TABLE 20-2.

Unadjusted Analyses of Radiological and Clinical Respiratory System Findings by Group

Variable	Statistic	Group: Ranch Hand Number	Ranch Hand Percent	Group: Comparison Number	Comparison Percent	Est. Relative Risk (95% C.I.)	p-Value
Thorax and Lungs	n	1,015		1,293			
	Abnormal	61	6.0	61	4.7	1.29 (0.90,1.86)	0.17
	Normal	954	94.0	1,232	95.3		
Asymmetrical Expiration	n	1,015		1,293			
	Abnormal	2	0.2	3	0.2	0.85 (0.17,4.45)	0.86
	Normal	1,013	99.8	1,290	99.8		
Hyperresonance	n	1,015		1,293			
	Abnormal	30	3.0	35	2.7	1.09 (0.67,1.79)	0.72
	Normal	985	97.0	1,258	97.3		
Dullness	n	1,015		1,293			
	Abnormal	2	0.2	1	0.1	2.55 (0.31,17.62)	0.43
	Normal	1,013	99.8	1,292	99.9		
Wheezes	n	1,015		1,293			
	Abnormal	24	2.4	21	1.6	1.47 (0.82,2.63)	0.20
	Normal	991	97.6	1,272	98.4		
Rales	n	1,015		1,293			
	Abnormal	6	0.6	7	0.5	1.09 (0.38,3.15)	0.86
	Normal	1,009	99.4	1,286	99.5		
X Ray	n	1,012		1,289			
	Abnormal	102	10.1	149	11.6	0.86 (0.66,1.12)	0.26
	Normal	910	89.9	1,140	88.4		

TABLE 20-3.

Adjusted Analyses of Respiratory Variables by Group*

Variable	Group: Ranch Hand Total	Group: Comparison Total	Adj. Relative Risk (95% C.I.)	p-Value	Covariate Remarks**
Asthma	1,012	1,290	1.16 (0.76,1.75)	0.57	PACKYR (p=0.023) GRP*PACKYR (Borderline: p=0.068)
Bronchitis	1,011	1,290	0.97 (0.76,1.25)	0.83	None
Pleurisy	1,012	1,289	****	****	GRP*PACKYR (p=0.0026)
Pneumonia	1,012	1,289	1.02 (0.82,1.26)	0.93	AGE (p=0.0001)
Tuberculosis	1,011	1,290	****	****	GRP*PACKYR (p=0.034)
Thorax and Lungs	1,011	1,291	1.27 (0.87,1.84)	0.19	AGE (p<0.0001) PACKYR (p<0.001)
Asymmetrical Expiration	1,011	1,291	0.81 (0.14,4.85)	0.85	AGE*PACKYR (p=0.036)
Hyperresonance	1,011	1,291	1.04 (0.63,1.73)	0.80	AGE (p<0.0001) PACKYR (p<0.0001)
Dullness	1,011	1,291	2.56 (0.31,17.66)	0.47	None
Wheezes	1,011	1,291	1.46 (0.80,2.64)	0.22	PACKYR (p<0.0001)
Rales	1,011	1,291	****	****	GRP*AGE (p=0.046) GRP*PACKYR (Borderline: p=0.070) AGE*PACKYR (Borderline: p=0.090)
X Ray	1,008	1,287	0.85 (0.65,1.11)	0.22	AGE (p<0.0001) PACKYR (p=0.0019) GRP*PACKYR (Borderline: p=0.060)

*Group-by-covariate interactions are described in Table 20-4.

**Abbreviations

PACKYR: Lifetime smoking history (pack-years)
GRP: Group

****Group-by-covariate interaction--relative risk, confidence interval, and p-value not presented.

TABLE 20-4.

Summary of Group-by-Covariate Interactions for Respiratory Variables

Variable	Interaction	Stratification	Statistic	Group: Ranch Hand Number	Group: Ranch Hand Percent	Group: Comparison Number	Group: Comparison Percent	Adj. Relative Risk (95% C.I.)	p-Value
Asthma	Group-by-Pack-Year	0	n	291		367			
			Abnormal	11	3.78	5	1.36	2.84 (1.00,7.89)	0.05
			Normal	280	96.22	362	98.64		
		>0-10	n	284		397			
			Abnormal	16	5.63	15	3.78	1.52 (0.75,3.10)	0.25
			Normal	268	94.37	382	96.22		
		>10	n	437		526			
			Abnormal	17	3.89	30	5.70	0.67 (0.37,1.23)	0.19
			Normal	420	96.11	496	94.30		
Pleurisy	Group-by-Pack-Year	0	n	291		366			
			Abnormal	8	2.75	8	2.19	1.27 (0.48,3.32)	0.64
			Normal	283	97.25	358	97.821		
		>0-10	n	284		397			
			Abnormal	18	6.34	8	2.02	3.29 (1.43,7.49)	<0.001
			Normal	266	93.66	389	97.98		
		>10	n	437		526			
			Abnormal	21	4.81	41	7.79	0.60 (0.35,1.02)	0.06
			Normal	416	95.19	485	92.21		

TABLE 20-4. (continued)

Summary of Group-by-Covariate Interactions for Respiratory Variables

Variable	Interaction	Stratification	Statistic	Group: Ranch Hand Number	Group: Ranch Hand Percent	Group: Comparison Number	Group: Comparison Percent	Adj. Relative Risk (95% C.I.)	p-Value
Tuberculosis	Group-by-Pack-Year	0	n	290		367			
			Abnormal	1	0.34	4	1.09	0.31 (0.06,2.26)	0.28
			Normal	289	99.66	363	98.91		
		>0-10	n	284		397			
			Abnormal	4	1.41	0	0.00	--	0.02
			Normal	280	98.59	397	100.00		
		>10	n	437		526			
			Abnormal	2	0.46	2	0.38	1.20 (0.21,2.04)	0.86
			Normal	435	99.54	524	99.62		
Rales	Group-by-Age	>1942	n	384		509			
			Abnormal	1	0.26	1	0.20	1.33 (0.14,12.78)	0.84
			Normal	383	99.74	508	99.80		
		1922-1942	n	600		741			
			Abnormal	5	0.83	6	0.81	1.03 (0.33,3.25)	0.96
			Normal	595	99.17	735	99.19		
		<1922	n	27		41			
			Abnormal	0	0.00	0	0.00	--	--
			Normal	27	100.00	41	100.00		

TABLE 20-4. (continued)

Summary of Group-by-Covariate Interactions for Respiratory Variables

Variable	Interaction	Stratification	Statistic	Group: Ranch Hand Number	Group: Ranch Hand Percent	Group: Comparison Number	Group: Comparison Percent	Adj. Relative Risk (95% C.I.)	p-Value
Rales	Group-by-Pack-Year	0	n	291		367			
			Abnormal	1	0.34	2	0.54	0.63 (0.09,5.26)	0.71
			Normal	290	99.66	365	99.46		
		>0-10	n	283		398			
			Abnormal	0	0.00	2	0.50	--	0.23
			Normal	283	100.00	396	99.50		
		>10	n	437		526			
			Abnormal	5	1.14	3	0.57	2.02 (0.51,7.67)	0.33
			Normal	432	98.86	523	99.43		
X Ray	Group-by-Pack-Year	0	n	290		365			
			Abnormal	15	5.17	36	9.86	0.50 (0.27,0.93)	0.03
			Normal	275	94.83	329	90.14		
		>0-10	n	282		398			
			Abnormal	28	9.93	32	8.04	1.26 (0.74,2.14)	0.39
			Normal	254	90.07	366	91.96		
		>10	n	436		524			
			Abnormal	59	13.53	81	15.46	0.86 (0.60,1.23)	0.40
			Normal	377	86.47	443	84.54		

-- No abnormals present in Comparison group.

TABLE 20-5.

Exposure Index Analysis Results for Officers
p-Values of Dependent Variable-by-Covariate Association[a,b]

Variable	D*EXP	D*AGE	D*PACKYR	D*EXP *AGE	D*EXP *PACKYR	D*AGE *PACKYR	D*EXP* AGE*PACKYR	Overall Abnormal	Overall Total	Overall Percent
Asthma								16	380	4.2
Bronchitis					0.08		0.009	52	380	13.7
Pleurisy		0.02						16	380	4.2
Pneumonia							0.04	75	380	19.7
Tuberculosis	(No Analysis; Only 3 Abnormal)									
Thorax and Lungs	0.05		0.02			0.06		17	380	4.5
Asymmetrical Exp.	(No Analysis; Only 2 Abnormal)									
Hyperresonance					0.07			9	380	2.4
Dullness	(No Analysis; Only 2 Abnormal)									
Wheezes								5	380	1.3
Rales	(No Analysis; Only 3 Abnormal)									
X Ray	0.09	0.01	0.06				0.08	34	380	8.9

[a]Dependent variable indicated by D in column headings.

[b]Abbreviations:

EXP: Exposure index.
PACKYR: Pack-years.

TABLE 20-6.

Exposure Index Analysis Results for Enlisted Flyers
p-Values of Dependent Variable-by-Covariate Association*

Variable	D*EXP	D*AGE	D*PACKYR	D*EXP *AGE	D*EXP *PACKYR	D*AGE *PACKYR	D*EXP* AGE*PACKYR	Overall Abnormal	Overall Total	Overall Percent
Asthma		0.07						6	175	3.4
Bronchitis							0.005	21	174	12.1
Pleurisy							0.08	9	175	5.1
Pneumonia				0.08				39	175	22.3
Tuberculosis	(No Analysis; Only 2 Abnormal)									
Thorax and Lungs	0.04							19	175	10.9
Asymmetrical Exp.	(No Analysis; 0 Abnormal)									
Hyperresonance	0.04				0.08			9	175	5.1
Dullness	(No Analysis; 0 Abnormal)									
Wheezes								7	175	4.0
Rales	(No Analysis; Only 1 Abnormal)									
X Ray	0.04							19	175	10.9

*Dependent variable indicated by D in column headings.

TABLE 20-7.

Exposure Index Analysis Results for Enlisted Groundcrew: p-Values of Dependent Variable by Covariate Association*

Variable	D*EXP	D*AGE	D*PACKYR	D*EXP* *AGE	D*EXP *PACKYR	D*AGE *PACKYR	D*EXP* AGE*PACKYR	Overall Abnormal	Overall Total	Overall Percent
Asthma				0.08			0.02	22	457	4.8
Bronchitis	0.08							55	457	12.0
Pleurisy				0.03				22	457	4.8
Pneumonia			0.01					81	457	17.7
Tuberculosis	(No Analysis; Only 2 Abnormal)									
Thorax and Lungs		0.06						25	456	5.5
Asymmetrical Exp.	(No Analysis; 0 Abnormal)									
Hyperresonance		0.007						12	456	2.6
Dullness	(No Analysis; 0 Abnormal)									
Wheezes				0.009	0.02			12	456	2.6
Rales	(No Analysis; Only 2 Abnormal)									
X Ray		0.0005						49	456	10.8

*Dependent variable indicated by D in column headings.

Two sets of analyses were performed on enlisted groundcrew data. In the first set of analyses, all three year-of-birth categories (born after 1942, born between 1922 and 1942, born before 1922) were used. In the second set of analyses, only those born between 1922 and 1942 and after 1942 were used, since only one enlisted groundcrew Ranch Hand was born before 1922. All testing results in the two sets of analyses were the same, except for the asthma-by-age interaction shown in Table 20-6.

Each of the dependent variable-by-exposure category interactions are noted by occupation category in Appendix R, Tables R-4 through R-18. These data are considered too sparse for meaningful interpretation.

SUMMARY AND CONCLUSIONS

A summary of the results on the analyses of reported history of respiratory illness and of radiological and clinical findings is given in Table 20-8.

Based on the 31 December 1986 mortality data, there were seven deaths from respiratory conditions in the Comparison group and none in the Ranch Hand group.

TABLE 20-8.

Overall Summary Results of Unadjusted and Adjusted Analyses of Pulmonary Disease

Pulmonary Disease	Unadjusted	Adjusted
Reported History of Respiratory Illness		
Asthma	NS	NS
Bronchitis	NS	NS
Pleurisy	NS	****
Pneumonia	NS	NS
Tuberculosis	NS	****
Radiological and Clinical Findings		
Thorax and Lungs	NS	NS
Asymmetrical Expiration	NS	NS
Hyperresonance	NS	NS
Dullness	NS	NS
Wheezes	NS	NS
Rales	NS	****
X Ray	NS	NS

NS: Not significant ($p>0.10$)
****Group-by-covariate interaction.

There were no group differences found for reported history of asthma, bronchitis, pleurisy, or tuberculosis based on the unadjusted analyses. Adjustments for age and lifetime smoking did not alter the findings of group similarity, although there was a significant group-by-pack-year interaction for pleurisy and for tuberculosis.

Similarly, there were no significant group differences in the unadjusted analyses for the radiological and clinical respiratory findings of thorax and lungs, asymmetrical expiration, hyperresonance, dullness, wheezes, rales, and x-ray interpretations. These findings were supported by the adjusted analyses, although there was a group-by-age interaction for rales.

The exposure index analyses revealed no consistent dose-response pattern.

Analyses of past history of respiratory illness and the clinical and radiological examination of the chest and lungs did not reveal any statistically significant differences between the Ranch Hand and Comparison groups suggestive of herbicide related disease. Several group-by-covariate interactions did exhibit statistical significance, but these findings did not indicate any consistent patterns suggesting different disease experience in the two groups.

REFERENCES

CHAPTER 20

1. Suskind, R.R., and V.H. Hertzberg. 1984. Human health effects of 2,4,5-T and its toxic contaminants. JAMA 251:2372-2380.

2. Lathrop, G.D., P.M. Moynahan, R.A. Albanese, and W.H. Wolfe. 1983. An Epidemiologic Investigation of Health Effects in Air Force Personnel Following Exposure to Herbicides--Baseline Mortality Study Results. Epidemiology Division, Data Sciences Division, USAF School of Aerospace Mecicine, Brooks Air Force Base, Texas.

CHAPTER 21

INTERPRETIVE CONSIDERATIONS

This chapter reviews several scientific issues that should be considered when attempting to reach conclusions on a study of this size and complexity. These issues are critical to the interpretation of the data analyses in this report. Data patterns observed in many clinical chapters of this report are also summarized so that hypothesis testing of group differences may be placed in better perspective.

DIOXIN ENDPOINTS

Based upon data in this report, final conclusions on herbicide causality must consider results of the various clinical areas, reflected in the separate chapters. Each chapter introduction has attempted to highlight the major organ systems that are known or suspected to be significantly affected by the ingredients of Agent Orange with particular emphasis on the effects of dioxin. Categories of clinical endpoints and their generally accepted degree of association with dioxin are presented in Table 21-1. These associations are based on the scientific literature.

TABLE 21-1.

Summary Associations of Adverse Health Effects to TCDD Exposure Reported in the Literature

Degree of Association by Clinical Chapter

Confirmed	Highly Suspected	Moderately Suspected	Negative or Weakly Suspected
Dermatology Neurology Hepatic	Malignancy	General Health Immunology	Psychology Cardiovascular Hematology Endocrine Renal Pulmonary

It is recognized that alternative conclusions based on these patterns of association are possible within the framework of current knowledge, particularly for the highly and moderately suspected areas (malignancy, general health, immunology). However, for illustrative purposes, two extremes are presented: multiple adverse findings in the Ranch Hand group for the areas

of dermatology, neurology, hepatic (discussed in Chapter 13), and cancer would suggest a case for TCDD causality, whereas multiple adverse findings in the weakly suspected areas, and not in any of the confirmed areas, would be difficult to ascribe to an overall TCDD causation.

The aspects of biological plausibility and specificity require balanced interpretation across clinical chapters, with careful attention placed on nonsignificant findings as well as significant findings. The chapters in this report should be viewed as artificial boundaries for convenience of presentation, and should not discourage consideration of their relatedness, or of the individual variables within them.

EXPOSURE

Approximately 600 exposure index analyses have been conducted in this study, underscoring attempts to associate increasing proportions of various abnormalities to estimates of increasing exposure.

To determine whether the results of the exposure analyses varied by chance, several perspectives were taken. Of the 255 adjusted exposure analyses (excluding 39 with interactions), 13 were statistically significant, a figure which is the expected number (based on $\alpha = 0.05$). It is recognized that this contrast is a crude yardstick, considering the relatedness of the dependent variables, statistical power, disproportionate representation of chapter variables, and the presence of interactions. The six possible patterns of exposure response (increasing, decreasing, V-shaped with fewer abnormalities at the low exposure level than the high exposure level, V-shaped with more abnormalities at the low exposure level than at the high exposure level, inverted V-shaped with fewer abnormalities at the low exposure level than the high exposure level, and inverted V-shaped with more abnormalities at the low exposure level than at the high exposure level) were tabulated (regardless of statistical significance) for the clinical chapters of dermatology, neurology, psychology, and renal. As noted in Table 21-1, two of these chapters contain clinical variables that have had confirmed associations to TCDD exposure, and two chapters have had negative or weakly suspected associations to TCDD. Of the 126 exposure analyses in these four chapters, 21 (or one-sixth) showed the primary pattern of interest, an increase--exactly the number expected. Taken together, these analyses suggest that statistically significant exposure analyses may have occurred due to chance among the data set, and that the pattern of dose-response may also have been random. These inferences, or that the exposure index was unrelated to actual exposure, together with the acknowledged limitations of the exposure index, indicate that estimated exposure may only be weakly relied upon to assert a causal relationship. Based upon the current exposure index calculations, either of the above inferential alternatives is possible.

The use of serum dioxin levels (see Chapter 23, Future Directions) in the next report will clarify the exposure calculations of this report and the Baseline Report. Thus, from an interpretive context, final conclusions on dose-response, and the implications to herbicide causation are based on current knowledge available for this report. These conclusions could change with future analyses using a factual exposure concept.

TYPES OF MEASUREMENTS

This report includes all types of measures traditionally used in morbidity followup epidemiologic studies, e.g., self-reports, structured interview responses, medical record data, physician findings, scalar measurements, biopsy results, laboratory determinations, morbidity indices, and mortality results. At many points in this report, various terms have been used to qualitatively describe the data and analyses arising from the measurement processes. In particular, the terms "subjective," "objective," "continuous," and "categorical," and "constructed indices" have been used to connote differences in data or data sets that are important in making statements of inference.

From the perspective of the Study Protocol, significant group differences for subjective historical variables, not mirrored by significant group differences in medical record findings or physician/laboratory testing, may be viewed as preliminary evidence of over-reporting by a group. The opposite finding of significant group differences for physical examination variables in the absence of reported symptoms may support the primary conclusion of significant subclinical group differences. Either of these alternatives may greatly affect an overall inference of herbicide causality. Hence, the descriptive phrases "subjective data" and "objective data" have not been used as value judgments of the worth of the data, but simply as inferential qualifiers.

This report contains numerous comments on the differences in results between analyses of continuous versus categorical data from the same variable (exclusively laboratory data). Because the statistical power is stronger for detecting mean shifts than categorical differences, it was anticipated that very small mean shifts might be more easily discerned than differences in proportions of abnormalities between the two groups. Both methods of examining the data reveal important aspects of the distribution. Inferentially, when both types of analyses were done, greater weight has been given to significant group differences when analyses of both data forms agree. Lesser weight was given to significant differences seen in only one analysis, and least weight to significant shifts in means if both group means were within normal range, and the mean difference was not supported by other statistical findings in related variables (e.g., hepatic test battery). Consistent patterns of findings within an organ system, or between related organ systems, is required to strongly suggest an inference of causality.

Several summary indices were constructed in this report, e.g., dermatology index, cranial nerve function index, and anatomic categories of abnormal peripheral pulses, and are similar to some indices in the 1984 Baseline Report. They were formed by summing or grouping related abnormalities for the purposes of assessing increased numbers and/or showing group directionality of overall results. They should not be strongly considered in final inferences because they are artificially derived.

BASELINE-FOLLOWUP EXAMINATION DIFFERENCES

A common difficulty of followup studies is the inherent variation in measurement systems from one observation period to the next. To the maximum extent possible, the USAF has restricted clinical variation by requiring the use of identical laboratory equipment for most clinical chemistries, by the

use of 50 samples from the Baseline serum bank to evaluate interexamination laboratory differences, and by the use of carefully prescribed written clinical procedures that allow little room for variation. Nonetheless, some interexamination variability must be expected, but in the presence of blindness to group membership, there is no reason to expect biases in the results with respect to either the Ranch Hand or Comparison groups.

This report has cited classical longitudinal analyses to assess changes in variables between the examinations by group. Of 21 variables examined, 5 showed statistically significant group differences in the changes between examinations. Four of these significant results were attributed to actual changes over time, while the other (e.g., sedimentation rate) was believed due to a change in laboratory methodology.

Other less refined longitudinal contrasts consisting of narrative discussions of Baseline results versus followup results have been presented in all chapters. Interpretive caution is required in assessing examination similarities or differences because of the slight changes in cohort composition between the examinations (see Chapter 2, Population), the use of slightly different statistical models and modeling strategy (see Chapter 7, Statistical Methods), and sometimes the use of the Original Comparison group. The relative contribution of these changes was not explored mathematically, but is believed to have played a minimal role in accounting for any large group shifts between examinations.

In the context of comparing results between examinations, there has been a subtle but consistent observation that group differences have substantially narrowed over the 3-year period, either by decreased findings in the Ranch Hands, increased findings in the Comparisons, or a combination of both mechanisms. In general, several broad interpretations are possible: any bona fide herbicide effect decreases over time, that the convergence is largely attributable to unquantifiable factors, that both examinations have produced chance results, or that these observations have been affected by the slight shifts in cohort composition and modeling strategy.

Several segments of this report have noted marked differences in the prevalence rates of abnormalities found at the Baseline and followup specialty examinations, e.g., the dermatology and neurology clinical assessments. The followup dermatological examination detected substantially more abnormalities than the Baseline examination, whereas far greater numbers of neurological abnormalities were noted at the Baseline examination than at the followup for some variables. These examination variances were affected by differences in "clinical sensitivity" between the examining teams, although clearly other factors (such as a true change in disease-abnormality status or slight cohort differences) contributed. The phrase "clinical sensitivity" refers to the inherent differences in clinical styles and interpretations of possible abnormalities that often prevail. Because of examiner blindness to exposure status, and because of the judgment that the interexamination variation was within the artful bounds of accepted medical practice, no bias was thought to have resulted from this inherent variation.

STUDY BIASES

Each reviewer of this report must reach a conclusion on whether the results of this study have been seriously flawed by the design, the operation

of significant biases, or both. The Protocol authors believe that the comprehensive multifaceted design is the chief strength of this study, although it is recognized that each and every published phase of the study must invite renewed inspection of fundamental scientific aspects of the study design.

It is believed that, with the exception of skin test readings, all data in this study were collected accurately and validly, and that blindness to group membership was well maintained throughout the collection process. This opinion is important from an inferential perspective in that both misclassification of data (tending to dilute true group differences) and bias in data (creating a false group effect) most likely did not occur appreciably in this study. Thus, it is believed that both the magnitude and direction of the group results found in this study reflect truth to the maximum degree possible, within the inherent boundaries of statistical models to account for all important adjusting variables.

GROUP INTERACTIONS: PATTERN RECOGNITION

Many of the adjusted analyses in this report have demonstrated significant group-by-covariate interactions, requiring stratified analyses to determine the nature of significant group differences. All significant two- and three-factor interactions have been included in the main text or in appendices. The analysis of followup data has found substantially more interactions than the analysis of Baseline data, due primarily to the larger number of covariates used in the followup analyses.

Several related viewpoints have aided in the overall interpretation of group-by-covariate interaction in the report. In the presence of a significant interaction, a direct conclusion on main group effects cannot be made, and the focal point of interpretation resides with the covariate stratum containing the significant group effect (or a reversal in nonsignificant group effects across strata). Past this point, however, there appears to be little consensus in how to best place the interaction into inferential context. Further interpretations appear to be largely individualistic.

No consistent pattern has emerged to support a finding of impairment in the Ranch Hands for any specific stratum of one or more covariates. In fact, of all the two- and three-factor interactions encountered, only one was thought to have possible biologic relevance. Other interactions may have such relevance, but the reason was not apparent. As with tests of group differences, significant interactions may occur by chance, but the method to calculate an expected number of group-by-covariate interactions, unfortunately, remains an open research question.

Because of the possible diverse interpretations of interactions, all significant two- and three-factor interactions involving group with statistically significant strata are presented in Table 21-2 for detailed inspection. No particular covariate or group pattern is noted, although the variables in psychology and gastrointestinal showed Ranch Hands at a relative detriment, while the interactions in the cardiovascular chapter indicated detrimental findings in the Comparisons.

Most variables without interactions in this report have shown remarkable concordance between unadjusted and adjusted results, both in terms of absolute value of relative risk and of statistical significance.

TABLE 21-2.

Summary of Significant Covariate Strata (or Covariate Level Difference) Found Within Significant Two- and Three-Factor Group-by-Covariate Interactions by Clinical Chapter and Dependent Variable (Group Direction and p-Value)

Clinical Chapter	Dependent Variable	Covariate Stratum	RH>C	C>RH	p-Value
General Health	Self-Perception of Health	Enlisted Groundcrew	*		0.003
Malignancy	Basal Cell Carcinoma (Verified Interval)	Enlisted Flyer	*		0.019
	Systemic Cancer (Verified plus Suspected, Interval)	Enlisted Flyer	*		0.042
	Basal Cell Carcinoma (Verified plus Suspected, Lifetime)	Intermediate Skin Reaction to Sun	*		0.038
	Systemic Cancers (Verified, Lifetime)	Enlisted Flyer	*		0.019
	Systemic Cancer (Verified plus Suspected, Lifetime)	Enlisted Flyer	*		0.004
Neurology	Pin Prick	Impaired (Diabetic Class)		*	0.021
Psychology	Paranoia	Born Before 1942	*		0.027
	Schizophrenia	High School	*		0.033
	Social Introversion	Combat Index--Low	*		0.002
	Validity	Black		*	0.038
	Total CMI	High School	*		<0.001
Gastrointestinal	SGOT	1-4 Drinks per Day	*		0.010
	Alkaline Phosphatase	Exposed to Ind. Chems.	*		<0.001
	Direct Bilirubin	Exposed to Ind. Chems.	*		0.035
	Triglycerides (cont.)	Born In or Before 1922	*		0.039
	Triglycerides (disc.)	Officer	*		0.035
	Uroporphyrins	BUN≤14		*	<0.001

TABLE 21-2. (continued)

Summary of Significant Covariate Strata (or Covariate Level Difference) Found Within Significant Two- and Three-Factor Group-by-Covariate Interactions by Clinical Chapter and Dependent Variable (Group Direction and p-Value)

Clinical Chapter	Dependent Variable	Covariate Stratum	RH>C*	C>RH	p-Value
Dermatology	Dermatology Index	Pre-SEA Acne: 1 vs. 0		*	0.004
Cardiovascular	Systolic Blood Pressure	Black/53 Yrs Old		*	0.006
	ECG (Overall)	0 Pack-years		*	0.038
	ECG (Arrhythmia)	7 Pack-years/10% Body Fat	*		0.018
	Posterior Pulses (Manual)	Enlisted Flyer		*	0.032
	Leg Pulses (Manual)	Officer/21% Body Fat		*	0.026
	Peripheral Pulses (Manual)	Officer		*	0.030
Hematology	WBC	Nonblack/30 Pack-years/ 35 Yrs Old	*		<0.001
	WBC	Black/Officer/35 Yrs Old		*	0.003
	WBC	Black/EFL/35 Yrs Old		*	0.050
	PLT	Nonblack/30 Pack-years and 1 pack/day	*		0.014
	PLT	Black/30 Pack-years and 1 pack/day	*		0.007
Renal	Urinary Protein	Normal (Diabetic Class)	*		0.018
	Urinary WBC	Nonblack/Born In or After 1942	*		0.001
	BUN	Black		*	0.017
	Urine Specific Gravity	Nonblack/Enlisted Groundcrew	*		<0.001
Endocrinology	Testosterone	<10% Body Fat		*	0.012
	Testosterone	10-25% Body Fat	*		0.023
	Differential Cortisol	Black/Born In or After 1942		*	0.003

TABLE 21-2. (continued)

Summary of Significant Covariate Strata (or Covariate Level Difference) Found Within Significant Two- and Three-Factor Group-by-Covariate Interactions by Clinical Chapter and Dependent Variable (Group Direction and p-Value)

Clinical Chapter	Dependent Variable	Covariate Stratum	RH>C*	C>RH	p-Value
Immunology	Total T Cells	Black		*	0.039
	B Cells	Nonblack/0 Pack-years		*	0.004
	Monocytes	Enlisted Groundcrew/ 4 Drinks/Day	*		0.003
Pulmonary	Pleurisy	1-10 Pack-years	*		<0.001
	Tuberculosis	1-10 Pack-years	*		0.020
	X-ray	0 Pack-years		*	0.030
	Total Interactions: 43		26	17	

*Relative risk greater than one, or Ranch Hand mean greater than Comparison mean.

CLASSICAL COVARIATES

Many of the dependent variables in this report are known to be significantly affected by risk factors also measured in this study. The use of these covariates in the adjusted analyses has served to clarify Ranch Hand-Comparison group differences in the presence of significant covariate group differences. Such adjustments, whether by a single covariate, multiple covariates, or covariate interactions, have given results on group differences generally quite similar to the unadjusted analyses both in terms of relative risk and statistical signficance. In fact, in only one instance in this report has an unadjusted result of $p \geq 0.10$ changed to a value of $p \leq 0.05$ in the adjusted analysis. The covariates used in this study were not effect modifiers (which may be synergistic with exposure and also be equally distributed between groups). Consistent effects were observed for almost all of the classical covariates of age, race, occupation, education, alcohol, smoking, percent body fat, and glucose tolerance. In only a few instances were unexpected effects noted, e.g., personality type, wine consumption, and a few smoking and alcohol "inversions."

The overall covariate effects observed in this study indeed reflect the mainstream of results found in well-conducted epidemiologic studies, and lend credence to the validity of the clinical endpoints and covariate values in this report.

MULTIPLE COMPARISONS

As noted in Chapter 7, Statistical Methods, the problem of multiple comparisons is complex and not easily adjudicated because of the total number of statistical tests, the number of tests performed on each dependent variable, and the biologic relatedness of many of the variables. A conscious effort has been made to expand inferential interest to borderline group associations ($0.05 < p \leq 0.10$) thereby increasing the probability of the acceptance of a false association. Each chapter summary has carefully flagged all borderline associations to provide expanded summary statements for possible inclusion in deriving final conclusions. Additional confidence in the final acceptance or rejection of an overall herbicide effect would be warranted if the majority of borderline associations were in the same consistent direction as the significant associations.

Multiple analyses on the same variable have been conducted in this report. Continuous and categorical data have been subjected to both unadjusted and adjusted analyses, and multiple adjusted analyses were sometimes conducted with different covariates or slightly different covariate sets. The question arises as to which results best reflect the truth when different results are found. In general, the following approach has been followed: the statistical significance of both continuous and categorical analyses is convincing, while significance for only the continuous analysis must be viewed in terms of the biologic relevance of the mean shift detected.

Overall, the multiple comparison issue is due to repeated hypothesis testing for group, exposure, and interaction strata differences. The calculation of expected numbers of significant associations for these tests is difficult (if not impossible) because of the relatedness of the dependent variables, the relatedness of the covariates, and the often difficult analytic decisions that arise in a "step-down, best model" strategy. Thus,

the final assessment of whether the frequency of significant associations does not meet, or exceeds expectation, must remain an interpretive judgment of each reader.

CAUSALITY

The AFHS is an inferential assessment of observed group differences. The inference of herbicide causality will be determined by a balanced judgment of the following factors: biological plausibility, consistency, specificity, coherence, time relationships, and strength of association. Except for aspects of association strength, most of these causality factors have been discussed in the preceding sections of this chapter. Nearly every statistically significant group difference in this report has only been of moderate to weak strength. Highly significant p-values ($p<0.001$) were not found for main group associations, but were observed for covariate tests. A few strata in the group interactions were highly significant. Most of the statistically significant estimated relative risks were below the value of 2.0 (a traditional boundary of interest in epidemiology). The few relative risks above 2.0 generally had very wide confidence intervals due to low proportions of detected abnormalities. Weakly significant associations, in particular, are cause to reassess the element of chance and the possible presence of other causality factors before a final conclusion of cause and effect is determined.

CHAPTER 22

CONCLUSIONS

INTRODUCTION

This chapter summarizes the conclusions drawn from the statistical analyses that have been conducted on the Air Force Health Study data base. The followup study, which began in 1985, was the logical extension of the 1982 Baseline study, building upon the strengths of the Baseline study and utilizing the data collected at both the Baseline and the followup. The high level of Government support and outstanding participation of the study subjects that characterized the Baseline study were maintained through this first followup.

STUDY PERFORMANCE ASPECTS

Of the living Baseline study participants, 99.2 percent were located and asked to participate in the followup. Participation in the followup physical examination and questionnaire was very high. Of the fully compliant Baseline participants, 971 of the 1,045 Ranch Hands (92.9%) and 1,139 of the 1,224 Comparisons (93.1%) participated in the followup. Thus, there was no group difference in compliance of the Baseline participants at the followup. Overall, the 2,309 participants in the followup (1,016 Ranch Hands and 1,293 Comparisons) represented a loss of 159 individuals and a gain of 199 since Baseline. One percent of the fully compliant Baseline population died between 1982 and the 1985 followup examination.

The bias/compliance analyses suggested that there had been no change between Baseline and the followup in the way replacements volunteered for entry into the study, and that no additional bias had been introduced at the followup due to scheduling differences. Although replacements were not health-matched at Baseline as they were at the followup, they were similar to refusals with respect to reported health, medication use, and income level. The results supported the use of the total Comparison group in the main analyses presented in this report.

POPULATION CHARACTERISTICS

Overall, the Ranch Hands and Comparisons reported similar social and behavioral characteristics. No significant differences were found in age, educational background, religious preference, current military status, and income level. Significantly more Ranch Hands smoked cigarettes at the time of the followup examination than did Comparisons, but there was no significant difference between groups on past cigarette, cigar, and pipe use and on recent and past use of marijuana. A much higher percentage of participants

reported past marijuana use at the followup than at Baseline. This difference was most likely due to a greater level of confidentiality afforded by the questionnaire technique. Risk taking behavior, assessed by questions on potentially dangerous recreational activities, revealed borderline significance. Slightly more Comparisons were scuba divers and more Ranch Hands raced motor vehicles. The difference in scuba diving was also significant at Baseline.

Patterns of Results

Both the chapter conclusions and the final conclusions of this report have been predicated upon concepts of consistency, specificity, coherence, strength, and plausibility as they apply to the interpretation of group differences. In particular, careful consideration has been given to a variety of data and patterns of results that have emerged from the clinical evaluations. Specifically, there were few differences in the proportions of abnormalities between groups; the positive associations have not aggregated in the clinical areas of prime dioxin concern, nor have they been of serious clinical importance; the unadjusted results have been remarkably concordant with the adjusted results, both in terms of relative risk and p value; the analyses using the Original Comparison set have largely mirrored the results found with the total Comparison group; many of the group differences noted at Baseline have disappeared at the followup examination, and only a few new associations have emerged; almost all of the covariates have acted as expected in the adjusted analyses; and the exposure index analyses and the group-by-covariate interactions have not demonstrated biological patterns of concern and appeared to be more likely due to chance than not. Due to the acknowledged limitations of the exposure index used in this report (and considering the potential use of dioxin body burden levels at the next followup), dose-response relationships have not been emphasized in reaching final conclusions.

The overall pattern of these findings indicates that this followup study cannot be viewed as alarming from the traditional perspectives of clinical medicine or epidemiology. This study, in fact, demonstrates similarity in current health status between the Ranch Hand and Comparison groups.

CLINICAL ASPECTS

General Health

The nonspecific assessment of general health showed relatively close similarity between the two groups. Ranch Hands rated their health as fair or poor more frequently, but this difference was found only in the enlisted groundcrew and not in the officers nor enlisted flyers. The perception of health in both groups had improved since Baseline. Physician-rated appearance of relative age was not found to be significantly different at the followup in contrast to the Baseline finding that a higher percent of Ranch Hands than Comparisons looked younger than their stated age. The categorical analysis of sedimentation rate showed that the Ranch Hands had more abnormalities than the Comparisons. These results were not supported by the continuous analysis of mean sedimentation rates and were opposite to the

Baseline results, which showed that younger Comparisons had elevated sedimentation rates. The categorical analysis of percent body fat showed no significant differences between the two groups, which was consistent with Baseline. However, the continuous analysis found that the Ranch Hands had a significantly lower mean percent body fat using age, race, and occupation as covariates. The detailed exposure analyses revealed no consistent exposure effects, and this result was consistent with the Baseline analysis. No longitudinal difference was found on perception of health. A significant group difference was found over time for the longitudinal analysis of sedimentation rate due to the change in the findings between the two examinations, possibly related to a change in laboratory methodology.

Malignancy

Skin and systemic cancers, both suspected and verified by medical records, showed no significant group differences for the Baseline-followup interval (1982-1985). However, for all neoplasms combined (malignant, benign, and uncertain), a borderline significant excess in the Ranch Hand group was noted in an unadjusted analysis. The analyses of interval cancers revealed group interactions for verified and verified plus suspected basal cell carcinoma and verified plus suspected systemic cancers. Nonsignificant findings were observed for verified and verified plus suspected sun exposure-related cancers. Verified systemic cancers did not differ significantly between groups.

The analyses of lifetime cancer found significant results for verified basal cell carcinoma and verified sun exposure-related skin cancers. Group interactions were noted for systemic cancer categories and for verified plus suspected basal cell carcinoma. The higher rate of basal cell carcinoma in the Ranch Hands versus the Comparisons found at Baseline was nonsignificant for the followup interval, but due to the effect of the larger number of Baseline cases and the significant confounding of average residential latitude, the adjusted analysis of lifetime basal cell carcinoma emerged as statistically significant.

There were several disparities in the distribution of testicular, colon, and smoking-related tumors in the groups. Further, one case of soft tissue sarcoma and one possible lymphoma (both in Ranch Hands) were diagnosed in the interval, balancing the two similar cases found in the Comparison group at Baseline. Considering that the systemic cancer curves are in their early stages for both groups, with perhaps insufficient latency, the cancer results of the followup examination should not be viewed as disturbing, but as cause for continued monitoring.

Neurological Assessment

None of the 27 neurological variables demonstrated a significant group difference, although several variables had relative risks which were greater than one. There was no group difference in reported neurological illnesses for the interval or for a lifetime history. Of the cranial nerve variables, speech and tongue position were marginally significant, with the Ranch Hands at a slight detriment. The analyses of peripheral nerve function showed no significant differences between the Ranch Hands and the Comparisons. In the analysis of central nervous system function, hand tremor was found to be of borderline significance, with the Ranch Hands faring slightly worse than the Comparisons. A borderline significant group interaction (Ranch Hand hand tremor by insecticide exposure) may have had biological and operational

significance. Overall, substantially fewer neurological abnormalities were detected at the followup examination than at the Baseline examination. The exposure analyses showed only occasional statistically significant results, although no consistent pattern with increasing exposure was evident. In the longitudinal analysis of the Babinski reflex, a significant change over time was observed. This was due to a nonsignificant finding in the Ranch Hands at the followup, which differed from the significant adverse finding at Baseline. The covariates of age, alcohol history, and diabetes showed classical effects with many of the neurological measurements. Overall, the followup examination results were quite similar to the Baseline findings.

Psychological Assessment

The reported and verified data on lifetime psychological illnesses showed no significant differences between groups. Distributional tests of the 14 Minnesota Multiphasic Personality Inventory (MMPI) scales, stratified by occupation, revealed that only 2 of the 42 results approached significance. For the total Cornell Medical Index (CMI), separate distributional tests were conducted with stratification by age, race, occupation, education, and current drinking status; a significant difference was found for one statum of each of the covariates. In all cases, the mean of the Ranch Hand distribution was greater than the mean of the Comparisons. The analysis of the 14 MMPI scales showed that there was a significant difference between the two groups for denial and masculinity/femininity, with more abnormalities in the Comparisons than the Ranch Hands. The results of the analyses for hysteria were of borderline significance, with more abnormalities in the Ranch Hands. There were more abnormalities in the Ranch Hands than the Comparisons for social introversion, which was of borderline significance. Differences in the total CMI and A-H area subscore were found to be significant, with more abnormalities in the Ranch Hands. There was no significant difference between the two groups on the Halstead-Reitan Battery impairment index, a measure of the functional integrity of the CNS. The exposure index analyses did not reveal any pattern consistent with a dose-response relationship. As expected, the effects of age, educational level, and alcoholic history showed profound effects on many of the psychological measurements.

Gastrointestinal Assessment

Although the followup gastrointestinal assessment disclosed more statistically significant findings than the Baseline examination, the abnormalities were distributed equally between the two groups, and there was no clinical, statistical, or exposure pattern consistent with an herbicide-related effect on health. No historical or biochemical evidence was found to suggest an increased likelihood of porphyria cutanea tarda (PCT) in the Ranch Hand group. Only sparse and nonsignificant liver disorders were reported for the interval between Baseline and followup. Also, for the lifetime history of liver disorders, there were no significant differences between groups. Further, there were no significant group differences in reported lifetime peptic ulcer disease. A review of digestive system mortality showed a relative excess in the Ranch Hands but a relative lack of malignant neoplasms. The results of the physical examination showed a borderline increase of hepatomegaly in the Ranch Hand group. There was a significantly lower mean serum glutamic-pyruvic transminase (SGPT) level, a greater mean alkaline phosphatase level, and a lower mean uroporphyrin level in the Ranch

Hand group. The analysis of coproporphyrin was of borderline significance, with the mean of the Ranch Hands in excess of the mean of the Comparisons. No group differences were found for serum glutamic-oxaloacetic transminase (SGOT), gamma-glutamyl transpeptidase (GGTP), total and direct bilirubin, lactic dehydrogenase (LDH), cholesterol, or triglycerides. The numerous group-by-covariate interactions did not disclose any consistent subgroup patterns detrimental to the Ranch Hands. These findings were generally consistent with the results of the 1982 assessment. The longitudinal analyses for SGOT, SGPT, and GGTP showed no significant differences between results by group over time.

Dermatological Evaluation

No significant group differences were identified in the dermatological evaluation. None of the questionnaire data showed an increased likelihood of past chloracne, as determined by anatomic patterns of acne, and no cases were diagnosed in the physical examination. Analyses were conducted on six dermatologic disorders (comedones, acneiform lesions, acneiform scars, inclusion cysts, depigmentation, and hyperpigmentation) and on a composite variable of 16 other minor conditions (the latter not generally associated with chloracne). Exposure index analyses did not reveal consistent patterns suggestive of a dose-response relationship. The longitudinal analysis, based on a composite dermatology index, showed no significant differences between the results over time. Substantially more dermatologic abnormalities were detected at the followup examination than at the Baseline examination. In general, however, the followup results were consistent with the findings at Baseline.

Cardiovascular Evaluation

Overall there was general similarity in the cardiovascular health of the Ranch Hands and the Comparisons. Of the 27 cardiovascular variables, there was a significant difference for only one, verified heart disease, with an excess in the Ranch Hand group. This finding was largely unsupported by other cardiac measurements. The cardiovascular assessment was based on reported and verified heart disease; the measurement of central cardiac function by systolic blood pressure, abnormal heart sounds, and ECG findings; and the evaluation of peripheral vascular function by diastolic blood pressure, funduscopic examination, presence of carotid bruits, and detailed manual and Doppler measurements of five peripheral pulses. Doppler recordings of five peripheral pulses were similar in both groups, a finding which was in marked contrast to the Baseline examination that found significant pulse deficits in the Ranch Hand group. This change was most likely due to a required 4-hour abstinence from tobacco prior to the pulse measurements. Overall, the exposure analyses were unsupportive of any meaningful dose-response relationship. The longitudinal analyses confirmed the change in pulse abnormalities in the Ranch Hand group over time, but showed no significant group change in overall ECG findings between the examinations.

Hematological Evaluation

The hematological evaluation found that neither group manifested an impairment of the hematopoietic system, consistent with similar findings at

the Baseline. The evaluation was based on eight peripheral blood variables: red blood cells (RBC), white blood cells (WBC), hemoglobin (HGB), hematocrit concentration (HCT), corpuscular volume (MCV), corpuscular hemoglobin (MCH), corpuscular hemoglobin concentration (MCHC), and platelet count (PLT). Both the discrete and categorical analyses revealed no significant group differences. The covariate effects of age, race, occupation, and smoking history were highly significant for many of the variables. Two group-by-covariate interactions in the analyses of mean differences did not appear to have a meaningful interpretation. The exposure index analyses did not support any plausible dose-response relationship. The longitudinal analyses of MCV, MCH, and PLT found significant differences only for PLT between the Baseline and the followup, with the Ranch Hands exhibiting a slight decline in mean level from Baseline and the Comparisons showing an opposite change.

Renal Assessment

None of the six renal variables of reported kidney disease, urine protein, occult blood, urine white blood cell count, blood urea nitrogen, and urine specific gravity showed a significant difference between the two groups based on the unadjusted analyses. In the adjusted analyses of the laboratory variables, however, there were significant group-by-covariate interactions that did not yield a consistent pattern to suggest a renal detriment to either group. The finding of group equivalence for past kidney disease was in contrast to the Baseline examination, which found significantly more reported disease in the Ranch Hand group. The difference in findings is more likely due to a change in questionnaire wording than to a true change in renal health. Like the Baseline findings, the exposure index analyses showed very little evidence of a dose-response relationship. In the longitudinal analyses of blood urea nitrogen, there was no significant group difference in the change between the examinations.

Endocrine Assessment

In general, the endocrine health status of the Ranch Hands and the Comparisons was reasonably comparable. The examination found no significant differences between the two groups for past thyroid disease, or thyroid and testicular abnormalities determined by palpation. In the analyses of the seven laboratory values (T_3 % Uptake; thyroid stimulating hormone [TSH]; testosterone; initial, second, and differential cortisol; and postprandial glucose), significant differences were found for TSH and testosterone, with higher mean levels in the Ranch Hands. These analyses were not supported by the categorical analyses. The thyroid test results were conflicting with respect to an assertion of hypothyroidism in the Ranch Hands (a possible dioxin effect). Mean levels of testosterone were significantly elevated in the Ranch Hand group as contrasted with the Comparisons in the 10-25 percent body fat category. The effects of personality score and percent body fat on the differential cortisol levels were not fully expected. Although tests of 2-hour postprandial mean values showed no significant group differences, comparable categorical tests revealed that significantly fewer Ranch Hands had impaired glucose levels, but conversely, had more (nonsignificant) diabetic levels of glucose. Analyses of the composite diabetes indicator (history plus 2-hour postprandial results) did not disclose significant group differences. The exposure index analyses suggested that the enlisted flyers in the medium exposure level were significantly different from those in the

low exposure level for differential cortisol, postprandial glucose, and testosterone. The corresponding high to low contrasts were not significant. The longitudinal analyses were based on T_3 % Uptake, TSH, and testosterone, and revealed only symmetrical and nonsignificant changes in the Ranch Hand and Comparison groups over the time interval.

Immunological Evaluation

Overall, there were no significant group differences or any indication of impaired immunological competence in either group based on comprehensive cell surface marker and functional stimulation studies. Six cell surface markers (total T cells, helper T cells, suppressor T cells, B cells, monocytes, HLA-DR cells, and a constructed helper/suppressor ratio variable) and three functional stimulation studies (PHA, pokeweed, and mixed lymphocyte culture) were conducted on 47 percent of the study population. No significant differences were revealed for five of these variables. In the analyses of the other five variables, there were significant group-by-covariate interactions, but no discernible pattern was identified to suggest a detriment in any subgroup of either group. Skin test assessments of delayed hypersensitivity were characterized by inter-reader variation and shifting diagnostic criteria for anergy. The skin test data were judged invalid and were not subjected to statistical testing for group differences. No consistent pattern of immunological deficits could be associated with increasing levels of herbicide exposure in the Ranch Hand group.

Pulmonary Disease

The pulmonary assessment did not reveal any statistically significant differences between the Ranch Hand and Comparison groups that were suggestive of an herbicide-related disease. The analyses consisted of group assessments of respiratory disease incidence, physical examination abnormalities, and the current prevalence of x-ray abnormalities. There were no significant differences between the Ranch Hands and Comparisons for history of asthma, bronchitis, pneumonia, or for six of seven clinical variables (excluding rales) determined by x-ray or auscultation. Analyses of history of pleurisy, history of tuberculosis, and rales showed significant but inconsistent group-by-covariate interactions. These findings did not indicate any patterns suggesting a different disease experience in the two groups. The exposure index analyses did not reveal any consistent pattern suggestive of an increasing dose response.

CONCLUSION

The results of the first followup study in 1985 have shown a subtle but consistent narrowing of medical differences between the Ranch Hands and Comparisons since the Baseline Study in 1982. The 1985 examination results provide reassuring evidence that the current state of health of the Ranch Hand participants is unrelated to herbicide exposure in Vietnam. Continued close medical surveillance of these military populations is strongly indicated. This followup report concludes that there is not sufficient plausible or consistent scientific evidence at this time to implicate a causal relationship between herbicide exposure and adverse health in the Ranch Hand group.

CHAPTER 23

FUTURE DIRECTIONS

The scope and complexity of the AFHS has required gradual refinement and correction to meet the challenges of changing technology and scientific direction, and to ensure continued participation of all enrolled members. This chapter outlines some of the changes incorporated in the fifth-year followup examination and identifies several areas of future work expected to significantly augment the study.

FIFTH-YEAR FOLLOWUP EXAMINATION

Since the fifth-year followup examination was initiated prior to the full analysis of the data from the third-year examination, most modifications were founded upon quality control issues and the desire to make the clinical content of the examination more responsive to the medical needs of the participants.

Clinical quality control enhancements were made to improve measurement techniques. The digit preference noted in systolic and diastolic blood pressure readings led to the use of automated blood pressure recording; all other parameters of the blood pressure readings (e.g., sitting position, three recordings, nondominant arm at heart level) were not changed.

The problem in skin test reading was met by a rigorous quality control plan that included the following elements: refresher training for readers; a required reading of the four skin tests of all participants by both readers, each blind to the results of the other; a required reread of 10 percent of all tests by each of the readers, each blind to the previous reading; and a required weekly report citing numbers and proportions of participants with possible anergy, reversal of induration-erythema measurements, and untoward skin reactions or other reading problems (e.g., participant refusal).

In addition, new skin test forms were developed to facilitate accurate recording and transcription; specific clinical criteria were formulated to require consultation by an allergist; and the skin test measurement criterion for possible anergy, consistent with current World Health Organization guidelines, was adopted for the clinical interpretation of all skin test readings. It is anticipated that this clinical quality control program will standardize both readings and interpretations, and will produce a uniformly superior data set.

EXPOSURE INDEX REFINEMENTS

Since the development of the Study Protocol and the analysis of the 1982 Baseline data, there has been concern among some scientists and the principal

investigators over the accuracy and validity of the exposure estimates. It is unclear whether statistically significant differences in some variables between the Ranch Hand and Comparison groups, unsupported by dose-response estimates, have been due to chance, or whether true differences are obscured by an inadequate exposure index or group misclassification.

In mid-1986, strong correlations between dioxin levels in fat tissue and serum were demonstrated by the CDC and other institutions. Because of these results, the Air Force is currently engaged in a collaborative study with CDC to determine whether serum dioxin levels vary significantly in the Ranch Hand population. Approximately 200 AFHS volunteers have supplied a pint of blood to be analyzed for dioxin at the CDC laboratories. If clear and meaningful exposure findings are evident from this study, several additional studies are feasible: testing can be expanded to the entire study population and a meaningful exposure index based on total current TCDD body burden may be developed; and by means of archived AFHS serum samples from the Baseline study, it may be possible to calculate a reasonably precise half-life of TCDD in humans. These expanded studies will allow the estimation of body burdens of TCDD at the time of departure from SEA (assuming the absence of intervening vocational and recreational exposures).

If, in fact, these potential studies become reality within the next 2 years, the fifth-year followup study data will be statistically analyzed using a more appropriate exposure index. In anticipation of this advance, the AFHS is currently collecting 280 to 350 ml of blood from all volunteers attending the fifth-year followup study.

ADDITIONAL ANALYSES AND STUDIES

As in the 1984 Baseline Report, not all of the measured dependent variables were subjected to statistical analysis (e.g., prothrombin, leutinizing hormone, follicle stimulating hormone), largely because they were not within the bounds of the Air Force-prescribed analyses. Exploration of many of the unanalyzed variables is contemplated as time and resources permit. Similarly, many analytic opportunities to define possible symptom-clinical sign clusters or syndromes by multivariate analysis of variance techniques were passed over due to time and charter. Particularly challenging as an area of future work may be the changing relationships of some immunological variables over time and the biological impact of these changes on the induction of diseases such as cancer. Likewise, future efforts to define shifting cardiovascular disease patterns are a logical extension of the rich longitudinal data base of the AFHS. Such efforts await future analysis and publication.

The assessment of possible selection and participation bias has been addressed in a comprehensive manner in this report (see Chapter 5). The analyses and discussion suggest that statistical use of the total Comparison group (versus the Original Comparison group) is justified in this report, and that the impact of selection and participation biases have been minimal. As the followup studies continue, it is anticipated that a wealth of data on

compliance-participation factors will be available for continued comprehensive bias analyses. In particular, it is hoped that more complete data will exist to examine the true differences in current health status between refusals and their replacements. As the data set grows over time, the bias analyses will become more complex and will have to deal with changing motivations of the participants to continue in this study. Such bias analyses and assessments will always be of great importance to this study as they ultimately set the bounds for an inference on herbicide causality.