A marginally significant association was found between reported use of medication and group (p=0.081). As shown in Table 5-13, a greater percentage of Ranch Hands (58.9%) reported medication use than did Comparisons (55.0%). Use of medication increased in both groups since 1992; however, that increase was parallel. In 1992, 44.1 percent of Ranch Hands and 40.4 percent of Comparisons reported medication use compared to 58.9 percent and 55.0 percent, respectively, in 1997. Table 5-14 shows reported work loss for fully compliant Ranch Hands and Comparisons. The difference between the two groups narrowed from 1992, and no significant association was found between work loss and group (p=0.968) in 1997.

5.7 CONCLUSION

Although more Comparisons than Ranch Hands refused to participate in the 1997 follow-up examination, there is no significant difference in the reasons for refusal among the two groups. The reasons for refusal differed with age and rank but did not differ significantly for race. Logistics and health reasons were the most common reasons for refusal, although there were a substantial number of veterans deemed hostile from whom a reason for refusal was not determined. In replacing noncompliant Original Comparisons, either compliant replacements were found or no replacement was necessary (e.g., the Original Comparison was deceased and no Replacement Comparison had been contacted previously) for approximately 91 percent of the cases.

Self-reported health status differed with group, age, rank, and compliance status, but not with race, among those reporting health status. Ranch Hands, older participants, enlisted men, and refusals were more likely to report fair or poor health. Ranch Hands reported fair or poor health more often than did Comparisons. In both groups veterans who refused were more likely to report fair or poor health than those who were fully compliant. This pattern of Ranch Hands reporting poorer health has been observed since the baseline examination. Using work loss and medication use as more objective indicators of health than health perception, Ranch Hands reported a slightly higher use of medications, but no difference was seen in reported work loss between Ranch Hands and Comparisons. A further analysis of self-perception of health, as reported by fully compliant participants at the 1997 follow-up examination, is given in Chapter 9, General Health Assessment.

In summary, the results of these analyses suggested that Ranch Hands may be reporting poorer health than Comparisons and that these group differences are present for both fully compliant participants and refusals. This holds true even after accounting for rank and age differences. In addition, the difference in the percentage of fully compliant participants and refusals reporting fair or poor health was similar for Ranch Hands and Comparisons.
Table of Contents

6 QUALITY CONTROL ............................................................................................................ 6-1

6.1 QUESTIONNAIRE QC ............................................................................................................. 6-1
  6.1.1 Design .............................................................................................................................. 6-1
  6.1.2 Data Collection ................................................................................................................. 6-2
  6.1.3 Processing and QA of Questionnaire Data ....................................................................... 6-2

6.2 PHYSICAL EXAMINATION QC ......................................................................................... 6-4

6.3 LABORATORY QC ................................................................................................................ 6-5
  6.3.1 QC Procedures for the Clinical Laboratory ....................................................................... 6-6

6.4 MEDICAL DATA QC ........................................................................................................... 6-7
  6.4.1 Overview of QC Procedures .............................................................................................. 6-7
  6.4.2 Data Processing System Design ....................................................................................... 6-8
  6.4.3 Design and Administration of Physical and Psychological Examination Forms ................. 6-9
  6.4.4 Data Completeness Checks .............................................................................................. 6-10
  6.4.5 Data Validation .................................................................................................................. 6-11

6.5 MEDICAL RECORDS CODING QC .................................................................................... 6-12

6.6 STATISTICAL ANALYSIS QC ........................................................................................... 6-12

6.7 ADMINISTRATIVE QA ....................................................................................................... 6-13

REFERENCES ......................................................................................................................... 6-14
List of Figures

Figure 6-1. Participant Evaluations of the 1997 AFHS Clinic Experience .............................................. 6-5
Figure 6-2. Physical Examination Form QC Process ............................................................................... 6-8
Figure 6-3. Conversion and Cleaning Process for Medical Data ........................................................... 6-10
6 QUALITY CONTROL

Quality control (QC) and quality assurance (QA) procedures were planned for and implemented throughout the 1997 Air Force Health Study (AFHS), from project initiation to final product delivery and acceptance by the Air Force. QC is defined as the procedures put in place to ensure the quality of the data collected. QA refers to the management of those procedures. This chapter provides an overview of the specific QC and QA measures developed and used by the project team, specifically in the areas of questionnaire and physical examination QC, laboratory QC measures, data management QC, statistical QC, and administrative QA. The Air Force, Science Applications International Corporation (SAIC), the National Opinion Research Center (NORC), and Scripps Clinic all participated in the formulation and implementation of the QC and QA procedures described in this chapter.

6.1 QUESTIONNAIRE QC

6.1.1 Design

For the baseline and subsequent follow-up examinations, the baseline and interval questionnaires were administered in person. In the examinations prior to 1997, the questionnaires were administered in hard copy, which was then key-entered into the final SAS® data set. For the 1997 follow-up, the interview responses were obtained electronically on laptop computers, using a computer-assisted personal interview (CAPI) system.

Effective CAPI design was the first step in QC of the data collection. By combining the two steps of data collection and data entry, the CAPI technique eliminated one possible source of recording error—key-entry of the data. Further, the logic checks, range checks, and intervariable consistency checks programmed into the CAPI system placed constraints on what the interviewer could type or select for any particular question during the interview. These constraints limited keystroke errors and data problems arising from the interview itself. The structure of the CAPI system ensured that skip patterns were followed correctly and that no questions were left unanswered. In certain sections of the questionnaire, CAPI offered significant enhancements to the flow and accuracy of the questionnaire over a paper-and-pencil execution. These enhancements included automatic unit conversions and elimination of multiform cross-references. These benefits were most notable in the calculations of alcohol and tobacco use and in updating information for children born prior to the last interview.

Using a process of reviewing, commenting, and concurring, Air Force researchers and NORC designers incorporated new questions and questions derived from the AFHS self-administered forms into the 1997 questionnaire. The goal was to create a cohesive instrument with questions grouped logically by subject because a cohesive questionnaire would enhance the participant’s focus on the subject being discussed and his understanding of the questions. In addition, the inclusion of the self-administered forms into the interval questionnaire decreased the participants’ frustrations with the study process by eliminating question redundancy, providing a logical sequencing of questions, and decreasing the time spent by the participant.

1 SAS and all other SAS Institute, Inc., product and service names are registered trademarks or trademarks of SAS Institute, Inc., in the USA and other countries.
An additional benefit of the CAPI questionnaire was the ability to print selected participant responses for the use of the debriefing physicians. These printouts were improved and refined during the physical examination period.

6.1.2 Data Collection

NORC recruited and trained eight interviewers and one field manager to administer the baseline and interval questionnaires. A minimum number of interviewers were selected to reduce variability between interviewing techniques. The interviewers were blind to the participants’ exposure status, thus avoiding bias.

The Field Manager, who supervised the interviewing at the examination site, observed the work of each interviewer and presented formal evaluations of their performance each quarter to the Air Force. Interviewers were evaluated on their ability to control the interview and to probe incomplete answers in a neutral manner. They also were graded on their vocal quality, reading quality, and on their use of associated forms and documents. The interviewers were graded on a scale of 1 to 4. A rating of 1 indicated an unacceptable performance and 4 an above-average performance with no errors. All interviewers performed at an above-average level and none required retraining.

Interviewers were required to regularly report questions or problems experienced while executing the questionnaires. “CAPI Problem Forms” and “Policy Decision Forms” were distributed for interviewers to complete, and these forms were faxed daily to the Data Collection Task Leader at NORC headquarters in Chicago, Illinois. Items reported on the forms included the following: (1) mistakes made and not corrected during the interview, (2) conditions reported by the participant after the interview was over, (3) technical problems with the CAPI instrument, and (4) problems with the printout for the debriefing physician. The Data Collection Task Leader corrected problems when necessary and provided assistance to interviewers in handling confusing or unusual situations.

6.1.3 Processing and QA of Questionnaire Data

Completed questionnaire data were transmitted daily via modem from the La Jolla, California, site to the receiving computer system in Chicago. As interviews were completed on the laptop computers at the site office, the CAPI system selected the newly completed cases, encrypted the interview data, and transmitted the interview data to the NORC modem pool in Chicago. Once in Chicago, the interview data were unencrypted, archived on a devoted volume of the NORC UNIX computer, and copied to the NORC wide area network. Each CAPI interview consisted of one multiple-record ASCII file representing the participant’s answers to questions. Using a standard utility, the ASCII files were converted from their vertical format to the horizontal format readable by SAS®. Programmers then read the horizontal files into SAS® and printed frequencies of all variables. Case data received in Chicago were reconciled regularly with the completion log at the interviewing site.

Some of the QC steps used in converting CAPI files to the SAS® data files include the following:

1. The case IDs of all completed interviews in the SAS® file were compared to the log of completed interviews kept at the site office. This ensured that all completed cases were received and that there were no duplicates.

2. The SAS® variables were compared to a hard-copy representation of the CAPI to ensure that all questions in the interview were present in the SAS® data file.
3. The response frequencies were compared to a hard-copy representation of the CAPI to ensure that no data were truncated.

One of the goals in the conversion process was to replicate, to the maximum extent possible, the variable names, formats, and structures used in the 1992 SAS® data set. To accommodate this goal, additional "post-processing" programs were created. The post-processing included renaming variables, assigning the appropriate variable labels and value labels, creating variables based on values of answers to more than one question (such as calculations of cigarette use), and merging variables collected outside of the interview into the data set.

Several steps were taken to ensure that the SAS® data file created from the post-processing programs contained the correct information:

- A list was created that mapped CAPI variables to SAS® variables. This allowed the NORC staff to ensure that variables were named properly and that all required variables were included in the SAS® data set.
- Format statements and frequencies were proofed against three representations of the questionnaire (the CAPI form, the 1997 hard copy representation, and the 1992 hard copy) to detect problems.
- Cross-tabulations and printouts of data items at the case level were generated to investigate complicated questions, such as whether a calculation was working correctly or why there was a missing value in a certain variable.
- Continuous reviews of the frequencies were performed until no more errors were detected.
- A cumulative data set of all interviews completed to date, accompanied by a footnote file explaining any anomalies or errors still to be resolved, was delivered quarterly and then monthly to the Air Force for review. All errors identified by the Air Force were corrected by NORC, the data set was corrected and delivered a final time, and the corrections were accepted.

Response frequencies for all data fields were reviewed regularly to ensure that data for all variables were captured, answers made logical sense, and the skips and checks programmed by CAPI were operating correctly. These QC checks revealed a small number of problems in the questionnaire, all of which were corrected without significant loss of data. These problems, along with the solutions applied, were documented in the footnotes included with the data file.

One of these problems was discovered during processing of the first questionnaires. During a variable-by-variable review of the interval questionnaire, NORC discovered that a short series of questions concerning mental and emotional illness had been omitted from the CAPI program. Three steps were taken to correct this situation:

1. A hard-copy version of the questions was immediately distributed to NORC's interviewers at Scripps Clinic so that the information would be obtained for the remaining participants in the current physical examination group. These data were manually entered into the questionnaire database.

2. A revised version of the interval questionnaire, containing the omitted questions, was installed on the interviewers' computers within 6 calendar days of the problem discovery.
3. NORC schedulers telephoned the participants who were not asked the omitted questions during their in-person interview to retrieve the information. These data were manually entered into the questionnaire database.

6.2 PHYSICAL EXAMINATION QC

The Scripps Clinic selection process for all personnel who were to interact directly with the participants ensured a high-quality physical examination. 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 process. Further, the Air Force reviewed the credentials of all key staff members and approved their participation in the study.

A complete pre-examination test was held. Eleven volunteers completed the physical examinations, interviews, psychological tests, and laboratory analyses several weeks before the scheduled start of the study. All aspects of patient contact were reviewed: the initial inbriefing of the participants, the logistics of transportation and patient flow within the clinic, and the final outbriefing by the diagnostician.

During the actual examinations, refinements were made whenever operational problems were detected. Whether detected by the Scripps staff, the Air Force onsite monitor, or the participants, study problems were addressed during periodic clinical QA meetings of key Scripps staff. For instance, participant temperatures were not recorded for the first few physical examination groups. This error in protocol was addressed in one of these meetings. The Air Force reviewed the affected records, found no comments concerning elevated temperatures, and coded these records as normal.

During the physical examination, the identification of 27 chest x rays was found to be questionable because of incomplete or improper labeling. Although no data from the x rays were to be used in the analysis, the 27 participants whose x rays were in question were contacted and arrangements were made to reshoot their x rays. All but six x rays were retaken; two participants refused.

Following examination of each participant group, the Scripps staff reviewed all physical examination forms for omissions, incomplete examinations, and inconsistencies. When issues were found, the examiners or technicians were contacted to correct the data. Special effort was made to complete this review while the participants were at the examination site. In all cases in which data were corrected, the form was initialed by the doctor or technician making the correction. (This subject is discussed in more detail in the Medical Data QC section of this chapter.) An optical scanner read all mark-sense physical examination forms as an ongoing QA of form completion.

The Air Force onsite monitor and the Scripps Clinic administrative team monitored compliance with the physical examination process. The Scripps Clinic Chief of Medicine and the SAIC Project Manager conducted additional periodic inspections. All such clinical reviews were performed unobtrusively and with the full consent of the participant; suggestions or corrections to the examination procedure always were discussed privately with the attending physician. These inspections emphasized aspects of clinical techniques, sequence, and completeness of the clinical data with respect to the examination forms and the blindness of the examinations. Of particular note were the detailed daily log entries of the Air Force monitors. These entries ensured continuity of knowledge (the monitors rotated approximately every 2 weeks) by documenting daily activities and, when needed, recording events requiring follow-up by either the Air Force or SAIC.

6-4
Establishing a rapport with each study participant was a primary goal of all the organizations involved in the study. Although this may not be a traditional QA parameter in most research studies, it is paramount in the AFHS. Maintaining participants' satisfaction encourages them to continue in the study, thus avoiding the need for significant participant replacement, which can reduce future statistical power or introduce bias, or both. Therefore, every staff member emphasized courtesy, empathy, assistance, and personalized treatment of each participant.

Participants were asked to fill out an evaluation form after completion of their 1997 follow-up physical examinations. The participant evaluations provided insight into the participants' experiences, including strong points of the programs and areas in need of improvement. These forms were reviewed by all study management staff.

Based on the participants' evaluation forms, 72.8 percent evaluated their overall clinic experience as excellent, and 25.0 percent classified it as good. One participant felt that the experience was unsatisfactory, and 2.1 percent of the participants rated it as satisfactory. Figure 6-1 charts those evaluations of the participants' clinic experiences.

![Figure 6-1](image-url)

**Figure 6-1. Participant Evaluations of the 1997 AFHS Clinic Experience**

### 6.3 LABORATORY QC

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

The following list outlines the tests performed and the methods and equipment used:

- Hematology assays were performed on Coulter STKS® equipment.
- Erythrocyte sedimentation rate determinations were performed using the large-tube Westergren method.
- Biochemical assays were performed using the Dade RxL® Automated Chemical Analyzer.
- Radioimmunoassays were performed with standard test kits.
- Electrophoresis and occult blood tests were performed manually.
- Hepatitis A, B, C, and D tests were performed using Abbott Commander® and Quantum® machines.
- Monospecific antibodies were used for immunoglobulin assays using the Beckman Array Protein System®.
- T & B lymphocytes were analyzed on BD FACSCAN equipment.
- Blood-cell counts were performed with standard microscopy.
- All urinalyses were performed using Bayer Atlas® equipment.
- All other assays were performed using industry-standard equipment and techniques.

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

Procedural QC included using the same instrument and reagents from the same lot numbers whenever possible throughout the study. If single lots were unavailable, analyses were conducted to calibrate subsequent lots and establish target levels and associated standard deviations. 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 created from every 10th participant sample and were analyzed separately to measure test reproducibility. In radioimmunoassays, all three control levels were run initially to validate the standard curve generated.

Scripps Clinic Medical Laboratory has defined quality requirements of accuracy above 99 percent and levels of precision above 97 percent. A variation of the Westgard Rules (1, 2) QC technique is routinely used in the Scripps Laboratory and was used for AFHS assays. In this variation, the \(1_s\) single rule and \(4_s\) multiple rule are used. The \(1_s\) rule indicates rejection of any run when the control value of any one of the three controls (low, mid, high) exceeds two standard deviations from the mean. The \(4_s\) rule indicates rejection of a run when four consecutive control measurements exceed one standard deviation on the same side. This approach ensures an effective system for reducing the probability of false rejection to the lowest acceptable level while maintaining error detection at more than 98 percent.
All QC data were analyzed and summarized in formal QC reports generated monthly. QC data were subjected to independent statistical analysis by the Air Force to produce and analyze time-dependent trends. For all equipment malfunctions or other exceptions, a formal QC exception report was prepared by the responsible individual and forwarded to the project management team. A summary of the coefficients of variation for each quantitative laboratory assay is presented in Appendix D. These coefficients of variation are given separately for each control level and lot.

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

6.4 MEDICAL DATA QC

6.4.1 Overview of QC Procedures

The QC procedures for the medical data consisted of multiple checks at all stages of the examination, data collection, and data processing cycle. A representation of the QC process is given in Figure 6-2. Although improvements were made throughout the physical examination period, QC procedures for data collection, conversion, and integration were developed before the clinical examinations began. All data collection instruments were tested at the pre-examination test conducted several weeks before the start of participant physical examinations. In addition, during the first 2 months of the clinical examinations, all data collection activities were routinely scrutinized to detect and correct procedural deficiencies. Other QC activities included the following:

- Automated QC techniques applied to laboratory data
- Clinical evaluations of all laboratory outliers
- Review of all physical examination findings by one of two diagnosticians
- Automated and manual data quality checking of hard copy against transcribed computer files.

Four interwoven layers of QC were instituted to ensure data integrity. These efforts focused on (1) data processing system design, (2) design and administration of all exams, (3) data completeness checks, and (4) data validation. In addition, Air Force investigators reviewed all physical examination forms and entries. Forms that were found to be questionable, inaccurate, or incorrect were returned to Scripps Clinic for adjudication.
6.4.2 Data Processing System Design

Standards were established for data element formats (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 so that data emanating from study sources (physical examination, questionnaire, and laboratory) were consistent in file format and structure. This approach was necessary to ensure that all data could be integrated into a single database for analysis.
Data collection forms were carefully designed to ensure that all required data elements would be collected in accordance with the study protocol and in a standardized format. These instruments were designed to reflect the order in which the examination itself would be administered and to provide for the sequential coding of information.

Completed clinical examination forms were converted from hard copy to machine-readable images using 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 translated into a SAS® data set, verified against original data sheets, and corrected as necessary. All corrections to the original data sets were saved in the processing program, which was delivered to the Air Force for verification.

Next, the SAS® data sets were subjected to validity checks. All potentially conflicting results, as well as any data values falling at the extremes of expected ranges, were manually reviewed. Extreme values were verified against the original data forms and either corrected or documented as valid results. Potentially conflicting results, either within one form or among forms, were returned to the examiners for review. These results were then documented as having been correctly recorded, corrected, or flagged for exclusion from analysis because of unresolvable examiner errors or omissions. This process was continued until all results were properly documented.

The validity checks were tested with the delivery of the first cumulative medical results data. At that time, it was discovered that some data were not properly cross-checked between collection forms. The discrepancies were adjudicated by the appropriate Scripps Clinic staff and corrected on the forms and in the database in accordance with the QC procedure. Additional QC steps were added to the procedures because of these discrepancies.

Once the edits were completed and the data verified, the “cleaned” files were transferred to the data analysis center for final inspection and integration into the study database. In this QC measure, 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. As an additional measure of QC, cumulative result data sets were delivered quarterly during the physical examination phase for Air Force review. The data sets were finalized following the close of the physical examinations and before the start of statistical analysis. The process for cleaning and converting the collected data into final data sets is found in Figure 6-3.

6.4.3 Design and Administration of Physical and Psychological Examination Forms

The examination forms were designed to elicit all required data while minimizing recording time, enhancing comprehension, and automating data input. Customized mark-sense forms were developed and optical mark recognition technology (OMR) was used to eliminate the risk of transcription errors. 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 both manual and electronic reviews of each form as it was completed. A mark-sense reader was used at Scripps Clinic to scan for completeness and to conduct some broad-based logic checks. Any forms containing missing, incomplete, or contradictory 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. In addition, any differences in interpretation among examiners were identified, and adjustments in recording protocols and programmed data extraction were made as necessary. All examination forms were signed by the examining physician, and the examiner identification number was coded in the database.

6.4.4 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 generate a control code for each rejected form. The control code identified the location of all verification checks failed for a given form.

When a data collection form was rejected, the reason for the rejection was determined. The exact data element was then corrected by comparing the rejected form to the values recorded in the data record created by the scanner. Some of the rejected forms did not contain actual data errors, but rather anomalies created in using mark-sense forms for data collection. For example, the scanner incorrectly counted incompletely erased responses and missed responses marked with too little carbon or graphite. In addition, examiners tended to mark responses clearly for abnormal findings and to mark responses lightly or to bypass responses for expected or desired findings. Failure of the form to provide the correct number of expected responses always resulted in rejection. These errors were resolved, as were the anticipated, more traditional errors.

Out-of-range results and data omissions were monitored to detect trends, possible bias situations, and other data-quality problems. This information was reviewed and relayed to examiners and internal auditors to assist in preventing or correcting chronic, but avoidable, problems. Refresher training was provided to examining physicians to avoid data omissions. Physicians were consulted to correct clinical data, and laboratory out-of-range results were reviewed for logical validity by an independent clinician.

6.4.5 Data Validation

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, errors, or omissions were listed and submitted for review to a physician. Data items that could not be definitively validated or recovered through consultation with the original examiner were assigned codes noting missing or invalid data values. Some reasons for unavailable data included the following:

- Participant refusal
- Unscoreable psychological tests
- Test not ordered (e.g., immunology tests, which were only ordered for a subset of the participants)
- Exemption from testing (e.g., exemption from postprandial glucose testing because of diabetes)

These unrecoverable data were excluded from subsequent analysis. The number of values not available for analyses is presented in each clinical chapter by variable.

In the validation process, transcription errors were found between the two dermatology data collection forms. Although these data were not to be analyzed for this report, all forms were manually checked and corrections were made by the dermatologist.

In validating the genitourinary data, SAIC found 14 participant records with inconsistent information. In all cases either the right testis, or both the left and right testes, were not indicated as normal or abnormal. Scripps Clinic physicians reviewed the records and concluded that the intention had been to code the testes normal. These results were recoded to reflect that finding. All changes were noted on the data collection forms.
All laboratory outliers were reviewed and adjudicated by an auditing physician. Each outlier was adjudicated using the following four codes:

1. Clinically explained or plausible (participant has single outlier)
2. Clinically explained or plausible (participant has multiple outliers)
3. Abnormal outlier not clinically explained but plausible
4. Abnormal outlier not clinically explained and not plausible.

These clinical judgments were included in the processing files. In the 1997 follow-up study, no laboratory outliers were coded as “4.”

6.5 MEDICAL RECORDS CODING QC

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

6.6 STATISTICAL ANALYSIS QC

Specific QC measures were developed for the statistical analysis efforts. The tasks requiring QC included construction of databases for the analysis of each clinical chapter, the statistical analysis itself, and the preparation of the clinical chapters.

Each specialized statistical database was constructed by defining and locating every variable within the many subparts of the composite follow-up database. Although the data had been subjected to QA 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., NORC, Scripps Clinic, or the Air Force).

QA largely depended on regular communication and general agreement among statisticians. Several meetings and consultations between the Air Force team and SAIC statisticians were held in conjunction with the development of the data analysis plan. In addition, many telephone conversations took place during the course of the physical examination. During the analysis, there were frequent telephone conversations, and any problems identified in the statistical analysis were resolved by team discussion. Specialized SAS® programs were developed by the task manager for each type of analysis (exposure, longitudinal, dependent variable-covariate associations) and form of the dependent variable (continuous, dichotomous, polytomous). The software was checked by comparing results from analyses on the same variable by different programs. These programs were adapted for use in all clinical areas by changing the data source, dependent variable, covariates, and exclusions, so that a consistent statistical methodology could be applied to all clinical areas. Modifications to the programs were made only as necessary (e.g., a sparse number of abnormalities that necessitated the exclusion of a particular covariate). Each analysis and the summary statistics reported for the analysis were replicated independently by a separate statistician. The analyses were conducted in accordance with the data analysis plan, which was reviewed
extensively by SAIC and the Air Force. Throughout the study, the Air Force and SAIC maintained duplicate databases. Upon completion of the analyses, SAIC delivered all analysis software and SAS® data sets for each clinical area to the Air Force for final review and archiving.

All tables and statistical results were checked against the computer output from which they were derived, and all statistical statements in the texts were checked for consistency with the results given in the tables. In addition, drafts of each chapter in this report were reviewed by the Air Force and SAIC investigators.

6.7 ADMINISTRATIVE QA

In recognition of the magnitude, complexity, and importance of the AFHS, SAIC created an internal Quality Review Committee (QRC). The QRC was established at the initiation of the 1985 follow-up and continued through the 1987, 1992, and 1997 follow-up studies. Its purpose was to provide general oversight to the AFHS program and advice on the appropriateness of program management and QC actions. The QRC comprised SAIC senior corporate personnel and consultants. These independent reviewers remained separate from the project management staff. The QRC met periodically to review study progress and any issues that either had an impact on study quality or were perceived as a potential problem. Members of the QRC also conducted first-hand evaluations of ongoing program operations.
REFERENCES


# Table of Contents

7 STATISTICAL METHODS

7.1 INTRODUCTION

7.2 MODELS AND ASSUMPTIONS

7.2.1 Model 1: Group and Occupation as Estimates of Exposure

7.2.2 Models 2 through 4: Serum Dioxin as an Estimate of Exposure

7.2.2.1 Prior Knowledge Regarding Dioxin

7.2.2.2 Fundamental Limitations of the Serum Dioxin Data

7.2.2.3 Model 2: Health versus Initial Dioxin in Ranch Hands

7.2.2.4 Model 3: Health versus Dioxin in Ranch Hands and Comparisons

7.2.2.5 Model 4: Health versus 1987 Dioxin in Ranch Hands

7.3 FACTORS DETERMINING THE STATISTICAL ANALYSIS METHOD

7.4 ANALYSIS METHODOLOGIES

7.4.1 Methods for Analyzing Continuous and Discrete Variables

7.4.2 Modeling Strategy

7.4.3 Longitudinal Analysis

7.5 INTERPRETIVE CONSIDERATIONS

7.5.1 Adjustments for Covariates

7.5.2 Multiple Testing

7.5.3 Trends

7.5.4 Interpretation of the Coefficient of Determination

7.5.5 Clinical Interpretation of Discrete versus Continuous Data

7.5.6 Power

7.6 EXPLANATION OF TABLES

7.6.1 Exposure Analysis

7.6.1.1 Continuous Variables

7.6.1.2 Discrete Variables

7.6.1.2.1 Discrete Variable with Two Categories

7.6.1.2.2 Discrete Variable with More Than Two Categories

7.6.2 Longitudinal Analysis

7.6.2.1 Continuous Variables

7.6.2.2 Discrete Variables with Two Categories

7.6.2.2.1 Discrete Variable with More Than Two Categories

REFERENCES
List of Tables

Table 7-1. Model 1: Assessing Health versus Group Status in Ranch Hands and Comparisons: Assumptions, Advantages, and Disadvantages ........................................................................................................ 7-3

Table 7-2. Model 2: Assessing Health versus Initial Dioxin in Ranch Hands: Assumptions, Advantages, and Disadvantages ........................................................................................................ 7-5

Table 7-3. Model 3: Assessing Health versus Categorized Dioxin in Ranch Hands and Comparisons 7-7

Table 7-4. Model 4: Assessing Health versus 1987 Dioxin in Ranch Hands: Assumptions, Advantages, and Disadvantages ........................................................................................................ 7-8

Table 7-5. Summary of Statistical Analysis Situations by Dependent Variable Form, Exposure Estimate, Analysis Cohort, and Analysis Type ..................................................................... 7-9

Table 7-6. Summary of Statistical Procedures ...................................................................................................................... 7-12

Table 7-7. Approximate Power To Detect a Group Effect at a 5-Percent Level of Significance (Discrete Dependent Variable) ............................................................................................ 7-19

Table 7-8. Approximate Power To Detect a Group Effect at a 5-Percent Level of Significance (Continuous Dependent Variable) ............................................................................................ 7-20

Table 7-9. Location of Table Results from Different Exposure Analysis Models ............................................................ 7-21

Table 7-10. Location of Table Results from Different Longitudinal Analysis Models ...................................................... 7-28
7 STATISTICAL METHODS

7.1 INTRODUCTION

This chapter summarizes the statistical methods used in the analysis of Air Force Health Study (AFHS) 1997 follow-up examination data to investigate relations between the health status of the 2,121 participants attending this examination and their corresponding group (Ranch Hand or Comparison) or serum dioxin estimates and measurements. Group contrast models were similar to analyses performed for the 1982 baseline and 1985, 1987, and 1992 follow-up examinations (1, 2, 3, 4). Models relating health to dioxin estimates and measurements were based on analyses performed for the Serum Dioxin Analysis Report for the 1987 Follow-up and 1992 follow-up examinations (4, 5).

The statistical methods used in this report encompassed four different forms of hypotheses or models applied to 266 study endpoints. Each of these models specified the study cohort or subset of participants included in the respective analyses together with the dioxin exposure or proxy estimates used in the analysis. The first model (Model 1) specified contrasts between Ranch Hands and Comparisons using group as a proxy for exposure, and it did not incorporate serum dioxin measurements. The remaining three models (Models 2, 3, and 4) all incorporated serum dioxin measurements. A summary description of each of the four models is provided in section 7.2, “Models and Assumptions.”

Each model and exposure estimate combination was implemented for study variables and type of analysis (unadjusted, adjusted, or longitudinal). The specific statistical procedures (e.g., analysis of variance or logistic regression) that were used are presented in section 7.3, “Factors Determining Statistical Analysis Method.” The relation between the factors and statistical procedures is presented in section 7.4, “Analysis Methodologies.” That presentation is followed by a discussion of “Interpretive Considerations” (section 7.5), and a description of the contents of tables used to report statistical analysis results throughout the report is given in the “Explanation of Tables” (section 7.6).

7.2 MODELS AND ASSUMPTIONS

The statistical analysis was based primarily on four models, each using a different estimate of exposure. The first model used group and military occupation (officer, enlisted flyer, and enlisted groundcrew) to assess health effects and dose-response relations related to herbicide exposure. Serum dioxin measurements were not used in this model. The other three models accounted for dioxin effects either through estimated initial dioxin levels for Ranch Hands or using current or recent serum dioxin levels for Ranch Hands and Comparisons to assess health endpoints (e.g., cholesterol, diabetes) and dose-response relations related to exposure. These analyses were accomplished with and without adjustment for covariates.

Throughout this report, dioxin levels are used as measures of both exposure to dioxin itself and exposure to dioxin-contaminated herbicides, including Herbicide Orange. Direct contrasts of Ranch Hand and Comparison veterans (Model 1) address the hypothesis of health effects attributable to any herbicide exposure experienced by Ranch Hand veterans during Operation Ranch Hand. Models involving dioxin measurements address the hypothesis that health effects change with the amount of exposure. Dioxin measurements are used as a measure of exposure to dioxin-contaminated herbicides because it is expected that as exposure to such herbicides increased, dioxin levels should increase. Therefore, the dioxin measurement serves as direct biomarker of exposure to dioxin-contaminated herbicides. No other
direct measure or estimate of herbicide exposure is available with which to address hypothetical dose-
response relations with health. Some indirect measures, such as self-report of skin contact among
enlisted groundcrew, or simply being a Ranch Hand enlisted groundcrew member, are valuable
alternatives because dioxin measures suggest that enlisted groundcrew experienced the heaviest
exposures. Reported skin exposure is not addressed in this report, but enlisted groundcrew status is
addressed in Model 1. The use of dioxin as a measure of exposure to dioxin-contaminated herbicides is
consistent with the goal of the study, which is to determine whether health effects exist and can be
attributed to occupational exposure to Herbicide Orange (6).

7.2.1 Model 1: Group and Occupation as Estimates of Exposure

This section describes the model that used the exposure group (Ranch Hand, Comparison) to assess the
relation between health status and dioxin exposure. Statistical analyses based on this model were termed
"Model 1" in the assessment of the clinical areas. Analyses of this type are straightforward, easy to
interpret, and well established in epidemiological studies. In this model, exposure was defined as “yes”
for Ranch Hands and “no” for Comparisons without regard to the magnitude of the exposure. As an
attempt to quantify exposure, three contrasts of Ranch Hands and Comparisons were performed along
with the overall Ranch Hand versus Comparison contrast. These three contrasts compared Ranch Hands
and Comparisons within each occupational category (officers, enlisted flyers, and enlisted groundcrew).
As described in the analyses performed for the Serum Dioxin Analysis Report for the 1987 Followup (5),
the average levels of exposure to dioxin were highest for enlisted groundcrew, followed by enlisted
flyers, then officers.

Table 7-1 provides a description of Model 1 and gives the assumptions, advantages, and disadvantages
for a continuously distributed health endpoint, y. The model presented in Table 7-1 is unadjusted for any
covariates—adjusted models are a straightforward extension.
Table 7-1. Model 1: Assessing Health versus Group Status in Ranch Hands and Comparisons: Assumptions, Advantages, and Disadvantages

<table>
<thead>
<tr>
<th>Model 1: ( y = \mu + G_i + e ) (All Ranch Hands and Comparisons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y = \mu + G_i + O_j + (GO)_{ij} + e ) (Ranch Hands and Comparisons by occupation)</td>
</tr>
</tbody>
</table>

where

\( y \) = health variable in group \( i \) and occupation \( j \)

\( G_i \) = effect due to group status \((i = 1, 2 - \text{Comparisons, Ranch Hands})\)

\( O_j \) = effect due to occupation \((j = 1, 2, 3 - \text{Officers, Enlisted Flyers, Enlisted Groundcrew})\)

\( (GO)_{ij} \) = interaction between group status and occupation \((i = 1, 2; j = 1, 2, 3)\); used to examine Ranch Hand and Comparison differences for each occupation

\( e \) = zero mean error.

Assumptions: Comparisons were unexposed and Ranch Hands were exposed.

For the purposes of investigating dose-response effects, enlisted groundcrew were more heavily exposed than enlisted flyers, and enlisted flyers were more heavily exposed than officers.

The error variance does not change with group status or occupation.

Advantages: Easily interpretable.

Disadvantages: Results are biased toward the null hypothesis of no dioxin effect if unexposed Ranch Hands are misclassified (i.e., remain in the analysis as exposed Ranch Hands). It is not possible to fully distinguish unexposed Ranch Hands from exposed Ranch Hands.

7.2.2 Models 2 through 4: Serum Dioxin as an Estimate of Exposure

Current dioxin levels in 1987 were determined by the Centers for Disease Control and Prevention from serum samples taken from approximately 2,000 Ranch Hands and Comparisons. Additional serum samples were taken from selected Ranch Hands and Comparisons at the 1992 and 1997 follow-up examinations to provide insight on dioxin levels and the elimination of dioxin from the body, and to provide measurements for new subjects and those who were not previously measured. A discussion of the details of dioxin measurement is found in Chapter 2, Dioxin Assay.

Investigation of the mechanics of dioxin elimination is currently under study by the Air Force. Based on samples collected in 1982, 1987, 1992, and 1997, issues such as half-life estimation and first-order pharmacokinetic assumptions are being investigated.

7.2.2.1 Prior Knowledge Regarding Dioxin

This section presents analytic strategies based on assumptions and models conceived after the Ranch Hand half-life study published in 1996 (7). Available data have suggested that the dioxin elimination process is first-order, based on measurements subsequent to the ingestion of dioxin by an individual (8). Data on 213 Ranch Hand veterans with dioxin measured in blood collected in 1982, 1987, and 1992 produced a half-life estimate of 8.7 years (7); this estimate was used in all calculations involving half-life.
The term “elimination” denotes the overall removal of dioxin from the body. Some of the analyses assumed that the amount of dioxin in the body (C) decreases exponentially with time according to the model $C = I \cdot \exp(-rt)$, where $I$ is the initial level, $r = \log(2)/h$ is the elimination rate, $h$ is the half-life, and $t$ is the number of years from the end of service in Southeast Asia (SEA) to the time of the blood measurement for dioxin. If a participant had measurements at more than one point in time, the measurement closest to the time of duty in SEA was used. This exponential elimination law is termed “first-order elimination.”

The first-order elimination assumption is equivalent to assuming a one-compartment model for dioxin distribution within the body. While a multicompartment model incorporating body composition and dioxin binding to tissue receptors would provide a detailed description of dioxin concentrations in different compartments, published multicompartment models for dioxin distribution within the body predict first-order elimination of dioxin, overwhelmingly because of fecal elimination (9).

The lipid-weight concentration of dioxin, expressed in parts per trillion (ppt) (10, 11), is a derived quantity calculated from the formula $\text{ppt} = \frac{\text{ppq} \cdot 102.6}{W}$, where ppt is the lipid-weight concentration, ppq (parts per quadrillion) is the actual whole weight of dioxin in the sample in femtograms, 102.6 corrects for the average density of serum, and $W$ is the total lipid weight of the sample (9).

The relation between the serum lipid-weight concentration of dioxin and lipid-weight concentrations in adipose tissue is a subject of continuing research. The correlation between the serum lipid-weight concentration and adipose tissue lipid-weight concentration of dioxin has been observed by Patterson, et al., to be 0.98 in 50 persons from Missouri (12). Using the same data, Patterson, et al., calculated the partitioning ratio of dioxin between adipose tissue and serum on a lipid-weight basis as 1.09 (95% confidence interval: [0.97,1.21]). On the basis of these data, a one-to-one partitioning ratio of dioxin between lipids in adipose tissue and the lipids in serum could not be excluded. Measurements of dioxin in adipose tissue generally have been accepted as representing the body burden concentration of dioxin. The high correlation between serum dioxin levels and adipose tissue dioxin levels in the Patterson, et al., study suggests that serum dioxin is also a valid measurement of dioxin body burden.

7.2.2.2 Fundamental Limitations of the Serum Dioxin Data

There are two evident limitations to the available data:

- While Ranch Hand data did not appear to violate a first-order elimination assumption, no serially repeated dioxin assay results, taken over many years and with which to evaluate directly the adequacy of the first-order elimination model in humans, were available.
- It was not known whether Ranch Hands with body burdens of dioxin at or below 10 ppt were exposed and their body burdens had decreased to these levels since their time of duty in SEA, or whether they were not exposed at all during their time of duty in SEA.

7.2.2.3 Model 2: Health versus Initial Dioxin in Ranch Hands

The relation between estimated initial dioxin levels and health was assessed in Ranch Hands using the model described in Table 7-2. Statistical analyses based on this model were termed “Model 2” in the assessment of the clinical areas. In this model, an initial dioxin level was estimated for a Ranch Hand from a current or recent lipid-adjusted dioxin measure, the length of time between the time of duty in SEA and the date of the blood measurement of dioxin, and an estimated half-life of 8.7 years. From studies conducted by the Air Force, body fat at the time of the blood measurement of dioxin appeared to be related to the dioxin half-life for a participant (7). Hence, this body fat measure was included in this
model as a covariate. Model 2 differs from Model 1 in that the estimate of exposure in Model 1 (group: Ranch Hand, Comparison) was not dependent upon extrapolation to an earlier date.

Table 7-2 also includes assumptions, advantages, and disadvantages of the model for a continuously distributed health variable, \( y \). The model presented in Table 7-2 is unadjusted for any additional risk factors, but extension to an adjusted model is straightforward.

**Table 7-2. Model 2: Assessing Health versus Initial Dioxin in Ranch Hands: Assumptions, Advantages, and Disadvantages**

**Model 2:**

\[
y = b_0 + b_1 \log_2(I) + b_2 BF + e
\]

where

\[
y = \text{health variable}
\]

\[
I = \text{extrapolated initial dose, assuming first-order elimination, } I = 4 + (C - 4) \cdot \exp(\log(2) \cdot t/h), \text{ where } 4 \text{ ppt is considered the median background level of lipid-adjusted dioxin;}
\]

\[
t = \text{length of time between the time of duty in SEA and the date of the blood measurement of dioxin in 1987, 1992, or 1997; } C = \text{lipid-adjusted dioxin, determined in 1987, 1992, or 1997; and } h = \text{dioxin half-life in Ranch Hands assuming first-order elimination (8.7 years assumed for analysis)}
\]

\[
BF = \text{body fat at the time of the blood measurement of dioxin, calculated from the formula shown below}
\]

\[
e = \text{zero mean error.}
\]

Body fat was calculated from a metric body mass index (13); the formula is

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

**Assumptions:**

- Ranch Hands received a single dioxin dose in Vietnam and background exposure thereafter.
- Ranch Hands experienced first-order dioxin elimination.
- The error variance does not change with health status or initial dioxin dose.

**Advantages:**

- Easily interpretable.
- Most efficient if first-order elimination and half-life are valid and \( y \) is linearly related to \( \log_2(I) \).
- The logarithm (base 2) of initial dioxin presents the dioxin data as a more symmetric distribution than the distribution of initial dioxin in its original units. In addition, the relative risk based on the logarithm (base 2) of initial dioxin is more meaningful than on the original scale (i.e., a doubling of initial dioxin rather than a 1 ppt increase in dioxin).

**Disadvantages:**

- Results are biased if first-order elimination or constant half-life assumptions are not valid.

In Table 7-2, the phrase “single dioxin dose” is a simplification of the process by which Ranch Hands accumulated dioxin during their time of duty in SEA. This process, which undoubtedly varied from individual to individual, is unknown; however, the time of duty in SEA for an individual Ranch Hand generally was short (1 to 3 years) relative to the time elapsed since his duty in SEA. Hence, additional knowledge regarding the accumulation of dioxin during an individual Ranch Hand’s time of duty in SEA, were it to become available, would not likely change conclusions drawn from any of the statistical analyses.
Analyses were performed on Ranch Hands who had lipid-adjusted dioxin levels greater than 10 ppt at either the 1987, 1992, or 1997 physical examinations. The value 10 ppt corresponds to the approximate 98th percentile of the Comparison lipid-adjusted dioxin distribution. Based on this Comparison dioxin distribution, it was believed that participants with greater than 10 ppt lipid-adjusted dioxin were definitely exposed. It was not known whether Ranch Hands with dioxin burdens at or below 10 ppt were exposed and their body burdens had decreased to these levels since their time of duty in SEA, or whether they were not exposed at all during their time of duty in SEA. Lipid-adjusted dioxin levels less than 10 ppt are subsequently called “background” levels.

7.2.2.4 Model 3: Health versus Dioxin in Ranch Hands and Comparisons

An assessment of the health consequences of dioxin above background levels was carried out with a model that was applied to both Ranch Hand and Comparison data. This model assessed health versus dioxin body burden categorized into four levels, given below:

- **Comparisons**—Comparisons with up to 10 ppt lipid-adjusted dioxin
- **Background**—Ranch Hands with up to 10 ppt lipid-adjusted dioxin
- **Low**—Ranch Hands with more than 10 ppt lipid-adjusted dioxin but at most 94 ppt estimated initial dioxin
- **High**—Ranch Hands with more than 10 ppt lipid-adjusted dioxin and more than 94 ppt estimated initial dioxin.

Statistical analyses based on this model were termed “Model 3” in the assessment of the clinical areas. The low and high Ranch Hand categories, of approximately equal size, were determined by the median estimated initial dioxin level (94 ppt) of the Ranch Hands with more than 10 ppt lipid-adjusted dioxin (i.e., the sample used in Model 2). In this model, an initial dioxin level was estimated for a Ranch Hand from a current or recent lipid-weight dioxin measure, the length of time between the time of duty in SEA and the date of the blood measurement of dioxin, and an estimated half-life of 8.7 years. From studies conducted by the Air Force, body fat at the time of the blood measurement of dioxin appeared to be related to the dioxin half-life for a participant. This body fat measure was included in this model as a covariate. Using this body fat measure in Model 3 for all Comparisons and Ranch Hands with dioxin measurements allowed body fat to act as a potential risk factor as well as an adjusting variable to explain half-life differences.

For a continuously distributed health variable, y, for example, the mean values of y within the background, low, high, and low plus high categories were contrasted with the mean values of y within the Comparison category. The mean value of y for the low plus high category was calculated as a linear combination of the low dioxin category and the high dioxin category, with weights based on the sample size in each of these categories. Relative frequencies were contrasted for discrete health variables. Table 7-3 shows this model and the assumptions, advantages, and disadvantages for the unadjusted analysis of a continuous variable; extension to an adjusted model is straightforward.
Table 7-3. Model 3: Assessing Health versus Categorized Dioxin in Ranch Hands and Comparisons

| Model 3: $y = b_0 + b_1 I_1 + b_2 I_2 + b_3 I_3 + b_4 I_4 + BF + e$ |
|---|---|---|---|---|---|
| $y$ | health variable |
| $I_1$ | indicator variable for categorized dioxin; $I_1 = 1$ if participant is a Comparison with a background level of dioxin, $I_1 = 0$ if participant is not a Comparison |
| $I_2$ | indicator variable for categorized dioxin; $I_2 = 1$ if participant is in background dioxin category, $I_2 = 0$ if participant is not in background dioxin category |
| $I_3$ | indicator variable for categorized dioxin; $I_3 = 1$ if participant is in low dioxin category, $I_3 = 0$ if participant is not in low dioxin category |
| $I_4$ | indicator variable for categorized dioxin; $I_4 = 1$ if participant is in high dioxin category, $I_4 = 0$ if participant is not in high dioxin category |
| $BF$ | body fat at the time of blood measurement of dioxin, calculated from the formula shown below |
| $e$ | zero mean error. |

Body fat was calculated from a metric body mass index (13); the formula is

$$Body Fat (in percent) = \frac{Weight (kg)}{[Height (m)]^2} \cdot 1.264 - 13.305.$$ 

Assumptions:
- Dioxin body burden has been eliminated with time.
- The error variance does not change with categorized dioxin body burden.

Advantages:
- Initial dioxin is probably a better measure for determining low and high exposure than current or recent lipid-adjusted dioxin measurements.
- Less dependent on the accuracy of the estimation algorithm for determining initial dioxin than Model 2.

Disadvantages:
- Makes no use of prior belief that some Ranch Hands received an unusually large dioxin dose in Vietnam; all Ranch Hands with high dioxin levels are treated similarly.
- "Background" Ranch Hand category is probably a mixture of exposed and unexposed Ranch Hands. Analysis may be biased toward the null hypothesis of no dioxin effect.
- "Low" and "high" Ranch Hand categories are based on initial dioxin model, which is based on valid half-life and first-order dioxin elimination. Bias is possible if model is incorrect. Also, a conditional null hypothesis is tested using these categories ("Is there a dioxin effect, given a specified level of exposure?").

7.2.2.5 Model 4: Health versus 1987 Dioxin in Ranch Hands

The relation between 1987 dioxin and health was assessed using the model described in Table 7-4. This measure of dioxin is termed "1987 dioxin" because most Ranch Hands were assayed for dioxin initially at the 1987 follow-up examination. This table also describes the assumptions, advantages, and disadvantages for the unadjusted analysis of a continuously distributed health variable, $y$. 

7-7
Ranch Hands with a dioxin measurement may have had their blood collected at the pilot study in April 1987, at the 1987 physical examination, at the 1992 physical examination, or at the 1997 physical examination. If an individual had measurements at more than one of these points in time, the measurement closest to the time of duty in SEA was used. If only a 1992 serum dioxin measurement was available, the level was extrapolated to the date of the 1987 physical examination. The model

\[ C_{1987} = 4 + (C_{1992} - 4) \cdot \exp(r t) \]

was used for extrapolation of lipid-adjusted dioxin to 1987 levels \( C_{1987} \), where \( C_{1992} \) is the lipid-adjusted dioxin level in 1992, 4 ppt is considered the median background level for lipid-adjusted dioxin, \( r = \log(2)/h \) is the elimination rate, \( h \) is the half-life (8.7 years), and \( t \) is the length of time between the physical examination in 1987 and the physical examination in 1992. This model was used only if the lipid-adjusted dioxin level in 1992 was greater than 10 ppt; otherwise, the 1992 measurement was used. A similar strategy was used for participants who had only a 1997 serum dioxin measurement. The estimate of exposure in Model 4 (1987 dioxin) was based on extrapolation to 1987 for only 39 out of the 863 Ranch Hands. Most measurements were based on 1987 dioxin measurements and extrapolation was not needed. Consequently, body fat at the time of the blood measurement of dioxin was not used in Model 4, which was different from the strategy used for Models 2 and 3.

### Table 7-4. Model 4: Assessing Health versus 1987 Dioxin in Ranch Hands: Assumptions, Advantages, and Disadvantages

**Model 4:** \( y = b_0 + b_1 \log_2(ppt + 1) + e \)

| where | \( y \) = health variable \n| \( ppt \) = lipid-adjusted dioxin = \( \text{ppq} \cdot 10^{2.6/W} \), where \( \text{ppq} \) = whole weight of dioxin in the sample in femtograms (102.6 corrects for the average density of serum) and \( W \) = total lipid weight of the sample \n| \( e \) = zero mean error. |

**Assumptions:** Ranch Hands received a single dioxin dose in Vietnam and background exposure thereafter. The error variance does not change with health status or 1987 dioxin.

**Advantages:** Using 1987 dioxia has less inherent variation than initial dioxin, which is extrapolated by a first-order elimination model across a 20- to 30-year time period. The logarithm (base 2) of (1987 dioxin + 1) presents the dioxin data as a more symmetric distribution than the distribution of 1987 dioxin in its original units. In addition, the relative risk based on the logarithm (base 2) of (1987 dioxin + 1) is more meaningful than on the original scale (i.e., a doubling of 1987 dioxin + 1, rather than a 1 ppt increase in dioxin).

**Disadvantages:** 1987 dioxin may not be a good surrogate for exposure if elimination rate differs among individuals. Individuals with measurements in 1992 only or 1997 only are extrapolated to 1987, and variation is increased with estimation using a first-order elimination model.
The relation between current health and dioxin was assessed using a model, termed “Model 4,” with lipid-adjusted 1987 dioxin as the estimate of exposure. Model 4 used the logarithm (base 2) of lipid-adjusted 1987 dioxin and is described in Table 7-4.

### 7.3 FACTORS DETERMINING THE STATISTICAL ANALYSIS METHOD

For a specified questionnaire-based or clinical measurement determined from the physical or laboratory examination, the selection of an analytical method depended on each of the following:

- **Dependent Variable Form:** Continuous or discrete
- **Exposure Estimate and Analysis Cohort:**
  - Model 1: Group—All Ranch Hands and Comparisons
  - Model 2: Initial dioxin—Ranch Hands having a dioxin body burden of greater than 10 ppt of lipid-adjusted dioxin, based on 1987 dioxin levels as defined in Section 7.2.2.5
  - Model 3: Categorized dioxin—Comparisons with a dioxin body burden of 10 ppt lipid-weight dioxin or less, based on 1987 dioxin levels, and all Ranch Hands with a dioxin measurement
  - Model 4: 1987 dioxin—All Ranch Hands with a dioxin measurement
- **Analysis Type:** Unadjusted, adjusted, or longitudinal.

Table 7-5 specifies 22 separate analysis situations based on dependent variable form, exposure estimate, analysis cohort, and analysis type. For each of the 22 situations, the statistical method is specified. For example, linear regression models were used for adjusted analyses of initial dioxin for continuous dependent variables.

<table>
<thead>
<tr>
<th>Exposure Estimate</th>
<th>Analysis Cohort</th>
<th>Analysis Type</th>
<th>Statistical Methods</th>
<th>Independent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: Group (Ranch Hands vs. Comparisons)</td>
<td>All RH &amp; C</td>
<td>Unadjusted</td>
<td>Analysis of Variance</td>
<td>Group</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adjusted</td>
<td>Analysis of Covariance</td>
<td>Group; Covariates</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Longitudinal*</td>
<td>Analysis of Covariance</td>
<td>Group; Age at the 1997 Follow-up Examination; 1982 Measurement</td>
</tr>
<tr>
<td>Model 2: Log(Initial)</td>
<td>RH &gt;10 ppt lipid-adjusted 1987 dioxin</td>
<td>Unadjusted</td>
<td>Linear Regression</td>
<td>Log(Initial); Body Fat at the Time of the Blood Measurement of Dioxin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adjusted</td>
<td>Linear Regression</td>
<td>Log(Initial); Body Fat at the Time of the Blood Measurement of Dioxin; Covariates</td>
</tr>
</tbody>
</table>

---

7-9
### Table 7-5. Summary of Statistical Analysis Situations by Dependent Variable Form, Exposure Estimate, Analysis Cohort, and Analysis Type (Continued)

<table>
<thead>
<tr>
<th>Exposure Estimate</th>
<th>Analysis Cohort</th>
<th>Analysis Type</th>
<th>Statistical Methods</th>
<th>Independent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal(^a)</td>
<td>Linear Regression</td>
<td>Log(_2)(Initial); Body Fat at the Time of the Blood Measurement of Dioxin; Age at the 1997 Follow-up Examination; 1982 Measurement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 3: Categorized Dioxin</td>
<td>All RH with a dioxin measurement, C ≤ 10 ppt lipid-adjusted 1987 dioxin</td>
<td>Unadjusted</td>
<td>Analysis of Covariance</td>
<td>DXCAT; Body Fat at the Time of the Blood Measurement of Dioxin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adjusted</td>
<td>Analysis of Covariance</td>
<td>DXCAT; Body Fat at the Time of the Blood Measurement of Dioxin; Covariates</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Longitudinal(^a)</td>
<td>Analysis of Covariance</td>
<td>DXCAT; Body Fat at the Time of the Blood Measurement of Dioxin; Age at the 1997 Follow-up Examination; 1982 Measurement</td>
</tr>
<tr>
<td>Model 4: Log(_2)(1987 Dioxin + 1)</td>
<td>All RH with a dioxin measurement</td>
<td>Unadjusted</td>
<td>Linear Regression</td>
<td>Log(_2)(1987 Dioxin + 1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adjusted</td>
<td>Linear Regression</td>
<td>Log(_2)(1987 Dioxin + 1); Covariates</td>
</tr>
<tr>
<td>Discrete</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1: Group (Ranch Hands vs. Comparisons)</td>
<td>All RH &amp; C</td>
<td>Unadjusted</td>
<td>Chi-Square Contingency Table, Logistic Regression</td>
<td>Group</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adjusted</td>
<td>Logistic Regression</td>
<td>Group; Covariates</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Longitudinal(^b)</td>
<td>Logistic Regression</td>
<td>Group; Age at the 1997 Follow-up Examination</td>
</tr>
<tr>
<td>Model 2: Log(_2)(Initial)</td>
<td>RH &gt;10 ppt lipid-adjusted 1987 dioxin</td>
<td>Unadjusted</td>
<td>Logistic Regression</td>
<td>Log(_2)(Initial); Body Fat at the Time of the Blood Measurement of Dioxin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adjusted</td>
<td>Logistic Regression</td>
<td>Log(_2)(Initial); Body Fat at the Time of the Blood Measurement of Dioxin; Covariates</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Longitudinal(^b)</td>
<td>Logistic Regression</td>
<td>Log(_2)(Initial); Body Fat at the Time of the Blood Measurement of Dioxin; Age at the 1997 Follow-up Examination</td>
</tr>
<tr>
<td>Model 3: Categorized Dioxin</td>
<td>All RH with a dioxin measurement, C ≤ 10 ppt lipid-adjusted 1987 dioxin</td>
<td>Unadjusted</td>
<td>Chi-Square Contingency Table; Logistic Regression</td>
<td>DXCAT; Body Fat at the Time of the Blood Measurement of Dioxin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adjusted</td>
<td>Logistic Regression</td>
<td>DXCAT; Body Fat at the Time of the Blood Measurement of Dioxin; Covariates</td>
</tr>
</tbody>
</table>
Table 7-5. Summary of Statistical Analysis Situations by Dependent Variable Form, Exposure Estimate, Analysis Cohort, and Analysis Type (Continued)

<table>
<thead>
<tr>
<th>Exposure Estimate</th>
<th>Analysis Cohort</th>
<th>Analysis Type</th>
<th>Statistical Methods</th>
<th>Independent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 4: $\text{Log}_2(1987 \text{ Dioxin} + 1)$</td>
<td>All RH with a dioxin measurement</td>
<td>Unadjusted</td>
<td>Logistic Regression</td>
<td>DXCAT; Body Fat at the Time of the Blood Measurement of Dioxin; Age at the 1997 Follow-up Examination $\text{Log}_2(1987 \text{ Dioxin} + 1)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adjusted</td>
<td>Logistic Regression</td>
<td>$\text{Log}_2(1987 \text{ Dioxin} + 1)$; Covariates</td>
</tr>
</tbody>
</table>

- Dependent variable usually paired difference score of (1997 to 1982) dependent variable values. For some clinical areas, paired difference scores were (1997 to 1985) differences.
- Analysis performed subject to the constraint that participant was normal at the 1982 baseline (or 1985) examination.

Note: $\text{Log}_2(\text{Initial}) = \text{Logarithm (base 2) of estimated initial dioxin level.}$
$\text{Log}_2(1987 \text{ Dioxin} + 1) = \text{Logarithm (base 2) of (1987 dioxin level} + 1).$
$\text{DXCAT} = \text{Categorized dioxin (incorporating group membership—three categories for Ranch Hands, one category for Comparisons).}$
$\text{RH} = \text{Ranch Hand.}$
$\text{C} = \text{Comparison.}$

7.4 ANALYSIS METHODOLOGIES

7.4.1 Methods for Analyzing Continuous and Discrete Variables

For analyses of continuous dependent variables, the general linear models approach was used for applying such techniques as simple and multiple linear regression, analysis of variance, analysis of covariance, repeated measures analysis, and survival time analysis. This approach permitted model fitting of the dependent variable as a function of group or dioxin and specified covariates. Continuous dependent variables were examined to ensure that assumptions underlying appropriate statistical methods were met. Transformations (e.g., square root, logarithmic) were used to enhance normality for specific continuous health variables. A further discussion of general linear models, as well as other methods used for the statistical analyses in this report, is found in Table 7-6.

For these continuous analyses, the SAS® general linear models analysis (PROC GLM) (14) was used.
After a model was fitted, tests of significance for a group or dioxin effect were developed. Associations with a p-value less than or equal to 0.05 were described as significant, and associations with a p-value greater than 0.05 but less than or equal to 0.10 were described as marginally significant.

The SAS® procedures LIFEREG and LIFETEST (14) were used for the time to diabetes onset variable in the endocrine clinical assessment. Statistical methods used to analyze measures of this type implemented $^{1}$

$^{1}$ SAS and all other SAS Institute, Inc., product and service names are registered trademarks or trademarks of SAS Institute, Inc., in the USA and other countries.
a technique known as "survival time" analysis. A further discussion of survival time analysis is found in Table 7-6.

For dichotomous discrete dependent variables, logistic regression was performed using SAS® PROC GENMOD (15). For dependent variables with more than two categories, polytomous logistic regression was performed using SAS® PROC CATMOD (14). Parameter estimation and model selection for polytomous logistic regression and ordinary logistic regression are similar. Both forms of regression use the maximum likelihood principle to obtain parameter estimates. For a model with k parameters for two equations, 2k parameters are estimated, k for each logit function. If ordinary logistic regression is applied twice (for example, once for abnormal low versus normal and then for abnormal high versus normal), 2k parameters are estimated; however, ordinary logistic regression maximizes two likelihood equations, each with k parameters, while polytomous logistic regression estimates all 2k parameters simultaneously with one likelihood equation. Polytomous logistic regression also can be used for dependent variables that have more than three levels and require more than two contrasts with a normal category. A further discussion of logistic regression and polytomous logistic regression is found in Table 7-6.

A chi-square statistic, adjusted for the continuity of the chi-square distribution, was used when a test of the relative frequency of abnormal measurements between Ranch Hands and Comparisons was performed, and the relative frequency of either the Ranch Hand or the Comparison group was zero. This test statistic yields p-values approximately equal to Fisher's exact test (16) for a two-sided alternative hypothesis.

Table 7-6. Summary of Statistical Procedures

<table>
<thead>
<tr>
<th>Chi-Square Contingency Table Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>The chi-square test of independence (17) is calculated for a contingency table by the following formula:</td>
</tr>
<tr>
<td>[ \chi^2 = \sum \frac{(f_0 - f_E)^2}{f_E} ]</td>
</tr>
<tr>
<td>where the sum is taken over all cells of the contingency table and</td>
</tr>
<tr>
<td>( f_0 ) = observed frequency in a cell</td>
</tr>
<tr>
<td>( f_E ) = expected frequency under the hypothesis of independence.</td>
</tr>
</tbody>
</table>

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

For 2x2 tables, the chi-square statistic above can be adjusted for the continuity of the \( \chi^2 \) distribution. This test statistic yields p-values approximately equal to Fisher's exact test (16) for a two-sided alternative and is as follows:

\[ \chi^2 = \sum \frac{\max (0, (|f_0 - f_E| - 1/2))^2}{f_E} \]
Table 7-6. Summary of Statistical Procedures (Continued)

Correlation Coefficient (Pearson's Product-Moment)

The population correlation coefficient \( \rho \) (18) measures the strength of the linear relation between two random variables \( X \) and \( Y \). A commonly used sample-based estimate of this correlation coefficient is

\[
\rho = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}
\]

where the sum is taken over all \((x, y)\) pairs in the sample. A student's t-test based on this estimator is used to test for a significant correlation between the two random variables of interest. For the sample size of approximately 2,121 in this study, a sample correlation coefficient of 0.04254 is sufficient to attain a statistically significant correlation at a 5-percent level for a two-sided hypothesis test, assuming normality of \( X \) and \( Y \).

Survival Time Analysis

The survival time model (19) permits a dependent variable with censored observations to be modeled in a general linear models framework. For example, if the time to diabetes onset is defined as an event, the time for participants for which this event has not occurred is right-censored. The survival time model is

\[
y = X\theta + \sigma e
\]

where
- \( y \) = vector of responses (e.g., time to diabetes onset), usually the logarithm of the survival times
- \( X \) = matrix of covariates, or risk factors (e.g., group status and age)
- \( \theta \) = vector of unknown regression parameters
- \( \sigma \) = unknown scale parameter
- \( e \) = vector of errors assumed to have a known distribution.

For a model with a dependent variable containing right-censored data, the log likelihood function is a combination of a probability density function for noncensored values and a survival distribution function for right-censored values. The model parameters can be estimated by maximum likelihood in SAS® PROC LIFEREG, using a Newton-Raphson algorithm, where the distribution of the random error term can be specified. The distributional assumptions regarding the error term can be tested by examining plots of the Kaplan-Meier survival functions using SAS® PROC LIFETEST.

PROC LIFEREG provides estimates, standard errors, and \( p \)-values associated with a chi-square test on each parameter (i.e., risk factor) in the model. These are used to test the significance of the group or dioxin term in the unadjusted and adjusted models. In this procedure, percentile estimates also can be produced for each group or each dioxin category in the unadjusted model. The percentile estimates are used to determine parameter estimates from the Weibull distribution. The Weibull distribution parameter estimates are then used in an iterative nonlinear estimation procedure (SAS® PROC NLIN [14]) to produce estimated means from a censored Weibull distribution. The loss function that is minimized in the estimation procedure is

\[
Loss = -\log\left( x \cdot \frac{\beta}{\theta^\beta} \cdot y^{\beta-1} \cdot e^{-\left(\frac{y}{\theta^{\beta}}\right)^\beta} \right) + (1-x) \cdot \left(1 - e^{-\left(\frac{y}{\theta^{\beta}}\right)^\beta}\right)
\]

where
- \( x = 1 \) if diabetic
- \( x = 0 \) if not diabetic

and
- \( y \) = time to onset of diabetes.
Table 7-6. Summary of Statistical Procedures (Continued)

**General Linear Models Analysis**

The form of the general linear model (18) for two independent variables is

\[ Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon \]

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 (e.g., group status and age)
- \( \beta_1, \beta_2 \) = coefficient indicating linear association between \( Y \) and \( X_1 \) and \( Y \) and \( X_2 \), respectively; each coefficient reflects the effect on the model of the corresponding independent variable adjusted for the effect of the other independent variable
- \( \epsilon \) = 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 more than two independent variables is immediate. Simple linear regression, multiple linear regression, analysis of variance, analysis of covariance, and repeated measures analysis of variance are all examples of general linear models analysis.

**Logistic Regression Analysis**

The logistic regression model (20) enables a dichotomous dependent variable to be modeled in a regression framework with continuous and discrete independent variables. For two risk factors, such as dioxin and age, the logistic regression model is

\[ \logit P = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon \]

where

- \( P \) = probability of disease for an individual with risk factors \( X_1 \) and \( X_2 \)
- \( \logit P \) = \( \ln (P/(1-P)) \) (i.e., the log odds for disease)
- \( X_1 \) = first risk factor (e.g., dioxin)
- \( X_2 \) = second risk factor (e.g., age).

The parameters are interpreted as follows:

- \( \alpha \) = log odds for the disease when \( X_1 = 0 \) and \( X_2 = 0 \)
- \( \beta_1 \) = coefficient indicating the dioxin effect adjusted for age
- \( \beta_2 \) = coefficient indicating the age effect adjusted for dioxin
- \( \epsilon \) = error term.

For a dichotomous measure, the term \( \exp(\beta_1) \) equals the adjusted odds ratio of abnormal versus normal for Ranch Hands (\( X_1 = 1 \)) compared to Comparisons (\( X_1 = 0 \)). If the probability of being abnormal is small compared to being normal for both the Ranch Hand and Comparison groups, the odds ratio is approximately equal to the relative risk of being abnormal between the two groups. If \( X_1 \) is a continuous covariate, \( \exp(\beta_1) \) represents the adjusted odds ratio of outcome 1 versus outcome 0 for a unit increase in \( X_1 \). If the risk factor is expressed in logarithmic (base 2) form, \( \exp(\beta_1) \) reflects the adjusted odds ratio for a twofold increase in the risk factor. Throughout this report and previous reports, the adjusted odds ratio was referred to as an adjusted relative risk. Correspondingly, in the absence of covariates (i.e., unadjusted analysis), the unadjusted odds ratio was referred to as an estimated relative risk.

This technique also was used for longitudinal analyses of dichotomous dependent variables to examine changes in health status between 1982 (or 1985) and 1997 in relation to the dioxin measures.
Polytomous Logistic Regression Analysis

Polytomous logistic regression (20, 21) allows a categorical dependent variable with more than two outcomes to be modeled in a regression environment with continuous and discrete independent variables. For polytomous logistic regression, the model equation depends on the scale of the dependent variable. This discussion focuses on nominal scaled dependent variables.

Suppose Y is a nominal scaled dependent variable with three outcomes labeled 0, 1, or 2 (normal, low, or high). Polytomous logistic regression models two logit functions, one for \( Y = 1 \) versus \( Y = 0 \) and the other for \( Y = 2 \) versus \( Y = 0 \). The zero outcome for \( Y \) is called the reference category. To model \( Y \) with two covariates such as group status and age, the polytomous regression model would be

\[
\logit P_i = \alpha_i + \beta_{i(1)} X_1 + \beta_{i(2)} X_2 + \varepsilon_i
\]

where

- \( P_i \) = probability that \( Y = i \) (outcome i) with covariates \( X_1 \) and \( X_2 \); \( i = 0, 1, 2 \)
- \( \logit P_i \) = \( \ln \left( \frac{P_i}{1-P_i} \right) \) (i.e., the log odds of outcome i versus outcome 0; \( i = 1, 2 \))
- \( X_1 \) = first effect (e.g., group status)
- \( X_2 \) = second effect (e.g., age).

The parameters are interpreted as follows:

- \( \alpha_i \) = log odds of outcome i versus outcome 0 when \( X_1 = 0 \) and \( X_2 = 0 \); \( i = 1, 2 \)
- \( \beta_{i(1)} \) = coefficient indicating the group status effect on the logit \( P_i \), adjusted for age; \( i = 1, 2 \)
- \( \beta_{i(2)} \) = coefficient indicating the age effect on the logit \( P_i \), adjusted for group status; \( i = 1, 2 \)
- \( \varepsilon_i \) = error term for logit \( P_i \); \( i = 1, 2 \).

This model assumes independent multinomial sampling.

Because the interpretation of each logistic modeling function is similar, consider the logit \( P_1 \) and suppose \( X_1 \) is a binary covariate (\( X_1 = 1 \) for Ranch Hands or \( X_1 = 0 \) for Comparisons). The term \( \exp(\beta_{1(1)}) \) equals the adjusted odds ratio of low versus normal for Ranch Hands (\( X_1 = 1 \)) compared to Comparisons (\( X_1 = 0 \)). If the probability of being low is small compared to being normal for both the Ranch Hand and Comparison groups, the odds ratio of low versus normal is approximately equal to the relative risk of being low between the two groups. If \( X_1 \) is a continuous covariate, \( \exp(\beta_{1(1)}) \) represents the adjusted odds ratio of outcome 1 versus outcome 0 for a unit increase in \( X_1 \).

The abnormal and normal categorizations for many of the discrete analyses were defined by categorizing laboratory and physical examination measures according to laboratory and clinic reference values. Cutpoints for the dependent variables erythrocyte sedimentation rate, cholesterol, and total testosterone were age-dependent. Consequently, normal and abnormal levels were constructed according to a participant’s laboratory value and age at the physical examination.
7.4.2 Modeling Strategy

In general, based on one of the adjusted analysis models described in Table 7-5, a model for dependent variables was based on the exposure effect (group or dioxin) and medically relevant covariates, as identified in Chapters 9 through 18 for each clinical category. As described previously, body fat at the time of the blood measurement of dioxin was included in Models 2 and 3.

The general modeling strategy did not remove any covariates from the model; however, the modeling strategy for the adjusted analysis of dependent variables in certain clinical areas was modified as necessary because of the large number of covariates or sparse number of participants with abnormal measurements. Stepwise elimination of covariates was conducted to allow for proper estimation of model parameters. When this strategy of removing covariates was necessary, the covariates removed from (or retained in) a model for a given health endpoint and model were specified in footnotes to the tables.

7.4.3 Longitudinal Analysis

Selected longitudinal analyses were performed to investigate changes in health status between 1982 and 1997 for Models 1, 2, and 3 as a function of dioxin exposure. Model 4 was not examined in longitudinal analyses because lipid-adjusted dioxin, the estimate of exposure in this model, changes over time and was not available for all participants in 1982 or 1997. All three models were adjusted for age at the time of the 1997 follow-up physical examination. Age was a well-known risk factor for nearly all clinical areas, and although Ranch Hands and Comparisons were matched on age, the estimates of dioxin exposure in Models 2 and 3 were not.

In the longitudinal analysis of discrete variables, only those participants whose health was classified as normal in 1982 were included in the analysis of the participants' health at the 1997 follow-up examination. Participants classified as "abnormal" in 1982 were excluded because the focus of the analysis was to investigate the temporal effects of dioxin exposure between 1982 and 1997. Participants classified as "abnormal" in 1982 were already abnormal before this period; consequently, only participants classified as "normal" at the 1982 examination were considered to be at risk when the effects of dioxin over time were explored. The rate of abnormalities under this restriction approximated the cumulative incidence rate between 1982 and 1997 (22).

The dependent variable in this type of analysis was the health of participants at the 1997 examination whose health was normal in 1982. The independent variables were the appropriate exposure estimate and age at the time of the 1997 follow-up physical examination. The analyses of Models 2 and 3 also were adjusted for body fat at the time of the blood measurement of dioxin. Tabular displays of the longitudinal analysis results of discrete dependent variables include summary statistics for 1982 and 1997, as well as 1985, 1987, and 1992 summaries, if available. The results of the statistical analysis restricted to those participants who were normal in 1982 also were provided.
In the longitudinal analysis of continuous variables, a general linear model approach, as explained in Table 7-6, was used. The dependent variable was the difference between the 1997 measurement and the 1982 measurement. This difference, measuring the change in the endpoint over this period of time, was modeled as a function of the estimate of exposure (group or dioxin), the participant’s age at the time of the 1997 follow-up physical examination, and the 1982 measurement of the continuous dependent variable. The analyses of Models 2 and 3 also were adjusted for body fat at the time of the blood measurement of dioxin. The reasons for using the health endpoint measurement in 1982 for longitudinal analysis of continuous variables were as follows:

- A linear relation between measurements of the dependent variable in 1982 and 1997 because of a difference in measuring devices was accounted for by using the 1982 measurement as an independent variable.
- The difference between two measurements taken over a period of time was generally correlated with the first measurement (23).
- The relation between the difference of the 1997 and 1982 measurements and the estimate of exposure may be confounded with the 1982 measurement, especially if the endpoint and the estimate of exposure were related.

Tabular displays of the results of longitudinal analysis of continuous dependent variables include summary statistics for 1982 and 1997, as well as 1985, 1987, and 1992 summaries, if available. Results of the statistical analysis relating the difference in the 1997 and 1982 measurements to the estimate of exposure also were provided. For some variables, 1985 clinical measurements were substituted for 1982 measurements because the variable was not analyzed at the 1982 examination or was inherently different from the 1997 variable due to differing clinical methods.

7.5 INTERPRETIVE CONSIDERATIONS

Several specific issues to consider when interpreting the results found in this report are discussed in this section. The issues discussed here include adjustments for covariates, multiple testing, trends in the results of endpoints within a clinical area, the proportion of variation explained by the model ($R^2$), interpretation of discrete and continuous analyses of a health endpoint, and statistical power to detect the effects of dioxin.

7.5.1 Adjustments for Covariates

In contrasts between all Ranch Hands and all Comparisons (Model 1), the matching variables age, race, and occupation were effectively eliminated as confounders. The initial and 1987 dioxin analyses within Ranch Hands (Models 2 and 4) and the categorized dioxin analysis within Ranch Hands and Comparisons (Model 3) did not benefit from the matched design. For example, military occupation was a strong confounder because it is highly correlated with dioxin levels in Ranch Hands and is related to some health variables through socioeconomic differences between officers and enlisted personnel. Education was highly associated with military occupation and certain psychometric results. Consequently, with the exception of a few analyses where the prevalence or history of abnormal results was sparse, all health endpoints were analyzed with and without adjustment for clinically relevant covariates.
7.5.2 Multiple Testing

Numerous dependent variables were considered because of the lack of a predefined medical endpoint. Each dependent variable was analyzed in many different ways to accommodate covariate information and different statistical models. Under the hypothesis of no relation between physical health and dioxin, approximately 5 percent of the many statistical tests (group or dioxin effects) in this report detected an association between group or dioxin and health \( (p \text{-values} \leq 0.05) \). Observing significant results because of multiple testing, even when there is no relation between dioxin and health, is known as the multiple-comparisons problem \((24)\) and is common in all large studies with multiple endpoints. It is generally difficult to distinguish between those statistically significant results that arise because of the multiple testing artifact and those that may be due to an actual dioxin effect. In order to weigh and interpret the findings, the strength of the association, consistency, dose-response patterns, and biologic plausibility were considered.

7.5.3 Trends

Assessing consistent and meaningful trends is essential when interpreting any comprehensive study with multiple endpoints, clinical areas, and covariates; however, caution must be used. Increased numbers of abnormalities or mean values with increased dioxin levels across medically related variables within a clinical area might indicate a group or dioxin effect. There may, however, be a moderate-to-strong correlation between these endpoints, where a change in one variable leads directly to a change in the other. Hence, the strength of the trends also was considered when assessing the suspected association.

7.5.4 Interpretation of the Coefficient of Determination

The coefficient of determination \((R^2)\) measures the proportionate reduction of the total variation in a continuously distributed health variable, \(y\), associated with the set of independent variables in a linear regression. A large value of \(R^2\) does not necessarily imply that the fitted model is a useful one. Large values of \(R^2\) would occur, for example, if \(y\) is regressed on an independent variable with only a few observed values. On the other hand, small values of \(R^2\) are generally seen in observational studies because little or no control has been applied in the assignment of the values of the “treatment” (dioxin) or the conditions under which the “treatment” has been applied. In this study, the dioxin measurements were taken many years after exposure and are subject to some measurement error. Thus, in most analyses, the values of \(R^2\) were small.

7.5.5 Clinical Interpretation of Discrete versus Continuous Data

Small but significant mean differences in a continuously measured health variable (e.g., alkaline phosphatase) between exposed and unexposed groups when there are no corresponding differences in the percentage of abnormal tests are difficult to interpret in any study. In this study, significant differences in the means between exposed and unexposed groups sometimes are observed without a corresponding difference between the groups in the percentage of participants with an abnormal measurement. Such contrasting situations may be interpreted as spurious outcomes of no clinical consequence, or as a subclinical dioxin effect. Significant trends in the mean with increasing levels of dioxin were interpreted as a dioxin-related effect if a corresponding trend was seen in the proportion above or below the normal range or if the trend was consistent with other findings.
7.5.6 Power

A type I error is making a false conclusion that an association (group or dioxin effect) exists when there is no association. The other possible inference error, a type II error, is the failure to detect an association when one actually exists. The power of a statistical test is 1 minus the probability of a type II error. The power of the test is the probability that the test will reject the hypothesis of no group or dioxin effect when an effect does in fact exist.

The fixed size of the Ranch Hand cohort limits the ability of this study to detect some group or dioxin associations if they exist. This limitation is most obvious for specific types of cancer, such as soft tissue sarcoma (STS) and non-Hodgkin’s lymphoma (NHL). These conditions are so uncommon that fewer than two cases are expected in this study, indicating that there is virtually no statistical power to detect low-to-moderate associations between dioxin and cancer. In an attempt to overcome the lack of power to detect group differences for specific types of systemic cancer, for example, all types of systemic cancer were combined into a single variable. It is still possible, however, that an increased risk could exist for a particularly rare type of cancer, allowing that increased risk to be missed in this study.

Table 7-7 and Appendix Tables E-1 through E-3 contain the approximate power at a significance level of 0.05 to detect specified relative risks for a given prevalence rate of a discrete dependent variable. Table 7-7 presents power calculations for Model 1 (group), and Appendix Tables E-1 through E-3 present power calculations for Model 2 (initial dioxin), Model 3 (categorized dioxin—low plus high Ranch Hand versus Comparison contrast), and Model 4 (lipid-adjusted 1987 dioxin). Power calculations were performed using the logarithm (base 2) of dioxin in Models 2 and 4, and consequently, the relative risk is for a twofold increase in dioxin. The power of a test for a discrete variable depends on the significance level, actual relative risk, prevalence of the condition, and the Ranch Hand and Comparison sample sizes (for Models 1 and 3) or the distribution of the dioxin data (for Models 2 and 4).

As an example, using age-adjusted incidence rates for all U.S. males (based on data from the Surveillance Epidemiology and End Results program of the National Cancer Institute), prevalence rates for all cancers, NHL, and STS were estimated as 0.07, 0.002, and 0.001, respectively. Thus, Table 7-7 shows a power less than 0.21 to detect a relative risk of 2.0 for the estimated prevalences of NHL and STS. For a disease with a prevalence of 0.05, the power to detect a relative risk of 1.5 would be 0.54.

### Table 7-7. Approximate Power To Detect a Group Effect at a 5-Percent Level of Significance (Discrete Dependent Variable)

<table>
<thead>
<tr>
<th>Prevalence of Condition</th>
<th>1.10</th>
<th>1.20</th>
<th>1.30</th>
<th>1.40</th>
<th>1.50</th>
<th>1.75</th>
<th>2.00</th>
<th>10.00</th>
<th>20.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005</td>
<td>0.05</td>
<td>0.06</td>
<td>0.07</td>
<td>0.09</td>
<td>0.10</td>
<td>0.15</td>
<td>0.21</td>
<td>0.92</td>
<td>0.97</td>
</tr>
<tr>
<td>0.01</td>
<td>0.06</td>
<td>0.07</td>
<td>0.09</td>
<td>0.12</td>
<td>0.16</td>
<td>0.26</td>
<td>0.36</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.02</td>
<td>0.06</td>
<td>0.09</td>
<td>0.14</td>
<td>0.19</td>
<td>0.26</td>
<td>0.45</td>
<td>0.62</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.03</td>
<td>0.07</td>
<td>0.11</td>
<td>0.18</td>
<td>0.26</td>
<td>0.36</td>
<td>0.60</td>
<td>0.79</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.04</td>
<td>0.07</td>
<td>0.13</td>
<td>0.22</td>
<td>0.33</td>
<td>0.45</td>
<td>0.72</td>
<td>0.89</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.05</td>
<td>0.08</td>
<td>0.15</td>
<td>0.26</td>
<td>0.40</td>
<td>0.54</td>
<td>0.81</td>
<td>0.94</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.10</td>
<td>0.10</td>
<td>0.24</td>
<td>0.44</td>
<td>0.64</td>
<td>0.80</td>
<td>0.97</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.15</td>
<td>0.12</td>
<td>0.32</td>
<td>0.58</td>
<td>0.79</td>
<td>0.92</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.20</td>
<td>0.14</td>
<td>0.38</td>
<td>0.67</td>
<td>0.87</td>
<td>0.96</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
The power calculations are defined in terms of the coefficient of variation (100 times the standard deviation of the dependent variable divided by the mean of the dependent variable) and the proportion mean change. The coefficient of variation relates the spread of the data relative to the magnitude of the data. In general, the power of a test is greater when the coefficient of variation is smaller. Table 7-8 presents power calculations for Model 1 (group), and Appendix Tables E-4 through E-6 present power calculations for Model 2 (initial dioxin), Model 3 (categorized dioxin—low plus high Ranch Hand versus Comparison contrast) and Model 4 (lipid-adjusted 1987 dioxin). Power calculations were performed using the logarithm (base 2) of dioxin in Models 2 and 4, and consequently, the relative risk is for a twofold increase in dioxin. The power of a test for a continuous variable depends on the significance level, actual difference in the true dependent variable means or slope of the dioxin coefficient, variation in the dependent variable data, sample size, and the distribution of the dioxin data if dioxin is the exposure estimate.

The proportion mean change in Table 7-8 and Appendix Table E-5 is defined as the difference in the true Ranch Hand and Comparison means, relative to the combined average of the two groups, assuming no transformation of the dependent variable. The proportion mean change in Appendix Tables E-4 and E-6 is defined as the change in the expected value (mean) of the dependent variable for a twofold increase in initial dioxin, relative to the dependent variable mean. The proportion mean change in Appendix Tables E-4 and E-6 corresponds mathematically to the slope of initial or 1987 dioxin divided by the dependent variable mean, assuming no transformation of the dependent variable. Analogous quantities can be derived based on transformed statistics. As an example, white blood cell count (on the natural logarithm scale) for all participants has a coefficient of variation of approximately 15 percent. With this coefficient of variation, for the 870 Ranch Hands and 1,251 Comparisons in Model 1, the power is approximately 0.86 for detecting a 2-percent increase in the mean white blood cell count of Ranch Hands relative to the mean white blood cell count of Comparisons (mean change = 0.02).

<table>
<thead>
<tr>
<th>Mean Change</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>25</th>
<th>50</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005</td>
<td>0.62</td>
<td>0.21</td>
<td>0.12</td>
<td>0.07</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>0.01</td>
<td>0.99</td>
<td>0.62</td>
<td>0.33</td>
<td>0.15</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td>0.02</td>
<td>1.00</td>
<td>0.99</td>
<td>0.86</td>
<td>0.44</td>
<td>0.15</td>
<td>0.09</td>
</tr>
<tr>
<td>0.03</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.78</td>
<td>0.27</td>
<td>0.15</td>
</tr>
<tr>
<td>0.04</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.95</td>
<td>0.44</td>
<td>0.23</td>
</tr>
<tr>
<td>0.05</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.62</td>
<td>0.33</td>
</tr>
<tr>
<td>0.10</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.86</td>
</tr>
</tbody>
</table>
In summary, this study has good power to detect relative risks of 2.0 or more with respect to diseases, such as heart disease and basal cell carcinoma, occurring at a prevalence of at least 5 percent in unexposed populations. In addition, the study size is sufficient to detect small mean shifts in the continuously distributed variables. The detection of significant mean shifts without a corresponding indication of increased Ranch Hand abnormalities or disease may be an artifact of multiple testing, could represent a subclinical effect, or could be of little or no medical importance.

7.6 EXPLANATION OF TABLES

This section explains the contents of the tables used to report the results of the analyses for continuous and discrete dependent variables (two levels and more than two levels). Selected tables from the General Health Assessment (Chapter 9) and the Hematology Assessment (Chapter 15) will be referenced throughout this discussion. The contents of each table depend on the form of the health status endpoint (i.e., whether the dependent variable under analysis is a continuous or discrete variable). A discussion of the contents of exposure analysis tables is discussed first, followed by an explanation of the longitudinal analysis tables.

7.6.1 Exposure Analysis

The results of the exposure analysis are displayed in subpanels within each table as specified in Table 7-9. The specification of the subpanels is applicable whether the dependent variable is continuous or discrete.

<table>
<thead>
<tr>
<th>Model</th>
<th>Exposure Estimate</th>
<th>Subpanel in Table</th>
<th>Type of Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Group</td>
<td>a</td>
<td>Unadjusted</td>
</tr>
<tr>
<td>2</td>
<td>Initial Dioxin</td>
<td>b</td>
<td>Adjusted</td>
</tr>
<tr>
<td>3</td>
<td>Categorized Dioxin</td>
<td>c</td>
<td>Unadjusted</td>
</tr>
<tr>
<td>4</td>
<td>1987 Dioxin</td>
<td>d</td>
<td>Adjusted</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Exposure Estimate</th>
<th>Subpanel in Table</th>
<th>Type of Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Group</td>
<td>e</td>
<td>Unadjusted</td>
</tr>
<tr>
<td>2</td>
<td>Initial Dioxin</td>
<td>f</td>
<td>Adjusted</td>
</tr>
<tr>
<td>3</td>
<td>Categorized Dioxin</td>
<td>g</td>
<td>Unadjusted</td>
</tr>
<tr>
<td>4</td>
<td>1987 Dioxin</td>
<td>h</td>
<td>Adjusted</td>
</tr>
</tbody>
</table>

*Ranch Hands and Comparisons.
*Ranch Hands only.

7.6.1.1 Continuous Variables

Table 9-8 in the General Health Assessment chapter presents an example of the results of the analysis when the dependent variable was continuous. Subpanels (a) and (b) show the results of unadjusted and adjusted Model 1 analyses that compared the Ranch Hand and Comparison means of a dependent variable. Contrasts between Ranch Hands and Comparisons also are presented within each occupational category (i.e., officer, enlisted flyer, and enlisted groundcrew).
For the unadjusted analysis in subpanel (a), a sample size (n) and a mean are presented for all occupational categories combined and separately for each occupational category. If the dependent variable was transformed for the analysis, the means of the transformed values were converted to the original scale and the column heading is footnoted. For each contrast of Ranch Hands versus Comparisons, the difference of means on the original scale and the associated 95-percent confidence interval are reported. The 95-percent confidence interval was constructed by adding and subtracting 1.96 multiplied by the standard error (for the upper and lower bounds, respectively) to the estimated mean. If the analysis was performed on a transformed scale, the 95-percent confidence interval on the differences of means is not presented and the column is footnoted. When presenting results from analyses of means based on log-transformed (or square root-transformed) data, means were converted back to original units. Conversion of the standard deviation from log units to original units is not recommended (25); therefore, confidence intervals for mean differences in original units are not presented. A p-value also is reported to determine whether a difference in means on the scale used for analysis for a specified contrast was equal to zero. The confidence interval and p-value for each occupational category were determined using analysis of variance techniques from a group-by-occupation interaction in the model. The group-by-occupation interaction was used to determine the model coefficients and standard errors simultaneously for officers, enlisted flyers, and enlisted groundcrew. The respective coefficients and standard errors from the group and group-by-interaction terms in the model, along with the covariances between the estimates, were combined as appropriate to construct the confidence intervals and p-values for the three occupational strata.

For an adjusted Model 1 analysis, subpanel (b) includes a sample size, an adjusted mean, a difference of Ranch Hand and Comparison adjusted means on the original scale, the associated 95-percent confidence interval (if the analysis was performed on the original scale), and a p-value for each contrast. Sample sizes for corresponding panels of unadjusted and adjusted analyses may differ because of missing covariate information. The confidence interval and p-value for each occupational category were determined using analysis of covariance techniques from a group-by-occupation interaction in the model.

Subpanel (c) of Table 9-8 reports summary statistics from the analysis that assessed the association between the continuous dependent variable and initial dioxin (Model 2) without adjusting for covariate information. The sample size and mean of the dependent variable (transformed to the original units, if necessary) are presented for low, medium, and high categories of initial dioxin. The low, medium, and high categories were determined by dividing all Ranch Hands with initial dioxin estimates into three approximately equal-sized categories based on their initial dioxin estimate. The numerical values defining these categories are specified in a table subpanel footnote. Means of the dependent variable, adjusted for percent body fat at the time of the blood measurement of dioxin, also are presented for the low, medium, and high categories of initial dioxin. Based on a linear regression analysis, adjusted for percent body fat at the time of the blood measurement of dioxin, the coefficient of determination (R^2), the estimated slope, and its associated standard error are reported. If the dependent variable was transformed for the regression analysis, the transformation is identified in the footnote. The p-value associated with testing whether the slope was equal to zero also is presented. The summary statistics that are reported were based on initial dioxin divided into three categories, whereas the R^2, slope, standard error, and p-value were based on log2 (initial dioxin) in its continuous form.

Based on analyses that incorporate covariate information, subpanel (d) reports summary statistics from the analysis that assessed the association between the continuous dependent variable and initial dioxin (Model 2). Similar to the unadjusted analysis, a sample size and adjusted mean of the dependent variable (transformed to the original units, if necessary) are presented for low, medium, and high categories of initial dioxin. The numerical values defining these categories are specified in a table subpanel footnote.
Sample sizes for corresponding panels of unadjusted and adjusted analyses may differ because of missing covariate information. Based on the multiple linear regression of the dependent variable on log2 (initial dioxin) and covariate effects, including percent body fat at the time of the blood measurement of dioxin, the coefficient of determination \( R^2 \), the adjusted slope for log2 (initial dioxin), and its associated standard error are reported. If the dependent variable was transformed for the regression analysis, the adjusted means, adjusted slope, and standard error are footnoted and the transformation is identified in the footnote. The p-value for testing whether the slope was equal to zero also is presented.

Subpanels (e) and (f) of Table 9-8 show the results of unadjusted and adjusted Model 3 analyses that contrasted the means of a continuous dependent variable for Ranch Hands with background, low, high, and low plus high dioxin levels with Comparisons having lipid-adjusted dioxin levels less than or equal to 10 ppt. The low and high Ranch Hand categories were determined by dividing all Ranch Hands with lipid-adjusted dioxin estimates greater than 10 ppt into two approximately equal-sized categories based on their initial dioxin estimate. The low plus high Ranch Hand category is a combination of the low and high categories. The note at the bottom of the table subpanels defines the dioxin categories. The mean for the low plus high category is a weighted average (transformed to the original units, if necessary) of the low Ranch Hand and high Ranch Hand categories' means on the scale used for transformation, where the weights were based on the low and high Ranch Hand categories' sample sizes. Sample sizes for corresponding panels of unadjusted and adjusted analyses may differ because of missing covariate information.

For the unadjusted analysis in subpanel (e), a sample size and dependent variable mean are presented for each category. If the dependent variable was transformed for the analysis, the means of the transformed values were converted to the original scale and the column heading is footnoted. The mean of the dependent variable adjusted for percent body fat at the time of the blood measurement of dioxin also is presented for each dioxin category. For each individual contrast of the Ranch Hand category versus the Comparison category, the difference of means on the original scale and the associated 95-percent confidence interval are reported. If the analysis was performed on a transformed scale, the 95-percent confidence interval on the differences of means is not presented and the column is footnoted. A p-value also is reported to determine whether a difference in means for a specified contrast was equal to zero. The p-value was based on the difference of means on the scale used for analysis. The adjusted mean, confidence interval, and p-value for each contrast was determined from an analysis of covariance model with adjustment for percent body fat at the time of the blood measurement of dioxin.

For the adjusted analysis in subpanel (f), the table includes a sample size, an adjusted mean (adjusted for percent body fat at the time of the blood measurement of dioxin and covariates), a difference in adjusted means on the original scale, and a 95-percent confidence interval on the difference in adjusted means (if the analysis was performed on the original scale). The p-value for testing whether the difference in adjusted means for a specified contrast was equal to zero also is presented.

Subpanel (g) of Table 9-8 reports summary statistics from Model 4 analyses, which assessed the association between the continuous dependent variable and 1987 dioxin without adjusting for covariate information. The sample size and mean of the dependent variable (transformed to the original units, if necessary) are presented for low, medium, and high categories of 1987 dioxin. The low, medium, and high categories were determined by dividing all Ranch Hands with 1987 dioxin levels into three approximately equal-sized categories based on their 1987 dioxin measurement. The numerical values defining the low, medium, and high categories of 1987 dioxin are specified in a table subpanel footnote. Based on a linear regression of the dependent variable on log2 (1987 dioxin + 1), the coefficient of determination \( R^2 \), the estimated slope, and its associated standard error are reported for each model. A
value of 1 was added to each measurement because of the presence of 1987 dioxin measurements of 0 ppt. If the dependent variable was transformed for the regression analysis, the means, slope, and standard error are footnoted and the transformation is identified in the footnote. The p-value associated with testing whether the slope was equal to zero also is presented.

Based on analyses that incorporate covariate information, subpanel (h) reports summary statistics for Model 4 analyses that assessed the association between the continuous dependent variable and 1987 dioxin. The sample size and adjusted mean of the dependent variable (transformed to the original units, if necessary) are presented for low, medium, and high categories of 1987 dioxin. The numerical values defining these categories are specified in a table subpanel footnote. Sample sizes for corresponding panels of unadjusted and adjusted analyses may differ because of missing covariate information. Based on the multiple linear regression of the dependent variable on log₂ (1987 dioxin + 1) and covariates, the coefficient of determination (R²), the adjusted slope for log₂ (1987 dioxin + 1), and its associated standard error are reported for each model. If the dependent variable was transformed for the regression analysis, the adjusted means, adjusted slope, and standard error are footnoted and the transformation is identified in the footnote. The p-value for testing whether the slope was equal to zero also is presented.

7.6.1.2 Discrete Variables

7.6.1.2.1 Discrete Variable with Two Categories

Table 9-3 in the General Health Assessment chapter presents an example of the results of analysis when the dependent variable is discrete and dichotomous. Subpanels (a) and (b) display the results of unadjusted and adjusted Model 1 analyses that compared the percentage of Ranch Hands and Comparisons that were considered abnormal for the dependent variable of interest (the abnormal classification for self-perception of health in Table 9-3 is “fair or poor”). Contrasts between Ranch Hands and Comparisons also are presented within each occupational category (i.e., officer, enlisted flyer, and enlisted groundcrew). For the unadjusted analysis in subpanel (a), a sample size and the number and percentage of participants considered abnormal are presented for each group within each occupational category. For the contrasts of Ranch Hands versus Comparisons, an estimated relative risk, an associated 95-percent confidence interval on the relative risk, and a p-value for testing whether the risk was equal to 1.0 are presented. The normal distribution was used to calculate an approximate 95-percent confidence interval. Results for each occupational category were determined from a group-by-occupation interaction that was included in the model.

For the adjusted analysis of Model 1, as presented in subpanel (b), the table presents an adjusted relative risk, a 95-percent confidence interval on the relative risk, and a p-value for testing whether the risk was equal to 1.0. The adjusted relative risk, confidence interval, and p-value were determined from a multiple logistic regression model that used the appropriate covariates for the clinical area and dependent variable of interest. Results for each occupational category were determined from a group-by-occupation interaction that was included in the model.

Subpanel (c) of Table 9-3 reports summary statistics for analyses that assessed the association between the dependent variable and initial dioxin (Model 2) without adjusting for covariate information. Sample sizes are presented for low, medium, and high categories of initial dioxin. The numerical values defining these categories are specified in a table footnote. The number and percentage of Ranch Hands considered abnormal are presented for the low, medium, and high initial dioxin categories. Based on a logistic regression model, adjusted for percent body fat at the time of the blood measurement of dioxin, an estimated relative risk and its 95-percent confidence interval are reported. The p-value associated
with testing whether the relative risk was equal to 1.0 also is presented. The normal distribution was used to determine an approximate 95-percent confidence interval. The summary statistics that are reported were based on initial dioxin divided into three categories, whereas the relative risk, confidence interval, and p-value were based on log₂ (initial dioxin) in its continuous form.

Subpanel (d) of Table 9-3 reports summary statistics for analyses that assessed the association between the discrete dependent variable and initial dioxin (Model 2), adjusted for percent body fat at the time of the blood measurement of dioxin and covariate information. The sample size given is based on a multiple logistic regression of the discrete dependent variable on log₂ (initial dioxin), percent body fat at the time of the blood measurement of dioxin, and covariates. Sample sizes for corresponding panels of unadjusted and adjusted analyses may differ because of missing covariate information. The adjusted relative risk for log₂ (initial dioxin) and its associated 95-percent confidence interval are reported and are based on this multiple logistic regression model. The normal distribution was used to determine an approximate 95-percent confidence interval. The p-value for testing whether the relative risk was equal to 1.0 also is presented.

Subpanels (e) and (f) of Table 9-3 show the results of unadjusted and adjusted Model 3 analyses that contrasted Ranch Hands having background, low, high, and low plus high dioxin levels with Comparisons having lipid-adjusted dioxin levels less than or equal to 10 ppt. The percentage of participants that were considered abnormal for the dependent variable of interest was contrasted between the four categories of Ranch Hands and Comparisons. The low and high Ranch Hand categories were determined by dividing all Ranch Hands with lipid-adjusted dioxin estimates greater than 10 ppt into two approximately equal-sized categories based on their initial dioxin estimate. The low plus high Ranch Hand category is a combination of the low and high Ranch Hand categories. The note at the bottom of the table subpanel defines the dioxin categories. The percentage of Ranch Hands in the low plus high category is a weighted average of the low Ranch Hand and high Ranch Hand categories, where the weights are based on the low category and high category sample sizes. Sample sizes for corresponding panels of unadjusted and adjusted analyses may differ because of missing covariate information.

For the Model 3 unadjusted analysis in subpanel (e), the sample size and the number and percentage of participants considered abnormal is presented for each dioxin category. For the individual contrasts of the Ranch Hand categories versus Comparisons, an estimated relative risk, a 95-percent confidence interval for the relative risk, and a p-value associated with testing whether the risk was equal to 1.0 are presented. The relative risk, confidence interval, and p-value were determined from a logistic regression model, adjusted for percent body fat at the time of the blood measurement of dioxin. The normal distribution was used to determine an approximate 95-percent confidence interval.

For the Model 3 adjusted analysis, subpanel (f) of the table presents an adjusted relative risk, a 95-percent confidence interval for the relative risk, and a p-value associated with testing whether the risk was equal to 1.0 for the individual contrasts of the Ranch Hand categories with Comparisons. The normal distribution was used to determine an approximate 95-percent confidence interval.

Subpanels (g) and (h) of Table 9-3 present summary statistics from Model 4, which assessed the association between the dependent variable and 1987 dioxin. For the unadjusted analysis, the sample size and the number and percentage of participants considered abnormal is presented for each 1987 dioxin category. The low, medium, and high categories were determined by dividing all Ranch Hands with 1987 dioxin levels into three approximately equal-sized categories. The numerical values defining these categories are specified in a table footnote. Based on a logistic regression model, an estimated relative risk and its 95-percent confidence interval are reported. The p-value associated with testing
whether the relative risk was equal to 1.0 also is presented. The normal distribution was used to
determine an approximate 95-percent confidence interval. The summary statistics are reported for 1987
dioxin divided into three categories, whereas the relative risk, confidence interval, and p-value were
based on \( \log_2(1987 \text{ dioxin} + 1) \) in its continuous form.

Incorporating covariate information, subpanel (h) reports summary statistics from analyses that assessed
the association between the dichotomous dependent variable and 1987 dioxin. The sample size is
presented for a multiple logistic regression of the discrete dependent variable on \( \log_2(1987 \text{ dioxin} + 1) \)
including covariates in the final adjusted model. Sample sizes for corresponding panels of unadjusted
and adjusted analyses may differ because of missing covariate information. Based on the multiple
logistic regression model, the adjusted relative risk for \( \log_2(1987 \text{ dioxin} + 1) \) and its associated 95-
percent confidence interval are reported. The normal distribution was used to determine an approximate
95-percent confidence interval. The p-value for testing whether the relative risk was equal to 1.0 also is
presented.

7.6.1.2.2 Discrete Variable with More Than Two Categories

Polytomous regression techniques were used to analyze discrete dependent variables having more than
two levels (e.g., abnormal low, normal, abnormal high—see Table 15-4 in the Hematology Assessment
chapter). Results were presented in a similar fashion to discrete variables with only two categories;
however, the number and percentage of participants for each dependent variable category (including
normal) are given. Therefore, the relative frequencies sum to 100 percent across the dependent variable
categories and the number of participants in each of the dependent variable categories adds to the total
number of participants in each exposure group or dioxin category. In addition, a relative risk, a 95-
percent confidence interval, and a p-value were presented for each contrast with the normal level of the
dependent variable (e.g., abnormal low versus normal and abnormal high versus normal).

In Table 15-4, subpanels (a) and (b) display the results of unadjusted and adjusted Model 1 analyses that
compared Ranch Hands and Comparisons on the relative frequencies of each abnormal level for a
specified discrete dependent variable. For example, the percentage of participants with an abnormally
high red blood cell count was contrasted to participants with a normal red blood cell count, and the
percentage of participants with an abnormally low red blood cell count was contrasted to participants
with a normal red blood cell count. Contrasts between Ranch Hands and Comparisons also are presented
within each occupational category (i.e., officer, enlisted flyer, and enlisted groundcrew). For the
unadjusted analysis in subpanel (a), a sample size is presented for each exposure group (Ranch Hand,
Comparison) across all occupational categories and within each occupational category. The number and
percentage of participants are presented for each level of the dependent variable for each group. For the
contrasts of Ranch Hands versus Comparisons, an estimated relative risk, a 95-percent confidence
interval for the relative risk, and a p-value associated with testing whether the risk was equal to 1.0 are
presented for each contrast against the normal level of the dependent variable (e.g., abnormal low versus
normal and abnormal high versus normal). The normal distribution was used to calculate an approximate
95-percent confidence interval. Results for each occupational category were determined from the group­
by-occupation interaction that was included in the model.

For a Model 1 analysis adjusted for covariate information and shown in subpanel (b), the table presents
an adjusted relative risk, a 95-percent confidence interval on the relative risk, and a p-value associated
with testing whether the risk was equal to 1.0 for each occupational category and each contrast. The
normal distribution was used to calculate an approximate 95-percent confidence interval. Results for
each occupational category were determined from the group-by-occupation interaction that was included in the model.

Subpanels (c) and (d) of Table 15-4 summarize the unadjusted and adjusted Model 2 analyses relating discrete dependent variables having more than two categories to initial dioxin. Both unadjusted and adjusted analyses are adjusted for percent body fat at the time of the blood measurement of dioxin. In subpanel (c), the sample size and the number and percentage of Ranch Hands in each category of the dependent variable are presented for each initial dioxin category (i.e., low, medium, and high initial dioxin). The relative risk, the 95-percent confidence interval for the relative risk, and the p-value associated with testing whether the risk was equal to 1.0 are presented for each abnormal level of the dependent variable (e.g., abnormal low versus normal and abnormal high versus normal). The summary statistics that are reported were based on initial dioxin divided into three categories, whereas the relative risk, confidence interval, and p-value were based on log₂ (initial dioxin) in its continuous form.

In subpanel (d), after adjustment for covariate information, the sample size, the adjusted relative risk, the 95-percent confidence interval for the relative risk, and the p-value associated with testing whether the risk was equal to 1.0 are presented for each abnormal level of the dependent variable. Sample sizes for corresponding panels of unadjusted and adjusted analyses may differ because of missing covariate information.

Subpanels (e) and (f) of Table 15-4 present unadjusted and adjusted Model 3 analyses of categorized dioxin versus a discrete dependent variable having more than two categories. Both unadjusted and adjusted analyses are adjusted for percent body fat at the time of the blood measurement of dioxin. Results are presented in a similar fashion to the group analysis (Model 1), except that contrasts involve the four Ranch Hand categories (background, low, high, and low plus high) versus Comparisons, and contrasts are not performed for each occupation. For the unadjusted analysis, a sample size is presented for each dioxin category. The low plus high Ranch Hand category is a combination of the low and high Ranch Hand categories. The percentage of Ranch Hands in the low plus high category is a weighted average of the low Ranch Hand and high Ranch Hand categories, where the weights are based on the low category and high category sample sizes. Sample sizes for corresponding panels of unadjusted and adjusted analyses may differ because of missing covariate information.

The number and percentage of participants for each level of the dependent variable are presented for each dioxin category in subpanel (e). For each contrast of a Ranch Hand category versus the Comparison group, an estimated relative risk, a 95-percent confidence interval for the relative risk, and a p-value associated with testing whether the risk was equal to 1.0 are presented. These results are given for each contrast against the normal level of the dependent variable (e.g., abnormal low versus normal and abnormal high versus normal). For an adjusted Model 3 analysis in subpanel (f), the table presents an adjusted relative risk, a 95-percent confidence interval on the relative risk, and a p-value for each contrast of Ranch Hands versus Comparisons for each abnormal level of the dependent variable.

Similar to the polytomous regression analysis using initial dioxin, unadjusted and adjusted analyses of discrete dependent variables with more than two categories were performed using 1987 dioxin in Model 4. In Table 15-4, summaries of the analyses are given in subpanels (g) and (h). For the unadjusted analysis in subpanel (g), sample sizes are presented for each 1987 dioxin category (i.e., low, medium, and high 1987 dioxin). The number and percentage of Ranch Hands for each dependent variable category for each 1987 dioxin category are presented. An estimated relative risk, a 95-percent confidence interval on the relative risk, and an associated contrast p-value are reported for each abnormal level of the dependent variable (e.g., abnormal low vs. normal and abnormal high vs. normal). The summary statistics that are
reported were based on 1987 dioxin divided into three categories, whereas the relative risk, confidence interval, and p-value were based on $\log_2 (1987 \text{ dioxin} + 1)$ in its continuous form.

Adjusted analysis results in subpanel (h) include a total sample size, an adjusted relative risk, a 95-percent confidence interval on the relative risk, and an associated contrast p-value for each abnormal level of the dependent variable. Sample sizes for corresponding panels of unadjusted and adjusted analyses may differ because of missing covariate information.

7.6.2 Longitudinal Analysis

The results of the longitudinal analysis are displayed in subpanels within each table as specified in Table 7-10. The specification of the subpanels is applicable whether the dependent variable is continuous or discrete.

<p>| Table 7-10. Location of Table Results from Different Longitudinal Analysis Models |
|--------------------------------------|-------------------------------|</p>
<table>
<thead>
<tr>
<th>Model</th>
<th>Exposure Estimate</th>
<th>Subpanel to Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Group&lt;sup&gt;a&lt;/sup&gt;</td>
<td>a</td>
</tr>
<tr>
<td>2</td>
<td>Initial Dioxin&lt;sup&gt;b&lt;/sup&gt;</td>
<td>b</td>
</tr>
<tr>
<td>3</td>
<td>Categorized Dioxin&lt;sup&gt;b&lt;/sup&gt;</td>
<td>c</td>
</tr>
</tbody>
</table>

<sup>a</sup>Ranch Hands and Comparisons.
<sup>b</sup>Ranch Hands only.

Most of the longitudinal analyses in this report are based on a comparison of data from the 1982 baseline examination and the 1997 follow-up examination, and the discussion of tables below is based on the comparison of the 1982 and 1997 examinations. Some analyses, however, are based on a comparison of data from the 1985 follow-up examination and the 1997 follow-up examination (e.g., neurological indices in Chapter 11, Neurological Assessment, or Doppler pulses in Chapter 14, Cardiovascular Assessment). The 1985 follow-up examination data were used because of methodological differences in the measurements between the 1982 baseline examination and the 1985 follow-up examination, or because the measurement was not obtained at the 1982 baseline examination. In addition, spirometry measurements were not taken at the 1985 follow-up examination, and Doppler pulse measurements were not made at the 1987 follow-up examination; therefore, summary statistics based on data from the respective examinations are not provided for these variables.

7.6.2.1 Continuous Variables

Table 9-15 in the General Health Assessment chapter presents an example of a longitudinal analysis when the dependent variable was continuous. In subpanel (a), a mean and a sample size (n) are provided for all occupational categories combined and separately for each occupational category (i.e., officer, enlisted flyer, and enlisted groundcrew). The mean and sample size are provided for data from the 1982 baseline examination and the 1985, 1987, 1992, and 1997 follow-up examinations. Summary statistics for the 1982 baseline examination and the 1997 follow-up examination were based on participants that attended both examinations. Summary statistics for the 1985, 1987, and 1992 follow-up examinations were based on participants that attended the 1982 baseline examination, the 1997 follow-up examination,
and the respective follow-up examination that was summarized. The summary statistics for the 1985, 1987, and 1992 follow-up examinations are provided for reference purposes. If the dependent variable was transformed for the analysis, the means of the transformed values were converted to the original scale and the transformation is specified in a footnote.

Subpanel (a) shows the Ranch Hand and Comparison difference in means between the 1997 follow-up examination and 1982 baseline examination. The Ranch Hand and Comparison difference in means between the 1997 follow-up examination and 1982 baseline examination is presented for all occupations combined and separately for each occupational category. The difference between Ranch Hands and Comparisons in the change between the 1997 follow-up examination mean and the 1982 baseline examination mean also is reported in subpanel (a). The p-value that was used to determine whether the difference in the examination mean change between Ranch Hands and Comparisons was equal to zero is given. This p-value was based on the difference in Ranch Hand and Comparison examination mean changes on the scale used for analysis. The p-value for each occupational category was determined using analysis of covariance techniques from a group-by-occupation interaction in the model. The longitudinal analysis performed in subpanel (a) was adjusted for the 1982 measurement of the dependent variable and age at the 1997 physical examination.

Subpanel (b) of Table 9-15 reports summary statistics on the continuous dependent variable of interest. The sample size and mean of the dependent variable (transformed to the original units, if necessary) are presented for low, medium, and high categories of initial dioxin. The low, medium, and high categories were determined by dividing all Ranch Hands with initial dioxin estimates into three approximately equal-sized categories based on their initial dioxin estimate. The numerical values defining these categories are specified in the table subpanel footnote. The mean and sample size are provided for data from the 1982 baseline examination and the 1985, 1987, 1992, and 1997 follow-up examinations. Summary statistics for the 1982 baseline examination and the 1997 follow-up examination were based on participants that attended both examinations. Summary statistics for the 1985, 1987, and 1992 follow-up examinations were based on participants that attended the 1982 baseline examination, the 1997 follow-up examination, and the respective follow-up examination that was summarized. If the dependent variable was transformed for the analysis, the transformation is specified in a footnote.

For each participant who attended both the 1982 and 1997 physical examinations, a difference between the dependent variable as measured at the 1997 follow-up examination and as measured at the 1982 baseline examination was created. The difference in these two measurements was on the scale used for analysis. The association between the difference in the examination measurements and initial dioxin was determined and adjusted for the 1982 measurement of the dependent variable, age at the 1997 physical examination, and percent body fat at the time of the blood measurement of dioxin. The estimated slope, its associated standard error, and the p-value associated with testing whether the slope was equal to zero are reported in subpanel (b). If the dependent variable was transformed for the regression analysis, the transformation is identified in the footnote. The summary statistics that are reported were based on initial dioxin divided into three categories, whereas the slope, standard error, and p-value were based on \( \log_2 \) (initial dioxin) in its continuous form.

Subpanel (c) of Table 9-15 shows the results of Model 3 analyses that contrasted the means of a continuous dependent variable for Ranch Hands with background, low, high, and low plus high dioxin levels with Comparisons having lipid-adjusted dioxin levels less than or equal to 10 ppt. The low and high Ranch Hand categories were determined by dividing all Ranch Hands with lipid-adjusted dioxin estimates greater than 10 ppt into two approximately equal-sized categories based on their initial dioxin estimate. The low plus high Ranch Hand category is a combination of the low and high categories. The
note at the bottom of the table subpanel defines the dioxin categories. The mean for the low plus high category is a weighted average (transformed to the original units, if necessary) of the low Ranch Hand and high Ranch Hand category means on the scale used for transformation, where the weights were based on the low and high Ranch Hand category sample sizes.

In subpanel (c), a mean and a sample size are provided for all Ranch Hand and Comparison dioxin categories. The mean and sample size are provided for data from the 1982 baseline examination and the 1985, 1987, 1992, and 1997 follow-up examinations. Summary statistics for the 1982 baseline examination and the 1997 follow-up examination were based on participants that attended the 1982 baseline examination and the 1997 follow-up examination. Summary statistics for the 1985, 1987, and 1992 follow-up examinations were based on participants that attended the 1982 baseline examination, the 1997 follow-up examination, and the respective follow-up examination that was summarized. The summary statistics for the 1985, 1987, and 1992 follow-up examinations are provided for reference purposes. If the dependent variable was transformed for the analysis, the means of the transformed values were converted to the original scale and the transformation is specified in a footnote.

Subpanel (c) shows the Ranch Hand and Comparison difference in dioxin category means between the 1997 follow-up examination and 1982 baseline examination. The Ranch Hand and Comparison difference in dioxin category means between the 1997 follow-up examination and 1982 baseline examination is presented for all occupations combined and separately for each occupational category. The difference between Ranch Hands and Comparisons in the change between the 1997 follow-up examination mean and the 1982 baseline examination mean also is reported in subpanel (c). The p-value that was used to determine whether the difference in the examination mean change between the Ranch Hand dioxin category and Comparisons was equal to zero is given. This p-value was based on the difference in Ranch Hand and Comparison examination mean changes on the scale used for analysis. The p-value for each occupational category was determined using analysis of covariance techniques. The longitudinal analysis performed in subpanel (c) was adjusted for the 1982 measurement of the dependent variable, age at the 1997 physical examination, and percent body fat at the time of the blood measurement of dioxin.

7.6.2.2 Discrete Variables with Two Categories

Table 9-10 in the General Health Assessment chapter presents an example of the longitudinal analysis when the dependent variable was discrete and dichotomous. In subpanel (a), the number and percentage of participants defined as abnormal and a sample size (n) are provided for all occupational categories combined and separately for each occupational category (i.e., officer, enlisted flyer, and enlisted groundcrew). The summary statistics are provided for data from the 1982 baseline examination and the 1985, 1987, 1992, and 1997 follow-up examinations. Summary statistics for the 1982 baseline examination and the 1997 follow-up examination were based on participants that attended both examinations. Summary statistics for the 1985, 1987, and 1992 follow-up examinations were based on participants that attended the 1982 baseline examination, the 1997 follow-up examination, and the respective follow-up examination that was summarized. The summary statistics for the 1985, 1987, and 1992 follow-up examinations are provided for reference purposes.

Subpanel (a) also shows the number of Ranch Hands and Comparisons and the number and percentage of participants considered abnormal at the 1997 examination (the abnormal classification for self-perception of health in Table 9-10 is "fair or poor"). These summary statistics are presented for all occupations combined and separately for each occupational category, and are restricted to participants that were considered normal in 1982 (the normal classification for self-perception of health in Table 9-10 is