

CHAPTER 6

QUALITY CONTROL

During the 1987 Air Force Health Study (AFHS) followup, stringent adherence to quality assurance (QA) was planned for and upheld throughout the study, from project initiation to final product delivery and acceptance by the Air Force. A quality program plan was developed for this study cycle, outlining all contract activities requiring periodic and/or systematic QA and quality control (QC) monitoring. The purpose of this chapter is to provide an overview of the specific QA measures developed and used by the project team, specifically in the areas of administrative QA; questionnaire, physical, and psychological examination QC; laboratory QC measures; data management QC; and statistical QC.

ADMINISTRATIVE QUALITY ASSURANCE

In recognition of the magnitude, complexity, and importance of the AFHS, a Quality Review Committee (QRC) was established, at the contractor's initiative, at the initiation of the 1985 followup and continued through the 1987 followup for the purpose of providing general oversight to the AFHS QA Program and advice on the appropriateness of program management and QC actions. The QRC was composed of senior corporate personnel from the prime contractor. These independent reviewers remained separate from the project management staff. The QRC met formally each quarter to review recent study progress and any issues that either had an impact on study quality or were perceived as a potential problem.

Assisting the QRC in day-to-day oversight responsibilities was a QA secretary. As part of the monitoring function, the QA secretary received exception reports from project task managers whenever an incident occurred that could affect study quality. Monthly reports were also prepared for the Air Force, documenting project compliance with project QA criteria and noting any instances of noncompliance.

The remainder of this chapter describes the specific QC procedures followed for the individual tasks.

QUESTIONNAIRE QUALITY CONTROL

The National Opinion Research Center (NORC) used both onsite and home-office QC procedures to produce a comprehensive data set. All AFHS questionnaires were pretested to evaluate their completion time and participant acceptability before they were used at the Scripps Clinic and Research Foundation (SCRF). Onsite QC procedures included observing and rating interviewers, review of every questionnaire at the completion of the interview, and monitoring participant evaluations. The Air Force also continuously conducted QA observations of all onsite activities. QC of data processing included manually editing each questionnaire, including verifying critical items (10% of total items) for each questionnaire, computerized cleaning (with both single item and interitem review for range and consistency), identifying values out of range, and reviewing the actual questionnaire copy to reconcile or correct detected errors.

NORC recruited and trained 12 interviewers according to the procedures described in Chapter 3. A minimum number of interviewers was selected to reduce interviewer variability. Additionally, these individuals were blinded to the participants' exposure status to avoid bias. Interviewers were required to ask questions exactly as recorded, and in the order in which they appeared. No personal interpretation was allowed.

An onsite field manager closely supervised each interviewer's work, observing individual interviews weekly during the examination schedule. The field manager reported directly to the NORC Project Director weekly, and was in turn evaluated by the Project Director during quarterly site visits, to ensure direct accountability by the home office and the field manager for promptly resolving any issues.

Specifically, interviewers were checked for accuracy in questionnaire skip patterns, probing, circling of the correct code, control of the interview, voice quality, reading, and use of associated documents. When called for, the onsite manager gave immediate retraining after each error and documented the content of this training. At weekly meetings, held with all interviewers, the field manager used generalizations from individual interviewer performance observations to train the entire group of interviewers.

The NORC field manager also monitored participant evaluations of the study closely and used the information gathered to plan and implement retraining. The manager and staff reviewed each completed questionnaire, attempting to retrieve missing data while the study participant was at the physical examination site. In addition, a second review of the questionnaires for completeness was conducted by a reviewer who was independent of the interviewing staff. Missing or ambiguous data were also retrieved by telephone when necessary.

Once the participant questionnaires were received for data processing, they were reviewed for completeness by a coding supervisor and staff dedicated to the AFHS for the entire project. Resolution of inconsistencies was accomplished by staff members, who coded all responses prior to keypunching. Questionnaires were then coded, and a 10-percent recode was done on open-ended items. When a batch failed the 10-percent recode, the entire batch was recoded and the coding staff was retrained.

During data entry, range validity checks were performed and 10 percent of the most important items in each questionnaire was verified. Data were then passed through a computer program that checked for inter- and intra-column errors. When errors were detected, the questionnaires were reviewed and the errors corrected. The process continued until no errors were detected by the cleaning program. Then, frequencies were reviewed and any anomalies or errors previously undetected were corrected by reviewing the questionnaires on a case-by-case basis. All corrections were documented and entered into the data base, but no changes were made to the original data recorded in the questionnaires. QA reports were generated monthly, detailing the summary statistics on the number of questionnaires reviewed, the number and types of transcriptions failing QC checks, and the average number of coding errors per batch processed. The data review process continued until no errors or discrepancies were discernible.

PHYSICAL EXAMINATION QUALITY CONTROL

QC was emphasized in the physical examination, as this data source provided most of the medical information for clinical and epidemiologic analyses.

Initial concern for a high-quality physical examination was addressed by a stringent SCRF selection process for all personnel who were to directly interact with the participants. Each staff member was hand-selected for the AFHS on the basis of expertise, experience, and a commitment to remain with the study throughout the examination cycle. Further, the Air Force reviewed the credentials of all key staff members and approved their participation in the study.

A complete pretest physical examination, interview, psychological test, and laboratory workup was done for 11 volunteers several weeks before the scheduled start of the study. Refresher training was given to the dermatologists to enhance their skill in diagnosing chloracne, techniques for detecting specific heart sounds were reviewed with the internists, and diagnosticians were reminded of the need to review Baseline and 1985 examination data as they formulated all diagnoses. Additionally, automatic monitors to measure blood pressure were instituted for more accurate readings. Further, all aspects of patient contact were reviewed: the initial inbriefing of the participants, the logistics of transportation and patient flow within the clinic, and the final outbriefing by the diagnostician.

During the examinations, refinements continued whenever operational problems were detected by the SCRF staff and the Air Force onsite monitor, or when participants identified areas requiring improvement. Both of these types of information were addressed during the weekly clinical QA meeting of key SCRF staff, chaired by the SCRF Medical Project Director and attended by an Air Force representative. In addition, written critique forms submitted by all participants were reviewed in detail at the SCRF weekly meetings, providing additional insight to both temporary shortcomings of the entire logistic process as well as the numerous strong points of the programs.

Following examination of each participant group, all physical examination forms were reviewed by the SCRF staff for omissions, incomplete examinations, and inconsistencies. The examiners or technicians were quickly contacted to correct the data. Special effort was made to complete this review while the participants were at the examination site. In all cases of data correction, a complete audit trail was maintained. Finally, all mark-sense physical examination forms were read by an optical scanner. (This subject is discussed in more detail in the Data Management Quality Control section of this chapter.)

Compliance with all aspects of the physical examination was monitored daily by the Air Force onsite monitor and the SCRF Medical Project Director. Additional periodic inspections were conducted by the SCRF Chief of Medicine and the Science Applications International Corporation (SAIC) Principal Investigator. All such clinical reviews were done unobtrusively, and with the full consent of the participant; suggestions or corrections to the examination procedure were always discussed privately with the attending physician. These inspections emphasized aspects of clinical techniques,

sequencing and completeness of the clinical data with respect to the examination forms, and the total blindness of the examinations. Of particular note were the detailed daily log entries of the six Air Force monitors. These entries ensured continuity of knowledge (the monitors rotated approximately every 2 weeks) by documenting examination procedural changes and recording events requiring followup by either the Air Force or the prime contractor.

Establishment of rapport with each study participant was a primary goal of all organizations involved in this study. Although "rapport building" may not be a traditional QA parameter in most research studies, it is paramount in the AFHS because maintaining the satisfaction of participants encourages them to continue in the study, and thus a significant reduction in future statistical power or bias, or both, is avoided. Therefore, every staff member, from the initial telephone recruiter to the nurse coordinator and the Project Manager, emphasized courtesy, empathy, assistance, and personalized treatment of each participant. Based on the evaluation forms, 67 percent of the participants evaluated their experience in the 1987 followup as excellent and 27 percent classified it as good. Five percent of the participants rated the experience as satisfactory and only 1 percent felt that it was unsatisfactory.

LABORATORY QUALITY CONTROL

Before the study was begun, specific QC laboratory procedures were designed, developed, and implemented to rapidly detect problems related to test/assay performance, validity of reagents, analysis of data, and reporting of results. All laboratory assays for the study were done with state-of-the-art laboratory equipment and techniques. Laboratory facilities all had the equivalent of National Institutes of Health Biosafety Level 2 approval ratings and were certified by the College of American Pathology.

Quality Control Procedures for the Clinical Laboratory

Hematology assays were performed on Coulter 5-Plus[®] equipment; sedimentation rate determinations were performed using the large-tube Westergren method. The Dupont Automated Chemical Analyzer[®] was used to perform the biochemical assays; radioimmunoassays were done with standard test kits. Electrophoresis and occult blood tests were performed manually. Hepatitis B tests were performed using Abbott Diagnostic kits. Monospecific antibodies were used for immunoglobulin assays using the Beckman Array Protein System[®]. Blood-cell counts were performed with standard microscopy, and Clinitek[®], a reflectance spectometry urinalysis, was used for all urinalyses. All other assays were done using industry-approved equipment and techniques.

All laboratory operations were controlled with the use of an integrated medical laboratory management information system that incorporated direct device to data base interfaces for automated testing equipment, and data entry for manual tests was performed by the laboratory technologists. An automated audit trail and a set of comments for technologist remarks were kept for each test so that any QC results could be retraced.

Procedural QC included using instrumentation and reagents from the same lot numbers throughout the study. Strict standards of calibration for all automated laboratory equipment were maintained at all times.

Trilevel or bilevel controls were used as the primary means for monitoring the quality of all tests. On every group of participant samples, one control (low, medium, or high) was run at the start, after every ninth sample, and at the end of each test run. Each trilevel control was used before repeating it in the run, when more than 18 experimental samples were analyzed. In addition, split aliquots were made from every tenth patient sample and were analyzed separately to measure test reproducibility.

All QC data were analyzed and summarized in formal QC reports generated weekly. QC data were subjected to independent statistical analysis to produce and analyze time-dependent trends. For all equipment malfunctions or other exceptions, a formal QC exception report was prepared by the responsible individual and forwarded to the QA officer and the project management team.

An additional measure of quality control introduced during the study was the cumulative sum (CUSUM) tests run with trilevel controls.¹ In particular, the fast initial response (FIR) CUSUM QC technique was used. It has an advantage in detecting long-term, subtle drift that could have substantial adverse analytical consequences.² FIR is a special case of the CUSUM QC scheme that increases the overall effectiveness of the QC procedure. Unlike QC procedures using standard control charts, which compare each observation to designated limits, these tests utilize the cumulative sum of deviations from a target value.

CUSUM statistics were accumulated for each of the trilevels to quickly detect instrument calibration problems as identified by excessive drift. If an out-of-control situation was indicated, the graph showed when the change first occurred. When CUSUM indicated an out-of-control situation, all adjacent patient samples were reanalyzed after the equipment was thoroughly checked and fresh controls were run. Coefficient of variation (CV) requirements were established before the study for each test.

FIR CUSUM generally has been applied to QC in industry, particularly in high-volume, high-precision applications. It is believed that FIR CUSUM has not generally been applied in a biomedical setting. This procedure has proven to be effective and is now being used regularly in the SCRF clinical laboratory.

As the examination portion of this study ended, laboratory outliers were analyzed for logical validity by an independent clinician. All out-of-range test results were examined and scored as clinically explainable, clinically possible, or clinically unexplained. No clinical laboratory data were excluded because all out-of-range results were found to be clinically explainable or clinically possible.

Quality Control Procedures for the Immunology Laboratory

The QC procedures for the Cellular Immunology section of the AFHS were structured to rapidly detect any problems in four major test parameters:

(1) assay performance, (2) reagent validity, (3) data analysis, and (4) results reporting. The QC measures were detailed in the Quality Procedures Plan and documented before testing started. Compliance was monitored daily by the Cellular Immunology laboratory supervisor. Key aspects of the program included instrument and equipment calibration and maintenance, assay controls, accuracy and precision determination, and system failure checks.

QC measures followed in all Cellular Immunology assays included:

- Testing of a blood sample from a normal, healthy control individual with each group of AFHS patient samples
- Duplicate testing of one random patient sample in each assay
- Quadruplicate testing of each patient sample for each variable in each of the functional assays (e.g., phytohemagglutinin [PHA] stimulation, natural killer cell, and mixed lymphocyte culture)
- Parallel testing and monitoring reactivity of various lots of reagents when appropriate
- Verification of patient and specimen identification by at least two individuals before final reporting to the data base
- Note codes attached to any data point with a detected deviation due to procedural setup error, assay malfunction, equipment malfunction, or assay technical error
- Note codes attached to any data point outside the range of expected values as identified by the Cellular Immunology laboratory supervisor
- Review of all final assay reports by the Cellular Immunology laboratory supervisor prior to entry into the data base.

QC for each functional assay including PHA, mixed lymphocyte culture, and natural killer cell consisted of monitoring assay controls, duplicate sample reproducibility, and trends in reagent reactivity. Assay precision was determined by calculating the CV of the quadruplicates for each variable tested. Also, a mean value of the CV for each assay was calculated. Individual CV's of 15 percent or less were the target values for the stimulated samples in the mitogen and natural killer cell assays. The Student's t-test was applied to duplicates to determine if there was a significant difference in sampling for the functional assays. Critical t-values at the 0.05 significance level were used to determine if duplicate sample results varied significantly. Positive and negative values were assigned, arbitrarily subtracting the second duplicate value from the first, to determine if there was a systematic bias in one direction. Grubbs' statistical test³ was used to identify any statistically significant outlier. This test was applied only to samples whose CV's were greater than 20 percent at a p-value of 0.01. The mitogen stimulation (PHA) effect was followed by daily evaluation of the radioactive counts in counts per minute. When counts fell below expected values, suggesting that reagent deterioration had occurred, new aliquots were used.

QC measures for the cell surface marker assays included: calculation of $(CD4 + CD8)/CD2$ (formerly $[T_4 + T_8]/T_{11}$) cell ratios, evaluation of flow cytometer computer outputs (cytograms and histograms), and duplicate sample testing. The cellular ratios should approximate the value 1.0 for a normal population. Validity of cytogram and histogram distributions generated by the flow cytometer was confirmed by the Cellular Immunology laboratory supervisor for each sample analyzed. The proportional difference between duplicate samples was calculated and monitored for significant differences.

On completion of this followup effort, the entire cellular immunology data base was reviewed by the Air Force team, laboratory staff, and an immunology consultant. Comments attached to the data points were also reviewed. Any data point that appeared to be a significant outlier was reviewed and coded as an unexplained outlier. Unexplained outliers were deleted from the data base as errors of an unknown nature. This review was conducted without knowledge of exposure status. The results of this review are presented in Chapter 19.

DATA MANAGEMENT QUALITY CONTROL

Overview of Quality Control Procedures

The QC program for the data management activity consisted of multiple checks at all steps of the examination, data collection, and data processing cycle. Data QC procedures for data collection, conversion, and integration were developed before the clinical examinations began. Pretesting of all forms, procedures, and logistic arrangements was conducted 3 weeks before the examinations actually began. Additionally, during the first 2 months of the clinical examinations, all data collection activities were intensely scrutinized to detect and correct procedural deficiencies.

QC activities also included automated QC techniques applied to laboratory data; clinical evaluations of all laboratory outliers; review of all physical examination findings by one of two diagnosticians who was not involved in the conduct of the physical examinations; and automated and manual data quality checking of hard copy against transcribed computer files for all questionnaire, physical examination, and medical coding data streams.

Five interwoven layers of QC were instituted to ensure data integrity. Efforts focused on (1) data processing system design, (2) design and administration of all exams or questionnaires, (3) data completeness checks, (4) data validation techniques, and (5) quality control of medical records coding. In some cases, the QC procedures described in this section were implemented throughout the data management task rather than assigned to a particular activity. These comprehensive QC procedures will be mentioned where appropriate throughout the remainder of this section.

Data Processing System Design

For each data stream, standards were set to establish data element format (character or numeric), data element naming conventions, data element

text labels, numeric codes for qualitative responses and results, QC range checks for continuous data elements, and QC validity checks for categorical data. A data dictionary provided detailed information on each data element.

A systems integration approach was applied to the design and implementation of data collection procedures and techniques so that data emanating from the various study sources (physical examination, questionnaire, laboratory) were consistent in file format and structure. This was necessary to ensure that all data could be integrated into a single data base management system for analysis. Figure 6-1 provides an overview of the QC activities used in the data management process.

Forms and questionnaires were carefully designed to ensure that all required data elements would be collected in accordance with the Study Protocol and in a standardized format. The design of these instruments was such that they reflected the order in which the examination itself would be administered and provided for the sequential recoding of information to streamline remaining data management activities.

Completed medical records and questionnaires were converted from hard copy to machine-readable images using customized data-entry systems or state-of-the-art optical mark reading equipment. Verification procedures were performed to ensure that a uniquely identified participant record existed within each data file, and that the appropriate number of responses for each applicable field was provided. Data files were then verified against original data sheets and corrected as necessary.

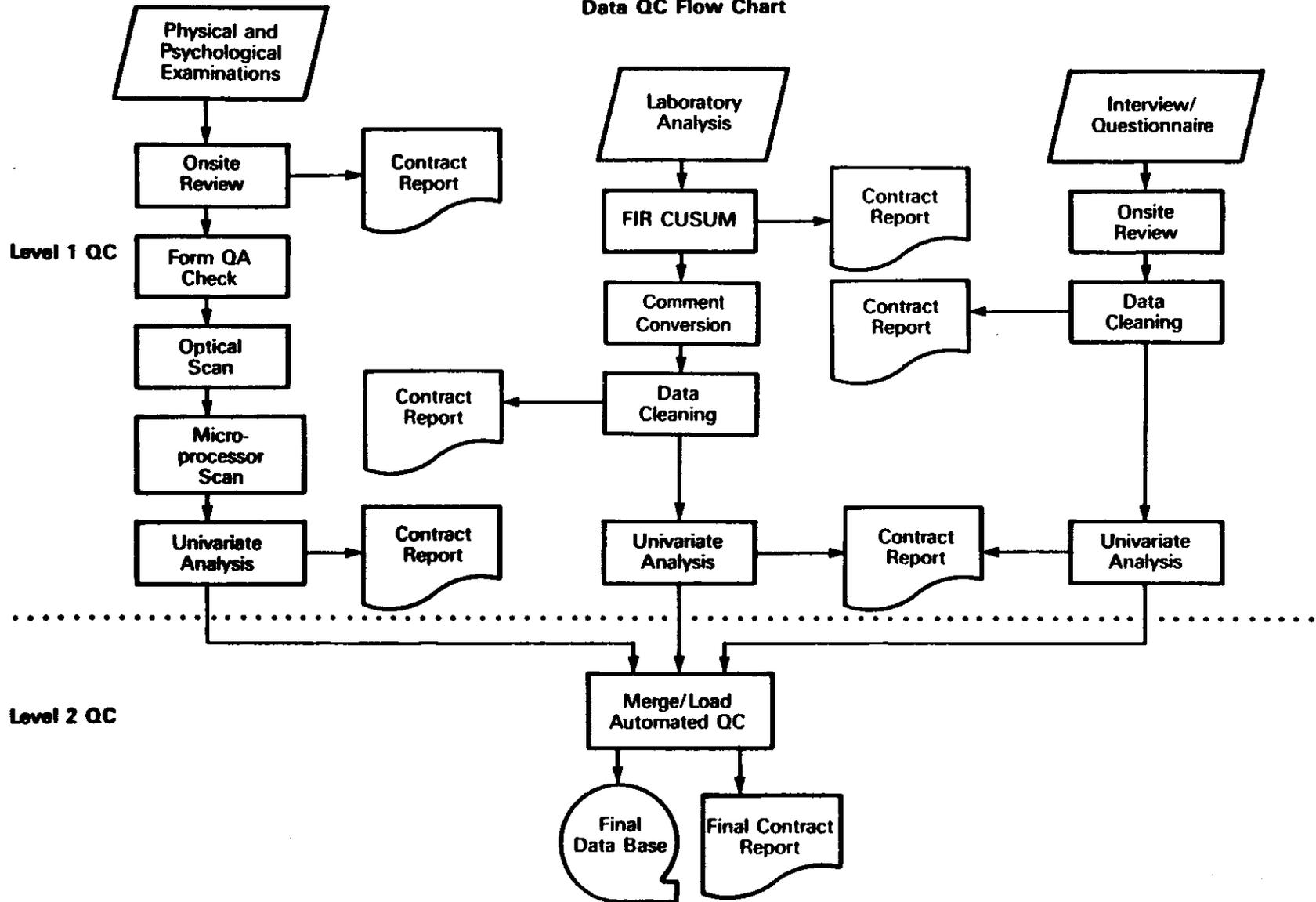
Data files were then subjected to validity checks. Any potentially conflicting results as well as any data values falling at the extremes of expected ranges were manually reviewed. Extreme values were reverified against the original raw data copies and either corrected or documented as valid results. Potentially conflicting results were returned to the examiners for review. These results were then documented as correctly recorded, corrected, or flagged for exclusion from analysis because of unresolvable examiner errors or omissions. This process was continued until all results were properly documented.

Once the edits were completed and the data reverified, the "cleaned" files or tapes were transferred to the data analysis center for final inspection and integration into the study data base. For this QC measure, each data file was loaded into a SAS® data set, and descriptive analyses were run. The validation, correction, transmission, and analysis QC procedures were repeated as necessary to ensure that all extreme or suspicious values had been validated.

Design and Administration of Physical and Psychological Examination Forms

As mentioned, the examination forms were designed to solicit all required data such that recording time was minimized, comprehension was enhanced, and data input could occur with a minimum of transcription errors. Optical Mark Recognition (OMR) technologies were selected to eliminate the risk of transcription errors and were applied to all psychological tests. Customized mark-sense forms were also developed and OMR technology was used

Data QC Flow Chart



6-9

Figure 6-1.
Two Levels of Quality Control Applied to All
Collected Data Prior to Statistical Analysis

to achieve these same objectives for segments of the physical examination and the self-administered questionnaires. The use of mark-sense forms allowed the creation of computerized data files directly from the raw data recorded on these forms.

QC procedures for all data collection instruments began with a review of all forms as they were completed. Any forms containing missing examination results were returned to the examining physician for completion before the participants left the site. Any questionable results or "hard-to-diagnose" conditions (such as heart sounds or peripheral pulses) were verified by the diagnostician at the outbriefing. All examination forms were signed by the examining physician, and the examiner identification number was coded in the data base. Detailed QC records were maintained, which indicated the examining physician and the type of deficiency detected. Deficiency reports were reviewed by the study coordinator to detect any patterns of physician data entry error. A final level of QC audit was accomplished by Air Force statisticians, who conducted a detailed screening of the data and checked for errors.

Data Completeness Checks

Customized programming of the OMR allowed for the identification of those forms (and their corresponding data records) with missing responses, as well as those with multiple responses to questions that required a single response. The OMR scanner was programmed to reject forms that failed completeness and multiple response checks and to output a control code for each rejected form. The control code identified the location of the first three verification checks failed for a given form.

When a raw data form was rejected, the reason for the rejection was determined and the exact data element was corrected by comparing the rejected raw data form to the values recorded in the data record created by the scanner. A customized set of rejection and resolution codes was developed for the study to describe all the reasons for a form's rejection and any subsequent reasons for changing a data value. Various codes identified values recovered from light marks, missing marks explained by examiner comments, and missing comment flags resolved by the presence or absence of text in the comment areas. These codes ensured data completeness by accounting for all questionable or missing responses.

Some of the rejected forms did not contain actual data errors but rather anomalies created in using mark-sense cards for data collection. For instance, incompletely erased responses and responses marked with too little carbon or graphite were incorrectly counted or missed, respectively, by the scanner. Examiners also tended to clearly mark responses for abnormal findings while bypassing or lightly marking responses for expected or desired findings. Failure of the form to provide the correct number of expected responses always resulted in rejection. These technology-based errors were resolved, as were the anticipated, more traditional errors.

The rejection code, data location code, resolution code, data inspector's initials, and correct data value were directly posted to a participant's data record. This innovative technique not only effectively

maintained a comprehensive audit trail of all record manipulations, it also provided a mechanism for measuring the frequency of specific errors.

Statistics were compiled on out-of-range results and data omissions that had been accepted in the previous QC audits. The results were monitored to detect trends, possible bias situations, and other data quality problems. This information was reviewed and relayed to examiners and internal auditors to assist in preventing or correcting chronic, but avoidable, problems. Refresher training was provided to examining physicians to avoid data omissions. Physicians were consulted to recover missing data, and out-of-range results were reviewed for logical validity by an independent clinician.

Data Validation Techniques

QC activities also included data validation techniques. As mentioned earlier, data files were examined in a series of verification and validation procedures developed to check the results within each participant's record for logical consistency and abnormal findings. Any records noted to have ambiguous findings, incongruent observations, extreme results, or errors or omissions were listed and submitted for review to a physician.

Again, clinical judgments were made by the auditing physician in assigning a validation code for each extreme or questionable data result. The validation codes allowed for indicating that data were deciphered from examiner comments or from related findings from another specialty area, or were accurately recorded and logically consistent with other findings for the participant. Data points that could not be definitively validated or recovered through clinical judgment and consultation with the original examiner were assigned codes noting missing or invalid data values. Some reasons for data not being available for analysis included participant refusal; incomplete, confusing, ambiguous, or unclassifiable information; contaminated samples; unscorable psychological examinations; use of data from previous Air Force studies at which the 1987 followup participant was not present; and an exemption from testing (e.g., exemption from delayed skin testing to prevent confounding of immunology panel results). These unrecoverable data points were excluded from subsequent analysis. The number of values that were not available for analyses is presented in Chapters 9 through 20 by variable and group.

Medical Records Coding Quality Control

After inventory, SAIC forwarded completed questionnaires and physical examination records to the Air Force at Brooks AFB, Texas, for diagnostic coding and verification of all subjectively reported conditions. The Air Force used the International Classification of Diseases, 9th Revision, Clinical Modification for morbidity coding; the International Classification of Diseases, 9th Revision, for mortality coding; the Systematized Nomenclature of Medicine for anatomic site coding; and the American Hospital Formulary Service for medication coding. Two coders independently processed each questionnaire and physical examination. Both codings were then subjected to a 100-percent QA and QC review, during which every posted code was checked against medical records. A third party adjudicated any discordances.

After QA and QC review and/or adjudication, information from the coding sheets was placed into the AFHS data base using a 100-percent double blind data entry and verification scheme. Any discordances were reviewed, corrected, and again subjected to double blind entry and verification. After coding and data entry, the Air Force batched the questionnaires and forwarded them to NORC in Chicago, Illinois, for data processing. The Air Force then obtained the NORC questionnaire data tape, matched this information to the Air Force data file, and resolved any differences. A single, final combined data base was produced by the contractor, and a copy was sent to the Air Force.

STATISTICAL ANALYSIS QUALITY CONTROL

Specific QC measures were developed for activities falling within the statistical analysis task: construction of data bases for the statistical analysis of each clinical chapter, the statistical analysis itself, and the preparation of the clinical chapters.

Each specialized statistical data base was constructed by defining and locating each variable within the many subparts of the composite followup data base. Although the data had been subjected to QC procedures during collection, statistical checks for outliers and other improbable values were conducted; anomalies identified by the statisticians were discussed with those responsible for the data collection, i.e., either NORC or SCRF.

The data base was frozen prior to starting the statistical analysis. However, during the data analysis, some discrepancies or data problems were identified. Each issue was investigated to determine the nature and impact on the outcome of the analyses and documented. For all but two issues, described below, the analyses were reaccomplished using revised data.

- 1) One Black Ranch Hand was inadvertently coded as a nonblack in the data base. Since all of the 1987 followup analyses had been completed before the error was identified, selected variables were reanalyzed to determine the impact of having one Ranch Hand misclassified on race. (Only the analyses that utilized race could be affected by this error.) Race was used in the adjusted analyses (group contrast and one stratum of the exposure index), interaction analyses, dependent variable associations, and unadjusted skin cancer analyses since Blacks were excluded. Variables were selected where (1) the result of the adjusted group contrast was significant, (2) the misclassified participant was abnormal, and/or (3) Blacks were excluded.

For group contrasts, race was used indirectly (i.e., exclusion or covariate). For most analyses, the effect was in the third decimal place of the p-value. Changes of this order of magnitude in the significance level could result from using two different statistical methods or different software manufacturers of the same analysis method. The change in the p-value was larger for stratified analyses and nonsignificant results but would not change the overall statistical conclusion. The change in the p-values for covariate associations was slightly larger (second decimal place). However, the dependent variable-race associations are strictly summary statistics and auxiliary information with no relevance to the statistical conclusion on group differences.

The misclassified Ranch Hand was an enlisted flyer. Since the sample size for the enlisted flyer cohort is smaller (171) than for group contrasts (2,294), the change in the p-value was also slightly larger, and the change followed the same pattern as group contrasts. However, minimal emphasis is placed on the results of the exposure analysis, and the change in results would not impact the overall statistical conclusions of a clinical area.

Thus, the effect of having one participant misclassified on race does not have a substantial effect on the analysis results and did not warrant reanalysis of the data.

- 2) In reviewing the medical records for diabetes, it was determined that 13 participants had been misclassified (11 participants were coded in error as having a verified history of diabetes, and 2 participants coded as normal actually have a history of diabetes as verified by medical record). Verified history of diabetes was used as a dependent variable in the endocrine assessment, a candidate covariate for neurological and renal analyses, an exclusion for 2-hour postprandial glucose in the endocrine assessment, and an exclusion in the cardiovascular assessment.

In the dependent variable analysis of verified history of diabetes, the classification of the 13 participants was corrected, and the analysis was reaccomplished. When verified history of diabetes was used as a covariate or exclusion, the misclassification of the 13 participants was judged to be negligible, and reanalysis using revised data showed little difference or was not deemed necessary.

QA largely depended on regular communication and general agreement among statisticians. Several meetings and consultations among the Air Force team, the SAIC Principal Investigator, the SAIC statisticians, and the University of Chicago staff members were held in conjunction with the development of the data analysis plan. During the course of the analysis there were frequent telephone conversations. Any problems arising in the statistical analysis were resolved by team discussion. The software was checked by comparing results from analyses on the same variable by different programs (for example, BMDP®-LR [logistic regression] and BMDP®-4F [log-linear model] will give the same results for dichotomous variables when the program options are appropriately chosen). The statisticians frequently checked that the number of observations used in an analysis was correct, and peer review ensured that the program code was appropriate for the chosen procedure. The analyses were conducted in accordance with the data analysis plan, which was reviewed extensively. Throughout the study, duplicate data bases were maintained by the Air Force and SAIC. Upon completion of the analyses, SAIC delivered all analysis software and SAS® data sets for each clinical area to the Air Force for final review and archiving.

All tables and statistical results were checked against the computer output from which they were derived, and all statistical statements in the text were checked for consistency with the results given in the tables. Additionally, drafts of chapters in the report were reviewed by the Air Force and SAIC investigators, and the QRC.

CHAPTER 6

REFERENCES

1. Bissell, A.F. 1969. CUSUM techniques for quality control. Appl. Stat. Vol. 18.
2. Lucas, J.M., and R. Crosier. 1982. Fast initial response for CUSUM quality control schemes: Give your CUSUM a headstart. Technometrics Vol. 24.
3. Grubbs, F.E. 1969. Procedures for detecting outlying observations in samples. Technometrics XI:1-21.