuals). The approach to the selection of appropriate cutpoints was to select all cutpoints on a case-by-case basis and, where indicated, use the norms of the SCRF laboratory.

Exclusions

Due to medical considerations, certain subjects were excluded from the analyses of selected clinical categories. The exclusions were generally defined in the Baseline study and are identified in the clinical chapters of this report. Other exclusions were the result of missing data.

OVERVIEW OF STATISTICAL PROCEDURES

This section summarizes the basic statistical approach used in the data analysis of the first followup of the AFHS. The approach consisted of four parts: (1) preliminary analysis of the dependent variables and covariates to check for data anomalies and to obtain a general overview of the data, (2) core analyses to carefully determine any possible effect of herbicide exposure, (3) analysis of the exposure index to investigate the dose-response relationship for the Ranch Hand group only, and (4) longitudinal analysis to examine changes over time. A summary of the statistical techniques utilized is provided in Table 7-4. This basic approach was utilized in the analyses for each clinical category.
### Chi-Square Contingency Table Test

The chi-square test of independence \(^2\) is calculated for a contingency table by the following formula:

\[
X^2 = \sum (f_o - f_e)^2 / f_e
\]

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.

### Fisher’s Exact Test

Fisher’s exact test \(^2\) is a randomization test of the hypothesis of independence for a 2x2 contingency table. This technique is useful for small samples and sparse cells. This is a permutation test based on the exact probability of observing the particular set of frequencies.

### General Linear Model Analysis

The form of the general linear model\(^1\) for two independent variables is:

\[
Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \epsilon
\]

where

- \(Y\) = dependent variable (continuous)
- \(\alpha\) = 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
- \(\beta_1, \beta_2\) = coefficient indicating linear association between \(Y\) and \(X_1\), \(Y\) and \(X_2\), respectively
- \(\beta_{12}\) = coefficient reflecting the linear interaction of \(X_1\) and \(X_2\)
- \(\epsilon\) = 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 multiple independent variables and interaction terms is immediate.
TABLE 7-4. (continued)

Summary of Statistical Procedures

Linear regression, multiple regression, analysis of variance, and analysis of covariance are all examples of general linear model analysis.

Kolmogorov-Smirnov Distribution Test

The Kolmogorov-Smirnov (K-S) test\footnote{1} is a nonparametric procedure which assesses differences between the distribution of two samples. Specifically, the K-S procedure tests the hypothesis that populations \( n_1 \) and \( n_2 \) are identical and is designed to detect all possible deviations from this hypothesis. The assumptions of the K-S test are that the observations from the two samples are mutually independent and that both sets of observations are samples from the same distribution.

Logistic Regression Analysis

The logistic regression model\footnote{2, 3} 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:

\[
\text{logit } P = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \epsilon
\]

where

- \( P \) = probability of disease for an individual with risk factors \( X_1 \) and \( X_2 \)
- \( \text{logit } P = \ln (P/1-P) \), 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:

- \( \alpha \) = log odds for the disease when both factors are at a 0 level
- \( \beta_1 \) = coefficient indicating the group effect adjusted for age
- \( \beta_2 \) = coefficient indicating the age effect adjusted for group
- \( \beta_{12} \) = coefficient indicating the interaction between group and age
- \( \epsilon \) = 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.

Homogeneity of the odds ratios across different strata was assessed by the method of Breslow and Day. ⁵

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

Proportional Odds Model

The proportional odds model ⁶ allows for the analysis of an ordered outcome 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 $X$ are in constant ratio. For example, if the response takes one of the four values "excellent," "good," "fair," or "poor," and $X$ is a simple indicator variable designating group (Ranch Hand versus Comparison), then the proportional odds model states that the odds for responding "excellent" versus "good," "fair," or "poor" in the Ranch Hand group are a multiple, $\exp(\beta)$, of the corresponding odds in the Comparison group. Likewise, the odds for responding "excellent" or "good" versus "fair" or "poor" 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 "excellent," "good," or "fair" versus "poor" 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 $X$ 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 $\pi_i(X)$, $i=1,\ldots,K$, denote the probability of responding at level $i$ for an individual with covariate vector $X$. Let $\kappa_j(X)$ be the odds that $Y \leq j$ given $X$, i.e.,

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

The proportional odds model specifies that

$$
\kappa_j(X) = \kappa_j \exp(\beta'X), \text{ for constant } \kappa_j
$$
TABLE 7-4. (continued)

Summary of Statistical Procedures

Thus the ratio of odds for individuals at covariate levels $X_1$ and $X_2$ is

$$\frac{\kappa_j(X_1)}{\kappa_j(X_2)} = \exp\{B'(X_1 - X_2)\}$$

and depends only on $X_1 - X_2$ and not on $j$.

Log-linear Analysis

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

$$\ln(Z_{ijk}) = U_0 + U_{1(i)} + U_{2(j)} + U_{3(k)} + U_{12(ij)} + U_{23(jk)} + U_{13(ik)} + U_{123(ijk)}$$

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.

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.

McNemar's Test

McNemar’s test effectively considers discordant pairs in which only the Ranch Hand or only the Comparison member in each pair experiences the abnormality. Using a chi-square approximation with continuity correction, the following statistic is used to test whether the off-diagonal entries are evenly divided:

$$\chi^2 = \frac{(b-c-1)^2}{b+c}$$

Where $b$ and $c$ are the number of pairs in which only the Ranch Hand is abnormal or only the Comparison is abnormal, respectively. This test is compared to a chi-squared distribution with one degree of freedom.
TABLE 7-4. (continued)

Summary of Statistical Procedures

Test for Linear Trend

For a kx2 contingency table in which the k groups fall into a natural order, Armitage developed a test for a linear trend in the proportions. Let $\pi_i$ denote the proportion of individuals in the $i$th row possessing some attribute (e.g., proportion of individuals with abnormal values at each of the three exposure level categories). A score, $X_i$, is assigned to each of the $k$ levels of the row variable, and the regression coefficient, $\beta$, of $\pi_i$ on $X_i$ is estimated. The regression coefficient is estimated in the usual way except that $\pi_i$ is weighted by the sample size, $n_i$, in each row. $t_1SE(\beta)$ provides a normal deviate for testing the null hypotheses of $\beta = 0$. 

7-13
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 for the construction of composite variables, the plausibility of normal/abnormal limits and cutpoints, and the choice of possible transformations to enhance the normality of the distribution of continuous dependent variables.

Another purpose of the preliminary analysis was to examine the relationship between the covariates and the dependent variables and the relationships between and among the covariates. To accomplish this, cross tabulations of discrete variables were constructed and analyzed by the chi-square, or Fisher's exact test. For continuous variables, simple t-tests of group differences were done and product-moment correlation coefficients were computed. The preliminary analyses were accomplished with the use of the SAS®. Selected covariate tables are presented in the clinical chapters for illustration.

Core Analysis

The core analysis consisted of a series of steps taken to ascertain whether or not the data indicated a significant difference between the Ranch Hand and Comparison groups for each dependent variable.

Both unadjusted and adjusted analyses were performed and are presented for each clinical chapter. Unadjusted analyses are simple contrasts between the Ranch Hand and Comparison groups of the mean values, or proportion with abnormal values, of each dependent variable, by t-tests, one-way analysis of variance, Fisher's exact test, or chi-square tests, as appropriate. Adjusted analyses take into account important covariates in the assessment of possible group differences, i.e., the covariates are included in the general linear, logistic regression, proportional odds models, or log-linear models.

Continuous Dependent Variables

When the dependent variable was continuous, the general linear models (GLM) procedure of SAS® was used to fit a model of the dependent variable in terms of the group indicator (Ranch Hand or Comparison) and appropriate covariates, and interactions between covariates. The covariates could be continuous or categorical variables. If necessary, the dependent variable was transformed prior to analysis by a transformation (e.g., logarithm) to enhance normality of its distribution. When a "best" model was fitted, according to the strategy outlined below, the test for significance of the group difference was then done on the adjusted group means, provided there were no significant interactions between the group indicator 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 or estimation of group differences at selected covariate levels using the best model identified.
For some non-normally distributed dependent variables, the Kolmogorov-Smirnov (K-S) test of significant differences between the distributions of the variables in the two study groups was conducted. The K-S test is a nonparametric test for the equality of two distributions designed to detect broad classes of alternatives.

Categorical 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 program BMDP®-LR; for this analysis, the covariates could be either continuous or discrete. For polychotomous dependent variables, where the number of categories was three or more, log-linear modeling was performed by the use of the program BMDP®-4F, by incorporating the full (k)-factor interaction term involving the (k) covariates used in the model. For this type of analysis, all covariates had to be categorized. The models were fitted by the method of maximum likelihood.

To make the results parallel to those obtained by logistic regression, i.e., because of the distinction between dependent and independent variables, the marginals were fixed 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 study, that includes all statistically significant effects and the group indicator 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 forced 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 Appendix E.

An initial model including all two-factor interactions and all three-factor interactions involving group was examined. Global tests at the 0.15 level, or individual tests at the 0.05 level, were used to screen out unnecessary three-factor interactions. A hierarchical stepwise deletion strategy was then used, eliminating effects with $p > 0.05$ (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.

The analysis was carried out by different statisticians, and there are necessarily subtle differences between them in presentation and approach. This, however, should not affect any of the final conclusions as to group differences. In some chapters, for instance, adjusted group means are presented, and in others the differences between the adjusted group means are

7-15
presented. In each case, the same conclusion may be drawn since the statistic of relevance is the difference between the adjusted group mean and the associated p-value. Further, if an interaction of group with a continuous covariate was found, two equally valid methods were used to illustrate how the interaction was arising. One method was to categorize the continuous covariate and describe the group differences within each (covariate-defined) stratum. Another technique was to present group differences for several selected values of the covariate. 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 group and all the covariates but not including interaction terms).

It is recognized that, due to the large number of group-by-covariate interactions examined (up to 7 per dependent variable) for some 150 variables, some of the group-by-covariate interactions judged significant at the 0.05 level may be spurious, i.e., chance occurrences and not of biological relevance. This is analogous to the concept of Type I error for a two-sample adjusted contrast.

When several covariates are used in an adjusted analysis of the group contrast for a single dependent variable, and each group-by-covariate interaction is tested at the 0.05 level, the chance of finding at least one that is statistically significant is, of course, greater than 0.05; this is assuming that there is no group effect or group-by-covariate interaction. How much greater depends on the interrelatedness of the covariates and their association with the dependent variable.

For a study of this size, with many interrelated dependent variables being examined, it is not known how to estimate the number of group-by-covariate interactions that may be due to chance alone. However, this frequency clearly will be more than 5 percent. It is noted that this concept should be considered when significant group-by-covariate interactions are interpreted. Further, it is important that the size of the p-value associated with each group-by-covariate interaction be carefully weighed, as should be the pattern of the interaction findings for related dependent variables.

EXPOSURE INDEX ANALYSES

As described in Chapter 8, the exposure index was constructed to portray the level of dose of the herbicide for the Ranch Hand or exposed group only. Exposure index analyses were conducted on all dependent variables. The objective of the analyses was to determine if there was a difference in the levels of the dependent variable corresponding to the levels of the exposure index.

The exposure index was trichotomized as high, medium, and low, separately, for each of the three occupational groups: officer, enlisted flyer, enlisted groundcrew. Thus, separate analyses were conducted for each occupational cohort. Discrete dependent variables were evaluated using log-linear and logistic regression models, treating exposure level as a categorical 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 examined.

7-16
LONGITUDINAL ANALYSES

General

Another objective of the AFHS is to observe the Ranch Hand population and the Comparison group carefully over time for the emergence, or deleterious rate change, of symptoms, signs, laboratory parameters, or frank disease. This followup objective is not without scientific and logistic 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 categorical longitudinal data.

Continuous Data

A repeated measurements analysis of variance procedure\textsuperscript{10} was used to analyze the variables measured on a continuous scale. The model for the dependent variable (Y) measurement on the kth participant (\( \pi_k \)) in the ith group (\( \alpha_i \)) at the jth time (\( \beta_j \)) is as follows:

\[ Y_{ijk} = \mu + \alpha_i + \pi_k(l) + \beta_j + \omega\beta_{ij} + \epsilon_{ijk} \]

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

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group (Ranch Hand vs. Comparison)</td>
<td>1</td>
</tr>
<tr>
<td>Subject/Group</td>
<td>2,108</td>
</tr>
<tr>
<td>Time (Baseline vs. Followup)</td>
<td>1</td>
</tr>
<tr>
<td>Group-by-Time</td>
<td>1</td>
</tr>
<tr>
<td>(Subject-by-Time)/Group</td>
<td>2,108</td>
</tr>
</tbody>
</table>

*Based on 971 Ranch Hands and 1,139 Comparisons.

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

Care must be taken in the interpretation of the main effect, time (\( \beta_j \)) (i.e., overall Baseline mean versus overall followup mean). This effect is totally confounded with laboratory differences, and with over 2,000 participants, "significant differences" come easily.

The source of variation due to group (\( \alpha_i \)) 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

7-17
followup, provided the group changes were consistent (no significant group-by-time interaction). All available participants were used at each Baseline and followup analysis, while only the participants with both measurements are included in the repeated measurement analysis.

Covariates were not used in these analyses. Generally, time-independent (e.g., year of birth) and time-dependent (e.g., smoking) covariates can be used. Only the time-dependent covariates would affect the primary source of interest, namely the group-by-time interaction. Hence, all of the previously considered time-independent covariates would affect only the main group effect, a source not of primary interest since it is being considered in the separate cross-sectional analyses.

Categorical Data

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

<table>
<thead>
<tr>
<th></th>
<th>Abnormal</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

As with the McNemar test, only the Abnormal→Normal and Normal→Abnormal off-diagonal data were used in further contrasts. A conventional $X^2$ test was used to test the null hypothesis of a comparable change pattern for the two groups (unpaired data).

<table>
<thead>
<tr>
<th></th>
<th>Abnormal</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Abnormal</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormal</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Normal</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

This test is equivalent to testing no group-by-time-by-endpoint interaction in a matched pair analysis.
CHAPTER 7

REFERENCES


This chapter describes the development of the exposure index of the AFHS. Portions of this chapter are paraphrased from the Baseline Morbidity Report, 24 February 1984.

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 Ranch Hand cohort of the AFHS. However, due to a variety of biomedical mechanisms, there can be exceptions to the hypothesis of a consistently increasing dose-response relationship.

An index of potential exposure to any of four 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 cannot be made. 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 provides some degree of useful inference for groups of similarly exposed individuals. In summary, the exposure index in the AFHS is a surrogate indicator of TCDD exposure.

The exposure index developed for the Baseline study and used in this report is defined in Table 8-1.

The exposure index for the ith subject, denoted \( E_i \), is defined as the product of the TCDD weighting factor, the gallons of TCDD-containing herbicide sprayed in the Republic of Vietnam theater during the tour of the ith subject, and the inverse of the number of men sharing the subject’s duties during the tour of the ith 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 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

8-1
TABLE 8-1.

Algorithm for Exposure Index

\[
E_i = \left( \frac{\text{Gallons of TCDD-Containing Herbicide Sprayed in the RVN Theater During the Tour of the ith Subject}}{\text{Weighting Factor}} \right) \times \left( \frac{1}{\text{Number of Men with Subject's Duties in the RVN Theater During the Tour of the ith Subject}} \right)
\]

where \( E_i \) = Exposure Index for the ith Subject

TCDD Weighting Factor = \(
\begin{cases} 
24.0 & \text{if before 1 July 1965} \\
1.0 & \text{if on or after 1 July 1965} 
\end{cases}
\)

Since prior to 1 July 1965 a combination of Herbicides Green, Pink, and Purple with a mean concentration of 48.0 ppm was sprayed, and after 1 July 1965 only Herbicide Orange with a mean concentration of 2 ppm was sprayed, the ratio is then 48:2 or 24:1.

Gallons of TCDD-Containing Herbicide Sprayed in the RVN Theater During the Tour of the ith Subject = \( \frac{\{\text{Number of Gallons of Herbicides Orange, Green, Pink, and Purple Expressed in Herbicide Orange Equivalent Gallons Based on Mean Concentration of TCDD}\}}{\{\text{Mean Concentration (ppm) of TCDD}\}} \)

Using the following:

<table>
<thead>
<tr>
<th>Herbicide</th>
<th>Mean Concentration (ppm) of TCDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>66</td>
</tr>
<tr>
<td>Orange</td>
<td>2</td>
</tr>
<tr>
<td>Pink</td>
<td>66</td>
</tr>
<tr>
<td>Purple</td>
<td>33</td>
</tr>
</tbody>
</table>

Number of Men with Subject's Duties in the RVN Theater During the Tour of the ith Subject = \( \frac{\{\text{Number of Personnel in the Same Occupational Category}\}}{\{\text{in the Same Occupational Category}\}} \)

Pink, and Herbicide Purple was sprayed between January 1962 and 1965. The estimated mean concentration of TCDD for this time was 48.0 ppm, using available data on the number of gallons procured and sprayed.

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 (CHECO) 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 F-1 of Appendix F.

The dates and occupational category of each Ranch Hand's tour(s) in the Republic of Vietnam 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. This information is also presented in Table F-1 of Appendix F.

A numeric exposure index reflecting the effective number of gallons of Herbicide Orange to which each individual was potentially exposed was computed. For the purpose of analysis, 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 8-2), 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 8-2, along with the frequencies of Ranch Hand participants by occupation and exposure level.

The current exposure index is not specific to job 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.

Because of the acknowledged imprecision of the exposure index, Air Force efforts are under way to develop new perspectives of exposure. One effort is the construction of a new questionnaire for the 459 enlisted groundcrew personnel that may permit more accurate exposure analyses within this category. Another approach is the measurement of serum dioxin levels.
### TABLE 8-2.

Exposure Index Categorization of 1,016 Compliant Ranch Hands

<table>
<thead>
<tr>
<th>Occupational Group</th>
<th>Exposure Category</th>
<th>Herbicide Orange Gallons Corresponding to Exposure Category</th>
<th>Number of Ranch Hand Participants in Exposure Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Officer</td>
<td>Low</td>
<td>&lt;35,000</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>35,000-70,000</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>&gt;70,000</td>
<td>123</td>
</tr>
<tr>
<td>Enlisted Flyer</td>
<td>Low</td>
<td>&lt;50,000</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>50,000-85,000</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>&gt;85,000</td>
<td>57</td>
</tr>
<tr>
<td>Enlisted Groundcrew</td>
<td>Low</td>
<td>&lt;20,000</td>
<td>154</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>20,000-27,000</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>&gt;27,000</td>
<td>142</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>1,016</td>
</tr>
</tbody>
</table>

The Air Force currently is conducting a pilot study in conjunction with the laboratories of the Centers for Disease Control, Atlanta, Georgia, to determine levels of TCDD in serum and to establish the validity of exposure differential within the Ranch Hand and Comparison groups. This study is in accordance with the Study Protocol commitment to estimate dosage of TCDD as accurately as current technology permits. If successful, use of time-adjusted TCDD levels would permit more accurate exposure analyses within the Ranch Hand group. Perhaps of most importance, accurate TCDD levels within the Ranch Hand group could standardize exposure to a comparable baseline for all participants. Thus, the use of adjusted TCDD levels will place the exposure concepts on a firm scientific basis, and if herbicide effects exist, they can be discerned more accurately.
CHAPTER 8

REFERENCES

CHAPTER 9
GENERAL HEALTH

INTRODUCTION

The effects of heavy, acute exposure to TCDD 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." However, it is difficult to measure overall health objectively, and for this reason general health outcomes, as defined by this study, should be judged in context with other more specific clinical endpoints. (It should be noted that "general health" outcomes have not traditionally been considered in other dioxin morbidity studies.)

Baseline Summary Results

Five general health variables were included in the Baseline examination: self-perception of health, appearance of illness or distress, relative age, sedimentation rate, and percent body fat. In the analysis of the 1982 Baseline examination data, a statistically significant difference was found between the Ranch Hand and Comparison groups in self-perception of health, with a greater percentage of Ranch Hands reporting their health as fair or poor than Comparisons. This was true in both the younger and older age groups (p=0.017 and p=0.025 for individuals 40 or less and more than 40 years of age, respectively). The relative risk of the Ranch Hand group was also somewhat greater in the younger subgroup than in the older subgroup (1.8 and 1.4, respectively). 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, 8 of the 9 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% versus 2.5%, p=0.029). No overall differences in percent body fat or sedimentation rate were found, although a significant interaction between age, group, and sedimentation rate was noted; younger exposed group members had fewer sedimentation rate abnormalities than did their Comparisons, whereas no difference was found in participants more than 40 years old. No statistically significant dose-response relationships were detected in the Ranch Hand group.

Parameters of the 1985 General Health Assessment

Variables of the Baseline examination (self-perception of health, appearance of illness or distress, relative age, sedimentation rate, and percent body fat) were analyzed for the third year followup effort.
As an assessment of the general health status of each individual, three subjective measures were made as well as two more objective measures. During the health interview each study participant was asked, "Compared to other people your age, would you say that your health is excellent, good, fair, or poor?" This self-assessment of health is susceptible to varying degrees of conscious and subconscious bias. The examiner recorded the appearance of illness or distress (yes/no) and 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 study subject’s group membership (Ranch Hand, Comparison), their assessments were less subject to bias.

The two objective measures were percent body fat, calculated from the body mass index, and the erythrocyte sedimentation rate. Although both variables are rather indirect measures of the general state of health, they are accepted indicators of poor health.

The adjusted statistical analyses below accounted for differences associated with age, race, and occupation. In the analysis of self-perception of health and sedimentation rate, adjustment was also made for personality score, determined from the Jenkins Activity Survey. This is a continuous variable derived by means of a discriminant-function equation based on items that best discriminate men judged to be Type A from those judged as Type B. Positive scores reflected the Type A direction and negative scores the Type B direction. Table G-1 of Appendix G gives the distribution of the covariates in the Ranch Hand and Comparison groups. Age, race, and occupation were distributed similarly in the two groups (due to matching), and personality scores were also not significantly different.

Aside from the subjective nature and potential bias in the self-reported perception of health, no specific issues related to assessment methodology require further comment. No individuals were excluded from analysis, except those with missing data.

Chi-square tests and logistic regression models were applied to the categorical data. The sedimentation rate was normalized by logarithmic transformation. The proportional odds model was also used for ordinal data provided by the self-perception of health and relative age variables. Fisher's exact test was applied to the reporting of illness or distress by the examining physician because of the small number of cases who were classified as "ill." A two-sample t-test was used to assess differences in unadjusted group means, followed by multiple regression analysis to incorporate covariates, for percent body fat and sedimentation rate.

RESULTS AND DISCUSSION

Subjective Assessments

Self-Perception of Health

Each participant was asked to designate his health as excellent, good, fair, or poor. The frequency distributions of self-perception of health for each cohort are given in Table 9-1.
The summarized data in Table 9-1 show that a higher percentage of Ranch Hands perceived their health to be fair or poor (9.1%) than the Comparisons (7.3%), although this difference was not statistically significant (Est. RR: 1.25, 95% C.I.: [0.95,1.64], p=0.14). Of considerable interest is that the percentage of both groups perceiving their health as only fair or poor was lower than that reported at the Baseline examination 3 years earlier (20.4% and 15.9% for Ranch Hands and Comparisons, respectively). This shift was the opposite of that expected from an aging effect. The data collection technique was an in-home interview in 1982 versus an onsite clinic interview in 1985, but this was not judged to be the likely cause of the improvement in health perceptions for the 3-year period. Whatever the cause, the effects were similar in both groups.

A test of association between health perception (dichotomized as excellent/good and fair/poor) was performed with the covariates of age (born in or after 1942, born before 1942), race, occupation, and personality score (Jenkins score, trichotomized as low [less than -5], medium [between -5 and 5], and high [greater than 5]). These associations were examined both within the Ranch Hand and Comparison groups and pooled over the two groups. The findings were similar, and Table 9-2 shows the results after pooling.

These results indicated a significant effect of age, with a higher percentage of the older cohort than the younger cohort reporting their health as fair or poor, as well as a significant effect of occupation, with the percentage of enlisted personnel reporting fair or poor health nearly twice that of the officers. No significant associations were noted for race or personality score.
TABLE 9-2.

Association Between Self-Perception of Health and Age, Race, Occupation, and Personality Score in the Combined Ranch Band and Comparison Groups

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Covariate Category</th>
<th>Excellent/Good</th>
<th>Fair/Poor</th>
<th>Total</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Number</td>
<td>Percent</td>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Age</td>
<td>Born ≥1942</td>
<td>903</td>
<td>94.0</td>
<td>58</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>Born &lt;1942</td>
<td>1,220</td>
<td>90.5</td>
<td>128</td>
<td>9.5</td>
</tr>
<tr>
<td>Race</td>
<td>Black</td>
<td>130</td>
<td>90.9</td>
<td>13</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>Nonblack</td>
<td>1,993</td>
<td>92.0</td>
<td>173</td>
<td>8.0</td>
</tr>
<tr>
<td>Occupation</td>
<td>Officer</td>
<td>819</td>
<td>94.8</td>
<td>45</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>Enlisted Flyer</td>
<td>347</td>
<td>89.7</td>
<td>40</td>
<td>10.3</td>
</tr>
<tr>
<td></td>
<td>Enlisted Groundcrew</td>
<td>957</td>
<td>90.4</td>
<td>101</td>
<td>9.6</td>
</tr>
<tr>
<td>Personality</td>
<td>Low</td>
<td>827</td>
<td>92.2</td>
<td>70</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>716</td>
<td>91.2</td>
<td>69</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>573</td>
<td>92.6</td>
<td>46</td>
<td>7.4</td>
</tr>
</tbody>
</table>
Adjusted analyses of self-perception of health were done by logistic regression using the covariates of age, race, occupation, and personality type. (Self-perception of health was dichotomized and the covariates categorized as in Table 9-2.) These analyses revealed statistically significant age and occupation effects, as well as a significant group-by-occupation interaction (p=0.015). Exponentiation of linear combinations of relevant regression coefficients generated adjusted relative risks for each occupational stratum. These summary data are presented in Table 9-3.

TABLE 9-3.
Adj. Relative Risks of Self-Perception of Health by Occupation

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Adj. Relative Risk (95% C.I.)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Officer</td>
<td>0.78 (0.42,1.46)</td>
<td>0.441</td>
</tr>
<tr>
<td>Enlisted Flyer</td>
<td>0.75 (0.38,1.46)</td>
<td>0.395</td>
</tr>
<tr>
<td>Enlisted Groundcrew</td>
<td>1.90 (1.25,2.88)</td>
<td>0.003</td>
</tr>
</tbody>
</table>

These analyses showed significant group differences in the self-perception of health for the enlisted groundcrew category but not for the officers or enlisted flyers. This is perhaps more clearly seen in Table 9-4, which gives the frequency distribution of self-perception of health stratified by occupation.

Among officers and enlisted flyers, a lower percentage of Ranch Hands than Comparisons perceived their health as fair or poor. (These same Ranch Hands were also less likely to view their health as excellent.) In the enlisted groundcrew cohort, 12.7 percent of the Ranch Hands reported their health as fair or poor versus 7.2 percent of the Comparisons.

Because the logistic model does not account for the ordinal nature of the self-perception of health variable, a proportional odds model for ordinal responses was also fit to the data in Tables 9-1 and 9-4.

For the ordinal responses in Table 9-1, the proportional odds model yielded a statistically significant result (p=0.037), with poorer health estimated to be 1.18 times greater in the Ranch Hand group than in the Comparison group (95% C.I.: [1.01,1.39]). For the data in Table 9-4, a proportional odds model fit to each occupational stratum (adjusting for age) yielded p-values of 0.65 for officers, 0.43 for enlisted flyers, and 0.031 for enlisted groundcrew. Thus, only the enlisted groundcrew category reached statistical significance, with adjusted proportional odds of 1.30 (95% C.I.: [1.02,1.64]).
TABLE 9-4.

Frequency of Self-Perception of Health by Occupation and Group

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Excellent</th>
<th>Good</th>
<th>Fair</th>
<th>Poor</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent</td>
<td>Number</td>
<td>Percent</td>
<td>Number</td>
</tr>
<tr>
<td>Officer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ranch Hand</td>
<td>238</td>
<td>62.6</td>
<td>124</td>
<td>32.6</td>
<td>13</td>
</tr>
<tr>
<td>Comparison</td>
<td>314</td>
<td>64.9</td>
<td>143</td>
<td>29.6</td>
<td>23</td>
</tr>
<tr>
<td>Enlisted Flyer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ranch Hand</td>
<td>67</td>
<td>37.8</td>
<td>94</td>
<td>53.1</td>
<td>13</td>
</tr>
<tr>
<td>Comparison</td>
<td>94</td>
<td>44.8</td>
<td>92</td>
<td>43.8</td>
<td>19</td>
</tr>
<tr>
<td>Enlisted Groundcrew</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ranch Hand</td>
<td>185</td>
<td>40.3</td>
<td>216</td>
<td>47.1</td>
<td>48</td>
</tr>
<tr>
<td>Comparison</td>
<td>266</td>
<td>44.4</td>
<td>290</td>
<td>48.4</td>
<td>39</td>
</tr>
</tbody>
</table>

Similar results were obtained when the analyses were performed on the 1,016 Ranch Hands and 955 Original Comparisons completing the third-year health interview. These results are provided in Table G-2 of Appendix G. In the unadjusted analysis, the estimated relative risk for fair or poor health versus excellent or good health reached statistical significance (Est. RR: 1.43, 95% C.I.: [1.03, 2.00], p = 0.042). In the adjusted analysis, group membership, age, and occupation effects were all statistically significant with an adjusted relative risk of 1.48 (95% C.I.: [1.05, 2.07]). The group-by-occupation interaction, however, did not reach statistical significance (p = 0.23). Nevertheless, little difference was seen in the officers and enlisted flyers, whereas among the enlisted ground crew, 12.7 percent of the Ranch Hands versus 7.4 percent of the Original Comparisons reported their health as fair or poor.

Contrasts of the Ranch Hand and Original Comparison groups using the proportional odds model yielded only borderline significant results. For the unadjusted analysis applied to the overall data, the estimated proportional odds were 1.17 (95% C.I.: [0.99, 1.39], p = 0.073). Stratifying by occupation and adjusting for age gave p-values of 0.76, 0.11, and 0.078 for the officers, enlisted flyers, and enlisted ground crew, respectively. The adjusted proportional odds in the enlisted ground crew cohort were 1.26 (95% C.I.: [0.97, 1.62]).
Appearance of Illness or Distress

The recording of the appearance of acute ill health or physical distress at the examination was intended to capture significant subjective health data that might (though not likely) escape corroboration by other physical examination or laboratory data. In particular, examining physicians were requested to affirm the presence of acute distress when the sign of hippocratic facies was present, a sign not easily feigned by participants. Very few participants were diagnosed as being acutely ill; these data are summarized in Table 9-5.

<table>
<thead>
<tr>
<th>Group</th>
<th>Acute Illness or Distress</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Ranch Hand</td>
<td>4</td>
<td>0.4</td>
<td>1,010</td>
</tr>
<tr>
<td>Comparison</td>
<td>6</td>
<td>0.5</td>
<td>1,287</td>
</tr>
</tbody>
</table>

*Fisher's exact test, 1-sided.

These data were too sparse to permit further meaningful analyses. Descriptively, it was noted that 9 of the 10 ill individuals were in the older age group; 9 of 10 were nonblack; and 2 were officers, 4 were enlisted flyers, and 4 were enlisted groundcrew. The 6 ill Comparison individuals were all Original Comparisons, as can be seen in Table G-3 of Appendix G.

Further, these results were in substantial contrast to the Baseline findings that revealed a marginally significant excess (p=0.056) of acute distress among the Ranch Hands.

Appearance of Relative Age

The examining physicians scored each participant as appearing younger, older, or the same as his chronological age. These data are presented in Table 9-6.
These frequency distributions showed that a slightly higher percentage of Ranch Hands than Comparisons appeared younger than their stated age, and almost equivalent percentages in both groups appeared older. Overall, there was no significant difference in the two distributions. The unadjusted findings in Table 9-6, however, did not confirm the significant tendency (p=0.029) at the 1982 Baseline examination for a higher percentage of the Ranch Hands than Comparisons to appear younger than their stated ages. Table 9-7 presents the association between each of the covariates and relative age (dichotomized as older looking versus the same or younger looking) after combining the Ranch Hand and Comparison groups.

As noted from this table, age and race were not significantly associated with the appearance of relative age, whereas occupation did reveal a significant association, with about 6 percent of the enlisted personnel appearing older than their stated ages compared to 1 percent of the officers.

An adjusted analysis using logistic regression with the covariates age, race, and occupation showed a significant effect due to occupation as well as a significant group-by-occupation interaction (p=0.038). Adjusted relative risks for each occupational stratum are given in Table 9-8.

The adjusted relative risk was greater than 1 for the officers, i.e., the odds of appearing older were greater in the Ranch Hand group than in the Comparison group, but the relative risk was less than 1 for the enlisted flyers. However, the associated confidence intervals were rather broad and did not rule out a relative risk of 1 in each case. Again, because the logistic regression model does not account for the ordinal nature of the dependent variable, a proportional odds model was applied to the enlisted flyer cohort (data in the officer and enlisted ground crew strata did not fit the model properly). The estimated proportional odds for the enlisted flyer cohort were nonsignificant (estimated odds: 0.49, 95% C.I.: [0.22,1.11], p=0.087).
TABLE 9-7.
Association Between Appearance of Relative Age and Age, Race, and Occupation in the Combined Ranch Band and Comparison Groups

<table>
<thead>
<tr>
<th>Appearance of Relative Age</th>
<th>Younger/Same</th>
<th>Older</th>
<th>Total</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Covariate</strong></td>
<td><strong>Covariate Category</strong></td>
<td><strong>Number</strong></td>
<td><strong>Percent</strong></td>
<td><strong>Number</strong></td>
</tr>
<tr>
<td>Age</td>
<td>Born &gt;1942</td>
<td>914</td>
<td>95.2</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Born &lt;1942</td>
<td>1,301</td>
<td>96.5</td>
<td>47</td>
</tr>
<tr>
<td>Race</td>
<td>Black</td>
<td>138</td>
<td>96.5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Nonblack</td>
<td>2,077</td>
<td>95.9</td>
<td>88</td>
</tr>
<tr>
<td>Occupation</td>
<td>Officer</td>
<td>855</td>
<td>99.0</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Enlisted Flyer</td>
<td>362</td>
<td>93.5</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Enlisted Groundcrew</td>
<td>998</td>
<td>94.4</td>
<td>59</td>
</tr>
</tbody>
</table>

TABLE 9-8.
Adjusted Relative Risks of Appearance of Relative Age by Occupation

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Adj. Relative Risk (95% C.I.)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Officer</td>
<td>4.52 (0.94,21.9)</td>
<td>0.060</td>
</tr>
<tr>
<td>Enlisted Flyer</td>
<td>0.44 (0.23,1.27)</td>
<td>0.159</td>
</tr>
<tr>
<td>Enlisted Groundcrew</td>
<td>1.05 (0.62,1.78)</td>
<td>0.849</td>
</tr>
</tbody>
</table>
A contrast of the Ranch Hand group with the Original Comparisons gave similar results, as shown in Table G-4 of Appendix G. Overall, there was little difference, but the group-by-occupation interaction was of borderline significance in the adjusted analysis (p=0.052). Differences were largely confined to the enlisted flyers, where fewer Ranch Hands than Comparisons appeared older than their stated ages (Adj. RR: 0.47, 95% C.I.: [0.20,1.12], p=0.089) (see Table G-5 of Appendix G). A proportional odds model applied to the enlisted flyer stratum gave adjusted proportional odds of 0.45 (95% C.I.: [0.20,1.02], p=0.055).

**Objective Assessments**

Two objective but nonspecific indicators of general health, the erythrocyte sedimentation rate and percent body fat, were analyzed in both discrete and continuous forms. Because the sedimentation rate was a highly skewed variable, it was normalized by logarithmic transformation for the continuous analyses. The sedimentation rate dichotomy was set at 20 mm/hr or less (normal) and greater than 20 mm/hr (abnormal) by the large-tube Westergren method. Percent body fat was based on height and weight obtained during the examination and was calculated according to the following formula:

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

It is recognized that this formula will overstate the percent body fat for very muscular, large-boned men. Percent body fat was trichotomized into less than 10 percent (lean), 10 to 25 percent (normal), and greater than 25 percent (obese), consistent with the Baseline Report. Because of the sparseness of the lean category, it was often necessary to use a dichotomous variable of lean-normal versus obese.

**Erythrocyte Sedimentation Rate**

The unadjusted contrast of log sedimentation rate means revealed no significant group differences (mean±SE=1.62±0.026 in the Ranch Hand group versus 1.59±0.021 in the Comparison group, t=0.73, p=0.47). The geometric mean values were 5.05 and 4.93 for the Ranch Hand and Comparison groups, respectively. Tests of association of dichotomized sedimentation rate, with the covariates age, race, occupation, and personality score, pooled over both groups, were conducted; these summarized data are shown in Table 9-9.

These results showed significant effects of age, with older individuals having a higher frequency of abnormal sedimentation rates than younger individuals, and a significant effect of personality score, with Type B individuals (low personality score) having more sedimentation rate abnormalities. The effect of occupation was of borderline significance (p=0.080), with a slightly higher percentage of abnormal values among the enlisted flyers than among officers or enlisted groundcrew. There was no evidence of any association between race and abnormal sedimentation rate.

An analysis of the log sedimentation rate, adjusting for age, race, occupation, and personality score, detected significant effects for all of the covariates except race, as well as a significant age-by-personality score interaction. As in the unadjusted analysis, the adjusted analysis did not reveal any significant difference between the Ranch Hand and Comparison groups (p=0.412).
### TABLE 9-9.

Association Between Sedimentation Rate and Age, Race, Occupation, and Personality Score in the Combined Ranch Hand and Comparison Groups

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Covariate Category</th>
<th>Normal ≤20mm/hr</th>
<th>Abnormal &gt;20mm/hr</th>
<th>Total</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>Born ≥1942</td>
<td>941 97.9</td>
<td>20 2.1</td>
<td>961</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Born &lt;1942</td>
<td>1,263 93.7</td>
<td>85 6.3</td>
<td>1,348</td>
<td></td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td>Black</td>
<td>136 95.1</td>
<td>7 4.9</td>
<td>143</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>Nonblack</td>
<td>2,068 95.5</td>
<td>98 4.5</td>
<td>2,166</td>
<td></td>
</tr>
<tr>
<td><strong>Occupation</strong></td>
<td>Officer</td>
<td>828 95.8</td>
<td>36 4.2</td>
<td>864</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>Enlisted Flyer</td>
<td>361 93.3</td>
<td>26 6.7</td>
<td>387</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enlisted Groundcrew</td>
<td>1,015 95.9</td>
<td>43 4.1</td>
<td>1,058</td>
<td></td>
</tr>
<tr>
<td><strong>Personality Score</strong></td>
<td>Low</td>
<td>843 94.0</td>
<td>54 6.0</td>
<td>897</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>758 96.6</td>
<td>27 3.4</td>
<td>785</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>595 96.1</td>
<td>24 3.9</td>
<td>619</td>
<td></td>
</tr>
</tbody>
</table>

9-11
However, in the dichotomous form, sedimentation rate abnormalities were significantly more prevalent in the Ranch Hands than Comparisons (Est. RR: 1.63, 95% C.I.: [1.12,2.38], p=0.013); these results are given in Table 9-10.

Logistic regression analysis found significant effects for age and personality score, and the adjusted relative risk of 1.68 (95% C.I.: [1.13,2.49], p=0.011), was very similar to the estimated relative risk of 1.63.

**TABLE 9-10.**

Unadjusted Analysis for Sedimentation Rate by Group

<table>
<thead>
<tr>
<th>Sedimentation Rate</th>
<th>Normal &lt;20 mm/hr</th>
<th>Abnormal &gt;20 mm/hr</th>
<th>Total</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>Number</td>
<td>Percent</td>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Ranch Hand</td>
<td>957</td>
<td>94.2</td>
<td>59</td>
<td>5.8</td>
</tr>
<tr>
<td>Comparison</td>
<td>1,247</td>
<td>96.4</td>
<td>46</td>
<td>3.6</td>
</tr>
</tbody>
</table>

The mean log sedimentation rate in the Original Comparisons was 1.636 plus or minus 0.025, not significantly different from the Ranch Hand mean (t=-0.45, p=0.65). The regression analysis yielded results very similar to those reported above, with little difference in the adjusted group means. Logistic regression analyses also gave similar results, with significantly more abnormalities in the Ranch Hand group (p=0.037).

In summary, there was no difference between groups based upon mean values of the sedimentation rate, unadjusted or adjusted, but both unadjusted and adjusted discrete analyses showed a significantly higher prevalence of sedimentation rate abnormalities in the Ranch Hand group. This finding was opposite to the Baseline findings in which Ranch Hands age 40 or less had significantly fewer sedimentation rate abnormalities than Comparisons, with no group difference in individuals over the age of 40.

**Percent Body Fat**

The mean percent body fat of Ranch Hands was significantly lower than that of Comparisons (21.10%±0.15. versus 21.54%±0.14, respectively; p=0.037). Because there were only a few values in the lean category (6 in the Ranch Hand group and 4 in the Comparison group), percent body fat was dichotomized into at most 25 percent (lean and normal) and more than 25 percent (obese) for tests of association between percent body fat and the covariates age, race, and occupation. The results are given in Table 9-11.
### TABLE 9-11.

#### Association Between Percent Body Fat and Age, Race, and Occupation in the Combined Ranch Band and Comparison Groups

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Covariate Category</th>
<th>Number</th>
<th>Percent</th>
<th>Number</th>
<th>Percent</th>
<th>Total</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Born ≥1942</td>
<td>802</td>
<td>83.4</td>
<td>159</td>
<td>16.6</td>
<td>961</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>Born &lt;1942</td>
<td>1,060</td>
<td>78.7</td>
<td>287</td>
<td>21.3</td>
<td>1,347</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>Black</td>
<td>110</td>
<td>76.9</td>
<td>33</td>
<td>23.1</td>
<td>143</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>Nonblack</td>
<td>1,752</td>
<td>80.9</td>
<td>413</td>
<td>19.1</td>
<td>2,165</td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td>Officer</td>
<td>719</td>
<td>83.3</td>
<td>144</td>
<td>16.7</td>
<td>863</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enlisted Flyer</td>
<td>314</td>
<td>81.1</td>
<td>73</td>
<td>18.9</td>
<td>387</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>Enlisted Groundcrew</td>
<td>829</td>
<td>78.4</td>
<td>229</td>
<td>21.6</td>
<td>1,058</td>
<td></td>
</tr>
</tbody>
</table>

These data demonstrated the significant effects of age, with a higher percentage of obesity in older men, and occupation, with a higher prevalence of obesity in enlisted personnel than in officers. Race was a noncontributory covariate. The covariate of smoking was unexplored.

An adjusted analysis of percent body fat, with the same covariates, also showed the significant effects of age, occupation, and an age-by-occupation interaction. The adjusted results showed a small, but significantly lower mean level of body fat in the Ranch Band group (adjusted difference = 0.443 ± 0.210, p=0.035).

With percent body fat dichotomized into obese versus normal or lean, the percent obese was lower in the Ranch Hands than in the Comparisons (18.2% versus 20.2%), but the difference was not significant (Est. RR: 0.90, 95% C.I.: [0.71,1.08], p=0.25). Logistic regression analysis also failed to detect a significant group difference (Adj. RR: 0.87, 95% C.I.: [0.71,1.08], p=0.204).

Analysis of percent body fat in the Ranch Hands and Original Comparisons gave somewhat different results. The overall difference in means was significant as before: 21.10 plus or minus 0.15 in the Ranch Hand group versus 21.58 plus or minus 0.16 in the Original Comparison group (t=-2.15, p=0.032). However, the regression analysis detected a statistically significant group-by-race interaction (p=0.041). The adjusted difference in mean percent body fat (Ranch Hand versus Comparison) was greater in Black participants (-2.26%)
than in nonblack participants (-0.34%). Of the Original Comparisons (Table G-7 of Appendix G), 20.4 percent were obese, greater than, but not significantly different from, the percent obese in the Ranch Hand group (p=0.230). Logistic regression analyses again detected significant age and occupation effects, but it detected no significant interaction between these variables. There was no strong evidence of a group-by-race interaction (models including all two-factor interactions gave a Z-value of 1.19 for the group-by-race interaction). The group effect was not statistically significant (Adj. RR: 0.87, 95% C.I.: [0.70, 1.09], p=0.242).

In summary, the unadjusted and adjusted tests of mean percent body fat showed a significantly lower value for Ranch Hands; correspondingly fewer Ranch Hands than Comparisons were obese, although this difference was not statistically significant. Few individuals were lean (less than 10 percent body fat). The 1982 Baseline examination found no difference in group means (p=0.67), or proportion of abnormalities (p=0.89). Further, analyses based solely upon the Original Comparison cohort found the difference in mean percent body fat between the Ranch Hand and Comparison groups to be greater in Blacks than nonblacks.

**EXPOSURE INDEX ANALYSES**

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, and high. Because it is not possible to assess the relative exposure between occupational groups, and since different cutoff values were used in the three occupational categories, separate analyses were performed within each occupational cohort. A detailed description of the exposure index is found in Chapter 8. Exposure analyses were performed on four of the five general health variables. Only four Ranch Hands were recorded as appearing ill or distressed (two were officers, both in the low-exposure category, and two were enlisted flyers, both in the high-exposure category). Further analysis was not done on this variable.

**Self-Perception of Health**

Table 9-12 presents dichotomized self-perception of health data by exposure level for the 1,016 Ranch Hands. While these unadjusted contrasts did not reach statistical significance within any of the occupational strata, the linear trend from low to high exposure in the officer cohort of the fair/poor category was of interest, and was subjected to further testing. Although the numbers were small at each exposure level, a test for linear trend led to a borderline significant increase of 2.5 plus or minus 1.3 percent per unit (step) increase in the exposure level category (p=0.064).

Logistic regression analyses adjusted for age (dichotomized), race, and personality score (trichotomized) did not detect any significant exposure level effects. The only significant covariate effect found was for age in the enlisted ground crew cohort. The adjusted relative risk for each occupational stratum is given in Table 9-13.
### TABLE 9-12.

Unadjusted Exposure Index Analysis of Self-Perception of Health by Occupation

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Exposure Index</th>
<th>Excellent/Good</th>
<th>Fair/Poor</th>
<th>Total</th>
<th>p-Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent</td>
<td>Number</td>
<td>Percent</td>
<td></td>
</tr>
<tr>
<td>Officer</td>
<td>Low</td>
<td>124</td>
<td>97.6</td>
<td>3</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>124</td>
<td>95.4</td>
<td>6</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>114</td>
<td>92.7</td>
<td>9</td>
<td>7.3</td>
</tr>
<tr>
<td>Enlisted Flyer</td>
<td>Low</td>
<td>51</td>
<td>92.7</td>
<td>4</td>
<td>7.3</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>59</td>
<td>90.8</td>
<td>6</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>51</td>
<td>89.5</td>
<td>6</td>
<td>10.5</td>
</tr>
<tr>
<td>Enlisted Groundcrew</td>
<td>Low</td>
<td>134</td>
<td>87.0</td>
<td>20</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>146</td>
<td>89.6</td>
<td>17</td>
<td>10.4</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>121</td>
<td>85.2</td>
<td>21</td>
<td>14.8</td>
</tr>
</tbody>
</table>

*Chi-square tests, 2 d.f.

### TABLE 9-13.

Adjusted Relative Risk of Self-Perception of Health by Occupation and Exposure Contrast

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Exposure Contrast</th>
<th>Adj. Relative Risk (95% C.I.)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Officer</td>
<td>Medium vs. Low</td>
<td>2.00 (0.49,8.15)</td>
<td>0.334</td>
</tr>
<tr>
<td></td>
<td>High vs. Low</td>
<td>2.93 (0.76,11.3)</td>
<td>0.119</td>
</tr>
<tr>
<td>Enlisted Flyer</td>
<td>Medium vs. Low</td>
<td>1.30 (0.35,4.86)</td>
<td>0.700</td>
</tr>
<tr>
<td></td>
<td>High vs. Low</td>
<td>1.50 (0.40,5.64)</td>
<td>0.549</td>
</tr>
<tr>
<td>Enlisted Groundcrew</td>
<td>Medium vs. Low</td>
<td>0.95 (0.47,1.92)</td>
<td>0.882</td>
</tr>
<tr>
<td></td>
<td>High vs. Low</td>
<td>1.21 (0.62,2.35)</td>
<td>0.580</td>
</tr>
</tbody>
</table>
Appearance of Relative Age

The dichotomy of appearance of relative age was assessed for exposure effects in each occupational cohort. These unadjusted analyses, shown in Table 9-14, provided no evidence of a dose-response effect. As can be seen, the number of participants within each stratum appearing older than their stated ages was quite small. The adjusted analyses by logistic regression did not detect any significant exposure or covariate effects.

### TABLE 9-14.
Unadjusted Exposure Index Analysis of Appearance of Relative Age by Occupation

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Exposure Index</th>
<th>Younger/Same</th>
<th>Older</th>
<th>Total</th>
<th>p-Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent</td>
<td>Number</td>
<td>Percent</td>
<td></td>
</tr>
<tr>
<td>Officer</td>
<td>Low</td>
<td>125</td>
<td>98.4</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>127</td>
<td>97.7</td>
<td>3</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>121</td>
<td>98.4</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>Enlisted Flyer</td>
<td>Low</td>
<td>52</td>
<td>94.6</td>
<td>3</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>62</td>
<td>95.4</td>
<td>3</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>55</td>
<td>96.5</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>Enlisted Groundcrew</td>
<td>Low</td>
<td>146</td>
<td>94.8</td>
<td>8</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>151</td>
<td>93.2</td>
<td>11</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>134</td>
<td>94.4</td>
<td>8</td>
<td>5.6</td>
</tr>
</tbody>
</table>

*Chi-square tests, 2 d.f.
**Erythrocyte Sedimentation Rate**

The sedimentation rate was analyzed both continuously on a logarithmic scale and dichotomously (normal, abnormal). One-way analyses of variance were performed on the sedimentation rate means categorized by occupation and exposure level. These tests showed no significant differences in the officer and the enlisted flyer strata ($p=0.76$, $p=0.64$, respectively). In the enlisted groundcrew stratum the means were marginally different, with the mean sedimentation rate increasing with increasing exposure level, but the differences were not statistically significant ($p=0.12$). When these data were adjusted by an analysis of covariance for age, the difference in mean sedimentation rates in the enlisted groundcrew was less noteworthy ($p=0.33$). Age was positively associated with the mean sedimentation rate in all three occupational strata ($p<0.001$, $p=0.009$, and $p<0.001$, respectively). The adjusted tests are reflected in Table 9-15 (means and confidence limits have been transformed back to the original scale).

A categorical analysis of the sedimentation rate by exposure level for each occupational stratum was also conducted. Differing from the previous continuous analyses, the categorical contrasts revealed a significant exposure effect ($p=0.027$) in the enlisted flyer stratum, albeit with small numbers. These summarized data are shown in Table 9-16.

Adjustment for age, race, and personality score revealed a significant high versus low exposure contrast in the enlisted flyer stratum. The adjusted analysis is fully shown in Table 9-17.

**Table 9-15.**

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Low (95% C.I.)</th>
<th>Medium (95% C.I.)</th>
<th>High (95% C.I.)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Officer</td>
<td>5.40 (4.71, 6.19)</td>
<td>4.78 (4.17, 5.47)</td>
<td>4.69 (4.09, 5.37)</td>
<td>0.31</td>
</tr>
<tr>
<td>Enlisted Flyer</td>
<td>5.10 (4.11, 6.33)</td>
<td>6.00 (4.91, 7.32)</td>
<td>5.00 (4.04, 6.19)</td>
<td>0.41</td>
</tr>
<tr>
<td>Enlisted Groundcrew</td>
<td>4.66 (4.10, 5.29)</td>
<td>5.09 (4.49, 5.77)</td>
<td>5.35 (4.69, 6.12)</td>
<td>0.33</td>
</tr>
</tbody>
</table>