

CHAPTER 7

STATISTICAL METHODS

INTRODUCTION

This chapter summarizes the statistical methods used in this report to investigate relationships between the health status of the 2,233 participants attending the Air Force Health Study (AFHS) 1992 followup examination and their corresponding group (Ranch Hand or Comparison) or serum dioxin estimates and measurements. Group contrast models are similar to analyses performed for the 1982 Baseline and 1985 and 1987 followup examinations (1,2,3). Models relating health to dioxin estimates and measurements are based on analyses performed for the Serum Dioxin Analysis Report for the 1987 Followup (4).

The statistical methods presented in this chapter encompass six different forms of hypotheses or models applied to more than 300 study endpoints across clinical areas. Each of these models inherently specifies the study cohort or subset of participants to be included in the respective analyses together with the dioxin exposure or proxy estimates to be used in the analysis. The first model specifies contrasts between Ranch Hands and Comparisons using group as a proxy for exposure, and it does not incorporate serum dioxin measurements. The remaining five models all incorporate serum dioxin measurements. A summary description of each of the six models is provided in the section "Models and Assumptions."

Each model and exposure estimate combination is implemented for study variables and type of analysis (unadjusted, adjusted, or longitudinal). The implementation is carried out with specific statistical procedures (e.g., analysis of variance or logistic regression) depending on the analysis being conducted as presented in the section "Factors Determining Statistical Analysis Method." The relationship between the factors and statistical procedures is presented in the "Analysis Methodologies" section. That presentation is followed by a discussion of Interpretive Considerations and a review of conventions for display of analysis results in the "Explanation of Tables" section.

MODELS AND ASSUMPTIONS

The statistical analyses in this report are based primarily on six 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 relationships related to exposure. Serum dioxin measurements are not used in this model. The other five models account 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 effects and dose-response relationships related to exposure. Analyses based on these models were carried out both unadjusted and adjusted for covariates.

Model 1: Group and Occupation as Estimates of Exposure

This section describes models that use the group (Ranch Hand, Comparison) of a participant to assess the relationship between health status and dioxin exposure. Statistical analyses of these models are 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 discovered in the analyses performed for the Serum Dioxin Analysis Report for the 1987 Followup, the average levels of exposure to dioxin were highest for enlisted groundcrew, followed by enlisted flyers and then officers. While using occupation as a surrogate for exposure may be somewhat imprecise, it provides an estimate of exposure that cannot possibly be influenced by a health condition. Occupation as a surrogate for exposure is not subject to the possible biases based on health conditions that can occur with serum dioxin estimates. However, an implicit assumption underlying this model is that Comparisons were not exposed and Ranch Hands were exposed.

Table 7-1 shows these models, the assumptions, advantages, and disadvantages for a continuously distributed health variable y . The model presented in Table 7-1 is unadjusted for any covariates—adjusted models are a straightforward extension.

Models 2 through 6: Serum Dioxin as an Estimate of Exposure

Current dioxin levels in 1987 were determined by the Centers for Disease Control (CDC) 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 followup to provide further insight on current dioxin levels and the elimination of dioxin from the body.

Further investigation of the mechanics of dioxin elimination are currently under study by the Air Force at this time. Based on samples collected at the pilot study, 1987 followup, and 1992 followup, issues such as half-life estimation and first-order pharmacokinetic assumptions are being further investigated.

Prior Knowledge Regarding Dioxin

This section presents analytic strategies based on assumptions and models conceived in 1988 and after the publication of the Ranch Hand dioxin pilot study and half-life substudy. At that time, available data showed that dioxin elimination appeared to follow first-order mechanisms. This observation was based on measurements subsequent to the ingestion of dioxin by an individual (5). Data on 36 Ranch Hand veterans with dioxin measured in blood drawn in 1982 and in 1987 produced a median half-life estimate of 7.1 years (6), and this median was used in all calculations involving half-life in this report.

Table 7-1.

Model 1: Assessing Health versus Group Status in Ranch Hands and Comparisons: Assumptions, Advantages, and Disadvantages

Model 1: $y = \mu + G_i + e$ (All Ranch Hands and Comparisons)

$y = \mu + G_i + O_j + (GO)_{ij} + e$ (Ranch Hands and Comparisons by occupation)

where,

y = health variable

G_i = effect due to group status ($i = 1,2$ - Comparisons, Ranch Hands)

O_j = effect due to occupation ($j = 1,2,3$ - 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.

Occupation is an estimate of exposure that cannot possibly be influenced by a health condition, whereas the serum dioxin estimate can be influenced by a health condition.

Disadvantages: Results will be 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.

The term "elimination" denotes the overall removal of dioxin from the body. Some analyses in this report assume that the amount of dioxin in the body (C) decays exponentially with time according to the model $C = I \cdot \exp(-rt)$, where I is the initial level, $r = \log(2)/h$ is the decay rate, h is the half-life, and t is the length of time between the participant's time of duty in Southeast Asia (SEA) and the blood draw for dioxin performed at the pilot study in April 1987, the blood draw for dioxin performed at the 1987 physical examination, or the blood draw for dioxin performed at the 1992 physical examination. If a participant had measurements at more than one of these points in time, the measurement closest to the time of duty in SEA was used. This exponential decay law is termed "first-order elimination" in this report.

The first-order elimination assumption is not equivalent to assuming a one compartment model for dioxin distribution within the body. While a multicompartment model incorporating body composition and 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD, or dioxin) binding to tissue receptors would provide a detailed description of dioxin concentrations in

different compartments, published multicompartment models for TCDD distribution within the body predict first-order elimination of TCDD, overwhelmingly due to fecal excretion (7).

The lipid-weight concentration of TCDD, expressed in parts per trillion (ppt) (8,9), is a derived quantity calculated from the formula $ppt = ppq \cdot 102.6/W$, 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 (7).

The relationship 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 (10). 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% C.I. = [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 cannot 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.

Fundamental Limitations of the Serum Dioxin Data

There are two evident limitations to the available data:

- While Ranch Hand data and ingestion data do 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 are available yet.
- It is not known whether Ranch Hands with body burdens of dioxin at or below 10 ppt were exposed, and their body burdens had decayed 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.

Model 2: Health versus Initial Dioxin in Ranch Hands

The relationship between an estimated initial dioxin exposure and health was assessed within Ranch Hands using the model described in Table 7-2. Statistical analyses of these models are termed "Model 2" in the assessment of the clinical areas. In this model, an initial dioxin exposure 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 draw for dioxin, and an estimated half-life of 7.1 years. From exploratory studies conducted by the Air Force, body fat at the time of duty in SEA and change in percent body fat from the time of duty in SEA to the date of the blood draw for dioxin appear to be related to the half-life of a participant. These body fat variables were included in this model as explanatory, or independent, variables and were not removed during stepwise procedures,

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

<p>Model 2: $y = b_0 + b_1 \log_2(I) + b_2 \text{BFTR} + b_3 \text{BFCH} + e$</p>	
<p>where,</p>	
<p>y</p>	<p>= health variable</p>
<p>I</p>	<p>= extrapolated initial dose, assuming first-order elimination, $I = 4 + (C-4) \cdot \exp(\log(2) \cdot t/h)$, where 4 ppt is considered the median background level of lipid-adjusted current dioxin</p>
<p>t</p>	<p>= length of time between the time of duty in SEA and the date of the blood draw for dioxin in 1987 or 1992</p>
<p>C</p>	<p>= lipid-adjusted current dioxin, determined in 1987 or 1992</p>
<p>h</p>	<p>= dioxin half-life in Ranch Hands assuming first-order elimination (7.1 years assumed for analysis)</p>
<p>BFTR</p>	<p>= body fat at the participant's time of duty in SEA, calculated from the formula shown below.</p>
<p>BFCH</p>	<p>= change in body fat between the time of the participant's duty in SEA and the date of the blood draw for dioxin in 1987 or 1992, calculated from the formula shown below</p>
<p>e</p>	<p>= zero mean error.</p>
<p>Body fat will be calculated from a metric body mass index (11); the formula is</p>	
$\text{Body Fat (in percent)} = \frac{\text{Weight (kg)}}{[\text{Height (m)}]^2} \cdot 1.264 - 13.305.$	
<p>Assumptions:</p>	<p>Ranch Hands received a single dioxin dose in Vietnam and background exposure thereafter.</p> <p>Ranch Hands experienced first-order dioxin elimination.</p> <p>The error variance does not change with health status or initial dioxin dose.</p>
<p>Advantages:</p>	<p>Easily interpretable.</p> <p>Most efficient if first-order elimination and half-life are valid and y is linearly related to $\log_2(I)$.</p>
<p>Disadvantages:</p>	<p>Will be biased if first-order elimination or constant half-life assumption is not valid.</p>

which are explained subsequently. Table 7-2 also includes assumptions, advantages, and disadvantages 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.

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, likely would not change conclusions drawn from any of the statistical analyses.

Analyses were carried out on Ranch Hands who had lipid-adjusted current dioxin levels greater than 10 ppt at either the 1987 or 1992 physical examination. The value 10 ppt corresponds to the approximate 98th percentile of the Comparison lipid-adjusted current dioxin distribution. Based on this Comparison dioxin distribution, it is believed that participants with greater than 10 ppt lipid-adjusted current dioxin were definitely exposed. It is not known whether Ranch Hands with dioxin burdens at or below 10 ppt were exposed and their body burdens had decayed 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. Current dioxin levels less than 10 ppt are subsequently called "background" levels. No additional data or other information exist to determine whether any of the Ranch Hands with background levels (≤ 10 ppt) of current dioxin received a dose above background levels in SEA.

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 assesses health versus dioxin body burden categorized into four levels. The four levels of categorized dioxin are given below:

- Comparisons—Comparisons with up to 10 ppt current lipid-adjusted dioxin
- Background—Ranch Hands with up to 10 ppt current lipid-adjusted dioxin
- Low, High—Ranch Hands with more than 10 ppt current lipid-adjusted dioxin.

Statistical analyses of these models are 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 of the Ranch Hands with more than 10 ppt current dioxin (i.e., the sample used in Model 2). In this model, an initial dioxin exposure 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 draw for dioxin, and an estimated half-life of 7.1 years. From exploratory studies conducted by the Air Force, body fat at the time of duty in SEA and change in percent body fat from the time of duty in SEA to the date of the blood draw for dioxin appear to be related to the dioxin half-life of a participant. These body fat variables were included in this model as independent variables that were not removed during stepwise procedures, which are explained subsequently. Using these body fat measures in Model 3 for all Comparisons and Ranch Hands with dioxin measurements allows 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 unknown, low, high, and low plus high categories combined were contrasted with the mean values of y within the background category. Relative frequencies were contrasted for discrete health variables. Table 7-3 shows this model, 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: $y = b_0 + b_1I_1 + b_2I_2 + b_3I_3 + b_4I_4 + b_5BFTR + b_6BFCH + e,$

where

- y = health variable
- I₁ = indicator variable for current dioxin; I₁ = 1 if participant is a Comparison, I₁ = 0 if participant is not a Comparison
- I₂ = indicator variable for current dioxin; I₂ = 1 if participant is in background category, I₂ = 0 if participant is not in background category
- I₃ = indicator variable for current dioxin; I₃ = 1 if participant is in low category, I₃ = 0 if participant is not in low category
- I₄ = indicator variable for current dioxin; I₄ = 1 if participant is in high category, I₄ = 0 if participant is not in high category
- BFTR = body fat at the participant's time of duty in SEA, calculated from the formula shown below
- BFCH = change in body fat between the participant's time of duty in SEA and the date of the blood draw for dioxin in 1987 or 1992, calculated from the formula shown below
- e = zero mean error.

Body fat will be calculated from a metric body mass index (11); the formula is

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

Assumptions: Dioxin body burden has accumulated with time.

The error variance does not change with categorized current dioxin body burden.

Advantages: Requires no assumption regarding the time course of dioxin accumulation or elimination.

Initial dioxin is probably a better measure for determining low and high exposure than current dioxin.

Less dependent on the accuracy of the estimation algorithm for determining initial dioxin than Model 2.

Disadvantages: Makes no use of prior belief that 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 is 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?").

Models 4, 5, and 6: Health versus Current Dioxin in Ranch Hands

The relationship between current dioxin, as determined for most Ranch Hands at the 1987 followup, and health was assessed using the models described in Table 7-4. This table also describes the assumptions, advantages, and disadvantages for the unadjusted analysis of a continuously distributed health variable y .

Ranch Hands with a dioxin measurement may have had their blood drawn at the pilot study in April 1987, at the 1987 physical examination, or at the 1992 physical examination. If an individual has 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

$$C_{1987} = 4 + (C_{1992} - 4) \cdot \exp(rt)$$

was used for extrapolation of lipid-adjusted current dioxin to 1987 levels (C_{1987}), and

$$C_{1987} = 24 + (C_{1992} - 24) \cdot \exp(rt)$$

was used for extrapolation of whole-weight current dioxin to 1987 levels (C_{1987}), where C_{1992} is the current dioxin level (lipid-adjusted or whole-weight) in 1992, 4 ppt is considered the median background level for lipid-adjusted current dioxin, 24 ppq is considered the median background level for whole-weight current dioxin, $r = \log(2)/h$ is the decay rate, h is the half-life (7.1 years), and t is the length of time between the physical examination in 1987 and the physical examination in 1992. This model was only used if the lipid-adjusted current dioxin level in 1992 was greater than 10 ppt; otherwise the 1992 measurement was used.

Three models were analyzed with current dioxin used as the estimate of exposure. Statistical analyses of these models are termed "Model 4," "Model 5," and "Model 6" in the assessment of the clinical areas. There is scientific debate as to the appropriate current dioxin measure. For the Serum Dioxin Analysis Report for the 1987 Followup, a lipid-weight current dioxin measure was used. As described above, the lipid-weight current dioxin measure (ppt) is related to the whole-weight dioxin measure (ppq) from the formula $\text{ppt} = \text{ppq} \cdot 102.6/W$, where ppt is the lipid-weight concentration, ppq 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. Other researchers advocate the use of the whole-weight current dioxin measure.

The models are similar in form to Model 2 ($y = b_0 + b_1 \log_2(I) + e$, see Table 7.2), except that a current dioxin measure was used instead of an initial dioxin estimate. Model 4 used the logarithm (base 2) of lipid-weight current dioxin. Model 5 used the logarithm (base 2) of whole-weight current dioxin. Model 6 used the logarithm (base 2) of whole-weight current dioxin, with the logarithm (base 2) of the total lipid weight of the sample ($\log_2[W]$) as an independent variable that was not removed during stepwise procedures, which are explained subsequently.

Table 7-4.
Models 4, 5, and 6: Assessing Health versus Current Dioxin in Ranch Hands:
Assumptions, Advantages, and Disadvantages

<p>Model 4: $y = b_0 + b_1 \log_2(\text{ppt}) + e$ Model 5: $y = b_0 + b_1 \log_2(\text{ppq}) + e$ Model 6: $y = b_0 + b_1 \log_2(\text{ppq}) + b_2 \log_2(W) + e$</p> <p>where</p> <ul style="list-style-type: none"> y = health variable ppt = lipid-weight current dioxin = $\text{ppq} \cdot 102.6 / W$, ppq = whole-weight of dioxin in the sample in femtograms (102.6 corrects for the average density of serum) W = total lipid weight of the sample e = zero mean error. <p>Assumptions: Ranch Hands received a single dioxin dose in Vietnam and background exposure thereafter.</p> <p style="padding-left: 40px;">The error variance does not change with health status or current dioxin.</p> <p>Advantages: Using current dioxin has less inherent variation than initial dioxin, which is extrapolated by a first-order elimination model across a 15- to 25-year time period.</p> <p>Disadvantages: Current dioxin may not be a good surrogate for exposure if elimination rate differs for individuals.</p> <p style="padding-left: 40px;">Individuals with measurements in 1992 only will be extrapolated to 1987, and variation will be increased with estimation using a first-order elimination model.</p>
--

FACTORS DETERMINING STATISTICAL ANALYSIS METHOD

For a specified questionnaire-based or clinical measurement determined from the physical or laboratory examination, the selection of an analytical method depends 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 greater than 10 ppt of current lipid-weight dioxin
 - **Model 3:** Categorized dioxin—Comparisons with 10 ppt lipid-weight dioxin or less and all Ranch Hands with a dioxin measurement
 - **Models 4, 5, & 6:** Current dioxin—All Ranch Hands with a dioxin measurement
- **Analysis Type:** Unadjusted, adjusted, or longitudinal.

Appendix Table D-1 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 are used for adjusted analyses of initial dioxin for continuous dependent variables.

ANALYSIS METHODOLOGIES

Methods for Analyzing Continuous and Discrete Variables

Similar to the analyses conducted in previous AFHS reports, health endpoints, or dependent variables, were treated as either continuous or discrete. For unadjusted analyses of Model 1, t-tests were used for continuous dependent variables and chi-square tests were used for discrete dichotomous variables to test for differences between Ranch Hands and Comparisons.

For other analyses of continuous dependent variables, the general linear model approach was used for applying such techniques as simple and multiple linear regression, analysis of variance, analysis of covariance, repeated measures analysis, and failure time analysis. This approach permitted model fitting of the dependent variable as a function of group or dioxin, relevant covariates, group-by-covariate or dioxin-by-covariate interactions, and interactions between covariates. Continuous dependent variables were examined to ensure that assumptions underlying appropriate statistical methods were met. Transformations were used to enhance normality for specific continuous health variables. A general method for determining a transformation can be found in an article by Box and Cox (12), and this method was used as a guide in determining the appropriate transformation. A further discussion of general linear models, as well as other methods used for the statistical analysis in this report, is found in Table 7-5.

For these continuous analyses, the SAS[®] procedure GLM (13) was used. When a “best” model was fitted, tests of significance for a group or dioxin effect were made. 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. If there was a significant interaction between group or dioxin and any covariate, the effect of group or dioxin on the dependent variable was assessed using stratification by different levels of the covariate(s) involved in the interaction.

The SAS[®] procedures LIFEREG and LIFETEST (13) were used for the time to diabetes onset variable in the endocrine clinical assessment. This variable consisted of censored and noncensored data, and statistical methods used to analyze measures of this type implement a technique known as “failure time” analysis. A further discussion of failure time analysis is found in Table 7-5.

Discrete dependent variables were analyzed by methods parallel to those used for continuous variables. For dichotomous discrete dependent variables, logistic regression was performed using BMDP[®]-LR (14). For dependent variables with more than two categories, polychotomous logistic regression was performed using BMDP[®]-PR (14). Parameter estimation and model selection for polychotomous logistic regression and ordinary logistic

Table 7-5.
Summary of Statistical Procedures

Chi-square Contingency Table Test

The chi-square test of independence (15) is calculated for a contingency table by the following formula:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

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.

For 2x2 tables, the chi-square statistic above will be adjusted for the continuity of the χ^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:

$$\chi^2 = \sum \frac{\max(0, (|f_o - f_e| - \frac{1}{2}))^2}{f_e}$$

Correlation Coefficient (Pearson's Product-Moment)

The population correlation ρ (17) measures the strength of the linear relationship between two random variables X and Y. A commonly used sample-based estimate of this correlation coefficient is

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

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 2,233 in this study, a sample correlation coefficient of 0.0415 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.

Table 7-5. (Continued)
Summary of Statistical Procedures

Failure Time Analysis

The failure time (or survival time) model (18) 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 a "failure," the time for participants that have not "failed" is right censored. The failure time model is

$$y = X\beta + \sigma\epsilon$$

where,

- y = vector of responses (e.g., time to diabetes onset), usually the logarithm of the failure times
- X = matrix of covariates, or risk factors (e.g., group status and age)
- β = vector of unknown regression parameters
- σ = unknown scale parameter
- ϵ = vector of errors assumed to come from 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 the SAS[®] LIFEREG procedure, using a Newton-Raphson algorithm, where the distribution of the random error term can be specified. The distributional assumptions of the error term can be tested by examining plots of the Kaplan-Meier survival functions using the SAS[®] LIFETEST procedure.

The LIFEREG procedure will provide 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, and to step out the nonsignificant covariate terms. 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) to produce estimated means from a censored Weibull distribution. The loss function that is minimized in the estimation procedure is

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

where, x=1 if diabetic
 x=0 if not diabetic

and y=time to onset of diabetes.

**Table 7-5. (Continued)
Summary of Statistical Procedures**

Fisher's Exact Test

Fisher's exact test (15) is a randomization test of the hypothesis of independence for a 2 x 2 contingency table. This technique was used for small samples and sparse cells. This is a permutation test based on the exact probability of observing the particular set of frequencies, or of one more extreme.

General Linear Models Analysis

The form of the general linear model (17) for two independent variables is

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_{12} X_{1i} X_{2i} + \epsilon_i$$

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 (e.g., group status and age)
- β_1, β_2 = coefficient indicating linear association between Y and X_1, Y and X_2 respectively; each coefficient reflects the effect on the model of the corresponding independent variable adjusted for the effect of the other independent variable
- β_{12} = coefficient reflecting the linear interaction of X_1 and X_2 , adjusted for linear main effects
- ϵ_i = error term.

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

Log-linear Analysis

Log-linear analysis (15) is a statistical technique for analyzing cross classified data or contingency tables. A saturated log-linear model for a three-way table, for example, is

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

where,

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

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

**Table 7-5. (Continued)
Summary of Statistical Procedures**

Logistic Regression Analysis

The logistic regression model (19) 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 dioxin and age, the logistic regression model would be

$$\text{logit } P_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \epsilon_i$$

where,

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

- α = log odds for the disease when $X_1 = 0$ and $X_2 = 0$
- β_1 = coefficient indicating the dioxin effect adjusted for age
- β_2 = coefficient indicating the age effect adjusted for dioxin
- β_{12} = coefficient indicating the interaction between dioxin and age, adjusted for linear main effects
- ϵ_i = error term.

In the absence of an interaction ($\beta_{12} = 0$) for a dichotomous measure (e.g., Comparisons, Ranch Hands), $\exp(\beta_1)$ reflects the adjusted odds ratio for individuals in group 1 ($X_1 = 1$) relative to group 0 ($X_1 = 0$). If the probability of disease is small, the odds ratio will be approximately equal to the relative risk. In the absence of an interaction for a continuous risk factor (e.g., initial dioxin in its continuous form), $\exp(\beta_1)$ reflects the adjusted odds ratio for a unit increase in the risk factor. 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, the adjusted odds ratios will be referred to as adjusted relative risks. Correspondingly, in the absence of covariates (i.e., unadjusted analysis), the odds ratios will be referred to as estimated relative risks.

This technique also will be used for longitudinal analyses of dichotomous dependent variables to examine changes in health status between 1982 (or 1985) and 1992 in relation to the dioxin measures.

Two-Sample t-Test

A statistical test for determining whether or not it is reasonable to conclude that two population means are unequal utilizes the t-distribution (17). Tests can be performed when population variances are equal or unequal; however, different t-distributions are used.

**Table 7-5. (Continued)
Summary of Statistical Procedures**

Polychotomous Logistic Regression Analysis

Polychotomous logistic regression (19,20) allows a categorical dependent variable with more than two outcomes to be modeled in a regression environment with continuous and discrete independent variables. For polychotomous logistic regression, the model equation depends upon the scale of the dependent variable. This discussion will focus 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). Polychotomous 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 polychotomous regression model would be

$$\text{logit } P_1 = \alpha_1 + \beta_{1(1)}X_1 + \beta_{1(2)}X_2 + \beta_{1(12)}X_1X_2 + \epsilon_1$$

$$\text{logit } P_2 = \alpha_2 + \beta_{2(1)}X_1 + \beta_{2(2)}X_2 + \beta_{2(12)}X_1X_2 + \epsilon_2$$

where,

- P_i = probability that Y = i (outcome i) with covariates X_1 and X_2 , $i = 0, 1, 2$
- $\text{logit } P_i = \ln (P_i/P_0)$ (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:

- α_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, $i = 1, 2$
- $\beta_{i(12)}$ = coefficient representing the interaction effect of group status and age on the logit P_i , adjusted for the main effects, $i = 1, 2$
- ϵ_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). In the absence of interaction ($\beta_{1(12)} = 0$), $\exp(\beta_{1(1)})$ 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 will be approximately equal to the relative risk of being low between the two groups. If X_1 is a continuous covariate that does not interact with X_2 , $\exp(\beta_{1(1)})$ represents the adjusted log odds ratio of outcome 1 versus outcome 0 for a unit increase in X_1 .

regression are very 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 need to be 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 also will need to be estimated. However, ordinary logistic regression maximizes two likelihood equations, each with k parameters, while polychotomous logistic regression estimates all $2k$ parameters simultaneously with one likelihood equation. To select a final model, polychotomous logistic regression utilizes a stepwise regression procedure similar to the stepwise procedure used in BMDP[®]-LR. Polychotomous linear 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 polychotomous logistic regression is found in Table 7-5.

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 sedimentation rate, cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides, and free testosterone were age-dependent. Consequently, normal and abnormal levels were constructed according to a participant's laboratory value and age at the physical examination. Additionally, cutpoints for serum insulin, serum glucagon, serum proinsulin, and serum C peptide were dependent on whether the participant was fasting. Normal and abnormal levels for these variables were constructed according to a participant's laboratory value and fasting status at the physical examination.

Modeling Strategy

In each clinical category, many covariates were considered for inclusion in the statistical models relating specific health endpoints and group or dioxin. The large number of covariates, consequent interaction terms, and resulting difficulties of interpretation obligate the adoption of a strategy for identifying a moderately simple model using a stepwise strategy, as outlined below. Interpretation of possible relationships were then made in the context of this simpler model.

In general, based on one of the adjusted analysis models described in Appendix Table D-1, a starting model for continuous variables was constructed containing two-factor interactions. First, screening was performed at the 0.15 significance level to eliminate unnecessary two-factor interactions. A hierarchical stepwise deletion strategy then was applied at the 0.15 significance level on the set of main effect covariates (to address possible confounding effects between the covariates and group or dioxin) and at the 0.05 significance level for interactions.

The modeling strategy was refined slightly for adjusted statistical analyses of discrete dependent variables. In particular, the starting model included all main effects and excluded all interactions. Main effects were stepped out of the model if the associated p -value was greater than 0.15 and interactions were entered into the model if the associated p -value was less than or equal to 0.05. The alternative strategy was used to avoid overspecification of the model and minimize collinearity among terms that could lead to imprecise parameter and

standard error estimates, especially where a large number of covariates or sparse number of abnormalities were encountered.

In general, the only effects not subject to the deletion strategy were the group or dioxin variables of interest (that is, group, initial dioxin, or current dioxin). For specific clinical areas, certain covariates were entered into the model and were not subject to the deletion strategy. In particular, caloric intake was retained in one set of analyses for body fat in the General Health Assessment (Chapter 9). For the analysis of diabetic participants in the Endocrine Assessment (Chapter 18), diabetic severity was retained in the model and was not subject to the deletion strategy. Age was retained in all final models of verified medical records variables in the Neurology Assessment and the Gastrointestinal Assessment (Chapters 11 and 13 respectively).

As described above, body fat at the time of duty in SEA and change in percent body fat from the time of duty in SEA to the date of the blood draw for dioxin were included in Models 2 and 3 and were not removed during stepwise procedures. Also, in Model 6, the logarithm (base 2) of the total lipid weight of the sample ($\log_2[W]$) was not subject to the deletion strategy.

With the objective of producing the simplest model, other lower-order effects were retained in the model only if they were involved in significant higher-order interactions. Significant two-factor interactions between covariates were retained in the model. If necessary, the modeling strategy for the adjusted analyses of dependent variables in certain clinical areas was modified because of the large number of covariates and sparse number of abnormalities, which could cause problems in the model estimation. As appropriate, pairwise covariate interactions were not investigated. If estimation problems were still encountered, stepwise procedures began with main effects only models. Also, preliminary investigation of dependent variable-covariate associations was conducted to possibly reduce the number of candidate covariates in the adjusted analyses of some clinical areas (for example, investigations of the lifetime alcohol history and current alcohol use covariates were conducted for the Cardiovascular Assessment, and the lifetime alcohol history covariate was retained for use in adjusted analyses—see Chapter 15).

In the analysis of a particular health variable, when no group or dioxin-by-covariate interactions were significant at the 0.05 level, adjusted means (21) and slopes or adjusted relative risks were presented. If the interaction was significant at the 0.05 level, the behavior of the group or dioxin variable was explored for different levels (categories) of the covariate to identify subpopulations for which a relationship might exist or where the relationship differs between subpopulations. Further, if any group or dioxin-by-covariate interaction was significant at a level between 0.01 and 0.05, the adjusted means and slopes or adjusted relative risks also were presented, after dropping the interaction term from the model. Also, at the discretion of the analyst, adjusted results may be presented after dropping the interaction term from the model if a group-by-covariate interaction or a dioxin-by-covariate interaction was significant at a level less than or equal to 0.01.

In many instances the clinical importance of a statistically significant group-by-covariate or dioxin-by-covariate interaction is unknown or uncertain. The clinical relevance of a

statistically significant interaction is strengthened if the same interaction persisted among related endpoints. Due to the large number of these types of interactions examined for approximately 330 variables, it is recognized that some of the group-by-covariate or dioxin-by-covariate interactions judged significant at the 0.05 level were spurious; that is, chance occurrences not of biological or clinical relevance. This issue was considered when these significant interactions were interpreted. It is important that the size of the p-value associated with each of these interactions is weighed carefully; for this reason, if the p-value for a group-by-covariate or dioxin-by-covariate interaction was between 0.01 and 0.05, the adjusted means or relative risks (omitting the interaction) were reported.

For all models that included a group-by-covariate or dioxin-by-covariate interaction in the final adjusted model, the stratified results display adjusted means, adjusted slopes, or relative risks, confidence intervals, and associated p-values determined from a model that included the interaction term. On occasions where cell sizes were small, statistics were generated from separate models for each covariate stratum. In general, results based on an analysis stratified by the covariate(s) involved in a group-by-covariate interaction or a dioxin-by-covariate interaction are not discussed in the text of a chapter. Usually only the results based on analyses performed after the deletion of an interaction are discussed in the text of a chapter. Exceptions to this strategy include interactions judged to be clinically relevant and situations where no additional analyses were performed omitting the interaction ($p \leq 0.01$ for the group-by-covariate or dioxin-by-covariate interaction).

Specialized Analyses

Military occupation was used in specialized analyses of Model 1. In particular, occupation and a group-by-occupation interaction was investigated in the context of the final model for Model 1 analyses. A final model was developed for each dependent variable, with group contained in the final model. As an additional analysis, if occupation and the group-by-occupation interaction were not in the final model, then they were added to this model. Summary statistics and results for the group variable were reported, and statistics and results on the group variable were presented for each occupational stratum.

For all clinical areas, with the exception of neoplasia, additional analyses were performed when occupation was retained in the final model for the five models involving dioxin. Dioxin exposure and occupation are related due to the military occupational duties performed by the participants. With the exception of neoplasia, occupation also is considered to be a risk factor in assessing the health of the participants. Analyses were consequently performed with occupation in the final model when it was significant, and again with occupation removed from the model. The results of analyses without occupation in the final adjusted model are only discussed in the text if the level of significance (significant, marginally significant, nonsignificant) differs from the original final adjusted model.

For the Neurology, Cardiovascular, Renal, Endocrine, and Pulmonary clinical assessments, additional analyses were performed when certain covariates were retained in the final model for the five models involving dioxin. These data showed significant associations with dioxin for the 1992 followup data, and included diabetic class (Neurology, Cardiovascular, Renal, and Endocrine Assessments), percent body fat (Cardiovascular,

Endocrine, and Pulmonary Assessments), total cholesterol (Cardiovascular and Endocrine Assessments) and HDL cholesterol (Cardiovascular and Endocrine Assessments). These covariates are well-known risk factors and should be introduced into adjusted models; however, these covariates may have been affected by dioxin exposure. Adjustment for these covariates has the potential to “over-adjust” the model for the effects of dioxin exposure. Due to the association between these covariates and dioxin, both the statistical and clinical interpretations of other health variables can be affected. When these analyses were found to be significantly associated with a dependent variable and retained in the final model, the dioxin effect was evaluated in the context of two models. In particular, analyses were performed with and without these covariates in the model to investigate whether conclusions regarding the association between the health endpoint and dioxin differ. The results of the analyses without these covariates in the final adjusted model are only discussed in the text if the level of significance (significant, marginally significant, nonsignificant) differs from the original final adjusted model.

Longitudinal Analyses

Selected longitudinal analyses were performed investigating changes in health status between 1982 and 1992 for Models 1, 2, and 3 as a function of dioxin exposure. Models 4, 5, and 6 were not examined in longitudinal analyses because current dioxin, the estimate of exposure in these models, changes over time and is not available for all participants in 1982 or 1992. All three models were adjusted for age in 1992. Age is 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 1992 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 1992. 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 are explored. The rate of abnormalities under this restriction approximates the cumulative incidence rate between 1982 and 1992.

The dependent variable in this type of analysis was the health of participants at the 1992 examination whose health was normal in 1982. The independent variables were the appropriate exposure estimate and age in 1982. The analyses of Models 2 and 3 also were adjusted for percent body fat at the time of duty in SEA and change in percent body fat from the time of duty in SEA to the date of the blood draw for dioxin. Tabular displays of the longitudinal analyses results of discrete dependent variables include summary statistics for 1982 and 1992, as well as 1985 and 1987 summaries if available. The results of the statistical analyses restricted to those participants who were normal in 1982 also are provided.

In the longitudinal analyses of continuous variables, a general linear model approach, as explained in Table 7-5, was used. The dependent variable was the difference between the

1992 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 in 1982, and the 1982 measurement of the continuous dependent variable. The analyses of Models 2 and 3 also were adjusted for percent body fat at the time of duty in SEA and change in percent body fat from the time of duty in SEA to the date of the blood draw for dioxin. Use of the health endpoint measurement from 1982 has the following three purposes:

- A linear relationship between measurements of the dependent variable in 1982 and 1992 due to a difference in measuring devices will be accounted for by using the 1982 measurement as an independent variable. For many of the laboratory measurements in 1982, 1985, and 1987 the Dupont Automated Chemical Analyzer[®] (ACA) was used. The Baxter/Dade Paramax[®] was used extensively in the 1992 laboratory analyses.
- The difference between two values taken over a period of time is generally correlated with the first measurement (22).
- The relationship between the difference of the 1992 and 1982 measurements and the estimate of exposure may be confounded with the 1982 measurement, especially if the endpoint and the estimate of exposure are related.

Tabular displays of the results of longitudinal analyses of continuous dependent variables include summary statistics for 1982 and 1992, as well as 1985 and 1987 summaries if available. Results of the statistical analyses relating the difference in the 1992 and 1982 measurements to the estimate of exposure also are 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 1992 variable due to differing analytical methods. For example, to enhance comparability, the longitudinal analyses for the Neurological Assessment were based on changes between 1985 and 1992 because Scripps Clinic and Research Foundation (SCRF) conducted both of these examinations, whereas the Kelsey-Seybold Clinic conducted the 1982 examinations.

INTERPRETIVE CONSIDERATIONS

Several specific issues to be considered when interpreting the results found in this report are discussed in this section. The issues discussed here include adjusted analyses, multiple testing, trends in the results of endpoints within a clinical area, the proportion of variation explained by the model (R^2), interpretation of continuous and discrete analyses of a health endpoint, and the ability to detect a significant difference based on the data at hand (power of the analyses). Additional interpretive considerations can be found in Chapter 1, Introduction.

Adjustments for Covariates and Interactions

In contrasts between all Ranch Hands and all Comparisons (Model 1) the matching variables age, race, and occupation were effectively eliminated as confounders. The current dioxin and initial dioxin analyses within Ranch Hands (Models 2, 4, 5, and 6) and the categorized current dioxin analyses within Ranch Hands and Comparisons (Model 3) did not benefit from the matched design. For example, military occupation is a strong confounder because it is highly correlated with current dioxin levels in Ranch Hands and is related to some health variables through socioeconomic differences between officers and enlisted personnel. Education is 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 is sparse, all health endpoints were analyzed with and without adjustment for clinically relevant covariates.

In addition, some covariates (e.g., percent body fat) may themselves be associated with dioxin exposure and may be related to the dependent health variable through their relationship with dioxin. In this situation, analyses of covariance adjusted for such a covariate are not valid, because the assumed independence of the “treatment” (current or initial dioxin) and the covariate is not met (23). There is no recourse but to analyze the data with and without adjustment for the covariate (see Specialized Analyses section above); both analyses potentially are biased. Unadjusted analyses must be viewed with caution and circumspection and, because some covariates may act in an intervening manner relating the “treatment” to the dependent variable, some adjusted analyses of covariance are themselves subject to bias. Bias introduced by intervening covariates is unavoidable in an observational study.

The adjusted models assessed the statistical significance of interactions between group or dioxin and the covariates to determine whether the relationship between group or dioxin and the health endpoint differed across levels of the covariate. Many times, the clinical importance of a statistically significant group-by-covariate or dioxin-by-covariate interaction is unknown or uncertain. The clinical relevance of a statistically significant interaction would be strengthened if the same interaction persisted among related endpoints. Due to the large number of group-by-covariate or dioxin-by-covariate interactions that were examined for approximately 300 variables, some of the interactions found significant at the 0.05 level might be spurious (i.e., chance occurrences not of biological or clinical relevance). This issue should be considered when significant group-by-covariate or dioxin-by-covariate interactions are interpreted. It also is important that the size of the p-value associated with each interaction be weighed carefully. For this reason, models without the dioxin-by-covariate interaction were implemented to address the possibility that some interactions may arise from multiple testing (see Modeling Strategy section above and Multiple Testing section below). Also, implementing models without the group-by-covariate or dioxin-by-covariate interactions allows the reader to examine results for all participants combined, whereas the interaction analyses explore the different relationships between dioxin and the dependent variable, depending on the subgroups of participants examined.

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 relationship between physical health and dioxin, approximately 5 percent of the many statistical tests (group or dioxin effects and group-by-covariate or dioxin-by-covariate interactions) in this report would be expected to detect an association between group or dioxin and health (p -values ≤ 0.05). Observing significant results due to multiple testing, even when there is no relationship between dioxin and health, is known as the multiple-testing artifact and is common in all large studies. Unfortunately, there is no statistical procedure to distinguish between those statistically significant results that arise due to the multiple testing artifact and those that may be due to an actual dioxin effect. Instead, in order to weigh and interpret the findings, the authors have considered the strength of the association, consistency, dose-response patterns, and biologic plausibility.

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. However, there may 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 must be considered when assessing the suspected association.

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, very 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 themselves subject to some measurement error. Thus, in most analyses in this report, the values of R^2 are small.

Clinical Interpretation of Discrete versus Continuous Data

Small but significant mean differences in a continuously measured health variable (e.g., systolic blood pressure) between exposed and unexposed groups when there are no corresponding differences in the percentage of abnormal tests are difficult to assess in any study. In this study, significant mean differences are sometimes observed without a corresponding group difference in the proportion outside the normal range. 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 are interpreted as a dioxin-related effect if a corresponding trend is seen in the proportion above or below the normal range or if the trend is consistent with other findings.

Power

Conducting a statistical test using a type I error, also called an alpha or significance level, of 0.05 means that, on the average in 5 cases out of 100, a false conclusion would be made that an association (group or dioxin effect) exists when, in reality, 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 and non-Hodgkin's lymphoma. 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 7-6 and Appendix Tables D-2 through D-5 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 7-6 presents power calculations for Model 1 (group), and Appendix Tables D-2 through D-5 presents power calculations for Model 2 (initial dioxin), Model 3 (categorized dioxin—low plus high Ranch Hand versus Comparison contrast), Model 4 (lipid-adjusted current dioxin), and Models 5 and 6 (whole-weight current dioxin). Power calculations were performed using the logarithm (base 2) of dioxin in Models 2, 4, 5, and 6, 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, 4, 5, and 6).

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, non-Hodgkin's lymphoma (NHL), and soft tissue sarcoma (STS) were estimated as 0.07, 0.002, and 0.001 respectively. Thus, Table 7-6 shows at least a power of approximately 0.65 to detect a relative risk of 1.5 given an estimated prevalence of 0.07 for all cancers. For the estimated prevalences of NHL and STS, the power to detect a relative risk of 2.0 would be less than 0.20.

Table 7-7 and Appendix Tables D-6 through D-9 provide the same information as Table 7-6 and Appendix Table D-2 through D-5 at a significance level of 0.05 for continuous dependent variables in terms of coefficients of variation (100 times the standard deviation of

Table 7-6.
Approximate Power to Detect a Group Effect at a 5 Percent Level of Significance
(Discrete Dependent Variable)

Prevalence of Condition	Relative Risk						
	1.10	1.20	1.30	1.40	1.50	1.75	2.00
0.005	0.05	0.06	0.07	0.09	0.10	0.16	0.21
0.01	0.06	0.07	0.09	0.13	0.16	0.26	0.38
0.02	0.06	0.09	0.14	0.20	0.27	0.46	0.64
0.03	0.07	0.11	0.19	0.28	0.38	0.62	0.80
0.04	0.07	0.13	0.23	0.35	0.4	0.74	0.90
0.05	0.08	0.16	0.27	0.41	0.56	0.83	0.95
0.10	0.10	0.25	0.46	0.67	0.82	0.98	1.00
0.15	0.13	0.34	0.60	0.81	0.93	1.00	1.00
0.20	0.15	0.40	0.70	0.89	0.97	1.00	1.00

the dependent variable divided by the mean of the dependent variable) and the proportion mean changes. Table 7-7 presents power calculations for Model 1 (group) and Appendix Tables D-6 through D-9 presents power calculations for Model 2 (initial dioxin), Model 3 (categorized dioxin—low plus high Ranch Hand versus Comparison contrast), Model 4 (lipid-adjusted current dioxin), and Models 5 and 6 (whole-weight current dioxin). Power calculations were performed using the logarithm (base 2) of dioxin in Models 2, 4, 5, and 6, 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 7-7 and Appendix Table D-7 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 D-6, D-8, and D-9 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 D-6, D-8, and D-9 corresponds mathematically to the slope corresponding to initial or current 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, serum insulin (on the natural logarithm scale) for all participants has a coefficient of variation of approximately 22 percent. With this coefficient of variation, for the 952 Ranch Hands and 1,281 Comparisons in Model 1, the power is slightly greater than 0.80 for detecting a 13 percent increase in the mean serum insulin of Ranch Hands relative to the mean serum insulin level of Comparisons (mean change = 0.03).

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

Mean Change	Coefficient of Variation ($100\sigma/\mu$)				
	5	10	25	50	75
0.005	0.65	0.22	0.08	0.06	0.05
0.01	1.00	0.65	0.15	0.08	0.06
0.02	1.00	1.00	0.46	0.15	0.10
0.03	1.00	1.00	0.80	0.29	0.15
0.04	1.00	1.00	0.96	0.46	0.24
0.05	1.00	1.00	1.00	0.65	0.34
0.10	1.00	1.00	1.00	1.00	0.88

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 histories 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.

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 Gastrointestinal Assessment (Chapter 13) will be referenced throughout this discussion. The contents of each summary table depend on the form of the health status endpoint (i.e., whether the dependent variable under analysis is a continuous or discrete variable). Generally, the results of the various analyses will be summarized in subpanels within each table as specified in Table 7-8. The subpanel specifications may be slightly different when adjusted analyses are not performed. This section also provides an explanation of the information contained in these tables.

Continuous Variables

Table 13-12 presents an example of the results of analysis when the dependent variable is continuous. Subpanels (a) and (b) show the results of unadjusted and adjusted analyses that compare the means of a dependent variable between Ranch Hands and Comparisons. Contrasts between Ranch Hands and Comparisons also are presented within each occupational category (i.e., officer, enlisted flyer, and enlisted groundcrew).

**Table 7-8.
Location of Table Results from Different Analysis Models**

Models	Subpanel in Table	Exposure Estimate	Type of Analysis
1	a	Group ^a	Unadjusted
1	b	Group ^a	Adjusted
2	c	Initial Dioxin ^b	Unadjusted
2	d	Initial Dioxin ^b	Adjusted
3	e	Categorized Dioxin ^a	Unadjusted
3	f	Categorized Dioxin ^a	Adjusted
4,5,6	g	Current Dioxin ^b	Unadjusted
4,5,6	h	Current Dioxin ^b	Adjusted

^aRanch Hands and Comparisons.

^bRanch Hands only.

For the unadjusted analysis, continuous dependent variable samples sizes (n) and means 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 are 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. Confidence intervals are determined from analysis of variance models for all occupational categories combined and for each occupational category, assuming equal variances in each group (i.e., Ranch Hands, Comparisons). If the analyses were performed on a transformed scale, 95 percent confidence intervals on the differences of means are not presented and the column is footnoted. A p-value also is reported to determine whether a difference in means on the scale used for analysis for a specified contrast is different from zero. The p-values are determined from t-tests for all occupational categories combined and within each occupational category, assuming equal variances in each group, unless the test for equal variances is rejected and the significance (≤ 0.05 , > 0.05) of the t-test is dependent upon the equality of the variances.

For an adjusted analysis, the table is modified to include sample sizes, adjusted means, differences of adjusted means on the original scale and the associated 95 percent confidence interval (if the analysis was performed on the original scale), and any covariates and interactions retained in the final adjusted model along with their associated p-values. Sample sizes for corresponding panels of unadjusted and adjusted analyses may differ because of missing covariate information. Confidence intervals and p-values for each occupational category are determined from the group-by-occupation interaction in the final adjusted model using analysis of covariance techniques. Covariates with p-values less than or equal to 0.15 and interactions with p-values less than or equal to 0.05 retained in the final model after implementing the modeling strategy are presented in a covariate remarks section, along with the associated p-values.

Subpanel (c) of Table 13-12, for example, reports summary statistics assessing the association between the continuous dependent variable and initial dioxin without adjusting for covariate information. Sample sizes and means 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 footnote. The low, medium, and high categories are based on categorizing all Ranch Hands with initial dioxin estimates into three approximately equally-sized categories, based on their initial dioxin estimate. Means of the dependent variable adjusted for percent body fat at the time of duty in SEA and change in percent body fat from the time of duty in SEA to the date of the blood draw for dioxin also are presented for the low, medium, and high categories of initial dioxin. Based on the linear regression analysis, adjusted for percent body fat at the time of duty in SEA and change in percent body fat from the time of duty in SEA to the date of the blood draw for 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 means, slope, and standard error are footnoted and the transformation is identified in the footnote. The p-value associated with testing whether the estimated slope is equal to zero also is presented.

Based on analyses that incorporate covariate and interaction information, subpanel (d) reports summary statistics assessing the association between the continuous dependent variable and initial dioxin. Similar to the unadjusted analyses, sample sizes and adjusted means 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 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 \log_2 (initial dioxin), including percent body fat at the time of duty in SEA and change in percent body fat from the time of duty in SEA to the date of the blood draw for dioxin, and covariate and interaction effects, the coefficient of determination (R^2), the adjusted slope for \log_2 (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 adjusted slope is equal to zero also is presented. Covariates with p-values less than or equal to 0.15 and interactions with p-values less than or equal to 0.05 retained in the final model after implementing the modeling strategy are presented in a covariate remarks section, along with the associated p-values.

Subpanels (e) and (f) of Table 13-12, for example, show the results of unadjusted and adjusted analyses that compare the means of a continuous dependent variable for Ranch Hands having background, low, high, and low plus high dioxin levels with Comparisons having background current dioxin levels. The note at the bottom of the table defines the dioxin categories. The low and high Ranch Hand categories are based on categorizing all Ranch Hands with lipid-adjusted current 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. For the unadjusted analysis, sample sizes and dependent variable means are presented for each category. If the dependent variable was transformed for the analysis, the means of the

transformed values are converted to the original scale and the column heading is footnoted. Means of the dependent variable adjusted for percent body fat at the time of duty in SEA and change in percent body fat from the time of duty in SEA to the date of the blood draw for dioxin also are presented for each dioxin category. The mean for the low plus high category is a weighted average of the low Ranch Hand and high Ranch Hand category means, based on the low and high Ranch Hand category sample sizes. 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 analyses were performed on a transformed scale, the 95 percent confidence intervals on the differences of means are not presented and the column is footnoted. A p-value also is reported to determine whether a difference in means for a specified contrast is different from zero. The p-value is based on the difference of means on the scale used for analysis. Adjusted means, confidence intervals, and p-values for each contrast are determined from an analysis of variance model with covariate adjustments for percent body fat at the time of duty in SEA and change in percent body fat from the time of duty in SEA to the date of the blood draw for dioxin.

For an adjusted analysis, the table is modified to include adjusted means (adjusted for percent body fat at the time of duty in SEA, change in percent body fat from the time of duty in SEA to the date of the blood draw for dioxin, and covariates and interactions retained in the final model), differences in adjusted means on the original scale, 95 percent confidence intervals on the differences in adjusted means (if the analysis was performed on the original scale), and any covariates and interactions retained in the adjusted model along with their associated p-values. Covariates with p-values less than or equal to 0.15 and interactions with p-values less than or equal to 0.05 retained in the final model after implementing the modeling strategy are presented in a covariate remarks section, along with the associated p-values.

Subpanel (g) of Table 13-12, for example, reports summary statistics from three models (Models 4, 5, and 6) assessing the association between the continuous dependent variable and current dioxin without adjusting for covariate information. A lipid-adjusted current dioxin measurement is used for Model 4, and a whole-weight current dioxin measurement is used in Models 5 and 6. The linear regression model in Model 6 additionally adjusts for \log_2 (total lipids). Means of the dependent variable (transformed to the original units, if necessary) are presented for low, medium, and high categories of current dioxin. Dependent variable means for Model 6 are adjusted for \log_2 (total lipids). Samples sizes are presented immediately below the mean in each level of current dioxin for each model. The numerical values defining the low, medium, and high categories of current dioxin are specified in a table footnote. The low, medium, and high categories are based on categorizing all Ranch Hands with current dioxin levels into three approximately equal-sized categories, based on their current dioxin measurement. Based on a linear regression of the dependent variable on \log_2 (current 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 current dioxin measurements of 0 ppt or ppq. 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 estimated slope is equal to zero also is presented for each model.

Based on analyses that incorporate covariate and interaction information, subpanel (h) reports summary statistics assessing the association between the continuous dependent variable and current dioxin for Models 4, 5, and 6. Similar to the unadjusted analyses, sample sizes and adjusted means of the dependent variable (transformed to the original units, if necessary) are presented for low, medium, and high categories of current dioxin. The numerical values defining these categories are specified in a table 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 $\log_2(\text{current dioxin} + 1)$, including covariates and interactions (and $\log_2[\text{total lipids}]$ for Model 6), the coefficient of determination (R^2), the adjusted slope for $\log_2(\text{current dioxin} + 1)$ 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 adjusted slope is equal to zero also is presented for each model. Covariates with p-values less than or equal to 0.15 and interactions with p-values less than or equal to 0.05 retained in the multiple regression model after implementing the modeling strategy are presented in a covariate remarks section, along with the associated p-values, for each model.

For each of the six adjusted models, if the final model contains a significant group-by-covariate or dioxin-by-covariate interaction with an associated p-value less than or equal to 0.01, then the adjusted means, difference of means, 95 percent confidence interval, and p-value or adjusted slope, standard error, and p-value may not be reported. The entries for these statistics are reported as four asterisks (****) and are identified by a table footnote. Covariates and interactions retained in the model are, however, reported in a covariate remarks section. For some clinical assessments, an analyst may exercise discretion and report the adjusted means, difference of means, 95 percent confidence interval, and p-value from a model that excludes the interaction having a p-value less than 0.01. When these discretionary followup analyses are performed, the results are reported along with two asterisks (**) and are explained by a table footnote. If the final model contains a significant group-by-covariate or dioxin-by-covariate interaction with an associated p-value between 0.01 and 0.05, then the adjusted means, difference of adjusted means, 95 percent confidence interval, and p-value or the adjusted slope, standard error, and p-value are reported from a model that excludes that interaction. The entries for these statistics are reported along with two asterisks (**) accompanied by a table footnote. In either case (i.e., $p \leq 0.01$ or $0.01 < p \leq 0.05$), stratified analyses are undertaken and the results are reported in an associated appendix for each individual clinical area. The specific appendix table that presents the stratified analyses is referenced in a table footnote.

Discrete Variables

Discrete Variable With Two Categories

Table 13-3 presents an example of the results of analysis when the dependent variable is discrete and dichotomous in form. Subpanels (a) and (b) display the results of unadjusted and adjusted analyses that compare Ranch Hands and Comparisons on the relative frequency for a specified discrete dependent variable (e.g., percent of participants with an abnormal condition). 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, a sample size and relative frequency is presented for each group within each occupational category. For the contrasts of Ranch Hands versus Comparisons, estimated relative risks, associated 95 percent confidence intervals for the relative risks, and p-values associated with testing whether the risks equal 1.0 are presented. The normal distribution is used to calculate an approximate 95 percent confidence interval, and the continuity adjusted chi-square test is used to determine the corresponding p-value.

For an adjusted analysis, the table presents adjusted relative risks, 95 percent confidence intervals on the relative risks, and covariates and interactions retained in the adjusted model along with their associated p-values. Adjusted relative risks, confidence intervals, and p-values are determined from a multiple logistic regression model using the BMDP[®]-LR procedure, which utilizes the normal distribution for determining an approximate 95 percent confidence interval and the chi-square distribution based on a likelihood ratio statistic (17) for determining the p-value. Results from each occupational category are determined from the group-by-occupation interaction that is forced into the final model. Covariates (p-values less than or equal to 0.15) and interactions (p-values less than or equal to 0.05) retained in the multiple logistic regression model after implementing the modeling strategy are presented in a covariate remarks section, along with the associated p-values.

Subpanel (c) of Table 13-3, for example, reports summary statistics assessing the association between the dependent variable and initial dioxin 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 percentage of Ranch Hands with the specified dichotomous characteristic (as cited in the column heading) is calculated from the data and presented for the low, medium, and high initial dioxin categories. Based on the logistic regression model adjusted for percent body fat at the time of duty in SEA and change in percent body fat from the time of duty in SEA to the date of the blood draw for dioxin, an estimated relative risk and its associated 95 percent confidence interval are reported. The p-value associated with testing whether the relative risk is equal to 1.0 also is presented. The normal distribution is used to determine an approximate 95 percent confidence interval, and the chi-square distribution based on a likelihood ratio statistic is used for determining the p-value. The summary statistics are reported for initial dioxin divided into three categories, whereas the relative risk, confidence interval, and p-value are based on \log_2 (initial dioxin) in its continuous form.

Subpanel (d) reports summary statistics assessing the association between the discrete dependent variable and initial dioxin, adjusted for percent body fat at the time of duty in

SEA, change in percent body fat from the time of duty in SEA to the date of the blood draw for dioxin, and covariate and interaction information. The aggregate sample size (n) is presented for a multiple logistic regression of the discrete dependent variable on \log_2 (initial dioxin) including covariates and interactions in the adjusted model. Based on a multiple logistic regression model, the adjusted relative risk for \log_2 (initial dioxin) and its associated 95 percent confidence interval are reported. The p-value for testing whether the relative risk is equal to 1.0 also is presented. The normal distribution is used to determine an approximate 95 percent confidence interval, and the chi-square distribution based on a likelihood ratio statistic is used for determining the p-value. Covariates (p-values less than or equal to 0.15) and interactions (p-values less than or equal to 0.05) retained in the multiple regression model after implementing the modeling strategy are presented in a covariate remarks section, along with the associated p-values.

Subpanels (e) and (f) of Table 13-3, for example, show the results of unadjusted and adjusted analyses that contrast Ranch Hands having background, low, high, and low plus high dioxin levels with Comparisons having background current dioxin levels based on the relative frequency for a specified discrete dependent variable (e.g., percent of participants in a dioxin category with an abnormal condition). The note at the bottom of the table defines the dioxin categories. The low and high Ranch Hand categories are based on categorizing all Ranch Hands with lipid-adjusted current 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.

For the unadjusted analysis, a relative frequency and sample size is presented for each current dioxin category. For the individual contrasts of the Ranch Hand categories versus Comparisons, estimated relative risks, associated 95 percent confidence intervals for the relative risks, and p-values associated with testing whether the risks equal 1.0 are presented. The relative risks, confidence intervals, and p-values are determined from a logistic regression model, adjusted for percent body fat at the time of duty in SEA and change in percent body fat from the time of duty in SEA to the date of the blood draw for dioxin. The low plus high Ranch Hand versus Comparison contrast is based on a separate logistic regression model in which the low and high Ranch Hand categories are combined. The normal distribution is used to determine an approximate 95 percent confidence interval, and the chi-square distribution based on a likelihood ratio statistic is used for determining the p-value.

For an adjusted analysis, subpanel (f) of the table presents adjusted relative risks, associated 95 percent confidence intervals for the relative risks, and p-values associated with testing whether the risks equal 1.0 for the individual contrasts of the Ranch Hand categories versus Comparisons. Covariates (p-values less than or equal to 0.15) and interactions (p-values less than or equal to 0.05) retained in the multiple regression model after implementing the modeling strategy are presented in a covariate remarks section, along with the associated p-values.

Subpanels (g) and (h) of Table 13-3, for example, present summary statistics from three models assessing the association between the dependent variable and current dioxin. The current dioxin measurement in Model 4 is lipid-adjusted current dioxin. In Models 5 and 6,

the dioxin measurement is whole-weight current dioxin, where Model 6 also is adjusted for total lipids.

For the unadjusted analyses, the percentage of Ranch Hands with the specified dichotomous characteristic (as cited in the column heading) and sample sizes are presented for low, medium, and high categories of current dioxin for each of the three models. The low, medium, and high categories are based on categorizing all Ranch Hands with current dioxin levels into three approximately equal-sized categories, based on their current dioxin measurement. The numerical values defining these categories are specified in a table footnote. Based on each logistic regression model, an estimated relative risk and its associated 95 percent confidence interval are reported. The p-value associated with testing whether the relative risk is equal to 1.0 also is presented. The normal distribution is used to determine an approximate 95 percent confidence interval, and the chi-square distribution based on a likelihood ratio statistic is used for determining the p-value. The summary statistics are reported for initial dioxin divided into three categories, whereas the relative risk, confidence interval, and p-value are based on \log_2 (current dioxin + 1) in its continuous form.

Incorporating covariate and interaction information, subpanel (h) reports summary statistics assessing the association between the discrete dependent variable and current dioxin for each of the three models. The aggregate sample size (n) is presented for a multiple logistic regression of the discrete dependent variable on \log_2 (current dioxin + 1) including covariates and interactions in the final adjusted model. Based on the multiple logistic regression models, the adjusted relative risk for \log_2 (current dioxin + 1) and its associated 95 percent confidence interval are reported. The p-value for testing whether the relative risk is equal to 1.0 also is presented for each model. The normal distribution is used to determine an approximate 95 percent confidence interval, and the chi-square distribution based on a likelihood ratio statistic is used for determining the p-value. Covariates (p-values less than or equal to 0.15) and interactions (p-values less than or equal to 0.05) retained in the multiple regression model after implementing the modeling strategy are presented in a covariate remarks section, along with the associated p-values.

In each of the six adjusted models, if the multiple logistic regression model contains a significant group-by-covariate or dioxin-by-covariate interaction with an associated p-value less than or equal to 0.01, then the adjusted relative risk, 95 percent confidence interval, and associated p-value may not be reported. The entries for these statistics are reported as four asterisks (****) and are identified by a table footnote. Covariates and interactions retained in the model are, however, reported in a covariate remarks section. For some clinical assessments, an analyst may exercise discretion and report an adjusted relative risk, 95 percent confidence interval, and an associated p-value from a model that excludes the interaction having a p-value less than 0.01. When these discretionary followup analyses are performed, the results are reported along with two asterisks (**) and are explained by a table footnote. If the multiple logistic regression model contains a significant group-by-covariate or dioxin-by-covariate interaction with a p-value between 0.01 and 0.05, then the adjusted relative risk, 95 percent confidence interval, and associated p-value are reported from a model that excludes that interaction. The entries for these statistics are reported along with two asterisks (**) accompanied by a table footnote. In either case (i.e., $p \leq 0.01$ or

0.01 < p ≤ 0.05), stratified analyses are undertaken and the results are reported in an associated appendix for each individual clinical area. The specific appendix table that presents the stratified analyses is referenced in a table footnote.

Discrete Variable With More Than Two Categories

Polychotomous regression techniques were used to analyze discrete dependent variables having more than two levels (e.g., abnormal low, normal, abnormal high—see Table 13-48). Results are presented in a similar fashion to discrete variables with only two categories, except that percentages are presented for all levels of the dependent variable, including normal, and relative risks, confidence intervals, and p-values are presented for each contrast with the normal level of the dependent variable (e.g., abnormal low versus normal and abnormal high versus normal).

Subpanels (a) and (b) of Table 13-48, for example, display the results of unadjusted and adjusted analyses that compare Ranch Hands and Comparisons on the relative frequencies of each abnormal level for a specified discrete dependent variable (e.g., percent of participants with an abnormally high condition versus those with a normal condition and percent of participants with an abnormally low condition versus those with a normal condition). 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, a sample size is presented for each group within each occupational category. Relative frequencies are presented for each level of the dependent variable for each group within each occupational category. Therefore, for each group within each occupational category, the relative frequencies sum to 100 percent across the dependent variable categories. For the contrasts of Ranch Hands versus Comparisons, estimated relative risks, associated 95 percent confidence intervals for the relative risks, and p-values associated with testing whether the risks equal 1.0 are determined from the BMDP[®]-PR procedure and presented 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 analysis, the table presents adjusted relative risks, 95 percent confidence intervals on the relative risks, and covariates and interactions retained in the adjusted model along with their associated p-values. Covariates (p-values less than or equal to 0.15) and interactions (p-values less than or equal to 0.05) retained in the polychotomous regression model after implementing the modeling strategy are presented in a covariate remarks section, along with the associated p-values.

For the unadjusted and adjusted analyses relating discrete dependent variables having more than two categories to initial dioxin, subpanels (c) and (d) of Table 13-48, for example, present sample sizes, relative frequencies, relative risks, 95 percent confidence intervals for the relative risks, and associated p-values. For the adjusted analysis, any covariates and interactions retained in the model along with their associated p-values also are presented. One difference between the table presentations for dichotomous dependent variables and discrete dependent variables with more than two levels is that relative frequencies of Ranch Hands belonging to each of the dependent variable categories are summarized with respect to each initial dioxin category (i.e., low, medium, and high initial dioxin). Therefore, for each

initial dioxin level, the relative frequencies sum to 100 percent across the dependent variable categories. Also, associations with initial dioxin are presented for each abnormal level of the dependent variable (e.g., abnormal low vs. normal and abnormal high vs. normal).

Subpanels (e) and (f) of Table 13-48, for example, present unadjusted and adjusted analyses of categorized dioxin versus a discrete dependent variable having more than two categories. 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. Relative frequencies are presented for each level of the dependent variable for each dioxin category. Therefore, for each dioxin category, the relative frequencies sum to 100 percent across the dependent variable levels. For each contrast of a Ranch Hand category versus the Comparison group, estimated relative risks, associated 95 percent confidence intervals for the relative risks, and p-values associated with testing whether the risks equal 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 low plus high Ranch Hand versus Comparison contrast is based on a separate polychotomous logistic regression model in which the low and high Ranch Hand categories are combined. For an adjusted analysis, the table presents adjusted relative risks, 95 percent confidence intervals on the relative risks, and p-values for each contrast of Ranch Hands versus Comparisons under each abnormal level of the dependent variable. Covariates and interactions retained in the adjusted polychotomous model, along with their associated p-values, also are presented.

Similar to the polychotomous regression analysis using initial dioxin, unadjusted and adjusted analyses of discrete dependent variables with more than two categories were performed using current dioxin in Models 4, 5, and 6. Summaries of the analyses are given in subpanels (g) and (h) (see Table 13-48 for an example). The current dioxin measurement in Model 4 is lipid-adjusted current dioxin. In Models 5 and 6, the dioxin measurement is whole-weight current dioxin, where Model 6 also is adjusted for total lipids. For the unadjusted analysis, sample sizes are presented for each current dioxin level within each of the three models. Relative frequencies (within each current dioxin level) are presented for each dependent variable category. Estimated relative risks, 95 percent confidence intervals on the relative risks, and associated contrast p-values are reported for each abnormal level of the dependent variable (e.g., abnormal low vs. normal and abnormal high vs. normal) for all three models. Adjusted analysis results, including adjusted relative risks, 95 percent confidence intervals on the relative risks, and associated p-values for the abnormal dependent variable categories are presented on the following page of the table. Covariates and interactions retained in the adjusted polychotomous model, along with the associated p-values, also are presented for each of the three adjusted models.

In each of the six adjusted models, if the polychotomous regression model contains a significant group-by-covariate or dioxin-by-covariate interaction with an associated p-value less than or equal to 0.01, then the adjusted relative risk, 95 percent confidence interval, and associated p-value may not be reported. The entries for these statistics are reported as four asterisks (****) and are identified by a table footnote. Covariates and interactions retained

in the model are, however, reported under a covariate remarks section. If the polychotomous regression model contains a significant group-by-covariate or dioxin-by-covariate interaction with a p-value between 0.01 and 0.05, or when an analyst deems it appropriate to present results from a model with a group-by-covariate interaction having a p-value less than 0.01, then the adjusted relative risk, 95 percent confidence interval, and associated p-value are reported from a model that excludes that interaction. The entries for these statistics are reported along with two asterisks (**) accompanied by a table footnote. In either case (i.e., $p \leq 0.01$ or $0.01 < p \leq 0.05$), stratified analyses are undertaken and the results are reported in an associated appendix for each individual clinical area. The specific appendix table that presents the stratified analyses is referenced in a table footnote.

GRAPHICS

The analytic activities for the analyses were supplemented by data plots. These graphics were produced using the S-PLUS[®] graphics procedure (24).

As part of the analyses of current dioxin, bivariate scatterplots were produced describing the relationship between selected dependent variables and the logarithm (base 2) of lipid-adjusted current dioxin + 1. Both the dependent variable and current dioxin are displayed in continuous form. The dependent variable transformation used in the analysis also has been used in the scatterplots. Participants excluded from the analysis are not displayed on these scatterplots, and consequently the graphical displays parallel the Model 4 analyses. These scatterplots are presented in Appendix Q-2.

CHAPTER 7 REFERENCES

1. Lathrop, G.D., W.H. Wolfe, R.A. Albanese, and P.M. Moynahan. 1984. The Air Force Health Study: An epidemiologic investigation of health effects in Air Force personnel following exposure to herbicides: Baseline morbidity study results. NTIS: AD A 138 340. USAF School of Medicine, Brooks Air Force Base, Texas.
2. Lathrop, G.D., S.G. Machado, T.G. Karrison, W.D. Grubbs, W.F. Thomas, W.H. Wolfe, J.E. Michalek, J.C. Miner, and M.R. Peterson. 1987. Epidemiologic investigation of health effects in Air Force personnel following exposure to herbicides: First followup examination results. NTIS: AD A 188 262. USAF School of Medicine, Brooks Air Force Base, Texas.
3. Thomas, W.F., W.D. Grubbs, T.G. Karrison, M.B. Lustik, R.H. Roegner, D.E. Williams, W.H. Wolfe, J.E. Michalek, J.C. Miner, and R.W. Ogershok. 1990. An epidemiologic investigation of health effects in Air Force personnel following exposure to herbicides: 1987 followup examination results, May 1987 to January 1990. NTIS: AD A 222 573. USAF School of Medicine, Brooks Air Force Base, Texas.
4. Roegner, R.H., W.D. Grubbs, M.B. Lustik, A.S. Brockman, S.C. Henderson, D.E. Williams, W.H. Wolfe, J.E. Michalek, and J.C. Miner. 1991. The Air Force Health Study: An epidemiologic investigation of health effects in Air Force personnel following exposure to herbicides: Serum Dioxin Analysis of 1987 Examination Results, NTIS: AD A 237 516-24. USAF School of Medicine, Brooks Air Force Base, Texas.
5. Poiger, H., and C. Schalatter. 1986. Pharmacokinetics of 2,3,7,8-TCDD in man. *Chemosphere* 15:1489-94.
6. Pirkle, J.L., W.H. Wolfe, D.G. Patterson, L.L. Needham, J.E. Michalek, J.C. Miner, M.R. Peterson, and D.L. Phillips. 1989. Estimates of the half-life of 2,3,7,8-tetrachlorodibenzo-p-dioxin in Vietnam veterans of Operation Ranch Hand. *Journal of Toxicology and Environmental Health* 27:165-71.
7. Kissel, J.C., and G.M. Robarge. 1988. Assessing the elimination of 2,3,7,8-TCDD from humans with a physiologically based pharmacokinetic model. *Chemosphere* 17:2017-27.
8. Akins, J.R., K. Waldrep, and J.T. Bernert, Jr. 1989. The estimation of total serum lipids by a completely enzymatic "summation" method. *Clinica Chimica Acta* 184:219-26.
9. Patterson, Jr., D.G., L. Hampton, C.R. Lapeza, W.T. Belser, V. Green, L. Alexander, and L. Needham. 1987. High resolution gas chromatographic/high-resolution mass spectrometric analysis of human-serum on a whole-weight and lipid basis for 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Annals of Chemistry* 59:2000-05.

10. Patterson, D.G., L.L. Needham, J.L. Pirkle, D.W. Roberts, J. Bagby, W.A. Garret, J.S. Andrews, Jr., H. Falk, J.T. Bernert, E.J. Sampson, and V.N. Houk. 1988. Correlation between serum and adipose tissue levels of 2,3,7,8-tetrachlorodibenzo-p-dioxin in 50 persons from Missouri. *Archives of Environmental Toxicology* 17:139-43.
11. Knapik, J.J., A.R.L. Burse, and J.A. Vogel. 1983. Height, weight, percent body fat, and indices of adiposity for young men and women entering the Army. *Aviation, Space, and Environmental Medicine* 54:223-31.
12. Box, G.E.P., and D.R. Cox. 1964. An Analysis of Transformations. *Journal Royal Statistical Society* B26:211-252.
13. *SAS/STAT Guide for Personal Computers*. Version 6, 4th ed. 1990. Cary, North Carolina: SAS Institute, Inc.
14. *BMDP Statistical Software Manual*. 1992. Berkeley, California: University of California Press.
15. Bishop, Y.M.M., S.E. Feinberg, and P.W. Holland. 1975. *Discrete multivariate analysis: Theory and practice*. Cambridge: MIT Press.
16. Feinberg, S.E. 1981. *The analysis of cross-classified data*. Cambridge: MIT Press.
17. Neter, J., W. Wasserman, and M. Kuntner. 1975. *Applied linear statistical models*. Homewood, Illinois: Richard D. Irwin, Inc.
18. Cox, D.R. 1984. *Analysis of Survival Data*. London: Chapman and Hall.
19. Hosmer, D., and S. Lemeshow. 1989. *Applied Logistic Regression*. New York: John Wiley & Sons.
20. Agresti, A. 1990. *Categorical Data Analysis*. New York: John Wiley & Sons.
21. Searle, S.R., F.M. Speed, and G.A. Milliken. 1980. Population marginal means in the linear model: An alternative to least squares means. *The American Statistician* 34:216-21.
22. Fleiss, J.L. 1986. *The Design and Analysis of Clinical Experiments*. New York: John Wiley & Sons.
23. Rosenbaum, P.R. 1984. The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society* 147:656-66.
24. Statistical Sciences, Inc. 1993. *S-Plus for Windows User's Manual, Version 3.1*. Seattle: Statistical Sciences, Inc.