

Chapter VII

STATISTICAL METHODS

1. Statistical Study Design

Study data fall naturally into 3 classes: data addressing symptoms, as reported by the subject at questionnaire or in the medical history; data addressing medical signs, determined at physical examination or by review of medical records; and data addressing mortality. A fully expressed or overt herbicide effect would be characterized by increased mortality and more signs and symptoms in the Ranch Hand group as contrasted with the comparison group. These effects should increase with increasing exposure to herbicide. As defined in the study protocol, a subclinical herbicide effect should not be associated with increases in mortality or symptom reporting, but should be found as increases in abnormal findings on physical examination of exposed personnel. These abnormal findings should be more common in the subset of individuals most highly exposed.

Symptom reporting is subjective by definition and, thus, subject to influences that could significantly impair proper inference. For example, a stoic and/or highly patriotic individual might unconsciously or consciously suppress the expression of symptoms. Similarly, anxiety associated with middle or older age could prompt elaboration of symptoms. Association of increased symptom reporting with increasing herbicide exposure is suggestive of a true herbicide effect but is not strongly confirmatory as exposed personnel are at least partially aware of the degree of their exposure and could be suppressing or elaborating symptoms in terms of their perceived exposure.

The study design permits a specific check on the possibilities of overreporting or underreporting of symptoms. The technique involves contrasting the incidence of physical findings when symptoms are present, between the Ranch Hand and comparison groups. The policy followed in this report is that, if there are no group differences in the sign to symptom ratio, underreporting or overreporting is considered unlikely. If there are group differences in the sign to symptom relationship, underreporting or overreporting is possible, but medically, a real group difference may still exist. Overreporting can be assessed by contrasting reported illness with the results of the physical examination and by medical record reviews. However, this assessment is much more difficult for reported psychological symptoms, since a record of hospitalization, the most reliable indicator of verified illness, occurs only in the most severe forms of psychological illness.

2. The Need for Adjustment Procedures

When samples are drawn from a very large or potentially infinite population of individuals, 2 samples of equal size rarely display the same number of diseased individuals. Thus, when comparing 2 groups of individuals, one must ascertain whether the differences are or are not compatible with differences

due to random sampling. Two groups of individuals are said to be statistically significantly different when the differences between the groups cannot be accounted for by random sampling or chance mechanisms. If 2 groups are statistically significantly different and 1 of the groups has experienced a specific exposure, this is suggestive that the exposure and the disease may be causally related. However, great care must be exerted in this setting since other unevaluated factors may be the true cause of the observed group differences, and group difference is only 1 element in the causal chain.

Adjustment procedures are those statistical procedures which allow objective treatment of intervening variables which can distort the true herbicide effect, if one is, in fact, present. Failure to deal with an important intervening variable can either, induce a false effect or obscure a bona fide effect. Statistical procedures for ascertaining statistical significance and for adjustment used in this report are briefly outlined in a subsequent section of this chapter.

The presence of intervening variables occurs either because the sampling procedure used was not completely random or because, by chance, widely different cohorts have been drawn. Matching is a statistical procedure which can partially protect against intervening variables. In this study, the exposed and comparison cohorts were matched on age, race and military occupational category.

Intervening variables are also called covariables, risk factors, or substitution variables, depending on the literature consulted. There currently exists no objective method for ascertaining that all relevant intervening variables have been accounted for. When all known intervening variables have been examined, there is some degree of comfort that observed relationships are correct. Small sample sizes can, however, markedly inhibit study of intervening variables.

A type of intervening variable that is of special interest is the confounding variable (Kleinbaum et al, 1981; Anderson et al, 1980). A confounding variable is an intervening variable that is associated both with the disease under consideration and the exposure categories being used in the study. Failure to adjust for the confounding variable means that the estimated exposure-disease association may be biased. Nonconfounding intervening variables, on the other hand, affect the precision of estimated exposure-disease associations.

In the context of intervening variables or covariables, the concept of interactions is important (Kleinbaum et al, 1982). Interaction occurs when the statistical distribution of a random variable (such as a relative risk, or the difference between group sample means) is a function of a second variable (such as age or weight). The study of interactions in a data set is important for it may lead to the discovery of subpopulations at increased or decreased risk from the population taken as a whole. Confounding and interaction can occur together or separately.

The use of 1 or more measures of exposure (exposure indices) is an extremely useful addition to the study of group differences. Supplementing the analysis of group differences, the use of exposure indices looks within the exposed group to determine whether the more highly exposed individuals tend to exhibit more disease or abnormalities. The use of exposure indices provides a potentially tighter assessment of herbicide exposure. However, by working with the Ranch Hand group, primarily, sample size limitations also impact this technique. Also, use of exposure indices does not obviate the need to be concerned with confounding and other intervening variables. The construction of exposure indices for the Ranch Hand II study is described in another section of this report.

3. Overview of Specific Statistical Methods

In this report, log-linear models have been used when the dependent variable under consideration was categorical or made categorical. Covariables that are intrinsically continuous were stratified for use as adjusting variables in the analysis. Most of the analyses presented in this report are unpaired analyses and, thus, do not fully exploit the paired design of the study. Prior to performing a paired analysis that collapses over matching variables, it is important to determine that the matching variables do not interact with the exposure variable in affecting the dependent variable. The tests presented in this report include these assessments of interaction and, thus, are the early stage of a full paired analysis, as well as being useful for inference in their own right. When unpaired analyses are performed on paired data, there is a consequent loss of test power and less of a chance of detecting a herbicide effect, if one exists. However, an unpaired analysis can actually be more powerful than a paired analysis if study noncompliance or other causes of missing data have resulted in large numbers of broken pairs (Bishop, et al, 1975). The software package used to perform the log-linear analyses in this report is BMD-P4F. In all analyses, the hierarchical modeling procedure was used which starts by examining all covariates and collapses across covariates only when relevant interactions are noted to be null.

Whenever the dependent variable was a continuous variable and the covariables were a mixture of categorical and continuously distributed values, regression, multiple regression and/or general linear models were used (e.g., GLM of the Statistical Analysis System). In these analyses in the report, the covariables were always entered as linear terms only. Also, unless otherwise noted, all group-by-covariate terms (interaction terms) were used in all models.

When group comparisons were made without adjusting for intervening variables, simple parametric tests were used, such as the statistic assuming underlying normal distributions. When it was judged that parametric assumptions were not reasonable, the hypothesis of no difference between Ranch Hand and comparison distributions was tested by the Kolmogorov-Smirnov Two-Sample Test (Gibbons, 1971).

In this study, a very large amount of data has been collected on each participant. In this report more than 190 dependent variables were tested. Testing at the 0.05 level means that in 5 out of 100 instances where there has actually been no association, an association will be falsely inferred. The picture is more complex in this report, since as with many epidemiologic studies, measures are not independent but are highly associated. Those variables thought to be most associated with one another have been grouped into clinical categories and these are used for reporting; e.g., general health, psychology, neurology, etc. However, it cannot be assumed that the clinical categories are completely independent from one another. Within each clinical category, whenever possible, summary indices have been developed to provide an overall view of participant status and lessen the likelihood of false inference. Another important concept which protects against false attribution of herbicide effect is careful consideration of the pattern of statistically significant results. If a herbicide effect is being falsely inferred, it might be in a direction opposite to that expected from prior reports. On the other hand, if a test is found significant with a high degree of confidence, its credibility must be considerably enhanced.

The inverse of falsely attributing a herbicide effect is the problem of failing to detect an effect when one actually exists. This involves the questions of study power. Power is addressed at length in the study protocol but an overview is provided in this chapter. Under the condition of equal Ranch Hand and comparison group sizes, and assuming unpaired analyses, Table VII-1 provides the approximate sample sizes needed to detect specific relative risks with approximate probability 0.80 ($\alpha = 0.05$). The present study is able to detect (with probability 0.80) those relative risks enclosed below the heavy line drawn through the table. Study power for continuous variables is shown in Table VII-2. The mean shift refers to the displacement of the Ranch Hand mean relative to the control. The variables considered are normally distributed, and unpaired testing is assumed in the table. The present study has approximately an 80% chance of detecting mean shifts below the heavy line drawn through the table.

One thousand forty-five Ranch Handers complied to the physical examination in this study. With this size group, disease states with a cumulative incidence in the group of 1/500 or less have a 10% chance or greater of no cases at all being encountered. More detail on this point is given in Table VII-3 where the probability of seeing no cases at all is provided for other cumulative incidence values.

Another view of study power can be obtained through use of the P values reported in this volume. These observed probabilities permit a direct evaluation of study power against the alternative hypothesis defined by the observed statistic. For example, in categorical tables, the chi-square statistic can be inferred from the cited P value. This observed chi-square statistic can be used as the alternative hypothesis to the null hypothesis of statistical independence. Taking the observed chi-squared statistic as the noncentrality parameter in the appropriate chi-squared distribution, a calculation of study power against the observed effect is possible (Johnson and Kotz, 1970). Table

VII-4 provides a short summary of P-value power relationships. Using Table VII-4, if a P-value of 0.10 is reported from a 2X3 table categorical analysis, it may be inferred that study power against the observed effect was 47% (using the two degrees of freedom column in the table). This implies that, if the groups are really as different as they appear from the data, this difference would be detected as statistically significant 47 times out of 100 hypothetical repetitions of this study.

Table VII-4 can also be used to approximately assess the power of linear model analyses. The test statistic in these analyses is an F distribution associated with γ_1 and γ_2 degrees of freedom. The degrees of freedom, γ_2 associated with dependent variable mean squared error is usually quite large in this study. Thus the $F(\gamma_1, \gamma_2)$ distribution can be usually well approximated by a $\chi^2(\gamma_1)$ distribution. The degree of freedom, γ_1 , will be 1 when equality between 2 variables such as slopes or group means is under test, and will be the number 2 when equality between 3 variables is under test, as in the tri-level exposure index case.

Table VII-1

NEEDED SAMPLE SIZES TO DETECT EXPOSURE EFFECTS
IN TWO SAMPLE TESTING ASSUMING EQUAL SAMPLE SIZES*

DATE OF DISEASE IN CONTROL POP = P CONTROL	MULTIPLIES FACTOR IN EXPOSED GROUP = RELATIVE RISK											
	1.25	1.50	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	
$\frac{1}{10000}$	1,408,647	388,536	114,381	36,618	19,623	12,843	9,339	7,244	5,869	4,905	4,196	
$\frac{1}{5000}$	704,244	194,244	57,182	18,306	9,809	6,420	4,668	3,621	2,933	2,451	2,097	
$\frac{1}{1000}$	140,722	38,810	11,423	3,656	1,958	1,281	931	722	585	489	418	
$\frac{1}{500}$	70,282	19,381	5,703	1,824	977	639	464	360	291	243	208	
$\frac{1}{100}$	13,930	3,838	1,127	359	192	125	90	70	56	47	40	
$\frac{1}{50}$	6,886	1,895	555	176	94	61	44	34	27	22	19	

*This study has unequal sample sizes; therefore these tabled values are underestimates.

Table VII-2

NEEDED SAMPLE SIZES TO DETECT EXPOSURE EFFECTS
IN TWO SAMPLE TESTING ASSUMING EQUAL SAMPLE SIZES*

MEAN SHIFT	VARIABILITY (σ/μ)				
	.05	.10	.25	.50	.75
0.5%	785	3,140	19,628	78,510	176,647
1.0%	196	785	4,907	19,628	44,162
1.5%	87	349	2,181	8,723	19,628
2.0%	49	196	1,227	4,907	11,040
2.5%	31	126	785	3,140	7,065
5.0%	8	31	196	785	1,776
7.5%	4	14	87	349	785
10.0%	-	8	49	196	442

*This study has unequal sample sizes; therefore these tabled values are underestimates.

Table VII-3

PROBABILITY OF ZERO CASES AS A FUNCTION
OF CUMULATIVE INCIDENCE

<u>Disease Prevalence</u>	<u>Probability of Finding Zero Cases in a Group of 1045 Participants</u>
1/10,000	.901
1/5,000	.811
1/2,000	.593
1/1,000	.351
1/500	.123
1/200	.005

Table VII-4

STUDY POWER AGAINST OBSERVED EFFECTS

OBSERVED PROBABILITY (P- VALUE)	DEGREES OF FREEDOM			
	1	2	3	4
.001	.908	.924	.938	.948
.01	.730	.780	.816	.845
.05	.500	.583	.642	.689
.10	.376	.470	.536	.590
.25	.210	.300	.367	.425

Study power can be severely influenced by the analytical or statistical method brought to bear on the data. For example, in an evaluation of blood pressure, very small differences in group mean blood pressure can be detected using parametric or nonparametric testing of measures of location; however, if group differences in hypertension prevalence are analyzed, a lesser or no group difference might be found using categorical statistical methods such as log-linear models. In general, there is less power to detect a group difference in specific medical diagnoses of a disease state with categorical procedures, than with the underlying continuous variable. However, even in the absence of statistically significant differences in disease rates, group differences in means and variances are still indicative of differences in disease rates that might be detected if sample sizes were larger. Because of these considerations, analyses in this report of continuous variables and the associated normal-abnormal categories are both provided wherever possible.

4. Verification By Medical Records and Interpretive Precision

This report contains a retrospective morbidity element since both the questionnaire and physical examination inquire about illnesses or medical conditions that may have occurred in the participant prior to this study. These reports of illness are currently being verified by medical record. The study plan additionally includes verification of negative responses. In this report, some reported conditions have been verified by medical record but no verification of negative responses is currently available. This correction of false positives improves the hypothesis testing only if the false negative rate can be assumed negligible, perhaps a reasonable assumption in a military population. If the false negative rate is not negligible, significant bias and loss of precision remains in the hypothesis test.