

CHAPTER 7

STATISTICAL METHODS

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

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

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

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

STATISTICAL STUDY DESIGN

An overt herbicide effect would be characterized by more symptoms, signs, abnormal laboratory tests, syndromes, or diseases in the Ranch Hand group than in the Comparison group. If the disease(s) were fatal, increased mortality might also be observed. A subclinical herbicide effect would be detected as an increase in abnormal findings on the physical examination (particularly laboratory tests) that may or may not also be associated with symptom reporting or increased mortality. Thus, the basic objective of the statistical analysis is to test for differences between the Ranch Hand (exposed) group and the Comparison (nonexposed) group.

In general, two types of data are used in the analysis. First, there are subjective data on symptoms reported by the participant in the questionnaire and in the review-of-systems section of the physical examination. Second, there are objective data, which include medical findings or signs identified during the physical examination, or by reviews of laboratory results, medical records, and death certificates.

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

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

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

STATISTICAL ISSUES

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

Intervening Variables

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

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

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

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

Besides covariates, both confounding variables and interactions must also be considered. A confounding variable is an intervening variable associated with both disease and exposure. (This is in contrast with a covariate that is associated only with disease.) Adjustments must be made for confounding variables to avoid a biased estimate of the group-disease relationship. An interaction exists when the effect of one variable varies across the levels of another variable. For example, the group difference might be large in one occupation group and negligible in another. Incorporating interactions in the analysis allows for the identification of subpopulations at increased or decreased risk.

Power

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

Table 7-1 contains the approximate sample size required to detect specific relative risks with an approximate power of 0.8 ($\beta = 0.2$) using an alpha level of 0.05 for a two-sided test and assuming equal Ranch Hand and Comparison group sizes and unpaired analyses. Relative risk is the ratio of the disease prevalence rate of the Ranch Hand and Comparison groups. Conditions or diseases with comparison population prevalence rates and exposed group relative risks corresponding to those below the heavy black line on the table can be detected with an approximate 0.8 probability with the sample sizes used in this study.

Table 7-2 provides the same information for continuous variables in terms of percentage mean shift and variability, assuming unpaired testing of a normally distributed variable and equal sample sizes.

In the first followup of the AFHS, 1,016 Ranch Hands participated in the physical examination. In this size group, the chance of identifying zero cases of a disease with a prevalence of 1/500 or less is greater than 10 percent. Table 7-3 contains the probability of encountering no cases of disease states for cumulative prevalence rates of 1/200, 1/500, 1/1,000, 1/2,000, 1/5,000, and 1/10,000.

Multiple Endpoints and Comparisons

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

The large number of endpoints under consideration presents a difficult problem in the assessment of Type I error rates. More than 150 dependent variables were tested, not to mention tests for interaction and multiple contrasts among the low, medium, and high exposure-level categories in the exposure index analyses. Furthermore, the dependent variables were correlated to varying degrees, and this makes it even more difficult to assess the attained significance levels. To allow for multiple endpoints, Bonferroni's inequality,¹ which requires significance at the α/K level where K is the number of endpoints considered, may be used, but this procedure

TABLE 7-1.

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

Occurrence Rate of Disease in Control Population	Relative Risk (Multiplicative Factor of Occurrence Rate for Exposed Group)										
	1.25	1.50	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00
$\frac{1}{10,000}$	2,822,082	783,901	235,164	78,384	43,544	29,391	21,944	17,415	14,393	12,243	10,640
$\frac{1}{5,000}$	1,410,882	391,901	117,564	39,184	21,766	14,690	10,968	8,703	7,193	6,118	5,317
$\frac{1}{1,000}$	281,922	78,301	23,484	7,824	4,344	2,930	2,187	1,735	1,433	1,218	1,058
$\frac{1}{500}$	140,802	39,101	11,724	3,904	2,166	1,460	1,089	863	713	606	526
$\frac{1}{100}$	27,906	7,741	2,316	768	424	284	211	167	137	116	100
$\frac{1}{50}$	13,794	3,821	1,140	376	206	137	101	79	65	54	47

*This study has unequal sample sizes; therefore, the tabled values are understated. The similar table in the Baseline Morbidity Report, 24 February 1984, is in error because tabulated sample sizes were only one-half of their correct values.

TABLE 7-2.

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

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

*This study has unequal sample sizes; therefore, the tabled values are understated. The similar table in the Baseline Morbidity Report, 24 February 1984, is in error because tabulated sample sizes were only one-half of their correct values.

TABLE 7-3.

**Probability of Zero Cases as
a Function of Prevalence**

Disease Prevalence	Probability of Finding Zero Cases in a Group of 1,016 Participants
1/10,000	0.903
1/5,000	0.816
1/2,000	0.602
1/1,000	0.362
1/500	0.131
1/200	0.006

becomes increasingly more conservative as the correlation among the endpoints increases. For the analysis results in this report, an alpha level of 0.05 was used for each dependent variable. In addition, group contrasts in strata defined by levels of a covariate involving in a group-by-covariate interaction were assessed by an alpha level of 0.05. The same was true for exposure level strata.

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

Paired Versus Unpaired Analyses

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

The general approach in this report, however, was to conduct unpaired analyses using all available data, based on stratification and/or covariate adjustment. In an unpaired analysis, the matching still serves to improve

the comparability of the two groups, and precision is usually gained from the stratification and covariate adjustment.

Mortality and Morbidity Data

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

Cutpoints

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

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

Exclusions

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

OVERVIEW OF STATISTICAL PROCEDURES

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

TABLE 7-4.

Summary of Statistical Procedures

Chi-Square Contingency Table Test

The chi-square test of independence² is calculated for a contingency table by the following formula:

$$\chi^2 = \sum (f_o - f_e)^2 / f_e$$

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

f_o = observed frequency in a cell

f_e = expected frequency under the hypothesis of independence.

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

Fisher's Exact Test

Fisher's exact test² is a randomization test of the hypothesis of independence for a 2x2 contingency table. This technique is useful for small samples and sparse cells. This is a permutation test based on the exact probability of observing the particular set of frequencies.

General Linear Model Analysis

The form of the general linear model¹ for two independent variables is:

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

where

Y = dependent variable (continuous)

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

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

β_1, β_2 = coefficient indicating linear association between Y and X_1 , Y and X_2 , respectively

β_{12} = coefficient reflecting the linear interaction of X_1 and X_2

ϵ = error term.

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

TABLE 7-4. (continued)

Summary of Statistical Procedures

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

Kolmogorov-Smirnov Distribution Test

The Kolmogorov-Smirnov (K-S) test³ is a nonparametric procedure which assesses differences between the distribution of two samples. Specifically, the K-S procedure tests the hypothesis that populations π_1 and π_2 are identical and is designed to detect all possible deviations from this hypothesis. The assumptions of the K-S test are that the observations from the two samples are mutually independent and that both sets of observations are samples from the same distribution.

Logistic Regression Analysis

The logistic regression model^{2,4} enables a dichotomous dependent variable to be modeled in a regression framework with continuous and/or discrete independent variables. For two risk factors, such as group and age, the logistic regression model would be:

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

where

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

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

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

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

The parameters are interpreted as follows:

α = log odds for the disease when both factors are at a 0 level

β_1 = coefficient indicating the group effect adjusted for age

β_2 = coefficient indicating the age effect adjusted for group

β_{12} = coefficient indicating the interaction between group and age

ϵ = error term.

In the absence of an interaction ($\beta_{12} = 0$), $\exp(\beta_1)$ reflects the adjusted odds ratio for individuals in Group 1 ($X_1 = 1$) relative to

TABLE 7-4. (continued)

Summary of Statistical Procedures

Group 0 ($X_1 = 0$). If the probability of disease is small, the odds ratio will be approximately equal to the relative risk.

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

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

Proportional Odds Model

The proportional odds model⁶ allows for the analysis of an ordered outcome variable. The model assumes that the odds of falling below a certain level rather than above it for individuals at different levels of an independent variable X are in constant ratio. For example, if the response takes one of the four values "excellent," "good," "fair," or "poor," and X is a simple indicator variable designating group (Ranch Hand versus Comparison), then the proportional odds model states that the odds for responding "excellent" versus "good," "fair," or "poor" in the Ranch Hand group are a multiple, $\exp(\beta)$, of the corresponding odds in the Comparison group. Likewise, the odds for responding "excellent" or "good" versus "fair" or "poor" in the Ranch Hand group are the same multiple, $\exp(\beta)$, of the corresponding odds in the Comparison group, as are the odds for responding "excellent," "good," or "fair" versus "poor" in the two groups. Thus, the model is appropriate whenever one frequency distribution is "shifted left" relative to another distribution. Incorporation of other variables into X allows the estimation of proportional odds ratios adjusted for covariates.

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

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

The proportional odds model specifies that

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

TABLE 7-4. (continued)

Summary of Statistical Procedures

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

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

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

Log-linear Analysis

Log-linear analysis² is a statistical technique for analyzing cross-classified data or contingency tables. A saturated log-linear model for a three-way table is:

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

where

- Z_{ijk} = expected cell count
- $U_{1(i)}$ = specific one-factor effect
- $U_{12(ij)}$ = specific two-factor effect or interaction
- $U_{123(ijk)}$ = three-factor effect or interaction.

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

McNemar's Test

McNemar's test⁴ effectively considers discordant pairs in which only the Ranch Hand or only the Comparison member in each pair experiences the abnormality. Using a chi-square approximation with continuity correction, the following statistic is used to test whether the off-diagonal entries are evenly divided:

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

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

TABLE 7-4. (continued)

Summary of Statistical Procedures

Test for Linear Trend

For a $k \times 2$ contingency table in which the k groups fall into a natural order, Armitage⁷ developed a test for a linear trend in the proportions. Let P_i denote the proportion of individuals in the i th row possessing some attribute (e.g., proportion of individuals with abnormal values at each of the three exposure level categories). A score, X_i , is assigned to each of the k levels of the row variable, and the regression coefficient, $\hat{\beta}$, of P_i on X_i is estimated. The regression coefficient is estimated in the usual way except that P_i is weighted by the sample size, n_i , in each row. $\hat{\beta}/SE(\hat{\beta})$ provides a normal deviate for testing the null hypotheses of $\beta = 0$.

Preliminary Analysis

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

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

Core Analysis

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

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

Continuous Dependent Variables

When the dependent variable was continuous, the general linear models (GLM) procedure of SAS® was used to fit a model of the dependent variable in terms of the group indicator (Ranch Hand or Comparison) and appropriate covariates, and interactions between covariates. The covariates could be continuous or categorical variables. If necessary, the dependent variable was transformed prior to analysis by a transformation (e.g., logarithm) to enhance normality of its distribution.⁸ When a "best" model was fitted, according to the strategy outlined below, the test for significance of the group difference was then done on the adjusted group means, provided there were no significant interactions between the group indicator and any of the covariates. Group differences in the presence of interactions were assessed using stratification by different levels of the covariate(s) involved in the interaction or estimation of group differences at selected covariate levels using the best model identified.

For some non-normally distributed dependent variables, the Kolmogorov-Smirnov (K-S) test of significant differences between the distributions of the variables in the two study groups was conducted. The K-S test is a nonparametric test for the equality of two distributions designed to detect broad classes of alternatives.

Categorical Dependent Variables

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

To make the results parallel to those obtained by logistic regression, i.e., because of the distinction between dependent and independent variables, the marginals were fixed in the model, effectively converting the log-linear model into a logit model. The significance of the relative risk for group was determined by examination of the appropriate model, as determined by the study, that includes all statistically significant effects and the group indicator or by examination of the significant interactions. Adjusted relative risks were derived from the coefficients of the appropriate model.

Modeling Strategy

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

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

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

presented. In each case, the same conclusion may be drawn since the statistic of relevance is the difference between the adjusted group mean and the associated p-value. Further, if an interaction of group with a continuous covariate was found, two equally valid methods were used to illustrate how the interaction was arising. One method was to categorize the continuous covariate and describe the group differences within each (covariate-defined) stratum. Another technique was to present group differences for several selected values of the covariate. Further, in the presence of small frequencies of abnormalities, exposure index analyses were occasionally carried out using only the main effects model (i.e., using group and all the covariates but not including interaction terms).

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

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

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

EXPOSURE INDEX ANALYSES

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

The exposure index was trichotomized as high, medium, and low, separately, for each of the three occupational groups: officer, enlisted flyer, enlisted groundcrew. Thus, separate analyses were conducted for each occupational cohort. Discrete dependent variables were evaluated using log-linear and logistic regression models, treating exposure level as a categorical variable (by means of two indicator variables) and adjusting for covariates. For continuous dependent variables, a general linear model was fit, adjusting for covariates and using two indicator variables to designate exposure level. Contrasts between medium and low, and between high and low exposure levels, were also examined.

LONGITUDINAL ANALYSES

General

Another objective of the AFHS is to observe the Ranch Hand population and the Comparison group carefully over time for the emergence, or deleterious rate change, of symptoms, signs, laboratory parameters, or frank disease. This followup objective is not without scientific and logistic challenge, considering mobile populations, problems of loss to study, changing laboratory methods and diagnostic criteria, and the diversity of many changing factors over a period encompassing numerous followup examinations. The following sections describe the statistical procedures used for both continuous and categorical longitudinal data.

Continuous Data

A repeated measurements analysis of variance procedure¹⁰ was used to analyze the variables measured on a continuous scale. The model for the dependent variable (Y) measurement on the kth participant (π_k) in the ith group (α_i) at the jth time (β_j) is as follows:

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

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

<u>Source</u>	<u>Degrees of Freedom*</u>
Group (Ranch Hand vs. Comparison)	1
Subject/Group	2,108
Time (Baseline vs. Followup)	1
Group-by-Time	1
(Subject-by-Time)/Group	2,108

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

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

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

The source of variation due to group (α_i) reflects a difference between the overall Ranch Hand and Comparison means (averaged over both times). This source should complement the group difference findings at Baseline and at

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

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

Categorical Data

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

		<u>Followup</u>			
		Ranch Hand	Comparison		
		Abnormal	Normal	Abnormal	Normal
<u>Baseline</u>	Abnormal		✓		X
	Normal	✓		X	

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

		Change Pattern	
		Normal-Abnormal	Abnormal-Normal
Group	Ranch Hand	✓	✓
	Comparison	X	X

This test is equivalent to testing no group-by-time-by-endpoint interaction in a matched pair analysis.¹¹

CHAPTER 7

REFERENCES

1. Neter, J., and W. Wasserman. 1974. Applied linear statistical models. Homewood, Illinois: Richard D. Irwin, Inc.
2. Bishop, Y.M.M., S.E. Feinberg, and P.W. Holland. 1975. Discrete multivariate analysis: Theory and practice. Cambridge: MIT Press.
3. Hollander, M., and D. Wolfe. 1973. Nonparametric statistical methods. New York: John Wiley & Sons.
4. Fleiss, J.L. 1981. Statistical methods for rates and proportions. 2d ed. New York: John Wiley & Sons.
5. Breslow, N.E., and N.E. Day. 1980. Statistical methods in cancer research. Volume I, The Analysis of Case-Control Studies, Lyon, International Agency for Research on Cancer (IARC Scientific Publications No. 32).
6. McCullagh, P. 1980. Regression models for ordinal data. J. Royal Stat. Soc. B42(2):109-142.
7. Armitage, P. 1955. Tests for linear trends in proportions and frequencies. Biometrics 11, 375-386.
8. Box, G.E.P., and D.R. Cox. 1964. An analysis of transformations. J. Royal Stat. Soc. B26:211.
9. Feinberg, S.E. 1981. The analysis of cross-classified data. Cambridge: MIT Press.
10. Winer, B.J. 1971. Statistical principles in experimental design. 2d ed. New York: McGraw Hill.
11. Breslow, N.E. 1982. Covariance adjustment of relative risk in matched studies. Biometrics 38:661-672.