

## Chapter VI

### QUALITY CONTROL PROCEDURES

Quality control aspects of the Air Force Health Study have been of major importance since the inception of the study design. The focus of quality control concerns has been 1) to ensure the highest quality and validity of this study, 2) to reduce variability and bias in all data, 3) to validate all statistical methods and enhance statistical power wherever possible, and 4) to protect government resources. The purpose of this chapter is to present a categorical overview of the quality control procedures and to present representative data, where appropriate.

#### 1. Prestudy Considerations

The Study Protocol was formulated and refined in 1979-1980, during which time it underwent 4 independent peer reviews and a final review and approval by the Science Panel of the Agent Orange Working Group. Knowledge gained from visits to national and international herbicide dioxin experts was also instrumental in refining the Protocol.

Initial contract management aspects were handled on a scientific business basis. The Principal Investigators developed comprehensive statements of work with specific evaluation criteria. All contract proposals were evaluated without reviewer knowledge of the proposer and then scored independently on their scientific and business merits. Contracts were awarded on the basis of scientific and medical quality; price considerations were secondary. Fixed-price competitive contracts were written where feasible. During the conduct of the contracts, numerous scientific and business meetings were held with the contractors in an attempt to ensure quality and timeliness of the data. Scientific concerns continued as the primary emphasis throughout the periods of contract performance.

The population ascertainment process for both the Ranch Hand and comparison groups has continued for over 4 years. Extensive computer searches and a hand review of all available military personnel records have assured an almost complete and comparable identification mechanism. In addition, individual responses to the Ranch Hand Reunion Association and wide media coverage of the Agent Orange issue have greatly assisted both the ascertainment and address-update processes. A few potential study participants whose records were burned in the National Personnel Record Center remain uncategorized at this time. Both populations were subjected to a rigorous systematic location process (see Chapter III), resulting in a location efficiency of 99.5%; this achievement has eliminated population selection bias and has afforded each individual a maximum opportunity to participate in the study. The computer technique to match each Ranch Hand to a comparison individual by job category, race, and age to the closest birth month was exceptionally rewarding, as about 70% of the matches were exact to birth month and year, as well as to job and race. Such precision has enhanced the analytic flexibility of the statistical techniques cited in this report.

## 2. Questionnaire Data

The quality of questionnaire data was enhanced by 2 distinct mechanisms: 1) all questionnaire instruments were designed by nationally recognized survey research organization; and 2) the instruments were administered in an in-home setting by another outstanding survey research firm. A minimum number of highly qualified interviewers were used to reduce data variability, and the interviewers were blind to the exposure status of the respondent. In addition, the interviewers were specially trained and then race matched to the study participants, where possible. Spouse fertility data was obtained independently of the male interview but within the same interview setting.

The data collection verification process was conducted sequentially. The Louis Harris Associates Incorporated (LHA) field interviewer completed a questionnaire thoroughness edit, followed by a Central Office thoroughness check and appropriate editing. Participants were recontacted by phone, when necessary. LHA trained the United States Air Force interviewers and project staff to complete the identical sequential process. A double blind key punch system was used for both the LHA and USAF collected questionnaire data. Range checks identified outliers, and discrepancies were resolved. The contractor randomly validated completed interviews by phone; however, these interviews have not been analyzed for this report. An early USAF sampling review of the data revealed key punch error rates in specific sections of the questionnaire that ranged from 0 - 1.4%. The USAF systematic review and recoding of all medical areas included in this report have reduced these error rates. Further, subsequent to the questionnaire, each participant's military personnel record was hand reviewed, in order to provide exact data in the time and location of military assignments. These data have been used in this report in lieu of the memory-dependent military duty information obtained by the questionnaire.

Most study-participant questionnaire data were designed to be cross-referenced to review-of-systems data and physical examination findings. A notable exception, fertility birth defect data, will be validated by birth certificate or medical records, if retrievable. Female response data were used in all fertility/birth defect analyses, when available. In instances of multiple marriages and offspring, unexpected difficulty was often encountered in assigning a child to the correct spouse pair. Such discordant results were resolved by a hand review and computer input of the questionnaire data. Thereafter, this system supported all offspring data for analyses herein. Next-of-kin interview data will be verified by cross reference to the deceased's medical records. No attempt was made to validate the abbreviated noncompliant questionnaire because of the individuals expressed disinterest in the study.

## 3. Physical Examination Data

The bulk of scientific data of most concern to the public and veterans will stem from the physical examinations in this study. Consequently, great emphasis has been placed upon quality control of the physical examination and laboratory procedures.

All examinations were conducted at a single site by a contract medical organization of unquestioned reputation. The contractor was required to provide board certified physicians for the examination. Dermatologists were required to attend a 1-day intensive training session on the diagnosis of chloracne. A minimum number of physicians and paramedical staff was used to reduce data variability. The credentials of each physician and senior psychologist were submitted to the Air Force for approval. The contractor fulfilled the commitment to maintain a stable work force throughout the contract, best exemplified by the facts that (1) approximately 90% of the general physical examinations were conducted by one internist, (2) all electromyographic tests were performed by one technician using a single constantly calibrated machine, and (3) 90% of the final diagnostic assessments were made by 2 internists (master diagnosticians). All medical examiners were required to adhere strictly to the physical examination specifications as cited in the Study Protocol and were not permitted to evaluate a participant outside of his medical specialty area. Thus, each examiner was blind to examination findings outside his area of expertise, as well as to the exposure status of each participant. An Air Force physician, serving as an on-site physician monitor, conducted frequent inspections of all aspects of the physical, psychological, and laboratory examinations to ensure contract compliance and to approve further diagnostic workups for those participants exhibiting serious medical findings. Further, the Air Force monitor was periodically supplemented by Air Force consultant physicians in the areas of internal medicine, cardiology, dermatology, psychiatry, psychology, immunology, and laboratory medicine. For study participants crossing 2 or more time zones, 1 to 4 additional rest days were provided before the examination, in order to standardize psychological and laboratory parameters. All examination data were provided to the diagnostician who confirmed significant positive findings and formulated a diagnosis, if one was warranted. The diagnostician then carefully debriefed the participant and recommended follow-up medical action, if indicated. Electrocardiograms (ECG's) on all participants were sent to the Clinical Sciences Division, USAF School of Aerospace Medicine for cross-reference to the USAF ECG Repository. All data from the examination was collated and checked for completeness; this process was rechecked prior to submission to the data processors. Computer entry of all data was made by a single key-to-disk entry with hard copy verification; visual range checks were accomplished prior to transmittal. The Air Force data processors conducted a small sampling from the data set and detected sectional error rates ranging from 0.2 - 1.3%, with 6 of the 7 sectional rates ranging from 0.2 - 0.4%. Plausible ranges were established for most variables and all data outside this range were verified against the hard copy of the examination. All discordant transcription errors were corrected; otherwise, the data were accepted as correct. Inconsistent dates were corrected, where possible. All data sets or subsets were checked for reasonability and, in many cases, the information was verified by the hard copy of the examination.

#### 4. Laboratory Procedures

Because the thrust of the physical examination was to cast as wide a clinical net as possible, the importance and number of laboratory tests were substantially increased over an ordinary diagnostic or screening examination. Thus, all contract and subcontract laboratories were required to be licensed

and certified by the College of American Pathologists or by the Centers for Disease Control under the Clinical Laboratory Improvement Act of 1977. For the laboratory battery of 36 tests, each responsible contract or subcontract laboratory was required to maintain quality control data for audit. The bulk of nonradioassay procedures was accomplished at the contract clinic; a DuPont Automated Chemical Analyzer III (ACA) and Hemalogs 890 and D90 Automated Counters performed the majority of tests. For the ACA, reagents of the same lot number were used throughout the study period. Stringent research grade coefficients of variation (CV's) were required for most assays (see Appendix XV), often necessitating repeat runs to meet these standards. Where available for specific assays, trilevel controls were run at intervals of every 10th specimen, and 1 specimen set of every 15th was run in duplicate. These results were used to generate cumulative sum quality control charts to determine if test systems drifted significantly out of control over time since the CV's are relatively insensitive to trends over time. Of the 14 assays with CV requirement standards, 7 were significantly ( $P < .05$ ) out of standard at 1 or more levels. On-site visits and detailed power calculations with respect to detecting differences between means showed that these variances would not substantially or biologically alter group comparisons or conclusions. Adjustment of study participant clinical values for drift and other variations in laboratory control levels was considered, but was determined unnecessary. This decision was made by evaluating participant and laboratory quality control values for High-Density Lipoprotein (HDL). Deviations were computed from each overall tri-level mean and these were subtracted from each participant's value. The distributions with and without adjustment were then contrasted. The results are tabulated below:

Table VI-1

HDL VALUES ON 2227 PARTICIPANTS (mg/100 ml)

	<u>Original Value</u>	<u>Adjusted Value</u>
Mean	46.18	46.12
Standard Deviation	12.61	12.72

No increase in HDL precision is noted. In fact, a small increase in the standard deviation was found, clearly indicating that adjustment would not improve the ability to detect group differences.

Immunologic assessments were performed by subcontract on 592 participants. Participants were randomly selected (terminal digit of their random study number) midway through the physical examination contract. The subcontractor was blind as to the exposure status and group membership of each individual. The functional capacity of lymphocytes to respond to mitogens or antigens and the number of T and B lymphocytes were measured in isolated peripheral blood. An Immunologic Peer Review Group (see Appendix I) was convened on-site to review technical procedures and to develop analytic strategies. This panel determined

that 56 of the 592 samples were not processed due to technical errors in specimen handling. The procedure used for isolation of purified mononuclear cells was substandard. This resulted in cell populations which were depleted of adherent mononuclear cells and contaminated with polymorphonuclear leukocytes and red cells. Differential counts on purified cells were not accomplished so that the actual number of mononuclear cells used for each assay was not determined. A number of the lymphocyte function assays had excessive variation, manifested by a coefficient of variation (CV) greater than 15%, as reflected in Table VI-2.

Table VI-2

PERCENT OF GROUPED LYMPHOCYTE FUNCTION ASSAYS EXCEEDING A CV OF 15%

<u>Functional Test</u>	<u>Percent</u>
Concanavallin A	15.8
Phytohemagglutinin	20.3
Tetanus Toxoid	75.7
Pokeweed Antigen	10.2

Although CV's were excessive, these variations appeared to be randomly distributed since there were no observed trends over time and there were no differences in error distribution between groups. Only 11 duplicate specimens were received (1 per 50 specimens). Intraspecimen reproducibility was impaired and several split samples varied by more than 50%. Similarly, intraspecimen reproducibility was reduced and represented sporadically within the data set. Further, 54/432 specimens (12.5% of the total) had a ratio of concanavallin A to phytohemagglutinin less than 0.30, indicating mitogen dysfunction rather than failure of lymphocytes to respond to mitogen. The low levels of stimulation observed in many tetanus toxoid-stimulated cultures additionally suggested that caution should be used in the interpretation of the functional results. Accordingly, the Immunology Peer Review Group recommended that the lymphocyte function data not be used clinically to determine the immune status of an individual participant. Further, the panel recommended that the functional data set be used only to evaluate differences, if any, between the Ranch Hand and comparison groups.

The T and B lymphocyte enumeration studies demonstrated acceptable reproducibility and acceptable daily and long-term variations between the total T lymphocyte ( $T_3$ ) and the sum of lymphocyte subsets ( $T_4$  and  $T_8$ ). Criteria for exclusion of T and B lymphocyte data were (1) samples exhibiting greater than a 30% background fluorescence (11 samples or 2%), and (2) samples with a  $T_3$  or  $T_{11}$  proportion of less than 10% (7 samples or 1.3%). Although differential counts were not performed initially on the Ficoll-hypaque separated cells, sufficient paraformaldehyde-stored cells were available after conclusion of the contract to permit a 250 cell differential count on 525 of the 592 specimens.

This count permitted the calculation of absolute T and B lymphocyte numbers. After application of acceptability criteria, cell count data were available on 490 specimens.

##### 5. In-House Data Collection and Statistical Analysis

The complexity and time constraints of this study have made it impractical to hire a series of contractors and expect them to accomplish integrated and timely work. Thus, the Air Force investigators and technical staff have assumed major roles in the areas of population ascertainment and location, verification of eligibility in the study, medical record and personnel record validations, determination of replacements, examination scheduling, medical coding, repository formation, and statistical analyses. Where at all possible, in-house actions have been documented by coding schemes, decision rules, user manuals, and computer audit trails. It is our desire to submit duplicate unedited copies of all contractor data tapes to the Advisory Committee for storage and any possible later use.

The data repository task has been monumental. All medical coding has been accomplished in duplicate with resolution of disputes. All in-house gathered data have been subjected to 100% echo and consistency checking. Subsamples have been obtained to develop quality control error rates. Backup hard copies have been created for all data bases in the event of computer loss or malfunction.

The statistical approach to this study consists of a preset state-of-the-art framework. The statistical strategy was detailed before the data were reviewed or the group membership codes broken. Both external peer review and internal reviews (conducted by civilian consultants) have validated our approaches. Computer software have been extensively validated by using mock data sets.