

## **CHAPTER 1**

### **INTRODUCTION**

#### **AIR FORCE HEALTH STUDY**

The Air Force Health Study (AFHS) is an epidemiologic investigation to determine whether occupational exposure to Herbicide Orange in a group of U.S. Air Force personnel is associated with adverse health effects. During the Vietnam conflict, Herbicide Orange was the primary herbicide used in a military operation, code-named Operation Ranch Hand, which disseminated the herbicide through aerial spraying for purposes of defoliation and crop destruction.

As documented in prespecified analytical plans and predecessor reports, the AFHS is based on a cohort design in a nonconcurrent prospective setting. The study design consisted of a baseline morbidity assessment that is to be complemented by five followup morbidity evaluations over a 20-year period. The baseline morbidity evaluation, conducted in 1982, was performed by the Air Force. Followup evaluations were conducted in 1985 and 1987. The 1985 and 1987 evaluations (also known as the third- and fifth-year studies, respectively) were performed, under contract to the Air Force, by Science Applications International Corporation (SAIC), in conjunction with Scripps Clinic and Research Foundation (SCRF) and the National Opinion Research Center (NORC). Future evaluations are planned for 1992, 1997, and 2002 (i.e., the 10-year, 15-year, and 20-year followup studies, respectively).

For the Baseline and the 1985 and 1987 studies, the major focus of the analyses was to compare the health status of the Ranch Hands (i.e., the exposed cohort) with that of the Comparisons (i.e., the unexposed cohort). An ancillary analysis used an approximate estimate of exposure (low, medium, and high) that was constructed for each Ranch Hand using historical military record information with herbicide procurement and usage records. For the most part, the constructed exposure index failed to display consistent and/or meaningful dose-response relationships.

During the conduct of the 1987 physical examination, the Air Force initiated a collaborative study with the Centers for Disease Control (CDC) to measure dioxin levels in the serum of Ranch Hands and Comparisons. The purpose of this report is to perform a thorough statistical evaluation to assess dose-response relationships between various measures of dioxin and approximately 300 health-related endpoints in 12 clinical areas. The statistical analyses associated with the serum data will evaluate the association between a specified health endpoint and dioxin among the Ranch Hands, as well as contrast the health of various categories of Ranch Hands having differing serum dioxin levels with the health of Comparisons having background levels of dioxin in their blood. The analysis of dose-response relationships based on serum assays provides an important enhancement over the previous AFHS investigations. This research is the first large-scale study of dose-response effects based on an accurate measurement of current dioxin. The results of this study supplement the findings of previous AFHS reports, which have focused on group contrasts between exposed and unexposed cohorts, rather than on the dose-response relationships in this report.

Of the 995 Ranch Hands who were fully compliant to the 1987 physical examination, 932 had serum specimens analyzed by CDC; 64 of these 932 specimens were reported by CDC as not quantifiable by the analytical method. Two of the 932 participants provided blood but were not part of the 1987 examination. The Ranch Hand participants used for the statistical analyses of the serum data excluded the 66 Ranch Hands specified above. Thus, the serum levels of the remaining 866 Ranch Hands were candidates for evaluating the association between health status and level of dioxin. Current dioxin levels exceeded 5 ppt for 742 of the Ranch Hands, and exceeded 10 ppt for 521 Ranch Hands. These two Ranch Hand groups are the maximal and minimal cohorts, described later in this chapter.

Of the 1,299 Comparisons who completed the 1987 physical examination, 1,198 had serum specimens analyzed by CDC. Dioxin assay information on a randomly selected subset of 888 Comparisons was received from CDC by January 1990, at which time statistical analyses involving Comparison data began. Eighty-three of the 887 Comparisons who completed the physical examination had a current dioxin level reported by CDC as not quantifiable. Therefore, 804 Comparisons were candidates for use in the statistical analyses.

An additional 314 Comparison dioxin assay results were subsequently received. Of these results, 311 were based on Comparisons who had completed the physical examination, and 3 were reanalyses of specimens of 3 Comparisons who completed the examination but whose dioxin result was indeterminant.

Chapter 2, Dioxin Assay, contains a more complete discussion of the dioxin assay, the 888 and the subsequently received 314 Comparison assay results.

### **Questionnaire Methodology**

One source of information used in the statistical analyses for the AFHS was the participant questionnaire. For the 1982 Baseline study, the questionnaire was administered at the participant's home. The questionnaires of the 1985 and 1987 followup cycles were administered at the physical examination site. New participants or participants who refused to take part in the 1982 and 1985 examinations had the option of responding to the Baseline questionnaire either at their residence or at the physical examination site. The instruments provided baseline or updated information on such items as: demographic characteristics, education, occupation, medical history, study compliance, toxic exposures, reproductive experience, personality type, sleep disorders, and risk factors for skin cancer. For a detailed discussion of the development, expansion, and implementation of the questionnaire (i.e., interviewer training, scheduling of participants, data collection, and data processing), the reader is referred to Chapter 3, Questionnaire Methodology, AFHS 1987 examination (1).

### **Physical Examination Methodology**

Another major source of information for the analyses in the AFHS resulted from the various health evaluations performed at SCRf in 1987. The evaluations consisted of the following major elements:

- Review-of-systems questionnaire
- Psychological testing

- Physical examination
- Laboratory testing
- Specialized testing (e.g., phlebotomy for measurement of serum dioxin)
- Psychological and medical outbriefings.

The logistical efforts involved in contacting, transporting, and examining the study participants for the 1987 phase of the AFHS are described in Chapter 4, Physical Examination Methodology, of the AFHS 1987 examination report (1).

During the clinical examinations, data were collected in the laboratory and by a general and two subspecialty (dermatological and neurological) examinations. In the clinical laboratory, cutpoints between normal and abnormal measurements are in most cases well defined. In the physical examinations that were conducted by multiple examiners, however, some subjective variation in data collection would be anticipated. By adhering to a strict examination protocol and by blinding the examiners to the exposure status of all participants, a group bias was avoided.

The format of the physical examination was designed to address the wide range of body organ systems suggested by the scientific literature on both human and animal studies, the spectrum of health problems reported by Vietnam Veterans listed in the Agent Orange Repository of the Department of Veterans Affairs, and concerns expressed in the press. The examiners were kept strictly unaware of the exposure status of each participant and were required to conduct their examinations in a standardized and consistent manner. Each participant was provided with all of his examination results by a specialist in internal medicine and a clinical psychologist. Whenever a condition requiring prompt medical followup or further evaluation was identified by one of these debriefers, arrangements and appointments were made with a referral physician before the participant departed from the clinic. In this manner, continuing treatment of important medical conditions was not overlooked.

### **Quality Control**

Throughout the 1987 examination, a number of steps were taken to maintain stringent quality control (QC) and quality review standards. In general, quality assurance (QA) activities were defined and implemented in the areas of administrative QA; questionnaire, physical, and psychological examination QC; laboratory QC measures; data management QC; and statistical QC. Chapter 6, Quality Control, of the AFHS report on the 1987 examination contains detailed descriptions of these quality control efforts (1).

### ***Administrative Quality Control***

For the 1985 and 1987 examinations, and the associated serum dioxin analyses presented in this report, an internal Quality Review Committee (QRC) was convened by the prime contractor. QRC members provided independent reviews and comments on draft report materials submitted to the Air Force. The QRC also provided advice on issues that might affect study quality.

### ***Questionnaire, Physical, and Psychological Quality Control***

For administration of the 1987 questionnaires, interviewers were provided specific training and detailed instructions by NORC on conducting the interviews. In addition, schedulers were trained to perform initial contacts with individuals to invite them to participate in the 1987 examination cycle. Conversion specialists were used to contact refusals or to identify replacements for unwilling Comparisons. Site supervisors monitored a sample of interviews from each interviewer. If necessary, immediate onsite retraining was provided for interviewers to ensure proper administration of the questionnaire. A rigorous review process for monitoring the completeness and quality of responses to the questionnaire items was followed.

After the questionnaires were reviewed for completeness and data validity, the questionnaire and physical examination records were provided to the Air Force for medical coding of the reported information. Once the medical coding was completed, the questionnaire information was provided to NORC for data processing. Various edit and data verification procedures were performed and discrepancies were resolved on a case-by-case basis. All corrections were documented and entered into the data base. QA reports were generated monthly and the review process was continued until no errors or discrepancies were found.

The physical examination provided most of the health status information used for clinical and statistical evaluation. Hence, a number of steps were taken to guarantee the quality and completeness of the information generated during the physical examination. The steps included a stringent selection process for all personnel directly involved with the study participants; a complete pretest of the physical examination, interview, psychological test, and laboratory test procedures before the start of the study; refresher training for diagnostic procedures (e.g., to diagnose chloracne); weekly review of participant critique forms; timely review, and revision if necessary, of items reported on the physical examination forms; and daily monitoring of clinical examination activities by the onsite Air Force monitor and the SCRF Medical Project Director.

### ***Clinical Laboratory and Immunology Laboratory Quality Control***

Multiple actions were implemented in the area of QC for the clinical laboratory. An integrated medical laboratory management information system was used to provide direct device to data base interfaces for automated testing equipment; stringent calibration standards were maintained for all automated equipment; control samples were used to monitor test quality; formal analysis and review of QC data was performed on a weekly basis; and CUSUM and FIR CUSUM techniques were used to detect calibration problems. A stringent QC procedure was also implemented in the cellular immunology component of the AFHS to address problems in assay performance, reagent validity, data analysis, and results reporting. Chapter 6 of the 1987 examination report provides an indepth discussion of the clinical and immunologic QC procedures (1).

### ***Data Management Quality Control***

The QC program for the data management activity consisted of multiple checks at all steps of the examination, data collection, and data processing cycle. Data QC procedures for data collection, conversion, and integration were developed before the clinical examinations

began. Pretesting of forms, procedures, and logistical arrangements was conducted 3 weeks before the examinations actually began.

Five interwoven layers of QC were instituted to ensure data integrity: data processing system design; design and administration of all exams or questionnaires; data completeness checks; data validation techniques; and quality control medical records coding.

### ***Statistical Analysis Quality Control***

QC was exercised in the following areas addressing the statistical analysis: construction of data bases for the statistical analysis of each clinical chapter, the statistical analysis, and the preparation of the clinical chapters containing the results of the statistical analyses. Each clinical area data base was examined for extreme and improbable values. Discrepancies were resolved through contact with the organization responsible for the data item of interest (e.g., SCRF or NORC). Technical issues related to statistical analysis were discussed, and resolved through frequent telephone and/or written communications between the SAIC statisticians and the Air Force principal investigators. The content of the report was verified for accuracy and validity among the reported text and tables, and for consistency with the output results generated by the statistical software.

### **Statistical Models**

The serum dioxin measurements were used in three different ways to assess the relationships between current health status and dioxin. Within a specified clinical area, the results of three analyses performed for each dependent variable were described under sections titled:

- Model 1: Ranch Hands -  $\text{Log}_2$  (Initial Dioxin)
- Model 2: Ranch Hands -  $\text{Log}_2$  (Current Dioxin) and Time
- Model 3: Ranch Hands and Comparisons by Current Dioxin Category.

Models 1 and 2 used serum dioxin values for only the Ranch Hands. For model 1, the dependent variable for each Ranch Hand was regressed on an initial dioxin level. The initial dioxin value was estimated retrospectively from a first-order pharmacokinetic half-life model using the measured current dioxin, the estimated half-life of 7.1 years (2) and time since the end of each Ranch Hand's tour of duty in Vietnam. For model 2, regression relationships were developed between the dependent variable for each Ranch Hand and the measured current dioxin level and time since the end of the tour in Vietnam. The latter model was implemented as an alternative to model 1 which was based on assuming a particular half-life model. Both of these models were implemented with and without adjustment for covariate information. While the overall analysis in model 2 specifically assesses the effect of differences between time strata, a current dioxin effect can be seen in the time stratified portions of the analyses as well.

Models 1 and 2 were also applied under two assumptions concerning exposure: the minimal assumption and the maximal assumption. Under the minimal assumption, the analyses are based on those Ranch Hands with current dioxin levels above 10 ppt. The basis

for the minimal assumption is that Ranch Hands currently having dioxin levels at or below 10 ppt are assumed not to have been exposed to dioxin during their Ranch Hand tour. Under the maximal assumption, the analyses are based on Ranch Hands with current dioxin levels above 5 ppt. The maximal assumption presumes that Ranch Hands with levels between 5 ppt and 10 ppt were only exposed to such an extent that their body burden of dioxin has just recently decayed to levels equivalent to normal background. Ranch Hands with current dioxin levels at or below 5 ppt were excluded from the analyses because of concerns raised by the CDC regarding the validity of the half-life model to extrapolate initial dioxin levels using such low dioxin levels. The minimal assumption is an attempt to focus the analyses on Ranch Hands who are more likely to have been exposed during their tour. The maximal assumption focuses on those participants known to be part of Operation Ranch Hand but the analyses may include some participants who possibly may not have been exposed to dioxin during their tours. Each assumption defines the size of the Ranch Hand groups being analyzed. The use of the terms "minimal" and "maximal" should not be interpreted as identifying those participants with a particular level or magnitude of dioxin exposure.

The analyses identified under model 3 compare the health of Ranch Hands with current dioxin values categorized as unknown (current dioxin at or below 10 ppt), low (current dioxin above 15 ppt but not above 33.3 ppt), and high (current dioxin above 33.3 ppt) with Comparisons having background levels (current dioxin at or below 10 ppt). "Unknown" is used as a description for Ranch Hands with current serum dioxin levels at background. Ranch Hands with current dioxin levels at or below 10 ppt were placed in a separate category (i.e., unknown) because the exposure resulting from their Vietnam tour could not be differentiated from background levels. Separating the unknown and low exposure categories by 5 ppt reduces concerns about the assignment of a Ranch Hand to either of the categories when the current level is very near a defined cutpoint. To remove any doubt about possible exposure in the Comparison group, any Comparisons having a current dioxin level above 10 ppt were excluded. Eighteen Comparisons had a current dioxin level above 10 ppt. Chapter 3 graphically displays distributions of serum levels for Ranch Hands and Comparisons.

### **Organization of the Report**

This report is organized as follows:

- Chapter 1 (Introduction) provides summary background information on AFHS and the serum dioxin analysis; and discusses specific technical items/issues that may affect the results of the different clinical area assessments.
- Chapter 2 (Dioxin Assay) describes the blood draw procedure used to determine the serum dioxin measurements; the analytical method used to determine the dioxin level from the serum; and QC procedures associated with the serum dioxin data.
- Chapter 3 (Relationship of Estimates of Dioxin and Exposure Index) provides a comparison of the constructed exposure index used in previous reports to the estimates of dioxin body burden used in this report.
- Chapter 4 (Statistical Methods) documents the statistical methods used in the individual clinical area assessments; and the statistical procedures and results of the half-life analyses performed by the Air Force.

- Chapter 5 (Covariate Associations) examines the associations between dioxin and the individual covariates used in the different clinical assessments.
- Chapters 6 through 17 present the results and medical discussion for each clinical area from the statistical analyses of the dependent variables using the three models described earlier in this chapter. Each chapter contains a brief overview of pertinent scientific literature. More detailed summaries can be found in the report of the 1987 examination (1).
- Chapter 18 (Conclusions) summarizes the findings and medical discussion of the statistical analyses performed for each of the 12 clinical areas.
- Chapter 19 (Future Directions) summarizes the anticipated future activities, and possible modifications to the existing instruments and methodologies used to investigate the association between health status and dioxin exposure.

## **INTERPRETIVE CONSIDERATIONS**

When interpreting the data presented in this report, careful consideration must be given to bias, interactions, consistency, multiple testing, dose-response patterns, trends, power limitations, strength of association, and biological credibility. Problems in evaluating negative results, extrapolating to other populations, and summarizing results also should be considered.

### **Bias**

With the introduction of the dioxin assay as the measure of exposure, important sources of bias are reduced to violations of the underlying assumptions of the three models upon which all analyses in this report are based. Closely associated with violation of assumptions is the possibility that an important covariate may have been overlooked.

Biased results will be produced if the assumptions underlying any of the three statistical models are violated. Of the three models, model 1 (see Chapter 4, Statistical Methods) is the most vulnerable to this kind of bias, since it depends directly on two unvalidated assumptions: (a) that dioxin elimination is by first-order pharmacokinetics and (b) that all Ranch Hands have the same dioxin half-life (7.1 years). If dioxin elimination is first-order, but some Ranch Hands have a shorter half-life than others (as suggested by unpublished analysis of paired dioxin measurements on 36 Ranch Hands, see Chapter 4, pages 4-9 through 4-12), then there would have been misclassification of initial dioxin exposure. If the clinical endpoint is not associated with a factor (e.g., relative weight change) that affects the elimination rate, then estimates of the odds ratio for common diseases associated with low and high levels of initial dioxin will, in general, be biased toward unity. However, if the clinical endpoint is associated with a factor that affects the elimination rate, then the odds ratio will be biased away from unity.

The validity of the constant half-life assumption cannot be assessed until the half-life study is expanded to all 500 Ranch Hands with current levels above background (above 10 ppt). Paired dioxin measurements on each of these 500 Ranch Hands, one derived from frozen serum samples collected in 1982 and the other from serum collected in 1987, will permit investigation of half-life variability with changes in weight, percent body fat, and disease since exposure. Assessment of the first-order elimination assumption will be based

on up to five dioxin measurements collected serially on each of 20 males who were exposed during a factory explosion near Seveso, Italy (3). The additional Air Force and Seveso data will be available in 1991.

Estimates of health effects derived from model 2 also could be biased if, for example, some Ranch Hands were fast dioxin eliminators (have a short dioxin half-life) and some were slow eliminators (have a long half-life). If this phenomenon was associated with a covariate (e.g., relative weight change between 1982 and 1987), lack of adjustment for this covariate would bias estimates of the slope or relative risk toward the null values (slope=0 and relative risk=1). Further investigation of this possibility will occur during the expanded half-life study, which is scheduled to begin in early 1991. A similar concern arises regarding estimates of effect derived from model 3. If, for example, a health effect was expressed many years after exposure, such an effect would probably be apparent in contrasts in disease rates between the background group and Ranch Hands in the high current dioxin category with the earliest tours of duty. The categorized current dioxin analyses were not adjusted for time since tour, however. Hence, it might not be possible to detect such an effect with that model because time since tour was not used for adjustment. This shortcoming is partially overcome by analyses based on model 2, which are adjusted for time since tour and the interaction between current dioxin and time.

Information bias, represented by overreporting disease symptoms, was precluded by verifying all diseases and conditions with medical records. It is possible that Ranch Hand conditions may be more verifiable because they may have been seen by physicians more often than Comparisons; this would be revealed by group differences in the quantity and content of medical records. Because currently there is no way to quantify these aspects, this potential source of bias remains unexplored. This source, however, if it exists, would affect only estimates of health effects derived from model 3 because Comparison data were not used in the model 1 and model 2 analyses. Information bias due to errors in the data introduced through data entry or machine error is negligible. All laboratory results were subject to strict quality control procedures. Medical coding data were verified completely by medical record review.

### **Adjustments for Covariates and Interactions**

In previous reports, the focus was on overall group contrasts between all Ranch Hands and all Comparisons, which took advantage of the matched design. In those analyses, the matching variables age, race, and occupation were eliminated effectively as confounders. The present dioxin analyses within Ranch Hands and the categorized current dioxin analyses within Ranch Hands and Comparisons are not benefited by the matched design. 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.

In addition, some covariates (e.g., percent body fat) may themselves be associated with current dioxin level and, perhaps, through their relationship with dioxin, may be related to the dependent health variable. In this situation, analyses of covariance adjusted for such a covariate are not valid, since the assumed independence of the "treatment" (current or initial dioxin) and the covariate is not met (4). There is no recourse but to analyze the data with

and without adjustment for the covariate; both analyses potentially are biased. Thus, unadjusted analyses must be viewed with caution and circumspection. 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 dioxin and the covariates to determine whether the relationship between dioxin and the dependent variable (health-related endpoint) differed across levels of the covariate. In many instances the clinical importance of a statistically significant 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. It is recognized that due to the large number of dioxin-by-covariate interactions that were examined for approximately 300 variables, some of the dioxin-by-covariate interactions judged significant at the 0.05 level might be spurious (i.e., chance occurrences not of biological or clinical relevance). This should be considered when significant dioxin-by-covariate interactions are interpreted. It is important that the size of the p-value associated with each dioxin-by-covariate 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 Chapter 4).

### **Consistency**

Ideally, an adverse health effect in Ranch Hands attributable to herbicide or dioxin would be revealed by internally and externally consistent findings. An internally consistent finding does not contradict prior information, other findings, or medical knowledge. An externally consistent finding has been established either previously in theory or empirically as related to exposure.

The findings of positive trends of increasing abnormalities with increasing levels of current dioxin with regard to lipids, percent body fat, and diabetes are internally consistent. The observed associations between dioxin and Millon Clinical Multiaxial Inventory scale scores appear inconsistent and isolated. They are not consistent between themselves or with known patterns of psychological disorder.

### **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. In the hypothetical case when Ranch Hand physical health is not related to dioxin, about 5 percent of the many statistical tests of hypotheses (dioxin effects and dioxin-by-covariate interactions) shown in this report should be expected to detect an association between dioxin and health in Ranch Hands ( $p$ -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 large studies. Unfortunately, there is no statistical procedure available to distinguish between those statistically significant results that arise due to the multiple testing artifact and those that may be due to a bona fide 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 credibility.

## **Trends**

Assessing consistent and meaningful trends is essential when interpreting any large study with multiple endpoints, clinical areas, and covariates. However, caution must be used when assessing trends. Increased numbers of abnormalities or means with increased dioxin levels across medically related variables within a clinical area might indicate a dioxin effect. In this case, it is important to note that there is a moderate-to-strong correlation between some endpoints. Hence, the strength of the trends also must be considered when assessing the suspected association.

## **Power Limitations**

The fixed size of the Ranch Hand cohort limits the ability of this study to detect a dioxin association. This limitation is most obvious concerning specific types of cancer, such as soft tissue sarcoma and non-Hodgkin's lymphoma, which are so uncommon that fewer than two cases are expected in this study, indicating that this study has virtually no statistical power to detect low-to-moderate associations (relative risks less than 5) with dioxin. On the other hand, these sample sizes are sufficient to detect very small mean shifts in the continuously distributed variables (see Chapter 4). For example, with regard to IgG, this study has approximately 90 percent power to detect a mean shift of 1 percent. The detection of significant mean shifts without a corresponding indication of increased Ranch Hand abnormalities or disease is considered to be of little importance or it may be an artifact of multiple testing. 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 prevalences of at least 5 percent in unexposed populations.

In an attempt to overcome the lack of power to detect group differences for specific types of systemic cancer, 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.

## **Strength of Association**

Ideally, an adverse effect, if it exists, would be revealed by a strong association between categorized current dioxin and a disease condition; that is, by a statistically significant relative risk greater than 2.0 for Ranch Hands in the high current dioxin category relative to the unexposed Comparisons (5). Statistically significant relative risks less than 2.0 are considered to be less important than larger risks because the relative risks less than 2.0 can easily arise due to unperceived bias or confounding. Relative risks greater than 5.0 are less subject to this concern. The numbers 2 and 5 are rules of thumb regarding analyses of association between a dichotomous endpoint (disease, no disease) and dichotomized exposure (exposed, unexposed). No such rules have been published regarding the analysis of continuously distributed endpoints (such as cholesterol) versus continuously distributed exposure (such as initial or current dioxin in models 1 and 2).

## **Biological Credibility**

The assessment of biological credibility requires consideration of the following question. In biological terms, can it be understood how the exposure under study could produce the effect of interest? While a lack of biological credibility or even a contradiction of biological knowledge can lead to the dismissal of a significant result, the failure to perceive a

mechanism may reflect only ignorance of the state of nature. On the other hand, it is easy to ascribe biological mechanisms that relate almost any exposure to almost any cancer. Thus, while pertinent, the response to this question is not always convincing.

### **Interpretation of Negative Results**

A 1985 study (6) presents minimal sample-size criteria for proof of safety and hazard in studies of environmental and occupational exposures. The study was directed at rectifying widespread misconceptions about proof of safety in the medical and scientific establishments and in other groups involved in public health and safety. Thus, a lack of significant results relating dioxin to a particular disease only means that this study is unable to detect a relationship between dioxin and health. This does not imply that a relationship does not exist, but that, if it does exist, it was not detected. A lack of significant results does not mean that dioxin is safe or that there is no relationship between dioxin and health, because this study is not designed, nor was it intended, to establish safety. This study was designed to determine whether a hazard existed for the exposed personnel and not whether dioxin was "safe."

### **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 two 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" (initial or current 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 measurement error. Thus, in most analyses, the values of  $R^2$  in this study 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.

### **Minimal versus Maximal Results**

The minimal and maximal assumptions for Ranch Hands having background dioxin levels ( $\leq 10$  ppt) were imposed to address the unknown exposure history of this subgroup. There were 345 Ranch Hands in this "unknown" category. In the minimal analyses, all of these were excluded from the data set. In the maximal analyses, only those with less than or equal to 5 ppt ( $n=124$ ) were excluded. The intent of these two analyses was to "trap" the true dioxin versus health relationship between them. The results of the maximal analyses

appear to be statistically significant more often than those of the minimal analyses. This could be due to the larger sample size of the maximal cohort or it could be due to the uncertainty of true exposure in Ranch Hands between 5 ppt and 10 ppt. There are no additional data available at this time with which to resolve these two interpretations.

### **Graphics**

The histograms, scatter plots, and graphical descriptions of interactions were included as aids to interpretation. The graphics alone are not sufficient to assess the relationship between dioxin and health. For example, a trend may be seen in a plot, but it could be statistically nonsignificant because the number of abnormalities is small. On the other hand, a statistically significant result can be clarified by the graphics, especially if the result depends on a few data points that appear far from the main cluster. Such points are termed "outliers" by statisticians. Outside of the initial quality control review activities, no additional effort was made to identify statistically significant outliers in this report.

### **The Checkmark Pattern**

In many model 3 analyses, the "unknown" Ranch Hand group has the lowest percentage of abnormalities; this phenomenon is termed "the checkmark pattern." These patterns are interesting but are without explanation at this time. Some reanalyses were accomplished with adjustment for military rank (officers, enlisted personnel), but the checkmark pattern remained after adjustment. This effect will be a subject of continued focus in future reports.

### **Extrapolation to Army Ground Troops**

Extrapolation of the serum dioxin results to the general population of ground troops who served in Vietnam is difficult because Ranch Hand and ground troop exposure situations were quite different. Based on serum dioxin testing results done by CDC (7) and others (8), nearly all ground troops tested have current levels of dioxin similar to background levels. Even ground troops who served in herbicide-sprayed areas of Vietnam had current levels indistinguishable from levels in men who never left the United States (with means of 4.2 ppt and 4.1 ppt, respectively). The AFHS subgroup most like the ground troops in terms of current dioxin levels are Ranch Hands who currently have background levels of dioxin (10 ppt or less—designated as the "unknown" current dioxin category in the model 3 analyses). Therefore, if the results of the AFHS are applied to the general population of Vietnam veterans, the focus should be on the unknown Ranch Hand versus background Comparison contrast in the model 3 analyses. However, extrapolating the results of these analyses to Vietnam veterans should still be made cautiously. There may be demographic distinctions between the unknown group of Ranch Hands and other Vietnam veterans that may be related to health. Also, if Ranch Hands in the unknown current dioxin category showed a significant health detriment relative to Comparisons in the background category, but there was no significant detriment for Ranch Hands in the high current dioxin category, the biological plausibility of such an effect would be questionable because this would not indicate a dose-response effect. In general, the adjusted model 3 analyses found that Ranch Hands in the unknown current dioxin category did not show a significant health detriment relative to Comparisons in the background current dioxin category. This was particularly true for the variables that exhibited a significant high versus background contrast.

## **Summary of Results**

Many readers of this report will attempt to tally statistically significant results across clinical areas and study cycles. A study of this scope with a multitude of endpoints and no prescribed strength of association to declare an effect demands, and at the same time defies, meaningful summary tabulation. Such summaries can be misleading because they ignore correlations between the endpoints, correlations between study-cycle results, and the nonquantifiable medical importance of each endpoint. In fact, many endpoints are redundant (e.g., psychological scales and indices developed from combining multiple variables) so as not to miss a dioxin effect and some (such as those arising from measures of pulmonary function) were not suspected beforehand to be related to dioxin exposure.

In addition, such tabulations combine endpoints that medically are not comparable. For example, a diminished sense of smell is of less medical importance than the presence of malignant neoplasm. Statisticians have attempted to summarize multidimensional repeated measures data with growth curve analyses. Such methods were not used in this study because they apply to continuously distributed data only, do not account for medical importance, and reduce the data too much.

Nevertheless, given the lack of adequate summary statistics, the tally of significant results will occur. Such summaries can be misleading and must be interpreted carefully.

## **CONCLUSION**

The interpretation of the AFHS requires careful consideration of potential biases, interactions, consistency of results, the multiple-testing artifact, dose-response patterns, trends, power limitations, strength of association, and biological credibility.

## CHAPTER 1

### REFERENCES

1. 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. Epidemiologic investigation of health effects in Air Force personnel following exposure to herbicides: 1987 followup examination results, NTIS: AD A 222 573. USAF School of Aerospace Medicine, Human Systems Division, Brooks Air Force Base, Texas.
2. 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. *J. of Toxicol. and Environ. Health* 27:165-71.
3. Mocharelli, P., D.G. Patterson, Jr., A. Marochi, and L.L. Needham. 1990. Pilot study (Phase II) for determining polychlorinated dibenzo-p-dioxin (PCDD) and polychlorinated dibenzofuran (PCDF) levels in serum of Seveso, Italy, residents collected at the time of exposure: Future plans. *Chemosphere* 20:967-74.
4. 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.
5. Breslow, N.E., and N.E. Day. 1980. *Statistical methods in cancer research*. Lyon, France: International Agency for Research on Cancer.
6. Bross, I.D. 1985. Proof of safety is much more difficult than proof of hazard. *Biometrics* 41:785-93.
7. The Centers for Disease Control. 1988. Serum 2,3,7,8-tetrachlorodibenzo-p-dioxin levels in U.S. Army Vietnam-era veterans. *JAMA* 260:1249-54.
8. Kahn, P.C., M. Gochfeld, M. Nygren, M. Hansson, C. Rappe, H. Velez, T. Ghent-Guenther, and W.P. Wilson. 1988. Dioxins and dibenzofurans in blood and adipose tissue of Agent Orange-exposed Vietnam veterans and matched controls. *JAMA* 259:1661-67.