EPA-600/1-78-012
February 1978

Environmental Health Effects Research Series

DEVELOPMENT OF QUALITY ASSURANCE PLANS FOR RESEARCH TASKS- HEALTH EFFECTS RESEARCH LABORATORY/RTP, NC

Health Effects Research Laboratory
Office of Research and Development
U.S. Environmental Protection Agency
Research Triangle Park, North Carolina 27711

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U.S. Environmental Protection Agency, have been grouped into nine series. These nine broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The nine series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies
6. Scientific and Technical Assessment Reports (STAR)
7. Interagency Energy-Environment Research and Development
8. "Special" Reports
9. Miscellaneous Reports

This report has been assigned to the ENVIRONMENTAL HEALTH EFFECTS RESEARCH series. This series describes projects and studies relating to the tolerances of man for unhealthful substances or conditions. This work is generally assessed from a medical viewpoint, including physiological or psychological studies. In addition to toxicology and other medical specialities, study areas include biomedical instrumentation and health research techniques utilizing animals — but always with intended application to human health measures.

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

EPA-600/1-78-012
February 1978

DEVELOPMENT OF QUALITY ASSURANCE PLANS FOR RESEARCH TASKS HEALTH EFFECTS RESEARCH LABORATORY RESEARCH TRIANGLE PARK, NORTH CAROLINA

Guidelines for Taskmasters

Quality Assurance Document # 2

Valid Through Fiscal Year 1978

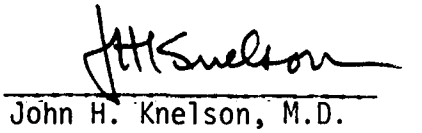
U.S. ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF RESEARCH AND DEVELOPMENT
HEALTH EFFECTS RESEARCH LABORATORY
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27711

DISCLAIMER

This report has been reviewed by the Health Effects Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

FOREWORD

The U.S. Environmental Protection Agency's Health Effects Research Laboratory located at Research Triangle Park, North Carolina conducts an extensive research program to evaluate the human health implications of environmental factors related to industrialized society. The purpose of this research is to provide information necessary to formulate environmental regulatory policies to protect or improve public health and welfare while at the same time enhancing the nation's productivity. To this end, the Laboratory conducts a comprehensive program in toxicology, epidemiology, and research on human subjects under controlled laboratory conditions. The quality of the data resulting from this research is an overriding factor in determining the usefulness of this information in EPA's regulatory activities. In recognition of the importance of data quality assurance, our Laboratory has initiated a comprehensive program to coordinate all the current activities in this area. Accordingly, the quality assurance guidelines presented in this document provide assistance to scientists in our Laboratory in the preparation of research protocols. Other guideline manuals in preparation will provide elements of quality assurance required for specific categories of measurements made in the Laboratory. I am confident that full implementation of our data quality assurance policy with the help of the guidelines manual and the increased awareness of the importance of good data acquisition and management procedures will enhance the scientific merit of our research program.



John H. Knelson, M.D.
Director
Health Effects Research Laboratory

ACKNOWLEDGEMENTS

This guideline document was completed through the auspicies of the Quality Assurance Committee, HERL-RTP, whose members reviewed and criticized the manuscript and made many helpful suggestions.

HERL Quality Assurance Committee
1977-78

Ferris Benson - Criteria and Special Studies Office
Robert Burton - Population Studies Division
Dorothy Calafiore - Population Studies Division
Walter Crider - Clinical Studies Division
George Goldstein - Clinical Studies Division
Ralph Linder - Environmental Toxicology Division
Margarita Morrison - Program Operations Office
Gerald Nehls - Statistics and Data Management Office
Edward Oswald - Environmental Toxicology Division
George Rehnberg - Experimental Biology Division
Ralph Smialowicz - Experimental Biology Division
David Svendsgaard - Statistics and Data Management Office

The preparation of the manuscript was supported through contract EPA-68-02-2725 with the Research Triangle Institute, Research Triangle Park, North Carolina.

TABLE OF CONTENTS

LIST OF FIGURES

1. SUMMARY

This document presents guidelines for the development of quality assurance (QA) plans for tasks at the Health Effects Research Laboratory, Research Triangle Park, North Carolina (HERL/RTP). These guidelines are designed to support taskmasters at HERL/RTP as they oversee the development and implementation of plans for specific intramural and extramural tasks. This document is second in a series of guidelines documents delineating the QA program at HERL/RTP: the first document [1] discussed quality assurance from the organizational and managerial perspective.

The responsibility for the development and implementation of an appropriate QA plan for a research task rests with the respective taskmaster. Hence, these guidelines include discussion both of quality control and quality assurance principles relevant to all HERL/RTP taskmasters. The discussion of various aspects of Quality Assurance is organized to parallel the sequence of events in the life of a research task. In this way it is intended to be comprehensive and to complement the scientific training and experience of the HERL/RTP investigator. Additional guidelines documents applicable to specific research areas are in preparation.

Following the introductory discussion, research quality control in the following areas is addressed:

§3.1	General approach to quality control in research.
§3.2-3.6	Planning (experimental design, personnel, facilities and equipment, recordkeeping, supplies).
§3.7, 3.8	Experimental (sample collection, sample analysis).
§3.9-3.15	Data quality activities (internal audits, preventive maintenance, calibration, documentation control, configuration control, data validation, feedback and corrective action).
§3.16, 3.17	Results (data processing and analysis, report design).

Suggestions for quality assurance activities (i.e., independent of task operating personnel) are then discussed.

2. INTRODUCTION

2.1 Purpose and Summary

This document is the second in a series designed to serve as the statement of the Quality Assurance (QA) Program at the Health Effects Research Laboratory, Research Triangle Park (HERL/RTP), U.S. Environmental Protection Agency (EPA). The purpose of the first document [1] is to outline the QA program in its entirety. It specifically addresses QA policies and delineates QA responsibilities throughout the functional and task management of HERL/RTP. This document logically follows Document No. 1 in that it outlines the aspects of quality assurance with which the taskmaster should be knowledgeable and which he may choose to implement in his ongoing or future projects.

As is stated in Document No. 1, the design and application of QA measures at the project level are the responsibility of the taskmaster or project officer with the support and approval of his functional management and, optionally, the QA organization. Inherent in this responsibility is the need for the taskmaster to:

a. become acquainted with the technical and administrative requirements of the HERL/RTP quality assurance policy;

b. include QA procedures and criteria with all documentation. This includes protocols, RFP's, proposals, work plans, and progress reports. The taskmaster may, at his option, consult with the QA representative from his Division or Office for QA recommendations;

c. supply requested data to the Quality Assurance Coordinator (QAC) for evaluation by the QAC and/or the QA Committee;

d. consult with the Division or Office QA representative regarding QA procedures (optional); and

e. evaluate the effects of quality assurance activities on project scope, quality, cost, and schedule.

The basis for the discussion that follows is the assumption that the research-trained taskmaster will automatically perform various QA functions in his area of major expertise. These guidelines are intended to describe principles that complement and document these functions for every aspect

of a research task that may be performed under the auspices of HERL/RTP; they do not provide solutions in detail. The purposes of these guidelines then are to:

a. support the taskmaster in his planning for comprehensive quality assurance appropriate to all areas of his research;

b. collect in one document general data quality checks for quick review;

c. document data quality checks currently in use at HERL/RTP for use by HERL/RTP professional and technical staff and other interested parties; and

d. provide a logical framework within which additional research relating to HERL/RTP data quality may be programmed.

In Section 3, as a major activity of the total QA program, aspects of internal data quality control (DQC) are discussed. These elements include:

- Experimental Design,
- Personnel,
- Facilities and Equipment,
- Recordkeeping,
- Supplies,
- Sample Collection,
- Sample Analysis,
- Internal Audits,
- Preventive Maintenance,
- Calibration,
- Documentation Control,
- Configuration Control,
- Data Validation,
- Feedback and Corrective Action,
- Data Processing and Analysis, and
- Report Design.

Section 4 outlines elements of data quality assurance (DQA) such as system audits and performance audits that may be incorporated into the task quality assurance program for verifying and documenting the level of data quality in a task.

In summary, this document is intended as a reference source for HERL/RTP taskmasters as they design and maintain project-specific quality control programs sufficient to insure that project quality objectives are realized in the most cost-effective manner. The document deals primarily with quality assurance aspects that will be applicable in varying degrees to the research performed within HERL/RTP.

The remainder of this section defines several terms used in connection with a quality assurance program.

2.2 Definitions

In order to effectively and efficiently integrate quality assurance practices into project work, the taskmaster must understand the fundamental concepts of quality assurance and quality control. The following definitions, drawn from references 1, 2, and 3, provide a review of these fundamental concepts.

2.2.1 Quality

The term quality means the totality of features and characteristics of a product or service that bears on its ability to satisfy a previously specified need. For measurement systems and research, the products are the reported data analysis results. The data characteristics of major importance are accuracy, precision, representativeness, and completeness.

2.2.2 Quality Assurance (QA)

The term quality assurance (QA) is used to describe a comprehensive system of plans, specifications, and policies that are designed to insure the collection, processing, and reporting of quality data in a cost-effective manner. Thus, the design of QA plans for particular tasks is the subject of this document. QA provides for total system data quality, resulting from data quality control and data quality assurance, from experimental design (e.g., measurement method) through final report production (e.g., statement of confidence limits and limits of applicability of results). In addition, it addresses possible needs for methods

development and independent value judgments of the relevance of a proposed project to prior specified HERL/RTP or EPA needs.

2.2.3 Data Quality Control

Data quality control (DQC) is a system of activities designed to achieve and maintain a previously specified level of quality in data collection, processing, and reporting. DQC is performed by the organization actually carrying out the task or project; i.e., it is executed by task personnel. DQC activities include control or correction for all variables suspected of affecting data quality. These variables are outlined in Section 2.1 and treated in detail in Section 3.

An important part of a complete data quality control program is the utilization of internal audits to insure that the desired level of data quality is being maintained. These audits consist of analyses similar to those discussed below for Data Quality Assurance (DQA), i.e., qualitative and quantitative checks on the system and/or procedures. The essential difference between internal audits and DQA is that internal auditing is performed within the task management (under the direction of the taskmaster) while DQA is performed independently of task management.

2.2.4 Data Quality Assurance

Data quality assurance (DQA) is a system of activities designed to provide management with an independent assurance that total system data quality control (DQC) is being performed effectively. DQA activities are both quantitative and qualitative and are performed by other than task personnel. To perform qualitative systems reviews, DQA personnel conduct onsite inspections of facilities, equipment, documentation, etc. A variety of techniques are available for quantitative audits. Blind samples, collaborative testing, and round-robin analyses are some of the usual techniques used to quantitatively verify DQC effectiveness. These DQA techniques, along with others to be developed, can be applied to the health-related research performed in the HERL/RTP.

2.2.5 Task

A task is an intramural or extramural project, or interagency agreement, the purpose of which is to produce technical research data for the HERL/RTP research program.

2.2.6 Protocol

As used in this document, the term protocol should be understood to include all planning documents used at HERL/RTP. Specifically included are task protocols, procedure statements, work plans, and scopes-of-work, irrespective of the nature of the task or organization actually performing the task.

3.0 ELEMENTS OF DATA QUALITY CONTROL FOR RESEARCH PROJECTS

3.1 General

In planning a QA program for a particular task, the taskmaster should attempt to account for all variables that are known or suspected to affect the data to be produced. Planning for such monitoring is not a simple task. Performing it with the necessary care is more difficult. However, it is becoming increasingly necessary to provide for such a QA program considering the number of reports that appear indicating that reagent quality and identity are not what the manufacturer claims them to be, instruments do not perform the function for which they are intended, electronic circuits are discovered to generate false signals due to mismatches, etc. Due to these general, and some specific, data quality problems, the EPA is currently developing comprehensive QA guidelines [4]; Federal standards for nonclinical laboratories [5] have also been proposed; and "Quality Assurance Practices in Health Laboratories" will be available late this year from the American Public Health Association [6]. Current research increasingly depends on sophisticated automated data collection systems, whether an isolated laboratory is involved or an entire monitoring system. The cost of this research is increasing at a corresponding rate. Efficient operation under such conditions requires carefully designed quality assurance plans for research tasks.

As performed at HERL/RTP, health-related research is frequently state-of-the-art, in concept as well as in technique. As such, it is not obviously susceptible to the normally available QA techniques. However, careful analysis indicates that virtually every research task within the HERL/RTP consists of two principle areas, whether the task is laboratory research or a monitoring program:

a. Data collection and processing - routine measurements performed by skilled technical personnel using well-characterized techniques (e.g., pH measurements, cell growth parameters, etc.).

b. Data analysis and reporting - nonroutine data analysis performed by the HERL investigator using physical models, statistical techniques, and other tools in a nonroutine, creative manner. This frequently involves collaborating with HERL support staff members and peers.

Each of these aspects of research is susceptible to the use of QA techniques by the taskmaster. Data collection techniques generally have adequately characterized quality control procedures associated with them that are quantitative in nature. The taskmaster uses professional judgment in determining the frequency, number, and specific reference materials to be used. Quality assurance of data analysis is less straightforward. Peer interaction, from the protocol stage to the report stage of a task, plays an important role. It is, therefore, important that effective mechanisms for peer review be officially recognized.

The production of research data is strongly affected by the "weak link" phenomenon. Thus, if experiment design, equipment maintenance, data analysis, etc., are excellent and sample analysis quality is poor, the overall task data quality is lowered. Similarly, no amount of competent technical skills, data analysis, etc., can compensate for poor experiment design.

In addition, there are aspects of a research task that affect data quality, but which are not easily quantitated or categorized. For example, technician fatigue and morale should be considered. Similarly, the tension between the need for quick response to unexpected developments and the need for strict accountability to funding agencies relates to planning for quality data. With these considerations in mind, these guidelines are designed to be supportive of taskmasters as they oversee the progress of their tasks.

As a research project progresses, it frequently becomes apparent that additional "nonessential" data (e.g., instrument settings, exact identity of the components of a buffer solution, etc.), which are not usually recorded, are useful for data interpretation. As a general rule, then, it is cost-effective to record well-organized, complete data from which an experiment can be properly reconstructed. Lab notebook (or station logbook) records of numerical as well as anecdotal data will frequently prove useful when experiment reconstruction becomes necessary.

The remainder of this section addresses the various elements of a research task. It should be realized that different research projects will involve different applications of these QA elements. The taskmaster, however, should be cautious in deleting considerations of any

element and should be certain that it will in no way affect the quality of the data that are produced by the task. If there is any uncertainty regarding the design of a task QA plan, the taskmaster may request the aid of his QA representative or the QA coordinator. It should be remembered that when properly used, quality assurance planning can be a very effective insurance policy against data of unacceptably poor quality.

3.2 Experimental Design

Adequate planning prior to the startup of a task is by far the most cost-effective program for task quality assurance. This planning should include a discussion of the experimental design including manpower; facilities, supplies, and equipment logistics; and detailed plans for data collection and analysis as well as statistical experimental design per se. The protocol that results from this type of planning serves at least three purposes: (a) it provides a planning focal point for obtaining answers to the not-so-glamorous questions of "who," where," "when," and "how," which are rarely considered during the initial brainstorming; (b) it documents for all interested parties that responsible planning has occurred; and (c) it provides criteria for making logical decisions when such decision points are reached in the later stages of the task life.

The contents of the task protocol were briefly mentioned in Document No. 1 [1]; the minimum contents of a research protocol have been proposed for nonclinical laboratory studies by DHEW/FDA [5]. These are shown, respectively, in Figures 1 and 2.

During initial phases of research planning, and during protocol development, the taskmaster should solicit advice from the various HERL support functions that will be involved. Specifically, the statistical design of the experiment, the data collection and analysis, and the animal care requirements should be planned in detail by the time the research protocol is drafted. (The ongoing collaboration of each of these functions should also be programmed, in order to successfully cope with the various unexpected difficulties that generally occur in research.) Each of these three areas is discussed below.

- Providing a clear statement of the hypothesis to be tested.
- Considering generally how results are to be demonstrated, particularly graphical presentations of data.
- Proposing analyses of:
  a. covariables (or covariates) considered,
  b. other possibly important covariables,
  c. controls to be used.
- Considering what data are needed to undertake this analysis and how they are to be processed.
- Considering what quality assurance plans and procedures will be implemented. If comprehensively treated in other sections of the protocol, they should be referenced.
- Determining whether new or old collection forms are needed.
- Determining the number and kinds of study subjects needed, and the statistical basis for the choice.
- Deciding upon a schedule for testing and other data collection (this includes scheduling the obtaining of exposure data and collection in particular areas when appropriate). Also, deciding upon the statistical basis for the sampling schedule and the number of sampling sites.
- Determining how to initiate the study and when subject selection and data collection are to begin.
- Determining the time span necessary for data collection and when data will be available for analysis.
- Determining the duration needed for data analysis: are analysis programs on hand?
- Deciding when draft reports and final report will be completed.
- Estimating anticipated problem areas in carrying out the study.

Figure 1. Example of major topics addressed in a task protocol [1].

1. A descriptive title and statement of the purpose of the study.
2. Identification of the test and control substance by name and/or code number.
3. The stability of the test and control substances in terms of the methods to be employed.
4. The name of the study director, the names of the other scientists or professional persons involved, and the names of laboratory assistants and animal care personnel.
5. The name of the sponsor and the name and address of the testing facility at which the study is being conducted.
6. The proposed starting date and date of completion of the study.
7. The proposed date for submission of the final study report to management or to the sponsor.
8. The number, body weight range, sex, source of supply, species, strain and substrain, age of the test system, and justification for selection.
9. The procedure for the unique identification, if needed, of the test system to be used in the study.
10. A description of the method of randomization, if any, of the test system with justification for the selected method.
11. A description and/or identification of the diet used in the study as well as solvents, emulsifiers, and/or other material(s) used to solubilize or suspend the test or control substance before mixing with the carrier.
12. The route of administration and the reason for its choice.
13. Each dosage level, expressed in milligrams per kilogram of body weight or other appropriate units, of the test or control substance to be administered and the method and frequency of administration.
14. Method by which the degree of absorption of the test and control substance will be determined if necessary to achieve the objectives of the study.
15. The type and frequency of tests, analyses, and measurements to be made.
16. The records to be maintained.
17. Nonroutine procedures required to assure personnel health and safety.
18. The date of approval of the protocol by the sponsor and the signature of the study director.

Figure 2. Proposed research protocol contents for nonclinical laboratories by DHEW/FDA [5].

3.2.1 Statistical Experimental Design

In any HERL task that involves the gathering and analysis of data, it is important to seek the aid of a competent statistician. The statistician should be consulted not only after the data have been gathered but during the planning phase of the study as well. No analysis plan, however ingenious, can compensate for a bad experimental design. Later, as the statistician is regularly involved in the daily execution of the plans, timely advice for cost-effective midcourse changes will be a valuable asset to the maintenance of task data quality.

In general, the statistician's support throughout the task will be most helpful as the taskmaster formulates, examines, and carries out the following phases of the task: (a) the objectives and hypotheses to be tested, (b) the design of a testing program to meet the objectives (i.e., the experimental design), (c) the data processing plans, and (d) the data analysis plan. These four phases and the statistician's role in them are discussed below.

a. Objectives and Hypotheses to be Tested

Determining the objectives and the hypotheses to be tested is obviously the first step that should be taken in designing any task. Precise formulation of the questions to be answered enables one to state the hypotheses to be tested in precise terms and thus to plan a task more effectively. The aim should be to make the statement lucid and specific, avoiding vagueness and excessive ambition. Often it is advisable to classify objectives as major and minor. This classification is particularly helpful in assigning priorities to objectives when the task involves cooperation among people of different interests.

b. The Design of a Testing Program to Meet the Objectives (the Experimental Design)

The testing program design should produce a clear definition of all the variables to be considered, the size of the testing program, the experimental units (e.g., animal models, cell cultures, humans, etc.) and exactly what data are to be collected. In designing the testing program,

the following questions should be answered:

1. Are all the relevant factors (e.g., temperature, subject age, etc.) being considered?

2. Are the effects of the relevant variables adequately distinguishable from the effects of other variables (e.g., would a factorial design be more appropriate)?

   In other words, one can consider an experiment as intended to determine the effects of one or more variables (factors) on measures of experimental outcome. From substantive considerations, the taskmaster determines the factors, and the levels of each, which should be varied in his experimental program. In experiments involving two or more factors, the "effect" of a specified level of a particular factor may depend on the levels of other factors in the experiment (the factors may "interact"). The "main effect" of a factor is determined by comparisons among the effects of various levels of the factor. In designing multifactor experiments, the taskmaster should carefully consider what effects--main effect and interaction effects--are of interest to him. The experimental plan should be such that it will result in all the data necessary to estimate the main effects and interactions of interest at the end of the experiment. Given the constraint of limited resources, the question must be answered as to which subsets of factors and levels will estimate the main effects and interactions of interest.

3. Is the plan as free from bias as possible (e.g., is randomization used correctly; are reasonable quality control procedures being employed)?

4. Does the plan use a historical measure of precision (experimental error) and if so is this precision sufficient to meet the objectives of the tests?

5. Is the scope of the testing plan consistent with the objectives given in Section 3.2.1.a (e.g., is the plan too limited)?

6. Is the testing plan cost-effective (e.g., would a more limited test plan provide equivalent information at a lower cost)?

7. Are the data collection plans appropriate to the test objectives (e.g., sample frequencies of every 5 minutes, every day, etc.; should additional, or fewer, variables be monitored)?

8. Are available resources adequate for collecting the quality and quantity of the data required?

Answering questions such as 1. through 8. allows the formulation of a statistically suitable testing program and alternative testing designs. It is important to note here that the analysis of data (Section 3.2.1.d.

below) can be made much easier if this phase (b.) is completed properly.

c. Data Processing

The data processing phase of a task is concerned with how the data are handled once they have been collected, and involves examining the following kinds of questions about the data gathered according to the testing program formulated in phase b. (above):

1. How is the data validated, i.e., what procedures are used to determine what data to include in the analysis? This question may involve developing a specific statistical procedure for rejecting outliers. Note here that treatment of possible outliers on the basis of a statistical evaluation of the data is a task that should usually be performed by the person or persons responsible for the analysis and interpretation of the data. Also, it should be made clear that experiments should not be repeated just because the results "don't look so good."

2. When are the data to be processed so that they can be analyzed, i.e., during the testing program or only at the end of data collection? This question is especially important if the test program extends over a long period of time, since preliminary analysis of the data may indicate that the testing program should be altered for the remaining tests.

3. If data from different instruments are to be compared, what is the comparability of outputs (e.g., one instrument may give continuous readings while another may only give discrete readings, or different detection principles may be utilized in different instruments)?

4. What (manual) data handling is required in order to convert "as recorded" raw data into the form in which it will be analyzed (e.g., copying from a lab notebook onto coding forms, keypunching cards from these forms, and reading the cards into a computerized data base)? Also what is a realistic estimate of the net error rate for this process?

d. Data Analysis

Initially, this phase involves reviewing any data analysis that has been proposed or has already been performed on the project, and also giving an outline of the analysis to be performed if no outline is available.

An outline of the data analysis should be prepared before the test design is completed or testing begins. If this outline is not prepared, it is quite likely that measurements that should be recorded for proper analysis will be overlooked or will not be recorded in the correct manner. For example, an outline of the analysis may reveal that it is essential

to record the level of an uncontrollable variable so that adjustments for the variable may be made when the data are analyzed. Conversely, unnecessary data may be identified and eliminated during this phase, thus conserving resources. In addition, if the project involves a large number of different types of measurements, it is important that an overall analysis plan be devised that insures that the objectives given in Section 3.2.1.a. are met in the most efficient manner. For example, a multivariate analysis may be preferable to several univariate analyses.

Once the data for the project have been gathered, the data analysis should be carried out with the close collaboration of a statistician. This is particularly important when the testing program has changed somewhat since the beginning of the project (which is frequently the case) and/or there is a large amount of missing data. In addition, the statistician and taskmaster should work closely together in presenting the results of the data analysis. In this regard the taskmaster should insure that the presentation is understandable to nonstatisticians. The statistician should make sure that the results are presented such that the reader is aware of the functional relationship linking the data and the tables or graphs. The statistician should also insure that statistical results are interpreted correctly based on the nature of the design and the statistical tests. Since any scientific study falls short of realism, useful conclusions usually require generalizations that tend to lie outside the realm of strict statistical justification. Thus, the reader of the technical report should be informed of the amount of statistical and physical justification supporting each conclusion.

3.2.2 Data Collection and Analysis

Once the production of raw data has begun, the manner in which they are collected and analyzed becomes important. Data validation--the technique of flagging data values that are suspected of having an excessive component of error--must be addressed. Manually collected data are frequently monitored by the person recording the data. However, computerized data acquisition systems do not have the potential for this treatment. They are known to pick up false voltage transients, and failure of one component of a system may seriously bias the data of major interest

in an experiment. In a system of reasonable complexity, a variety of warnings may be identified by careful analysis of the relationships and patterns of values of the incoming data.

The use of control charts, or the concept, should be considered for use in specific data validation procedures. Used properly, individual out-of-range points and data trends will be readily apparent, and informed response by the taskmaster will be possible. While control charts are not usually found in health-related laboratories, they may be adapted to virtually any operation that involves repetitive measurements of a parameter (e.g., subject body weight or total cell count) whose value should lie within a known range. Also, instrument calibration data may be plotted on a control chart as a means of detecting trends that document instrument drift characteristics and signaling impending failure.

Computerized data acquisition systems are being used increasingly. They frequently permit a statistically acceptable, cost-effective extension of the control chart concept for data validation. There are several advantages to using such a system. It accepts truly raw data to produce intermediate and final results in tabular or graphical form, thus minimizing human error. Similarly, the capability of rapidly and automatically comparing experimental data against recent values of similar data can serve as a "real-time" check on data validity.

Data analysis involves the matching of the experimental system with a model system and evaluating the differences. Since real-world data are never sampled exactly, one source of discrepancy between the data and the model is due to measurement error. Only rarely will the model exactly correspond to the test system, thus adding another component of data-model disagreements. The experiment should be designed such that data analysis will highlight the actual model-test differences, rather than masking the discrepancy as "error." Appropriate statistical design of the experiment is essential at this point. Care must also be taken that apparently irrelevant physical aspects of the test system do not produce data that lead to erroneous interpretations (e.g., diurnal fluctuations in serum enzyme levels are frequently larger than the response to the experimental stimuli on many biological systems). In order to maximize the quality

of data from a testing program, the taskmaster should routinely consult with researchers who have specialized in such related areas.

3.2.3 Biological Systems

The majority of the research and support associated with the HERL/RTP directly involves biological systems. While this is common knowledge among the HERL/RTP staff, it touches upon an important, and sometimes troublesome, difference between the experimental situation at the HERL/RTP and the situation at laboratories that do not perform research directly on biological systems. The implication of this difference is that, while the experimental variables analyzed and modeled in other laboratories present a complex challenge, the experimental variables associated with biological systems studied at the HERL/RTP are orders of magnitude more complex. The "simple" systems under study in most physical science research laboratories involve the effects of a few to a few dozen experimental variables, most of which are monitored, if not controlled. Biological systems, even the most simple, involve the interactions between several dozen recognizable molecular species. And if research trends continue, several hundred distinctly recognizable molecular interactions will soon be characterized in the most simple monocellular systems.

The challenge of such a large array of experimental variables can presently best be met by permitting variation of only a selected few of these variables. For this reason the taskmaster should exercise his best professional abilities to recognize and fix all but the experimental variables. This is the purpose of care in selecting, maintaining, dosing, and analyzing biological subjects, whether they be cell cultures, animals, or humans.

Human subjects come from diverse and largely unknown backgrounds. This variability among human subjects can be minimized (but not eliminated) by careful pretest screening and questioning. The results thus obtained are directly applicable to human health problems. On the other hand, cell culture lines that have been quite thoroughly characterized for several generations are available for research. And the results of

cell culture studies seldom, if ever, apply without interpretation to aspects of human health. Intermediate between these two extremes are animal subjects, some lines of which have been quite well characterized for several generations and which correlate closely with certain aspects of the human system. It is thus not surprising that a large proportion of health effects research is performed using animal subjects. Proper maintenance requirements of animals, however, are relatively more costly (in dollars and labor) than for cell cultures. Since careful characterization of animal subjects is no less important than for cell culture models, the balance of this subsection is devoted to a very brief discussion of animal care.

Comprehensive guidelines for animal care are presently being developed for HERL/RTP, and general guidelines are presently available [5,7]. A brief discussion of the basic aspects of animal care is included here, due to its importance to overall task data quality. The basic concept, common to all scientific research, is to attempt to control all but the experimental variables. Early, intensive, and consistent consultation with qualified professionals from the Lab Animal Support will maximize the quality of data that are generated using laboratory animals.

Animal selection should be based on awareness of the species' genetically determined immunities, etc., as well as the specific dose-response relationship to be investigated. The research protocol should clearly state the basis for selection of a particular species, the anticipated interferences with the experiment design, and any preliminary testing required for adequate characterization of the system unknowns (e.g., interfering antibodies).

Acceptance testing, or prescreening and surveillance, should be sufficiently comprehensive to insure that only suitable animals are included as experimental subjects and controls. While the added expense of such testing may limit the quantity of animals used, the increase in the quality will generally more than compensate for this loss.

Personnel assigned to animal care and dosing should have sufficient technical competency to provide reliable routine care to experimental animals. In addition, their training and responsibilities should permit their active participation in the research (e.g., to note unusual be-

havior or health of any of the test animals; to note abnormalities in the dosing formulation). As the individuals most intimately associated with cause-effect relationship, which produces the raw data, these individuals should be made aware of their role in the research and treated accordingly.

The dosing and vehicle matrix should be chosen carefully and should be well characterized with respect to the specific experimental animals. If the particular choice has not been well characterized, it should be changed, or detailed studies performed to characterize it prior to experimental work. Choice of the control group and the specific regimen should be made on the basis of acceptable data quality, excepting only those aspects of the control that are reliably documented (i.e., complete equivalency of the experimental and control group regimen should be routine, excepting only the test substance).

In short, the animal subjects should be treated as any nonbiological supply, i.e., they should be as well characterized as possible.

3.3 Personnel

As noted above, task operational personnel are intimately involved in one of the most crucial aspects of the particular research task: the generation and recording of the experimental cause-effect relationships that result in task raw data. The upper limit of the quality of the results is set during this phase of a research task. Statistical treatment may be used to estimate precision and accuracy; creative thinking may rationalize discrepancies; but the upper limit of data quality for the task cannot be improved beyond that produced by task personnel at the time of the (various phases of the) experiment. Two aspects of the personnel relationship to acceptable data quality are (a) technical qualifications and (b) intangible aspects.

The usual approach to technical qualifications is that personnel have the education, training, and experience to perform the assigned function. Similarly, training in good laboratory practice (generally and job-oriented) is recommended [5]. Such stipulations are certainly reasonable, and should be the documented practice of the taskmaster. Attempts should be made to insure that all task personnel keep abreast of contem-

porary developments in their fields of expertise. Adequate theoretical briefing should be provided to bench technicians so that they will be capable of recognizing and recording unusual and unanticipated events.

Another aspect relating personnel and data quality is far less tangible, but none the less important. It refers to the general mental state of task personnel. Appropriate work loads prevent excessive mental and physical fatigue. Useless effort is avoided with optimum laboratory and equipment configurations. Good interpersonal relationships support full productivity. Proper management techniques (neither too restrictive nor too permissive) result in maximum productivity and data quality. In addition, the complex issue of motivation [8, Section 18] is an important factor in total personnel performance and data quality. The taskmaster is in the position to recognize and address such aspects relating to task personnel that have a healthy atmosphere for research and a direct effect on overall task data quality.

3.4 Facilities and Equipment

The facilities and equipment selected for an investigation should be known to be capable of producing acceptable quality data at minimum risk to task personnel (and subjects).

Within HERL/RTP, the primary purpose of research conducted is to better model the responses of the human biological system. Frequently, nonhuman biological systems used for experimental purposes are selected with the intention of extrapolating results to characterize the human system. Due to the intentional similarity of the two systems, a significant risk of cross-contamination and infection is a constant threat to experimental results, as well as personnel health. While it may be impractical or undesirable for the HERL/RTP investigator to strictly follow the various published animal facility guidelines, deviations should be made only at the advice and with the approval of the professional staff of the Lab Animal Support.

Similarly, many nonbiological systems are used for health-related research, yet with potential risk to operating personnel. Insult to operating personnel by noxious fumes, electrical shock, etc., should be

anticipated and eliminated as conducive to the long-range, cost-effective maintenance of data quality.

The experimental facility should be examined carefully prior to the commencement of experimentation. If it is a new facility, it will be most cost-effective to properly design the facility for its intended purposes. Modification of an existing facility is the usual case. In either case, resource (i.e., dollars, manpower, time, etc.) limitations always exist that directly and indirectly affect data quality. The various options, and their effects on data quality, should be frankly evaluated and discussed with the management. When the task involves a new experimental design in a facility already used by the investigators, de novo evaluation should be the norm. For a variety of reasons, this is difficult and may not be carried out. However, if a complete evaluation of the requirements of the experimental design as well as of potential error sources is conducted at the outset of a research project, future invalidation of much or all of the experimental work may be prevented. (For example, reference 9 reports that under certain conditions, light from fluorescent fixtures has caused mutations in the hamster cell chromosomes. If substantiated, these findings may bring into question an entire body of research. Rigorous attention to such seemingly trivial detail can ameliorate this type of problem.)

In addition to the technical suitability of the facility for execution of the task, it is in the taskmaster's interest to evaluate and configure the facility with due care for the physical and mental comfort of the technical staff who will be using the facility. The discussion in Section 3.3 (Personnel) extends here to the human engineering of hoods (for poisonous and noxious gases), sinks, walkways, counters, etc. While there will be necessary trade-offs in facility configuration, its influence on traffic patterns, the environmental aspects (temperature, airflow, lighting, noise levels, etc.), and other fatigue- and confusion-producing aspects should be evaluated and related to the effect on data quality.

Depending on the type of research involved, facility security should be specifically considered. This will range, for a wide variety of reasons, from areas available for common use by even nontask personnel to

stringently restricted areas. Relating to data quality, the facility configuration should be carefully controlled (see Section 3.1.3, also ref. 3, Section 1.4.19). As is frequently the case, even routine instrument maintenance activities can have a profound effect on data quality--for example, a new design of a replacement emission source for a spectrophotometer may affect data adversely (or positively), which only becomes apparent during later analysis. If possible, authority to approve facility configuration changes should be limited to one professional staff member who is qualified to document and evaluate such changes.

As with the facility used for the task, the equipment should be evaluated for its applicability to the task research. The relationship of the measurement method and the variables to be monitored should be well characterized during the initial task activities if not before they have begun. Similarly, the subtleties of design and performance of different manufacturers' equipment should be thoroughly evaluated, preferably with the aid of a professional who has both a theoretical and practical understanding of the specific instrument operation. In this regard, it is not uncommon to learn that unadvertised features of one instrument will permit acquisition of significantly higher quality and/or quantity data. As discussed below in relation to supplies, acceptance testing for new equipment should be performed on an item-by-item basis and documented for comparison with future testing. This testing program should be designed in such a way that operation of the instrument at its extreme limits (i.e., "worst-case") as well as routine settings will be thoroughly characterized before it is made available for routine use.

In relation to equipment, the desirability of full- or part-time operator and/or maintenance support should be considered. Frequently, sophisticated instrumentation performs poorly or not at all when several occasional users have access to it. Similarly, minor but frequent maintenance often keeps an instrument operating at peak performance. In such cases, the cost of a dedicated operator is frequently justified.

3.5 Recordkeeping

Provision for a complete, permanent, easily accessible record of the raw experimental data should be made prior to, during, and following com-

pletion of task experimental work. This should include a written record (in ink, in a bound, page-numbered notebook) of equipment serial numbers, reagents and supplies used, animal identification and test data, as well as a record of equipment modifications and other seemingly inconsequential information that will permit more accurate analysis at later dates. Reference 5, Section J lists proposed rules for nonclinical laboratory reports and records, their generation, storage and retrieval, and retention on a long-term basis. When data are logged by computers, it is important that adequate provision be made for redundant and physically separate long-term storage of such records.

Recordkeeping of this quality serves at least two useful functions: (a) it makes possible the reanalysis of a set of data at a future time when the model has changed significantly--thus increasing the cost-effectiveness of the data, and (b) it may be used in support of the experimental conclusions if various aspects of the study are called into question. This latter point goes to the heart of scientific research: objectively, it is often possible to interpret data in more than one way--and the raw data should be available for evaluation by qualified professionals; subjectively, when recordkeeping habits are sloppy, suspicion is quickly aroused that (all) other aspects of the research are similarly of poor quality.

In addition to the issues discussed above, the taskmaster's investment in the design of suitable data logging forms for repetitively measured parameters will be repaid. This will occur in insuring data completeness, higher productivity of technical personnel, and later, ease of reading the raw data. Computerized data acquisition systems have many advantages. However, they must be closely monitored for false or erroneous signals that may not be easily detectable.

3.6 Supplies

As noted in Section 3.1.3 (Biological Systems), a basic premise of scientific research is that all but specified variables are controlled or held constant. However, reports regularly appear in the technical literature of impure and/or mislabeled supplies; e.g., supposedly "germ-free"

animal subjects are found to have been infected after the end of experiments in which they were used thus invalidating the entire experiment. An acceptance testing program for all incoming expendables/supplies--be they chemicals, biologicals, etc.--should be applied prior to, and (judiciously) during use. Resources are always limited, hence the design of a suitable testing program is important. This is facilitated by learning as much of the processing history of the supplies as possible, by anticipating possible experimental interferences using the existing model, and by conferring with other users of the same consumable.

The results of a successful acceptance test should (a) confirm that the substance fully corresponds to the label specifications, and (b) confirm that known or suspected interferents are absent. Especially when the acceptance testing is lengthy and/or costly, adequate amounts of a common lot should be purchased to permit completion of the tests. Sufficient excess to permit unanticipated testing, plus a specified amount for storage, should also be included.

Following successful completion of the acceptance test, an expiration date should be permanently marked on each container and it should be stored on a first-in-first-out basis. The shelf-life of many substances is known; in some cases, estimation of shelf-life may be necessary. In most cases, sample tests exist that can, to a first-approximation, rapidly document the strength and purity of a substance (or animal) immediately prior to use. Reliable estimates of strength as a function of time should be used to determine a conservative useable lifetime of solutions, mixtures, emulsions, etc.

In this latter instance, a well-designed central stockroom tracking system will facilitate rapid reference to the identity of other users of a substance. This will be useful for informal sharing of information of interest as well as for rapidly identifying and locating the users when a specific problem (e.g., purity or identity) has been detected with the particular substance.

3.7 Sample Collection

In sampling, one generates a new system, because as soon as a portion of material is removed from the whole its history becomes different

from the whole.* Primary consideration must be given to keeping the sample collection system as nearly representative of its condition when sampled as possible, regarding all the parameters under investigation. The processes involved in obtaining, holding, preserving, transporting, and resampling can potentially introduce significant direct and indirect changes in the material destined for analysis. Quality control measures must be specifically designed to quantitate and characterize any sample degradation or interaction with its particular container and environment.

Samples must be positively identifiable by those taking the sample and by others who are involved in subsequent analytical or handling steps. This does not preclude the use of blind samples, spiked samples, or other audit methods to assure the quality of the test system in part or as a whole.

The personnel-related requirements for the technical and support aspects of the sample collection program vary in type and number. All operating personnel need to know exactly what is required of them, how it is to be done, and when. Written instructions answering these questions for every phase of their involvement should be developed and provided as appropriate. Periodic "practice work" may be necessary in order to maintain the desired level of data quality. Each person should have a clear understanding of who will answer his questions on test protocol.

In complex sampling activities it may be helpful to discuss personnel roles relative to the total task. By this, operating personnel can obtain a more complete perspective of their respective tasks, their interaction with others, and an overview of the experiment design. The underlying theme of these discussions is the rationale for the sampling protocol and the quality control measures for sample validation, and the need to call for supervisory assistance whenever the test system is suspect. Periodic meetings during task implementation may help in information exchange, procedure standardization, and improved quality control of the project.

3.8 Sample Analysis

Sample analysis--whether it be a spectrophotometer reading or viable colony count--involves a repeated sequence of similar operations by tech-

*A corollary to this is that the existing system is also altered by sampling activities.

nical personnel and/or automated instrumentation. For this reason, sample analysis is susceptible to the use of quality control techniques. Adequate, correct, and available operating procedures used by suitably trained and motivated technical personnel are the norm in a laboratory research context. Quality control activities on sample analysis range from the nearly reflex use of a standard polymer film to calibrate an infrared spectrophotometer to the more visible use of split-sample aliquots, standard samples, and other techniques generally associated with calibration.

These latter require conscious and visible support and planning by the taskmaster if they are to succeed. Sample blanks should be analyzed on a regular basis. Samples spiked with known amounts of the analyte serve as a check on analytical bias. Split-sample aliquots can be analyzed by different analysts at different times using a different set of reagents (each as desirable) as another measure of data quality. Quality control measurements requiring highly developed subjective evaluations (e.g., pathological evaluation of tissue) may require side-by-side or round-robin analysis in order to establish the quality of the data. The taskmaster should choose the specific quality control activities appropriate to a given task in such a way as to emphasize the need for highest quality data commensurate with existing limitations.

3.9 Internal Audits

During the life of a task it is desirable to have an up-to-date evaluation of data quality. In this way, timely corrective action (see Section 3.15) is possible and data quality can be maintained at the desired level. Internal audits conducted by the operating group or organization are used to obtain data for this evaluation.

A variety of tools are available for use in internal audits, but they generally fall into four categories:

a. Reference materials are available from several sources, most notably, the National Bureau of Standards [10,11], i.e., NBS-SRM's. These may be included for analysis in various types of measurement systems at relatively low cost and with little interference to the normal laboratory routine.

b. Reference devices may be obtained (e.g., the reference flow (ReF) device for high volume samplers) for which the critical parameters are known to the auditor but not the analyst. These are somewhat more disruptive of laboratory operations, and there is no possibility of anonymity of the sample; however, the final result is still a measure of the performance of the total analytical system, including the operator.

c. Cooperative analysis, such as round-robin analysis, is useful for estimating the precision (not accuracy unless the analyte is a reference material) of measurement among several different operators and/or laboratories.

d. Side-by-side analysis, or collaborative analysis, may be necessary if important variables are not controllable in the sample.

These basic types of audit techniques may be applied to almost any measurement system. Both EPA and NBS are expanding their services to allow calibration of many audit substances and devices for which no NBS-SRM's previously were available. Frequently, however, cooperative or side-by-side analysis will be necessary for internal audits of HERL laboratory analyses due to the lack of suitable reference materials or devices and the complex nature of the evaluation. In these cases, the taskmaster (or project leader, for extramural tasks) will need to relate his responsibility to monitor and quantitatively document the task data quality with the various costs involved in this type of audit.

In either situation, the program and rationale for internal audits should be designed on the basis of individual components of the specific measurement process, and clearly planned for and budgeted into the task plans. By the use of internal audits, the taskmaster (or project leader) will be able to objectively evaluate data quality as his task progresses.

3.10 Preventive Maintenance

In order to insure long-term data quality in a cost-effective manner, a rational preventive maintenance (PM) program should be followed. This assumes importance roughly in proportion to the amount of instrumental data that is recorded. Reference 3 contains a good discussion of preventive maintenance, especially as related to routine measurements (Air Quality Monitoring). In particular, preventive maintenance will

increase the completeness of data from continuous monitoring systems, which is an important measure of quality for such systems.

In a laboratory research environment, PM has a less visible benefit; the effect on minimizing and controlling equipment downtime is nonetheless real. Preventive maintenance can be budgeted and scheduled based on failure analysis data available to (or developed by) the equipment manufacturer. Extended laboratory use of specific items can be scheduled with higher reliability, and shorter, more controllable, and less catastrophic interruptions than if maintenance occurs only when failures occur.

The laboratory equipment program should include: (a) scheduling, (b) performance, and (c) recordkeeping. Scheduling of PM should be developed based on the effect of equipment failure on data quality, any relevant site-specific effects, and equipment failure analysis (lacking this, the failure rate should be estimated). This schedule should be available to the person or group responsible for performing the maintenance, as well as the person or group using the particular item of equipment. In this way, use of the equipment may be scheduled appropriately.

Preventive maintenance should be performed by qualified technicians, by service contract, if necessary, according to a predetermined schedule. The specific service should be programmed based on the considerations noted in the preceding paragraph, and specified to both the user and maintenance groups. A predefined set of data should be obtained both before and after the maintenance activities to permit equipment performance evaluation. Calibration (see Section 3.11) should also be performed following all maintenance activities.

Documentation of maintenance--scheduled or not--is essential to monitoring and documenting data quality. A bound notebook should be kept with each instrument as a record of its maintenance history. A detailed description of adjustments made and parts replaced should be recorded in it. If the notebook is the multicopy type, one of the copies can be routed to the maintenance group for analysis. This analysis may include such considerations as mean time between failures (MTBF) for specific components, MTBF analysis for systems (individual and laboratory-wide), and indication of an onsite spare parts inventory appropriate to cost-effectively support minimum equipment downtime.

3.11 Calibration

Calibration is the process of establishing the relationship of a measurement system output to a known stimulus. In essence, calibration is a reproduceable point to which all sample measurements can be correlated. This process is a key element of any scientific measurement program, since without a valid calibration or reference system, the validity of the data from the measurement program will be questionable.

A sound calibration program should include provisions for calibration procedure documentation of frequency, conditions, standards, and records reflecting the calibration history of a measurement system (a monitoring network or a Spectronic 20).

Calibration procedures should be well-documented, step-by-step procedures for performing the needed referencing of a given system to a standard(s). Whether the procedure utilizes a specific standard (as in the calibration of a spectrophotometer) for the referencing procedure or visual analysis by trained personnel (e.g., a pathologist reading a microscope slide), a clearly written concise procedure is needed. A procedure of this type will help to minimize the bias that may be introduced into a system via operator technique. Calibration procedures for most systems can be obtained from NBS or ASTM. Other procedures may have to be developed in-house and must undergo extensive evaluation to determine, as nearly as possible, the accuracy, precision, replicability, repeatability, and reproducibility [3] of the procedure. To assure that the same calibration or reference point is being maintained for a measurement system, it is essential that a calibration schedule be initiated whether it involves simple daily checks or full-scale, multipoint calibrations. Provisions for action to be taken if an unforeseen circumstance occurs should be specified. Adherence to an exercise of this nature can minimize the generation of erroneous and/or indefensible data.

Environmental conditions are another type of reference point that must be dealt with when calibrating measurement systems. If the system is sensitive to environmental conditions (temperature, pressure, light, humidity, etc.), the calibration will not be valid unless the documented conditions are maintained as required.

The quality of the calibration standards is the most important aspect of any calibration program, for without high quality standards, the accuracy of the calibration cannot be demonstrated. Standards should be of the highest possible quality and should be referenced to a higher level primary standard such as an NBS-SRM. If no NBS-SRM exists for a particular system, cross-referencing of outside certification or the use of other primary standards or devices (such as ASTM standards) is acceptable. Calibration standards should also be obtained or prepared in the range for which the measurements are to be made. For example, a source concentration gas cylinder would not be used to calibrate an ambient monitor. Various organizations [10, 11, 12] list reference materials applicable to health-related research for use by HERL/RTP taskmasters.

Calibration history is the final point to be mentioned in this section. Each calibration and the full history of all calibrations performed on a measurement system must be recorded. This enables personnel to perform a systematic review of the data quality from a measurement system at a later date.

3.12 Documentation Control

Operating procedures for task measurement activities should be clearly documented and available to task operating personnel. A formal procedure for insuring that procedural and system changes are incorporated into existing documentation and that those changes result in corresponding changes in the habits of operating personnel is essential.

Section 1.4.1 of reference 3 clearly describes a comprehensive, practical document control indexing format appropriate for use within EPA laboratories. It has the advantage that only current versions of documentation are generally retained, and updating may occur at any time. An example of the information placed in the upper right-hand corner of each page is as follows:

Section No.	2.12
Revision No.	0
Date	September 27, 1977
Page	1 of 5

(Note that the date is the date of the revision.) A complete description of this system is given in reference 3.

3.13 Configuration Control

An adequate program of equipment/hardware configuration control will readily permit tracking all changes that are made to a data-producing system that may affect data quality. This applies to individual instruments as well as to entire data acquisition systems.

For extensive systems, such variables as sampling site changes, monitoring instrument replacements, etc., should be recorded similarly to calibration and maintenance (Sections 3.10 and 3.11), i.e., in a bound, page-numbered notebook reserved for this purpose. Major changes should require express approval of the responsible taskmaster. Treatment relevant to such systems is given in reference 3, specifically applied to air monitoring systems.

Configuration control for the laboratory environment is no less important. It includes instrument location in the laboratory as well as modifications (e.g., sample holder of different design) that affect measurement data. Equipment configuration changes should be made (permanent) only when the effect is well-characterized and demonstrated to improve data quality.

3.14 Data Validation

Data validation must be defined with reference to the requirements of each task. Frequently, laboratory data validation relies on the highly trained professional judgment of the investigator or technician. To rely on such capabilities in a monitoring network situation invites disaster. In both extremes, the data should be flagged but not discarded unless there is definitely identifiable error (e.g., an obvious and documented equipment malfunction).

Data validation may be defined as a systematic procedure whereby data are filtered and accepted or rejected based on a set of criteria for providing assurance of the validity (accuracy, precision, representa-

tiveness, completeness) of data prior to their ultimate intended use [3]. Criteria for each application of data validation techniques should be documented and implemented for all task data. Automated data acquisition systems are particularly suited for comparing reported data values with earlier stored values of the same parameter and establishing and updating such statistics as parameter mean and standard deviation. Similarly, checks for data completeness, calibration performance, signal levels within reliable measurement range (i.e., above minimum detectable and below saturation limits), etc., may be designed into data validation systems.

In a laboratory environment, operating personnel who are alert and adequately trained regularly perform this type of screening as they manually collect data. This requires particular attention that valid data are not rejected without adequate reason. Data should not be rejected "because they don't look right" or other similarly subjective reasons; it is generally the case that such data are frequently valuable as the particular model is developed to a higher level of sophistication.

In either the laboratory environment or the complex data acquisition system, provision should be made for regular analysis of the appropriateness of the specific validation criteria. This analysis should include both technical and professional inputs in order to keep a proper balance of theoretical and practical considerations in the setting of limits on the data. In all cases, data validation procedures should not be permitted to delete raw data, but only to flag it when a clearly stated validation criterion is exceeded.

3.15 Feedback and Corrective Action

For each task, a system for detecting, reporting, and correcting problems that may be detrimental to data quality must be established. As noted in reference 3, this system "...can be casual when the organization is small or the problems few. When this is not the case ... action documentation and status records are required." The exact system design should optimize the conflicting needs for quick response and thorough communication/documentation of the problem and its solution. More complex data acquisition systems, such as air monitoring systems, require a

formalized closed-loop system with standard forms for various stages of the problem and its solution. In a laboratory context, if a "fix" is not immediately apparent, direct contact between the taskmaster and the involved technician may be the most effective "system".

Additional feedback systems should, at least informally, be established. For example, the discovery of an impure substance by one investigator should be communicated to all other users of the particular substance as rapidly as possible. This can be facilitated by the use of adequate stockroom records.

A description of the problems, solution of the problems, and estimates of the effect of the problem incidents on data quality should be made available to appropriate management on a regular basis.

3.16 Data Processing and Analysis

Data from health-effects research are rarely, if ever, used in the form in which they are recorded. The initial phase of data processing (i.e., data reduction) processes the data into a form suitable for manipulating conceptually as well as for possibly performing preliminary statistical and other calculations. These intermediate results are then analyzed in terms of the particular model of interest to the investigator. Each of these transformations of the raw, observed data is made by a manually or electronically programmed series of manipulations. Hence, each transformation is a potential source of error in the final result. The automated, sophisticated analysis of large amounts of data thus carries the inherent potential for significant error due to the processing analysis functions quite apart from experimental errors.

The overall reliability of contemporary computer hardware systems is extremely high, due to various routine internal (to the machine) auditing checks. The major source of error may be traced to the software (i.e., programs), which provide the detailed instructions for operation of the hardware. Typical errors may generally be traced to insufficient testing of the program during the development stage, or improper application by the user. Either condition is difficult to detect due to the wide range of values that may be supplied to a program for processing and that cause

no hardware-detectable error. The only insurance currently available against the "Garbage In, Garbage Out" problem is for each user to exercise his or her best professional capabilities to estimate reasonable results. If such are not produced by the software system, a concerted effort should be made to determine the exact source of the discrepancy.

The potential for such software problems is greater with increased use of locally (i.e., within laboratory group) written programs for individual minicomputers and microcomputers. In addition to verification of the proper handling of "good data," extensive testing of the proper handling of "bad data" (i.e., data containing some representative, anticipated errors) should be performed over the complete range of possible values and thoroughly documented. Suggestions from the Data Management Staff for properly testing and debugging these programs will be cost-effective in terms of accurate and rapid computations.

3.17 Report Design

The most visible product of a research task is the document(s) that comprises the report of the important findings. Publication guidelines applicable to the HERL research reports are available [13,14]; minimum technical contents for nonclinical laboratory reports have been proposed [5] and are shown in Figure 3.

As in all scientific research reports, and within the indicated consistent style stipulations [13,14], the report should be concise and complete, with adequate discussion of the important technical aspects of the research to permit a qualified professional to duplicate the research. Adequate data should be included to permit at least partial calculation of important results. The conclusions, based on the data, and the reasoning to support those conclusions should be clearly stated. As much graphical and illustrative data correlation (with supporting tables, as appropriate) should be used as is feasible. Error estimates should be included with all quantitative and qualitative values reported, as well as the basis upon which the estimates were made.

Much of the research conducted under the auspices of HERL/RTP is highly specialized and frequently at the forefront of the technology, yet few of the individuals who make up the audience for the reports are specialists in the particular technical area. For this reason, the

1. Name and address of the facility performing the study and the dates on which the study was initiated and completed.

2. Objectives and procedures stated in the approved protocol, including any changes to the original protocol.

3. Raw data generated while conducting the study and any transformations, calculations, or operations performed on the data.

4. Statistical methods employed for analyzing the data.

5. The test and control substances identified by name and/or code number, strength, quality, and purity.

6. Stability of the test and control substances under the conditions of administration.

7. Methods used.

8. Test system used. When animals are used, include the number in study, sex, body weight range, source of supply, species, strain and substrain, age, and procedure used for unique identification of test system.

9. Dosage, dosage regimen, route of administration, and duration.

10. Any unforeseen circumstances that may have affected the quality or integrity of the nonclinical laboratory study.

11. The name of the study director.

12. A summary of the data, and a statement of the conclusions drawn from the analysis.

13. The reports of each of the individual scientists or other professionals involved in the study, e.g., pathologist, statistician. The dated signature of the study director and of all scientists and other professionals on their respective segments.

14. The location where all raw data and the final report are to be stored.

Figure 3. Proposed minimum report technical contents for nonclinical laboratories by DHEW/FDA [5]

purpose(s) and conclusion(s) of the research should be stated as clearly as possible (see Section 2.2). The estimated errors, as well as the limits of applicability of results, should be stated in such a way as to minimize misinterpretation. Application of the results to alternative theories (models) should be provided, with indication of the rationale used in reaching the stated conclusions rather than the alternative conclusions.

Data quality control and data quality assurance activities (Sections 2.2.2 and 2.2.3) should be discussed in as much detail as possible. This is especially true of in-house reports. This discussion should permit the specialist and nonspecialist alike to correctly assess the level of the quality assurance effort invested in the research. This should, in addition, permit subjective evaluation of the validity and accuracy of the reported results and conclusions.

4.0 DATA QUALITY ASSURANCE FOR RESEARCH PROJECTS

Discussion to this point has focused on aspects of the quality assurance plan that influence test data quality--from the perspective of operating technicians (or organization in the case of extramural research). In the following sections, the discussion focuses on QA aspects from the perspective of personnel separate from operating personnel (see Sections 2.2.3 and 2.2.4). The fundamental concept is that the taskmaster has at his or her disposal a variety of probes, or checks, on data quality quite independent of the functioning of the task research system. The choice of suitable probes, and their applications to the system (of research), is the taskmaster's, with the support of the QA organization within HERL/RTP.

4.1 Quantitative Estimates of Data Quality

Quantitative measurements and comparisons (i.e., quantitative audits) provide the best possible objective estimates of data quality--insofar as they are available. Recent efforts by the National Bureau of Standards to develop environmentally useful Standard Reference Materials (NBS-SRM's) are rapidly producing new NBS-SRM's. A current catalog of NBS-SRM's [10,11] may be obtained from:

Office of Standard Reference Data
National Bureau of Standards
Washington, D.C. 20234

In addition, the World Health Organization maintains information on worldwide sources of biological standards [12].

Appropriate use of the available reference materials by the taskmaster can provide an objective measure of specific parameter data quality. A variety of techniques, all of which should be designed as blinds (i.e., operating personnel unaware of the presence of the reference sample) are available. Direct analysis of the reference material and routine duplicate samples, one of which is "spiked" with a known amount of the reference material, are two possible uses of reference materials in analytical

systems for the evaluation of solution concentration, aerosol characterization, etc.

Unfortunately, NBS-SRM's do not exist for many measurements of interest. In such cases, there are still techniques for probing the quality of the task research system. Round-robin analysis of aliquots of a single sample may be performed by any number of laboratories. While accuracy (i.e., deviation from a "true" value) cannot be measured, an estimate of analytical variability (precision) is available. For labile samples, collaborative (side-by-side) analysis may be used (e.g., several technicians would distinguish and count normal cells contained on a set of plates). This is equivalent to the round-robin test, but is performed at one location and at approximately the same time. To give a measure of various research system components' variability, interlaboratory and intralaboratory analysis/measurement programs may be designed. In this case it is important that the statistical design of such testing recognize such aspects as operating shift changes, diurnal biological changes, and other nonrandom variability in the sample(s) and total measurement system.

4.2 Qualitative Estimates of Data Quality

In addition to the various quantitative probes available to a taskmaster, there are also qualitative probes of task research data quality. The comparison, rather than between two numerical values, is between the proposed and executed(ing) plans.

Thus the protocol (or work plan in the case of extramural support) is a statement of the reasoned plans of the operating organization. From qualitative measures of data quality (i.e., qualitative, or system, audit), an individual, independent of the operating organization or group, compares the planned activities with what is observed to occur. While complete agreement is no guarantee of high quality data, discrepancies are an indication that all is not well, that the task is not under the control of the taskmaster as it should be. Thus, the qualitative audit includes consideration of the execution of the points addressed in the protocol (which should be essentially the points covered in Section 3): Are data actually being collected according to the statistical design; are oper-

ating personnel properly qualified for their responsibilities; are records properly recorded and maintained, etc.

In summary, the taskmaster can use the various available probes to effectively demonstrate and document the quality of data being produced in a task by use of suitable quantitative and qualitative probes into the task research system.

REFERENCES

1. Health Effects Research Laboratory, Management Policy for the Assurance of Research Quality, Research Triangle Park, N.C., EPA/600-1-77-036, 1977.

2. The American Society for Quality Control, Glossary and Tables for Statistical Quality Control, Milwaukee, Wisconsin, 1973.

3. Environmental Protection Agency, Quality Assurance Handbook for Air Pollution Measurement Systems, Volume I, Principles, EPA-600/9-76-005.

4. Environmental Protection Agency, Quality Assurance Research Plan, FY 1978-81, EPA-600/8-77-008, 1977.

5. "Non-Clinical Laboratories Studies: Proposed Regulations for Good Laboratory Practice," Federal Register, Friday, November 19, 1976, pp. 51206-51230. (Also see revisions: Friday, January 7, 1977, p. 1486; and Friday, January 28, 1977, pp. 3367-8.)

6. Inhorn, S. L., ed., Quality Assurance Practices in Health Laboratories, American Public Health Association, 1977.

7. U.S. Department of Health, Education, and Welfare, Guide for the Care and Use of Laboratory Animals, US DHEW/PHS/NIH, DHEW Publication No. (NIH) 77-23, 1972.

8. Juran, J. M., F. M. Gryna, Jr., and R. S. Bingham, Jr., eds., Quality Control Handbook, McGraw-Hill, 1951, 1780 pp.

9. Bradley, M.O., and N. A. Sharkey, Nature, 266:724-25, 1977.

10. National Bureau of Standards, Special Publication 260, U.S. Department of Commerce.

11. National Bureau of Standards, NBS Standard Reference Materials for Environmental Research Analysis and Control, U.S. Department of Commerce.

12. World Health Organization, Biological Substances: International Standards, Reference Preparatons, and Reference Reagents, Geneva: World Health Organization, 1977.

13. Environmental Protection Agency, Handbook for Preparing Office of Research and Development Reports, EPA-600/9-76-001, 1976.

14. Health Effects Research Laboratory, "Health Effects Research Laboratory Procedures for Publishing Office of Research and Development Technical and Scientific Materials," Research Triangle Park, N.C., July 1977.

BIBLIOGRAPHY

American Council of Independent Laboratories, Quality Control System for Independent Laboratories, 1971.

Sherma, Joseph, Manual of Analytical Quality Control for Pesticides and Related Compounds in Human and Environmental Samples, EPA-600/1-76-017, U.S. Environmental Protection Agency, Health Effects Research Laboratory, Research Triangle Park, North Carolina, February 1976.

Thompson, J.F., ed., Analysis of Pesticide Residues in Human and Environmental Samples, U.S. Environmental Protection Agency, Health Effects Research Laboratory, Research Triangle Park, North Carolina, December 1974.

Whitehead, T.P., Quality Control in Clinical Chemistry, John Wiley and Sons, New York, 1977, 130 pp.

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-600/1-78-012		

4. TITLE AND SUBTITLE	5. REPORT DATE
DEVELOPMENT OF QUALITY ASSURANCE PLANS FOR RESEARCH TASKS - Health Effects Research Laboratory/RTP, NC	February 1978
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Environmental Protection Agency Office of Research and Development Criteria and Special Studies Office, HERL Research Triangle Park, N.C. 27711	10. PROGRAM ELEMENT NO. 1AA601 11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Health Effects Research Laboratory RTP,NC Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, N.C. 27711	13. TYPE OF REPORT AND PERIOD COVERED 14. SPONSORING AGENCY CODE EPA 600/11
15. SUPPLEMENTARY NOTES	

16. ABSTRACT

This document is designed to provide, in one location, a summary of details to be considered in the development of task-specific Quality Assurance plans for research tasks at the Health Effects Research Laboratory, Research Triangle Park, North Carolina. It is directed toward taskmasters as they design plans for "in-house" and contracted research tasks.

The logical structure of a research task is analyzed, from the initial planning stages through report preparation. The production of high quality data is dependent on consistently high quality efforts by all associated task personnel during all phases of task execution. Thus, guidelines for the taskmaster for planning and maintaining quality in each of those phases are presented. In addition, methods for monitoring and documenting data quality are discussed.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Quality Assurance Management Health Quality Quality Control	Health Effects Laboratory Quality Assurance Health Effects Research Quality	05 A 14 B

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
RELEASE TO PUBLIC	UNCLASSIFIED	48
	20. SECURITY CLASS *(This page)*	22. PRICE
	UNCLASSIFIED	

EPA Form 2220-1 (9-73)