NWRI CONTRIBUTION 89-99

A MANUAL FOR EFFECTIVE INTERLABORATORY QUALITY ASSURANCE

Coordinated by K.I. Aspila

Research and Applications Branch
National Water Research Institute
Canada Centre for Inland Waters
867 Lakeshore Road, P.O. Box 5050
Burlington, Ontario, Canada L7R 4A6

February 1989

A MANUAL FOR EFFECTIVE

INTERLABORATORY QUALITY ASSURANCE

Coordinator: K.I. Aspila

Contributors - H. Alkema
N. Arafat
V. Cheam
J. Clark *
H.B. Lee
J. Lee
Y.D. Stokker
K. Miles**
R. White *
A. Zingaro**

*International Joint Commission
Great Lakes Regional Office
Windsor, Ontario

**Computing and Programming
Services Section
Research Support Division
National Water Research
Institute

Research and Applications Branch
National Water Research Institute
Canada Centre for Inland Waters
867 Lakeshore Road, P.O. Box 5050
Burlington, Ontario, Canada L7R 4A6

February 1989

MANAGEMENT PERSPECTIVE

A successful strategy to ensure a high quality product in environmental programs includes the integrated influence of quality control, quality assurance and quality management. This manual's focus is on the "interlaboratory quality assurance". It has been prepared by staff of the Quality Assurance Group. The manual documents how interlaboratory studies are designed, prepared, carried out and evaluated. Included is a description of a data base management system to archive historical QA data and a narration on the role and value of certified reference materials for environmental research and monitoring programs.

This issue will form one part of a five section manual to be published shortly by the Water Quality Branch.

Dr. J. Lawrence
Director
Research and Applications Branch

PERSPECTIVE-GESTION

La coordination du contrôle, de l'assurance et de la gestion de la qualité est une stratégie qui permet aux programmes environnementaux de donner de bons résultats. Ce manuel porte sur le "contrôle de la qualité interlaboratoire". Il a été préparé par le Groupe de contrôle de la qualité. On y traite de la conception, de la préparation, de la réalisation et de l'évaluation des études interlaboratoires. Un système de gestion de base de données permettant d'archiver les données de contrôle de qualité accumulées au fil des ans est décrit. Ce manuel traite aussi du rôle et de la valeur des matériaux de référence homologués relativement à la recherche environnementale et aux programmes de surveillance.

Le présent texte constituera l'une des cinq sections du manuel qui sera bientôt publié par la Direction de la qualité des eaux.

Monsieur J. Lawrence
Directeur
Direction de la recherche et des applications

ACKNOWLEDGEMENTS

We acknowledge Mr. Pat Lively who in the early years of Environment Canada had insight and was instrumental in implementing the first interlaboratory studies program.

A sincere thanks is extended to Mr. R.E. White and Dr. John Clark of the International Joint Commission's Great Lakes Regional Office for their development of the computerized techniques for addressing bias in large round-robins. A similar thanks is provided to Don King, Quality Assurance Officer, Ontario Ministry of Environment for his valuable and indirect contributions and for his support of the external QA programs.

The authors appreciate the patience and support provided by Mr. A.S.Y. Chau, QA Project Chief and Dr. John Lawrence, Director, Research and Applications Branch, NWRI.

A special thanks is extended to Dr. J. Gaskin for the initial review and to Ms. Norma Snelling for patiently typing this manuscript.

The authors acknowledge A.S.Y. Chau, Chief, QA Project, for the outline of contents for this manual and his continual scientific input, discussions and suggestions provided during the course of writing and coordinating this manual.

PREFACE

A key element in the quality management strategies of environmental programs is "external quality assurance". The principle element of external QA is the "interlaboratory comparison study". This manual focuses on the interlaboratory study and provides insight on i) how such studies relate to quality management, quality assurance and quality control; ii) how such studies develop and employ reference materials, certified reference materials; and iii) how effective intercomparison studies are designed, prepared, distributed and interpreted. On the issue of interpretation, this manual provides an overview on an analytical quality control data base management system that is essential to the administration of large arrays of QA data.

PRÉFACE

Le recours à des observateurs de l'extérieur pour favoriser l'assurance de la qualité est un élément clé de la gestion de la qualité dans le cadre des programmes relatifs à l'environnement. Cette stratégie repose essentiellement sur les "études comparatives interlaboratoires". Le présent manuel porte sur les études interlaboratoires et donne un aperçu de i) la relation entre ces études et la gestion, l'assurance et le contrôle de la qualité; ii) de la création et de l'utilisation, dans le cadre de ces études, de matériaux de référence homologués ou non; et iii) de la conception, de la préparation, de la distribution et de l'interprétation de ces études. En ce qui a trait à l'interprétation, ce manuel donne un aperçu d'un système analytique de gestion de base de données relatives au contrôle de la qualité, système qui s'avère essentiel pour gérer de grandes quantités de données.

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 ON THE ADMINISTRATION OF QUALITY

1.1 The Management Process

The basic element contributing to a successful and effective environmental program[1] is a management structure receptive to "quality issues" and partaking actively in the management of quality. The organization will only be successful[2] if management is able to direct responsibility, be accountable and have traceability on all matters pertaining to quality. This process is referred to as "quality management", and to be successful, senior management must define and implement a quality management plan[3,4]. This plan includes the assignment of tasks, protocols and procedures that verify that their facility can achieve the level of product quality required, and that at all times verification exists that defined quality is indeed being achieved.

An effective quality management plan includes designated quality assurance officers whose primary role is ensuring management that a quality assurance program[5,6] is in place. This program, implemented by management through a quality assurance implementation plan (QAIP), is a planned schedule of activities that assures managers that a quality control program is in place and is being carried out effectively. This quality assurance program within an agency is external to the laboratory and the environmental monitoring and surveillance projects. In brief it is management's program that verifies through a neutral audit process that field, laboratory and data handling systems can and are at all times achieving their data quality objectives. One specific form of an audit inherent to a quality assurance implementation plan is the "interlaboratory study". Such studies are neutral third party evaluations of laboratory performance. This manual addresses this subject.

The third element of management's role and responsibility for quality issues is the quality control program[7]. This control program includes all "technical efforts" practiced in the field, laboratory and data handling systems that verify and document that data quality

objectives are being maintained continuously. Such control programs are monitored by the quality assurance office under directives of the quality assurance implemenation plan. In brief, quality control is simply those planned systems of technical activities within a laboratory that document product quality.

A flow diagram depicting the management mechanisms to implement, monitor, verify and document quality is given in Figure I.

1.2 The Interlaboratory Study (External Quality Assurance)

An interlaboratory study[8] normally consists of providing an identical set of several test samples to various laboratories for the analysis of specific constituents. Results reported are analyzed[9,10] and a report is prepared. If the objectives of the study, its design and the evaluation techniques are clear, concise, and well-thought-out, then these reports can be very informative to both the participants and the management they represent[11,12].

The design of a study must be carefully established in order to meet the requirements or objectives outlined in the quality management plan. If a laboratory or field method is not well established the study may be designed to obtain information on how well that method performs by one or more analysts. On the other hand the studies may be designed to verify if the same or different methods can produce the same results. (i.e. Are they comparable?) If the methods are comparable, then the design may be established to verify if application of the method is controlled. For complex methods such as those required for trace toxic organics, the design of a study may be specialized to address a certain area such as the calibration standard or extraction efficiency. These studies are collaborative studies[12] that evaluate methods and very often require a very detailed format[13].

A successful and well designed interlaboratory study can provide valuable feedback to analysts, lab managers and data users.

For instance the studies may identify:

- overall precision and bias within a laboratory or between laboratories;
- percent recovery of a constituent;
- erratic performance;
- measurement systems that are out of control;
- measurement systems that have significant baseline errors (poor blank corrections);
- complete failure of a method (not suitable for analysis of substrate);
- operational blunders;
- complete inadequacy of intra-lab QC;
- inadequacy of internal laboratory standards;
- complete adequacy (or inadequacy) of two or more laboratory measurement systems to allow inference that these 2 systems will produce compatible data bases which are adequate for interagency use;
- a neutral third party assessment of the overall performance of a laboratory (a vital statistic when contract laboratories are under review by management).

The interlaboratory study is an element of great significance in "quality assurance". Under the framework of a quality management plan, interlaboratory study reports must inform the appropriate levels of management on the status of quality and control with the clear understanding that authority and power exist to implement corrective action if performance is substandard. This management process must include the "project heads" and "project managers" whose use of data and whose whole data bases may be affected. Project leaders are assumed to retain close routine liaison with their laboratory data producers. The above relates to the QA management process. However, perhaps the most important information transfer is to the laboratory managers via the

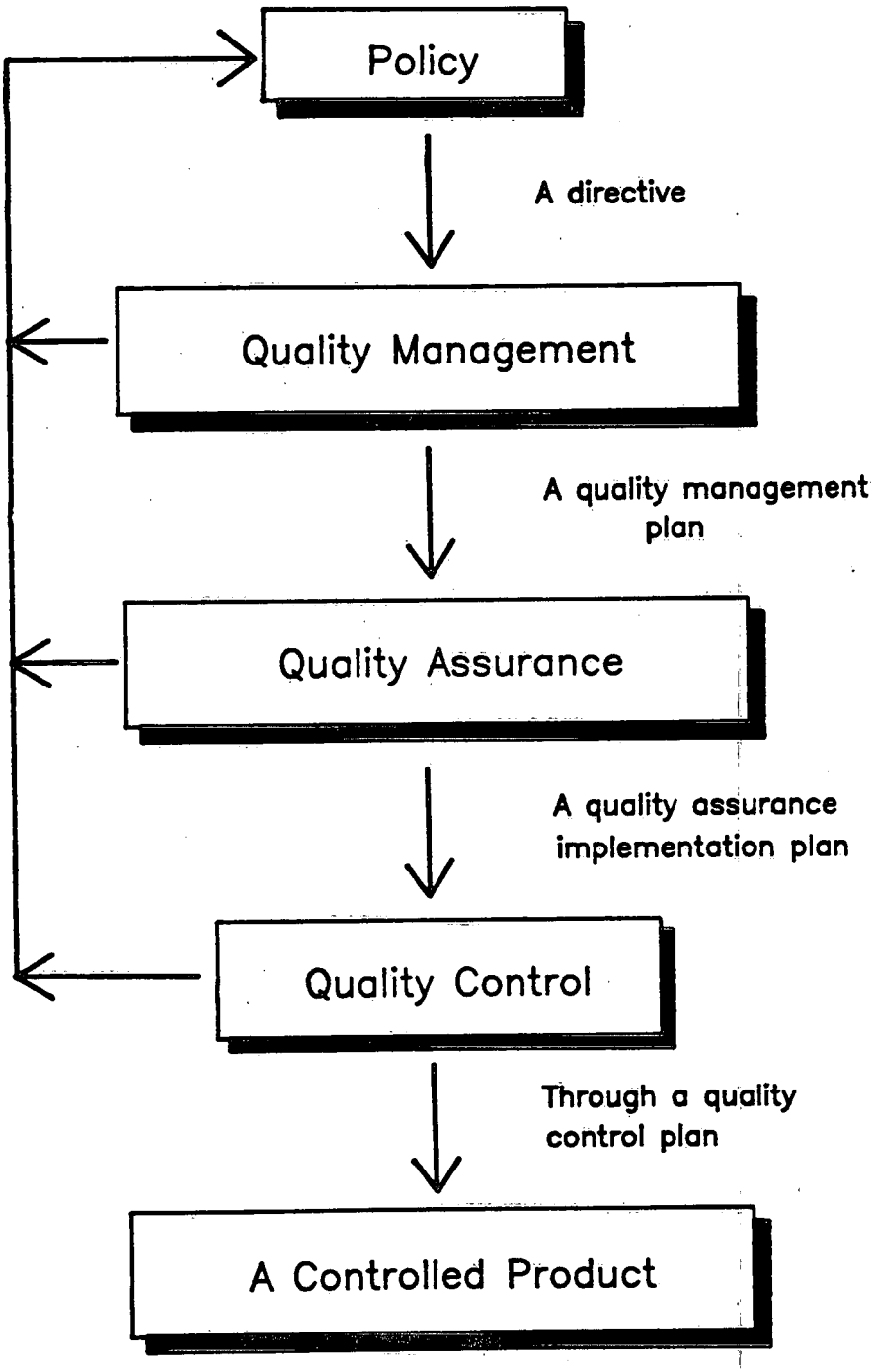


Fig. 1 A Conceptual Framework on Quality

interlaboratory study evaluation report. The transfer of performance statements must be swift since most issues are "laboratory measurement issues" and if serious must be quickly addressed to prevent the data base from being tainted with bad data. Various types of poor and excellent performance in an interlaboratory study are given in Figure 2. These examples present the situations that external QA programs must review, assess and report on. External QA programs at NWRI are highlighted in Table 1.

1.3 Limitations of Intralaboratory Quality Control

A single laboratory working in total isolation is simply unable to verify that its methods' applications, standards and product output are adequate for its data users. Moreover it is unable to show comparability with its peers when two or more different laboratories are to merge their data bases. Management at all levels of authority must accept this reality, and must react in supporting external quality assurance to authenticate the effectiveness of intralaboratory control measures and allow different laboratories the opportunity of merging their data bases for common use.

The most serious failure of an intralaboratory QC program is its inability to authenticate the validity of calibration standards. This is especially true for toxic organics which are prone to solvent evaporation, degradation and uncertainty of purity of stock standards. Variability of supply and instability of diluted standards are also in question. A major concern is that when the laboratory calibration is erroneous then so will be the resulting data. The magnitude of the

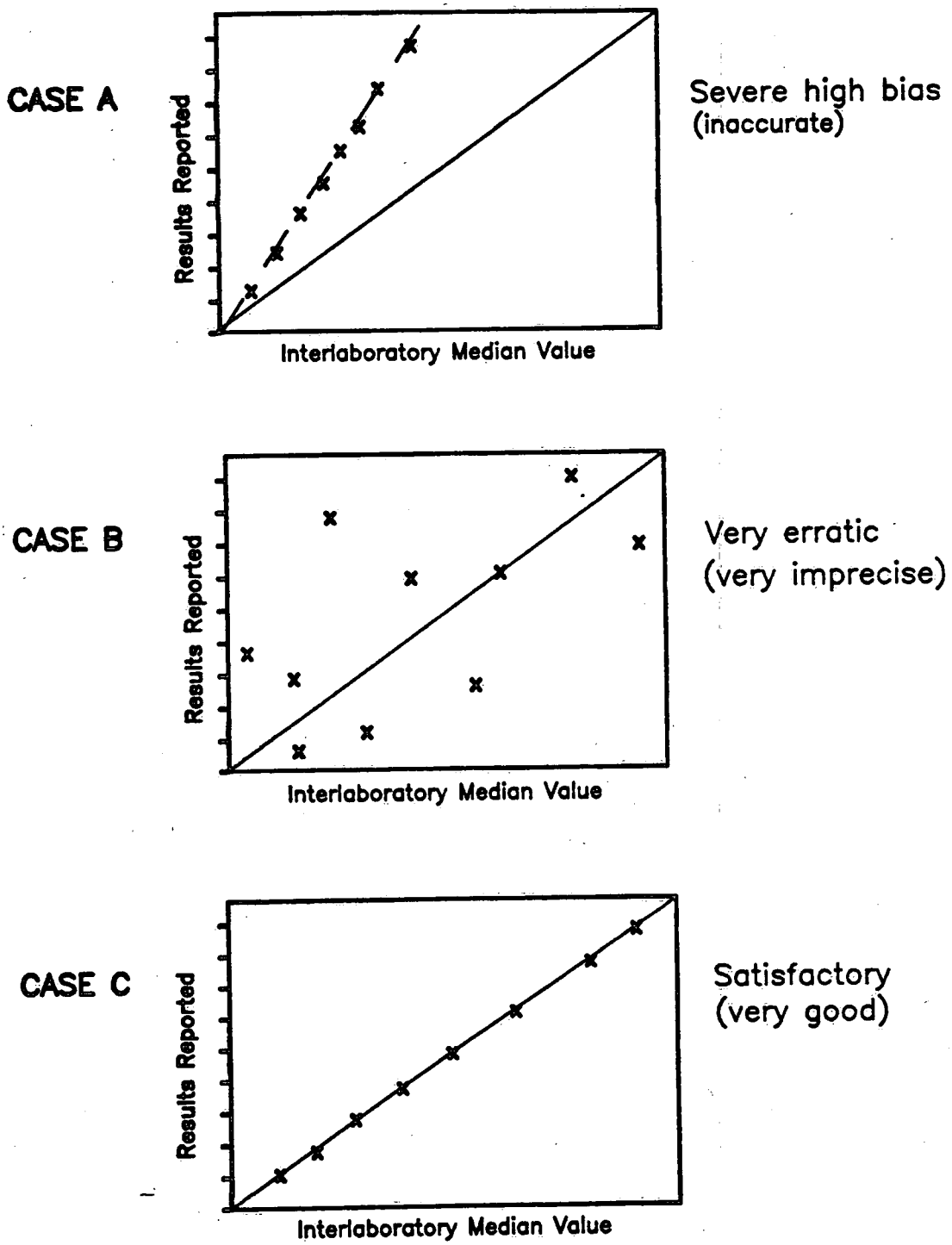
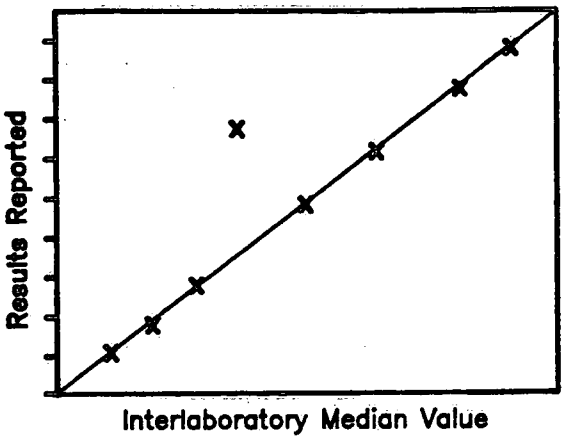


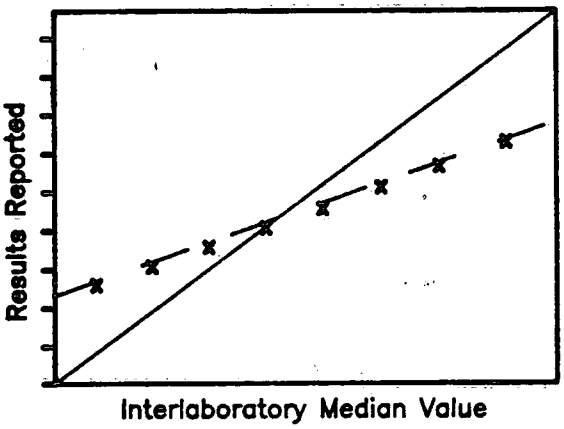
Fig. 2a Some typical types of Laboratory Performance revealed by External QA Studies

CASE D



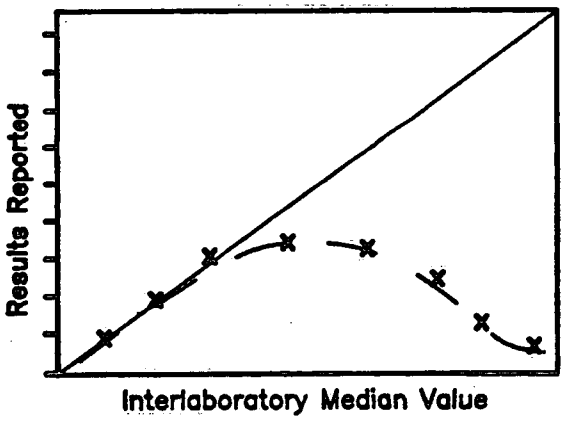
Out of Control

CASE E



Biased with
Blank Problem

CASE F



Methods Failure

Fig. 2b Some typical types of Laboratory Performance revealed by External QA Studies

TABLE 1

External Quality Assurance Programs
at the National Water Research Institute

Program	Number of Labs	Clients
LRTAP	102	US-Canada Acid Rain Labs
IJC	140	Great Lakes Surveillance
UGLCCS	16	Bi-National (connecting channels)
National	137	Canada (national program)
FP and PPWB	18	Federal-Provincial program
FICP	40	Pesticide Labs
Eulerian	8	US-Canada (Acid Rain)
National Dioxin QA	20	Commercial and Federal

calibration bias may be very small or extremely large but in either case, can be swiftly detected by a suitable intercomparison study.

A second serious deficiency of a within-lab QC program lies in its inability to address verification of the control on the test sample blanks. For example, if a laboratory addressing a colourimetric analysis of phosphorus uses dirty water to prepare calibration standards,its instrument zero baseline will appear right but in actuality will be technically wrong. A simple interlaboratory study using "clean waters" as blanks and very low level standards will quickly reveal any existing anomoly. A clean test sample in an impure matrix will actually yield a negative instrument response and, will be translated into a negative concentration on test samples. This case example, which sometimes occur, serves to (a) reveal the deficiency of relying only on intralab QC, (b) the merits of an intercomparison study and (c) the valuable contribution made if all laboratories were stimulated into reporting all calculated values whether positive, zero or negative.

A third major deficiency of the sole reliance on intralaboratory QC is that it is quite unable to provide management and program data users with any information on the "comparability" of data between different laboratories. Documented intralaboratory statistics may draw inferences on potential levels of comparability but it is the "demonstrated" peformances of laboratories that will provide the assurance.

Awarding of analytical contracts should not rely solely on intralaboratory QC information. Management who approve contracts for environmental measurements should remain vigilant and recognize that performance of a labortory should be assessed on both written documentation as well as demonstrated performance in the analysis of test samples through an ongoing external QA program. This is critical to the contract selection process. The multilaboratory, multilab, multisample intercomparison study can quickly reveal peer group performance and verify suitabiity of a contract lab. Relying solely on intralaboratory information is quite unacceptable.

1.4 References

1. J. Lawrence, A.S.Y. Chau, and K.I. Aspila. Analytical Quality Assurance: Key to Reliable Environmental Data. Canadian Research, p. 35-47, Nov. 1982.
2. J.S. Hunter. National System of Scientific Measurement. Science, 210, 869-874, 1980.
3. D. King. Quality Assurance Policy and Guidelines. Laboratory Services Branch, Ontario Ministry of Environment.
4. D. King. Quality Management Plan. Laboratory Services Branch, Ontario Ministry of Environment.
5. M.W. Brossman, T.J. Hoogleem and R.C. Splinter. Quality Assurance Project Plans: a Key to Effective Cooperative Monitoring Programs. Quality Assurnace for Environmental Measurement, Ed. J.K. Taylor and T.W. Stanley, ASTM Special Technical Publication 867, 1982, ASTM, 1916 Race Street, Philadelphia, PA 19103.
6. D. Hart. Quality Assurance Coordination of Great Lakes Surveillance. In a report prepared by Beak Consultants Ltd., Brampton, Ontario, L6T 5B7 for Environment Canada, Inland Waters Directorate, Ontario Region, Burlington, Ontario.
7. H. Agemian. Quality Assurance in the National Water Quality Laboratory. 1988. National Water Quality Laboratory, Water Quality Branch, 867 Lakeshore Road, P.O. Box 5050, Burlington, Ontario, L7R 4A6.
8. K.I. Aspila, R.E. White and J.L. Clark. Quality Assurance Aspects of the International Joint Commission Great Lakes Monitoring Program. In ASTM Special Technical Publication 867 (1985), a Symposium on Quality Assurance of Environmental Measurements, Aug. 8-12, 1983, Boulder, Colorado (published by ASTM, 1916 Race St., Philadelphia, PA 19103).
9. W.J. Youden. Ranking Laboratories by Round-Robin Tests. p. 165-9 to p. 169-13, in Precision Measurement and Calibration, Harry H. Ku, Editor, NBS Sepcial Publication 300, Vol. 1, U.S. Government Printing Office, Washington, DC, 1969.

10. J.L. Clark. Evaluation of Performance of Laboratories Determining Water Quality Constituents Through Natural Water Samples Whose True Values are Unknown. In Summary of Conference Presentations, Environmetrics 81, p. 54-55, 1981, Alexandria, Virginia, April 8-10, 1981.
11. (ASTM-D19), D4210-83. Intralaboratory Quality Control Procedures and a Discussion Reporting Low-Level Data. A Standard Practice. ASTM, 1916, Race St., Philadelphia, PA 19103.
12. W.J. Youden and E.H. Steiner. Statistical Manual of the Association of Official Analytical Chemists. AOAC, PO Box 540, Benjamin Franklin Station, Washington, DC, 20044 (1975).
13. J.K. Taylor. Quality Assurance of Chemical Measurements. Lewis Publishers Inc., Chelsea, Michigan, USA.

2.0 PREREQUISTES FOR AN INTERLABORATORY STUDY

Prior to the inception, development and implementation of an interlaboratory study, a number of factors must be addressed. Such factors as management issues technical issues are important.

2.1 Management Issues

When a small or large scale environmental program is conceived and implemented, management has the responsibility of assuring that the program yields credibile and traceable data[1]. One element of the management plan is the external quality assurance program and the associated interlaboratory studies[2,3,4]. To be successfully implemented, the external quality assurance program must have the complete support of management. Areas that must be addressed include:

- Assigning responsible and qualified quality assurance persons to develop the interlaboratory program.

- Providing resources and adequate lead time to develop the reference materials essential and pertinent for laboratory analysis.

- Providing guidance documents to outline the data quality objectives of the environmental program.

- Allowing the quality assurance project personnel developing the study to participate in the subgroups or committees when the environmental program is multi-jurisdictional.

- Defining all laboratories and associated managers involved in both the environmental program and the QA assessment process.

2.2 Technical Issues

When the quality management process has been implemented (through the policy directive) a need exists to develop the interlaboratory QA infrastructure and to implement the quality assurance plan. The plan must incorporate the ongoing production of essential reference materials suitable for developing certified reference materials; a suitable means for rapid analysis of all reference materials to verify stability; and as well, the necessary logistics to deliver intercomparative studies. When data are returned, the results must be swiftly and correctly interpreted in order to provide timely and essential advice to laboratories, managers and data users.

On a technical basis, the implementation of an effective QA program requires:

- an adequate budget to develop and deliver QA studies;
- suitable physical resources such as a cold storage for large volumes of water, soils, sediments, biota and vegetation;
- a well-equipped and suitable technical support facility to acquire, synthesize, and house large environmental reference materials;
- skilled chemists who are knowledgeable in various aspects of QA and the essential concepts of statistics, computers and chemistry;
- a well-equipped laboratory system of proven competence that is able to produce and process data in a timely manner;
- a computer facility and network with programming capability and procedures in place for rapid data input and retrieval and timely output of reports;
- secretarial support to assist in the routine output of reports.

2.3 References

1. D. King. Credibility, The Consequence of Quality Assurance. 1982. Laboratory Services Branch, Ontario Ministry of Environment, Rexdale, Ontario.
2. J. Lawrence, A.S.Y. Chau and K.I. Aspila. Analytical Quality Assurance: Key to Reliable Environmental Data. Canadian Research, p. 35-37, Nov. 1982.
3. K.I. Aspila, R.E. White and J.L. Clark. Quality Assurance Aspects of the International Joint Commission Great Lakes Monitoring Program. In ASTM Special Technical Publication 867 (1985), a Symposium on "Quality Assurance of Environmental Measurements", Aug. 8-12, 1983, Boulder, Colorado (published by ASTM 1916 Race St., Philadelphia, PA 19103).
4. J.S. Hunter. National System of Scientific Measurement. Science, 210, 869-874, 1980.

3.0 ON DESIGNING AN INTERLABORATORY QUALITY ASSURANCE STUDY

When management, through the Quality Management Plan, has authorized the development of an intercomparison studies program considerable care must be taken in designing the study program[1,2]. Optimizing maximum benefits to client and keeping the development cost at a minimum are often required. Studies may be designed simply to verify calibration standards or to address whole analytical systems. They may be designed to compare intensively. analytical methodologies, or simply to determine the level of comparability of a number of different laboratories which use different methods to supply data to a large program data base. Some studies may have a very simple design whereas others are most complex. The decision to select a particular study design is partially influenced by the study objective and the costs related to the study design and operational stages.

3.1 Simple Designs

A simple design is a study containing very few samples and requiring a minimum sample workup. Such studies may include:

a) a single ampule standard solution ready for direct analysis;
b) several ampule standard solutions for direct analysis or requiring appropriate dilution;
c) one or two natural samples;
d) one or two natural samples so highly characterized to be classified as certified reference materials.

Simple studies have merit in that as they are relatively inexpensive to prepare and inexpensive for the participating laboratories. Unfortunately, they are not always very informative, and may provide little information on the full characteristics of a laboratory's calibration program (e.g., providing information on such

issues as baseline corrections or curvature of the calibration). simple studies by techniques such as Youdens pairs are able to provide visual information implying the possible source of error (random or systematic error). To this end, they provide inexpensive information at low cost. Their value is improved when the samples are CRMs or true value calibration standards.

3.2 Multi-Sample Designs

The most popular design and the most effective is usually one which contains many samples. For these studies the samples selected cover the entire routine concentration range of the equipment used by clients and each sample contains many constituents. Such designs have been popular in the LRTAP, Eulerian and IJC Great Lakes QA programs. There are normally ten samples, with most of them having an extensive history of use in previous studies. For each study, several new samples are introduced. When possible, synthetic samples (waters) or fortified samples (e.g., fish homogenates or wet sediments) are included.

The interpretation of data for the multisample approach is more complex and is addressed in Section 8.

3.3 Frequency of Studies

When a QA program is being designed to meet the needs of an environmental plan, some thought needs to be given to the frequency of the external QA studies. If the environmental program has a well-developed QM-QA-QC structure in place, the question is mundane, as the QM-QA-QC process will provide the necessary assurance. For example, the management, managers and users of data would have already set their data quality objectives, and the level of protection and the need for verification of proven performance would already have been discussed and verified. Unfortunately, this ideal situation is seldom if ever achieved.

New programs, involving new labs (contract facilities) should always have a plan of action in place before external studies are conducted; this is to verify that the participating laboratories can indeed meet minimum specification. Too often this not done and program staff and users of data are left with less than adequate bench marks to work with.

When a laboratory or several laboratories are engaged in a long term monitoring program, it is important to verify performance on a continuous basis, and for an annual field program, the external lab audit performance should be made at the beginning, the middle and near the end of the analytical season. This may mean three studies per year and for critical programs perhaps even monthly audits. Clear evidence of how an external audit program (external QA) can demonstrate performance, is given in Chapter 8. The graphical element reveals major improvement, steady decline, a level of incompetence and very excellent performance in one very successful program. In this program all participants had been assumed to be very adequate.

3.4 References

1. J.K. Taylor. Quality Assurance of Chemical Measurements. Lewis Publishers, Inc., 128 South Main Street, P.O. Drawer 519, Chelsea, Michigan 48118.
2. Harry H. Ku. Precision Measurement and Calibration. NBS Special Publication 300, Vol. 1, US Government Printing Office, Washington, DC, 1969.

4.0 PREPARATION OF CONTAINERS

An essential pre-requisite to the successful preparation of interlaboratory study test samples is the use of sample containers, closures and equipment that will not influence or contaminate the substrate. For many constituents at ultra trace concentrations, the container can be a source of contamination and efforts must be made to prevent or limit this possibility. The following is a description of how sample containers can be prepared[1,2].

4.1 Aqueous Samples for Inorganic Parameters

Containers used for aqueous samples are normally plastic (conventional or high density polyethylene, polycarbonate, polystyrene or Teflon). For major ions, nutrients and physical parameters residing in moderate to hard waters, the need for rigorous cleaning is not always critical. New bottles must be subjected to quality control checks using distilled water (or deionized distilled water) blanks to verify the absence of contamination. Similarly, for high concentrations of trace metals (Fe, Mn, Zn, Cu, Ni, Co, Pb, Cd, As etc.) the use of new bottles, without cleaning, may be acceptable but caution must be exercised. High concentrations of metals are those over 250 µg/L.

Containers for soft surface waters and rainfall samples where the constituent concentrations are usually very low, require special attention in the control of "blank" values. Experience has shown that chromic acid for bottle washing can be adequate. If this powerful acid is used, one must be very meticulous in washing the closures since acid residues may adhere to the bevelled edges of the closures. A more cautious approach in the preparation of bottles for soft water samples is to use the bottles after they have been left filled with distilled deionized water for several months. Prior to use, a representative number of bottles must undergo quality control testing to ensure that no interfering substance is present. Since bottles are purchased in large

quantities, it is precedent to request a single production lot. This action ensures that all bottles have an equal chance of containing the same contaminant, should any be present.

4.2 Aqueous Samples for Organic Parameters

Glass bottles are generally used for studies involving toxic organics in water; the 40 oz. whiskey bottle is normally the bottle of choice. These bottles require rigorous cleaning. The following is a suitable procedure:

- fill bottles with a hot detergent solution and leave for two hours (a suitable detergentsolution is Liqui-Nox in hot tap water);
- brush inside of bottles to dislodge particulates;
- rinse three times with hot tap water to remove all traces of detergent;
- rinse three times with distilled deionized water;
- rinse inside and outside of bottles with reagent grade acetone.
- oven-dry the bottles for two hours at 200°C;
- plastic closures may be cleaned less rigorously because the test samples do not come in contact with the plastic due to the use of clean Teflon as a sealing liner.

4.3 Sediment Bottles

Bottles used for sediments that contain inorganic constituents may not require as thorough a cleaning as in the case of organic parameters. A quick check of a representative set of bottles using a clean acid wash solution may reveal no appreciable metal residue (compared to the potential metal content of the sediment). If there is uncertainty regarding the contamination of the bottles (with respect to the parameters of interest), all bottles should be acid-washed by filling the bottles with nitric acid and allow soaking to continue for a minimum period of 24 hours. Closures should be treated in the same manner.

The extent to which the washing is carried out will depend in part on the anticipated level(s) of the parameter of interest and on the requirements of the QA program.

Bottles used as containers for sediment samples destined for toxic organic studies need to be treated in a manner similar to that of the whiskey bottles used for water samples (Section 4.2). Polyethylene closured lined with solvent washed aluminum foil are used to prevent contamination from the plastic lining of the closures. The aluminum foil may be washed with two solvents; first acetone and then hexane. An alternative cleaning process for the foil is oven drying at 150°C for 12 hours to drive off the organic contaminants.

4.4 Bottles for Fish Homogenates

Bottles for fish (15 to 50 ml ointment jars) are treated in the same manner as sediments containers.

4.5 Ampules

Glass ampules (borosilicate-score break) used for toxic organic standards are generally required in 10 to 100 gross quantities. Their application with respect to injection-ready dilute standards require that care must be taken to ensure they are clean. A suitable cleaning procedure is as follows:

- stand 50 to 100 ampules upright in a large, clean glass beaker;
- carefully, fill each ampule with acetone;
- fill beaker with acetone to ensure all ampules are filled and fully submerged;
- place beaker into an ultrasonic cleaner for five mintues;
- empty ampuls of acetone, and dry in the oven at 200°C for two hours.

4.6 References

1. D.T.E. Hunt and A.L. Wilson. The Chemical Analysis of Water. 2nd Edition, The Royal Society of Chemistry, Burlington House, London W1V OBn.
2. Environment Canada. Sampling for Water Quality. Water Quality Branch, Inland Waters Directorate, Ottawa, Ontario K1A OH3.

5.0 THE ROLE OF CERTIFIED REFERENCE MATERIALS AND STANDARDS IN EXTERNAL QA PROGRAMS

Nearly every phase in environmental protection and pollution control depends on the identification and measurement of pollutants in the environment. Millions of dollars are involved in generating analytical data (expensive sampling trips, manpower, and equipment for analysis and for data interpretation). There are even greater financial implications when decisions such as sewage treatment process changes, plant modification or construction of new facilities, import and export of food (e.g., fish) are based on analytical data generated. Indeed, as pointed out by Uriano and Gravatt[1] that "never before have so many critical decisions involving health, safety and economics depended on the quality of chemical anaytical data". Therefore, assurance of environmental data quality is an extremely important aspect in the effort to ensure the quality of the environment and health of the public[1]. Quality assurance must be an integral part of analytical and data interpretation activities since questionable data result in questionable interpretation and subsequently in questionable decisions or conclusions.

5.1 The External Quality Assurance Program

An effective interlaboratory quality assurance program should involve various activities to assist analytical laboratories to generate reliable data. Distribution of test samples to participants and generation of data reports are two important areas of the program. Interlaboratory quality assurance programs consist of many research and investigative activities such as sample preservation, sample handling, and analytical methodology. Validated procedures and methods are used in analytical laboratories for interlaboratory quality control studies and a respository for reference analytical standards for distribution to regional analytical laboratories is necessary for the operation of the program. In addition, research is conducted to develop Certified

Reference Materials (CRMs) and Reference Materials (CRMs) and Reference Materials (RMs) for both organic and inorganic parameters in environmental substrates. This enhances the effectiveness of the intra- and interlaboratory quality control programs.

Development and application of CRMs constitutes some of the most important activities in an effective quality asurance program. Without standard or certified reference materials accuracy of an analytical method[3] cannot be determined.

5.2 Certified Reference Materials

CRMs are stable, homogeneous and well-characterized reference materials prepared in quantity and having essentially identical or very similar matrices to the field program materials in order to eliminate or minimize matrix effect between reference and test samples. The assigned values are obtained by repetitive analysis by several operators and different methodologies in one or more qualified laboratories of known precision and accuracy. Application of CRMs should enable the user to evaluate and calibrate the whole measurement process rather[4] than just a part of it.

RMs are similar to CRMs except they are less rigorously characterized. They are the forerunners of CRMs.

Ideally, all CRMs should be made from naturally contaminated samples to reflect actual environmental situations. Ideally, all CRMs should be made from naturally-contaminated sample materials to reflect actual environmetanl situations. In practice, it is not always feasible to prepare CRMs from naturally contaminated samples for the following reasons:

- resource and time restraints often limit the number of CRMs that can be developed and produced at a given time.
- sites of suitable natural samples may not be accessible.

Alternatives such as a fortified material using natural samples which are not contaminated with the parameters of interest may be used. However, spiking a matrix has several inherent weaknesses, notably:

- the spiked compounds may not become well mixed and integrated into the sample matrix.
- the recovery of the spike is often not a measure of the true recovery of the endogeneous compounds from the real samples. It is quite common for spiked recoveries to be higher than the corresponding recoveries from real samples.

Spiked CRMs using solvent or reagent grade samples such as distilled water, not only inherit all the pitfalls of spiking but do not take into account the influence of sample matrix: distilled water is significantly different from natural water in many aspects of water quality. The limitations of spiking have been discussed by many authors. For example, Trautmann[5] and Brownman[6] have respectively reviewed the problem and limitations of spiking, and recently, Albro[7] re-emphasized the weakness of "spiking" to obtain recovery information.

5.3 Importance of CRMs

The importance of suitable CRMs for the assurance of data reliability cannot be overly emphasized. The success of interlaboratory quality assurance programs, "depends upon the availability and use of high quality CRM and standard reference samples"[8]. Interlaboratory programs that are not based on CRMs can only give a measure of between-laboratory precision, not accuracy[9]. The need to identify both precision and accuracy in the control of a measurement system is well recognized. According to Hunter, no measurement system can be truly under statistical control without measures of both of its precision and its intralaboratory bias (a reflection of accuracy).

In an intralaboratory quality control program, the application of suitable CRMs as control samples can increase the program effectiveness. In the absence of CRMs, accuracy of the analysis is determined by spiking the analyte(s) into blank samples and determining their % recovery. As noted earlier, this widely practiced technique is a poor compromise because spike recovery is not a measure of the recovery of the endogeneous compounds. This alternative action results in misleading information, and generates false confidence in both the accuracy of the method used and the data obtained.

The value of interlaboratory quality control studies using suitable CRMs are important in data interpretation such as for trend analysis of pollutants and in monitoring and surveillance programs involving data from various agencies. Lacking quality assurance and the interlaboratory calibration, data users cannot correlate data between methods or between laboratories. Because of this, data users have found it difficult to identify trends even when such trends exist[11].

5.4 The CRM Program at the National Water Research Institute NWRI

Procedures for the preparation of CRMs depend on the type of sample matrix and parameters that are involved. A simplified scheme is outlined in Table 2. In the case of waters and sediments, sample sites are selected and test samples are analyzed to determine background concentration levels. A large sample is then collected. Depending on the requirements and physical characteristics of the sample, the sample is freeze dried, blended and/or mixed. Homogeneity of the bulk samples are checked. After sub-sampling, between and within sub-sample homogeneity is then checked by repetitive analyses. The "reference" values of the parameters to be certified are obtained by replicate analyses using multi-operators, multi-laboratories and a minimum of two independent methods. The precision and accuracy of the chosen analytical methods and analysts are predetermined and monitored during analysis. The total number of analytical data used to generate the design value varies but on the average is about 200.

TABLE 2

Development and Preparation of a CRM

1. Preliminary Investigations
   - spiking and homogenization techniques
   - long term storage facilities and stability studies
   - preservation requirements
   - selection of containers
   - freeze drying, grinding, screening and sub-sampling

2. Selection of Sampling Site
   - sample type, location
   - resource and space restraints
   - technical considerations
   - historical data
   - in-house analysis to confirm suitable parameters exist in small scale field study

3. Sample Collection
   - arranging
   - collecting 1000 litres of water; two tonnes of wet sediment

4. Sample Handling and Preparation
   - freezedrying, crushing, homogenization, bottling

5. Certification
   - in-house analysis using two or more standard methods
   - external analysis using different methods
   - QA studies
   - data reduction

6. Maintaining the Registry
   - inventory control
   - monitoring sample integrity
   - external QA studies
   - internal QC applications

Research into the development of CRMs (Table 3) is one of the more important activities in the Quality Assurance Program at the NWRI. Strategies and activities for organic parameters are different from those for inorganic parameters, and thus the program is divided into two key areas: inorganic CRMs and organic CRMs. Environmental substrates such as water, sediment and biota are the matrices of interest. Those CRMs that are available through the QA Program at NWRI are listed in Tables 3a, 3b and 3c.

5.4.1 Sediment CRMs for inorganics

The development and preparation of CRMs for inorganic parameters in sediment are less complex than those for organic parameters, mainly because there are less variables to affect the preparation and also there are literature analogies on the preparation of rock and soil CRMs and the recently developed NBS river sediment for several trace metals. Due to matrix differences, a rock or soil CRM is not completely suitable for sediment work even when the certified parameters are of the same type and similar levels. For the same substrate (e.g, sediment) there are considerable matrix variations in different locations resulting from different geological, biological and human activities. Some analytical measurement systems are sensitive to matrix variation and can also be sensitive to concentration levels. For quality assurance purposes, it is therefore desirable to have at least one CRM having characteristics similar to that encountered in the routine test samples.

5.4.2 Sediment CRMs for toxic organics

Unlike the situation for inorganic parameters, development and preparation of CRMs for organic parameters are very complex[12-20] and require considerable inhouse research to provide background information. There is very little literature on the subject. Furthermore, since samples from different locations may have different matrices, each sample needs to be dealt with individually.

The following are key areas that need to be investigated before the development of a CRM for organics can be initiated:

a) homogeneous techniques;
b) spiking technique (if needed);
c) long term storage conditions;
d) preservation techniques;
e) choice of methods and procedures for certification of levels.

Sediment reference materials[12-20], currently certified or undergoing certification are listed in Table 3.

5.4.3 Aqueous CRMs (inorganic parameters)

A list of several aqueous CRMs are given in Table 3c. The majority of these samples are from natural sources and each have been used in several large interlaboratory studies.

5.5 References

1. G.A. Uranio and C.C. Gravatt. The Role of Reference Materials and Reference Methods in Chemical Analysis. In CRC Critical Reviews in Analytical Chemistry, 361, 1977.
2. ACS Committee on Environmental Involvement. Anal. Chem., 53, 2242, 1980.
3. J. Sherma. Manual of Analytical Quality Control for Pesticides and Related Compounds, EPA-600/1-76-017, 1976/
4. W.E. May, J.M. Brown, S.N. Chesler, F. Guenther, L.R. Hilpert, H.S. Hertz and S.A Wise. National Bureau of Standards Special Publications 519, Proceedings of the 9th Materials Research Symposium, April 10 to 13, 1978 (issued April 1979).
5. W.L. Trantman. The Persistence and Analysis of Organochlorine Pesticides in Soils and Sediments. Ph.D. Thesis, University of Wisconsin Library, Madison, Wisconsin.

6. M.G. Bowman. Extraction and Analysis of Organochlorine Insecticides from Lake Sediments. Ph.D. Thesis, University of Wisconsin Library, Madison, Wisconsin.
7. P.W. Alrbo, The Referee: AOAC, 4, May 1981.
8. E.W. Bretthauer, G.B. Morgan and R.E. Jaquish. Standard Reference Materials and Environmental Monitoring. In Standard Reference Materials and Meaningful Measurements, R.W. Seward (Ed.) U.S. Government Printing Office, Washington, 1975.
9. H.T. Yolken. The National Standard Reference Materials Programs in the USA. In NBS special publication No. 408, US Government Printing, Office, Washington, 1975.
10. J.S. Hunter. The National System of Scientific Measurement. Science, 210, 869, 1980.
11. J.B. Kirch. Study of Users' Requirements for Water Quality Branch Data and Recommendations on Improvement of the Data Dissemination Service, draft, 1974.
12. A.S.Y. Chau. Standard Reference Materials in Environmental Quality Assurance. In "Water Around the World, Part 2", Water Quality Bulletin, 8, 26, 1983.
13. A.S.Y. Chau and R. Thomson. Analytical Reference Materials. I. Preparation and Purification of Photomirex. J. Assoc. Off. Anal. Chem. 62:1302, 1979.
14. A.S.Y. Chau, J. Carron and H.B. Lee. Analytical Reference Materials. II. Preparation and Sample Integrity of Homogeneous Fortified Wet Sediment for Polychlorinated Biphenyl Quality Control Studies. J. Assoc. Off. Anal. Off. Chem. 62:1312, 1979.
15. A.S.Y. Chau and H.B. Lee. Analytical Reference Materials. III. Preparation and Homogeniety Test of Large Quantities of Wet and Dry Sediment Reference Materials for Long Term Polychlorinated Biphenyl Quality Control Studies. J. Assoc. Off. Anal. Chem. 63:947, 1980.
16. V. Cheam and A.S.Y. Chau. Analytical Reference Materials. IV. Development and Certification of the First Great Lakes Sediment Reference Material for Arsenic, Selenium and Mercury. Analyst 109:775, 1984.

17. H.B. Lee, R.L. Hong-You and A.S.Y. Chau. Analytical Reference Materials. V. Development of a Sediment Reference Material for Chlorobenzenes and Hexachlorobutadiene. Analyst 111:81, 1986.
18. H.B. Lee, G. Dookhran and A.S.Y. Chau. Analytical Reference Materials. VI. Development and Certification of a Sediment Reference Material for Selected Polynuclear Aromatic Hydrocarbons. Analyst 112:31, 1987.
19. H.B. Lee and A.S.Y. Chau. Analytical Reference Materials. VII. Development and Certification of a New Great Lakes Sediment Reference Material for Total Polychlorinated Biphenls. Analyst 112: 37, 1987.
20. V. Cheam, K.I. Aspila and A.S.Y. Chau. Analytical Reference Materials. VIII. Development and Certification of a New Great Lakes Sediment Reference Material Eight Trace Metals. NWRI Contribution #88-71. April 1988.

Table 3a. Sediment CRM's for Toxic Organics

CRM/RM Identification	Sediment Source	Year Initiated	Parameters+
EC–1	Hamilton Harbour	1978	PCB , PAH
EC–2	Lake Ontario	1980	PCB*, PAH*, Chlorobenzenes*
EC–3	Niagara River Plume	1982	PCB*, PAH*, Chlorobenzenes*
EC–4	Toronto Harbour	1983	PCB*, PAH*
EC–5	Humber River	1983	PCB*, PAH*
EC–6	Lake Erie	1986	PCB*, PAH*
EC–7	Lake St. Clair	1987	PCB*, PAH*

+Concentration levels of organic parameters range from 0.01 to 25.0 ug/g and vary between the different CRMs & RMs.

*Certification in progress for some parameters.

Table 3b. Sediment CRM's for Inorganics

CRM/RM Identification	Sediment Source	Year Initiated	Parameters
WQB–1	Lake Ontario	1974	As, Se, Hg
WQB–2	Lake Ontario	1974	As, Se, Hg, Trace Metals*
WQB–3	Hamilton Harbour	1980	As, Se, Hg, Trace Metals
SUD–1	Sudbury	1982	Trace Metals*
TH–1	Toronto Harbour	1983	Trace Metals*
HR–1	Humber River	1983	Trace Metals*

*Certification in progress for some parameters.

Table 3c. Aqueous CRM's for Inorganics

CRM/RM Identification	Year Initiated	Initial Volume (L)	Parameters	Concentration Range (mg/L)
CM–ION–91	1981	1,000	Major Ions	0.05–13.5
CM–ION–92	1980	1,000	Major ions	0.03–106
CM–ION–93	1981	1,000	Trace Metals	0.01–1.0
CM–ION–94*	1984	1,000	SO_4	2.8
CM–ION–95	1987	1,000	Major Ions	

*Coloured Water

6.0 PREPARATION OF TEST SAMPLES

A study consisting of ten test samples and 70 laboratories means the preparation of almost 1000 individual bottles (or ampules). The primary task in this effort is to have a rapid procedure capable of producing identical aliquots so that each laboratory gets a very similar portion of the stock (see Fig. 3). For aqueous distilled water standards, organic standards, surface waters or precipitation (that have been properly prefiltered using a 0.45 µm filter), the issue of subsampling to yield homogeneous identical test samples is not critical since all constitutents are dissolved and normally are quite uniformly dispersed. Dry or wet sediments fish and unprocessed natural waters provide special problems. All efforts must be made to verify that every subsample prepared from a large stock is identical.

Large samples (1 to 10 kg of fish or sediments) and 25 to 200 litres of water are normally required to start the development of a reference material. Such a reference should yield 100 to 500 subsamples. It is wise to have a moderate excess since these test samples are required to monitor stability, homogeneity and can be used in subsequent following studies. Larger samples, 200 kg of sediments or 3000 litres of waters, should be considered if the substrate is slated for certification.

The following is a discussion on the procedures required.

6.1 Precipitation and Surface Waters (Lakes and Streams)

The collection of bulk rainwaters can be handled by large rain sampling devices (1 to 50 sq meter) be they large plastic covered greenhouse roof tops or 1 or 2 square meter buckets. In preparing large rainwater samples, it is more critical to be efficient in the collecton of a large saple (50 to 500 litres) than in being specific with respect to site selection or sampling protocols.

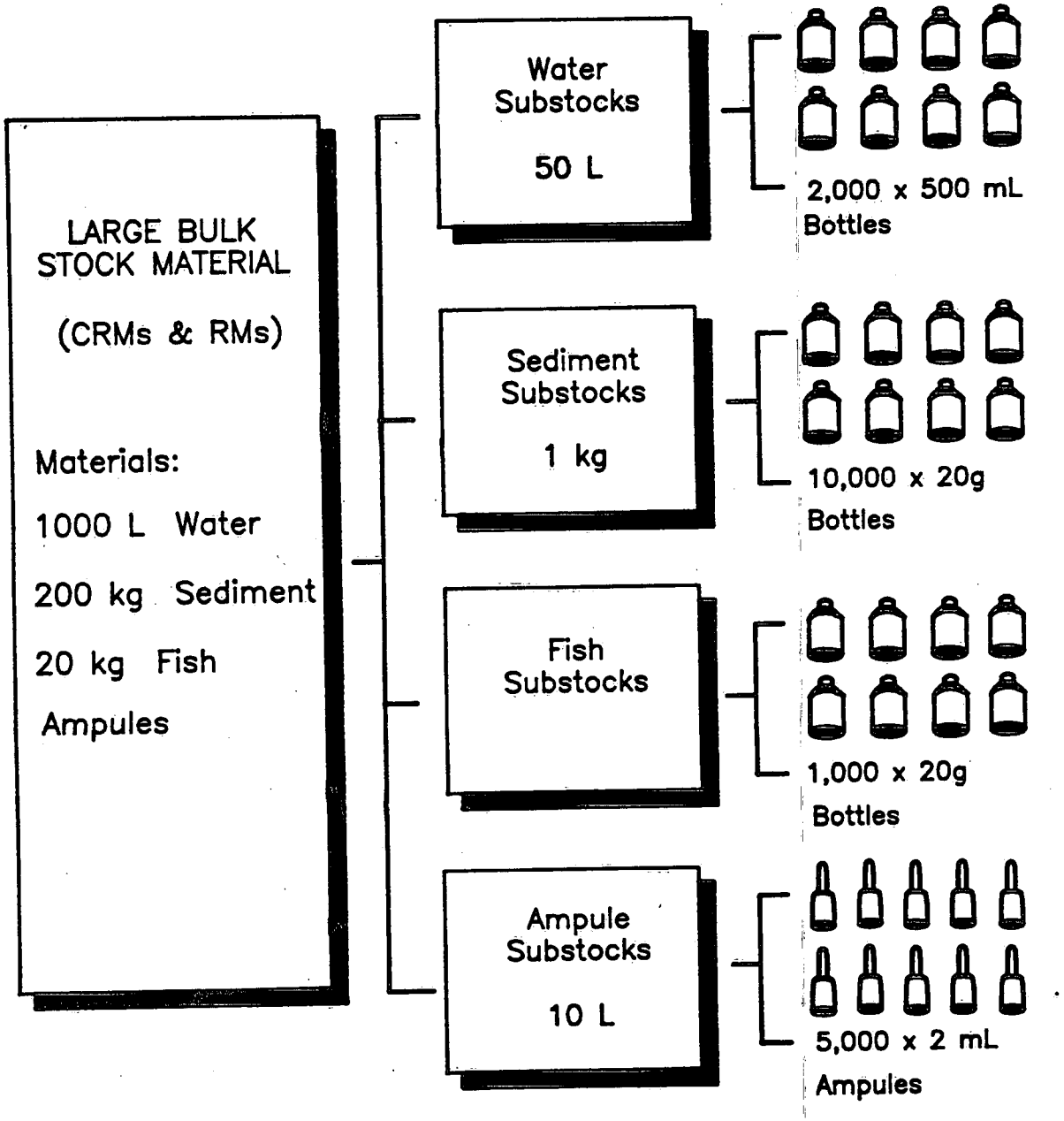


Fig. 3 Schematic Diagram for Sample Production

As with rain, lake or stream water can be collected in bulk by numerous methods, using such means as submersible pumps or plastic pails. Effort should be made to get collect the specific type of water that represents the environmental program being addressed.

Containment vessels (e.g., plastic bottles or tanks) for the storage of bulk reference mwaters can be large in size (100 to 200 litres) and in numbers. Volumes of 200 litres are suitable for use over a few years. The collection of 1000 or 2000 litre units can prove very useful when there are adequate resources and storage space available.

Once the bulk waters once acquired, a number of handling procedures are caried out. These include:

a) centrifuging the stock to remove particulate matter;
b) pasteurizing the water by autoclaving at 80 to 90°C (a temperature sensor is required to verify 80°C has been reached and maintained for about 10 minutes);
c) combining the specific aliquots of water into the large barrel or tank that will serve as the long term containment vessel;
e) establishing an analytical monitoring program to verify long term stability of the chemical constituents.

Step (b) above is not always a critical step but should be considered if prior knowledge exists that various bacteria exist, and would affect the nutrient equilibrium (ammonia to nitrate or vice versa). Experience in the National water Research Institute's various QA programs over a 15 year period has shown that many clean unpreserved lake waters may not always require pasteurizing. Some rainwaters with very high ammonia content and infested with insects may require pasteurizing. An alternative approach is to collect the water sample, centrifuge and then allow the sample to remain dormant about six months to a year. This storage time allows bacterial action to reach either completion or a steady-state equilibrium. The level of finesse on processing water prior to use relates to its intended use. Experience, knowledge and project objectives may often provide directional guidance.

Once the bulk waters are documented, stored, and characterized by internal analysis, they are then ready for subsampling. For most water studies the substock required is 40 to 50 litres (a clean 14 gallon container will suffice). There are two approaches that can be used in removing the sample, a) syphoning with a glass tube and clean plastic tubing, or b) rigorously agitating the bulk water with a high speed stirring device or with a high volume pump which pumps the water in and out of the tank or barrel. If option b) is used, the 40 L to 50 L substock can be slowly syphoned during the rigorous mixing process. The decision to use either a) or b) depends on the absence or presence of particulate matter in the water. A centrifuged water has most constituents of interest in the dissolved phase and subsampling can be handled by a simple syphon.

When the substock has been transferred and contained in a clean vessel, small aliquots are simply removed by gravity feed into clean empty bottles. This transfer is normally done while the 40 to 50 litre stock is mixed rigorously by a stirring device. It remains the choice of the QA person as to whether all test sample bottles are rinsed prior to filling.

The primary objective in subsampling is to have all test samples identical. To achieve this the analyst should always be aware of airborne contamination from such mundane sources as dandruff (Se, Zn) clothing dust (Zn), deodorants (Al) and floor dust, concrete dust, dry skin, tissue paper (Zn), etc.

Once prepared, the bottles must be labelled and appropriately stored if unstable. If the production run is large considerable care must be made to ensure that unlabelled bottles for one sample do not intermingle with unlabelled bottles of another sample.

Knowledge of sample stability is essential when producing aqueous reference samples. To ensure stability often requires preservatives. A list of some stabilizing agents is given in Table 4.

TABLE 4

Preservation of Samples (Aqueous)

Constituent	Bottle	Preservative
Trace Metals	0.50 litre polythylene	0.1 HNO_3
AS and Se	0.50 litre polythylene	0.2% H_2SO_4
Mercury	100 ml flint glass or Pyrex bottles	1% H_2SO_4 + 0.05% $_2CR_2O_7$
Silver	amber 0.5 litre polythylene	0.4% EDTA
Li, Be, Sb	0.5 litre polythylene	0.2% HNO_3
Total Phosphorus	0.1 l glass bottles	0.2 or 0.3% H_2SO_4
Orthophosphate	0.1 l glass bottles	sterilize (autoclave) store at 4°C
NTA	100 ml polyethylene bottles	1.85% formaldehyde (4°C)
Turbidity	100 ml glass bottles	store at 4°C
Major Ions, Nutrients and Physicals	0.5 litre polyethylene	store at 4°C

6.2 Preparation of Sediments

The preparation of dry reference sediments in large quantities can be a major undertaking. The following are guidelines:

- identify site location by reviewing all available environmental information;
- identify and assemble resources, such as sampling devices, vessels, crew, etc.;
- collect about 1000 kg of wet sediment (or as much as can be practically handled);
- deep-freeze the collection pails (a large freezer is required) for at least one week;
- drill a large number of small holes into each pail;
- allow each pail to thaw slowly (this can release about 40 to 80% of the water out of the sample and usually takes 3-5 days;
- air dry the sediment with large fans;
- freeze-dry the sediment (a large freeze drier is required);
- pass the dried sediment through a suitable crushing device (ball mill or roller mill);
- sieve through 200 mesh screens (higher mesh screens are preferred to enhance homogeneity of the final product);
- homogenize the stock sediment (a large sealed mixing device may be required);
- ensure homogeneity of the bulk material by analyzing various aliquots for the parameter of interest;
- bottle the sediment (an automated dispenser is essential if there are thousands of bottles.

The above relates to processing large reference sediments that have initial weights of 100 or more kilograms. Smaller samples (10 to 100 kg wet) are less difficult to handle and can be developed at the laboratory bench. The use of a programmable freeze dryer is useful to

finalize the drying since for silt or clay sediments it yields a fine powder requiring little crushing.

6.3 Preparation of Wet Sediments for Toxic Organics

Large numbers of wet sediments reference samples are more difficult to prepare than dry sediments. The reason for this is the separation of suspended sediments from the predominantly aqueous media.

When preparing wet samples some experimentation is required to adjust the percentage of water to allow a wet sediment to be mixed and maintained into a uniform and homogeneous slurry. A large blender with variable speed blades is suggested. If the sample materials cannot be maintained in a slurry state then water must either be removed or water must be added. When successful, one needs to rapidly scoop subsamples into glass vessels and do this until the stock sample is exhausted. Large commercial systems such as those used in the pharmaceutical industry are available for adding water and/or removing measured aliquots during the mixing process.

When wet or dry sediment reference samples are produced it is very important to ascertain that the bottled subsamples are all taken from a well-mixed homogeneous stock mixture.

7.0 DISTRIBUTION OF THE INTERLABORATORY STUDY

An efficient and cost effective distribution of a large intercomparison study (100 sample sets) requires a well-organized facility containing appropriate supplies.

7.1 General

Prior to any distribution of interlaboratory samples, a letter of introduction andquestionnaires should be sent to all prospective laboratories and management personnel to alert potential participants of a pending study. This initial work includes a brief description of the study and any special treatment to be applied to the samples. All efforts should be made to identify the appropriate date of distribution and a required completion date. For some studies, the delivery date may be firmly set as in cases of perishable test samples. A deadline for reporting data may also be firm.

7.2 Documentation

Documentation is essential to the interlaboratory study and very often must be prepared several days in advance of the sample distribution date. The paper work includes covering letters, instruction forms, report forms for results and for methods related comments. If the study is method specific it may include specific instructions and procedures that must be followed. For some programs, guidelines may be given on how to report low level concentrations and the need to report all calculated values.

For large studies, specific documentation for a givan laboratory may be replicated and provided with the box of test samples. In such cases, one set of documents is sent to the laboratory manager, one to the analyst, and other copies may be sent to appropriate and responsible program managers.

When test samples are distributed internationally, it is essential to prepare documentation and packaging suitable for import and export regulations.

Verification of the receipt of the shipment of test samples in good condition, will require the enclosure of a confirming letter and report form with a self-addressed envelope.

7.3 Packaging

The packaging of test samples must be secure and be able to withstand considerable abuse. A simple criteria to decide if the package is adequately secure is to design the package to allow it to be thrown several meters, bounced off walls or rolled down several flights of stairs. If it withstands these shocks it will withstand travel by most carriers. For toxic samples all packages must comply with Workplace Hazardous Materials Implementation System (WHMIS) and Transport of Dangerous Goods regulations.

7.3.1 Water samples

As a guideline, glass bottles should be packaged in blocks of styrofoam where holes have been cut out to contain individual glass bottles. The blocks of stryrofoam should be sealed in plastic. The plastic bag (heat sealed closed) will contain the liquid should the glass bottle break. The blocks of styrofoam should then be placed into a larger box and encased on all sides with 2 to 5 cm of loose fitting styrofoam chips. The chips serve as a buffer if the cardboard box is punctured.

Plastic bottles containing water can be packaged as above. All containers should be sealed tightly and verified as "leakproof". Secure air-tight plastic bags (heat sealed) are essential for containing any liquid due to closures that are accidently broken.

As a precaution, any package heat-sealed air-tight with a plastic bag should have a small puncture to release air and prevent the

package from exploding and disintegrating the box when "air transport" and depressurizaton are anticipated. Some cargo planes pressurize only to about 8000 feet altitude.

The shipment of water samples in the winter when heated carrier service is unavailable, may require special treatment. For instance, half full glass bottles may not break but the chemistry may change on freezing. These matters may require review in the design and development of the study. Likewise, in the heat of summer, delivery truck temperatures that may exceed 45°C is an issue that must be recognized and verified as not influencing tests samples. It is imperative that the QA chemist must consult the carrier on all aspects of the anticipated shipping environment to ensure the continual integrity of test samples.

7.3.2 Fish or wildlife reference samples

Whole fish homogenates must be shipped in the frozen state and must remain frozen during transit to reduce sample degradation. Large freezer boxes can be prepared from foam slabs (5 to 8 cm thick). The custom made inserts can be used to line large cardboard boxes on all sides. The fish bottles (25 to 100 ml bottles) need to be secured in boxes and placed in the large freezer box. A considerable mass of freezer bags are normally required to retain heat and to keep the small mass of fish frozen. It is advisable that a test run of a constructed box be made where the temperature of the frozen contents are monitored over 72 hours to verify the box used is adequately insulated and can prevent the product from thawing. A thin thermocouple and recording thermometer (-40 to +10°C) is required.

An alternative to home-made insulated boxes are the commercially available boxes. They are, however, much more expensive, and will require tests to verify that samples remain frozen over the anticipated travel time.

Coordination of the shipment of frozen biological tissue is essential. These studies must move on a specific date to be in total

transit time no more than three days and the recipient must be alerted in advance. Awareness by the recipient that the box is being sent is essenital since the box must not be allowed to remain unrefrigerated on receipt. These boxes also require special labelling as to "perishable goods" and should have the phone numbers of specific employees visibly recorded on the box, so that receivers in the shipping and receiving section can alert the laboratory to take special action.

7.4 Transport Procedures

The shipment of test samples (small boxes) may be by mail, commercial couriers or sometimes by personal delivery. The decision is normally a balance between cost and time. The mail service is the least expensive, but may not be the most appropriate since there are restrictions on the mailing of some chemicals.

Private couriers are by far the most preferable if resources are available. A single box (1 kg) may cost $5 for local service. Transcontinental service may exceed $200 per box for the larger boxes of fish homogenates. A single study distributed to 50 to 100 laboratories may have a shipping bill of $2000 to $6000. The fee structure is determined by weight and destination and whether ground transport or air transport is required. Perishables should always be sent by air and the shipment should normally be moved on a Monday or Tuesday to guarantee same week delivery.

Logistics involved with the transportation of test samples should be discussed with carrier service representatives, whether it is the postal service or private couriers. There are restrictions[1,2,3,4], and for some carriers very essential paperwork is required. Close consultation is indeed required since the shipper or agent is ultimately responsible for any damage. For example, inappropriate labelling such as "natural tap water - for cyanide analysis" causes undue concern and is very misleading. A package containing a steel cylinder encasing a

teflon insert should not be identified as a "Teflon Bomb". This label will unnecessarily bring in the bomb squad.

Transport of toxic chemicals even at trace concentrations requires review with all the appropriate authorities[4].

7.5 Storage and Inventory Control

The distribution of any study (e.g. 10 different samples to over 70 laboratories) will deplete a portion of the stock supply of reference materials and the sample supply prepared for the study. Control on the consumption of the supply of remaining materials must be exercised through cataloguing and the maintenance of inventories.

A study in progress very often requires spare samples as replacement for damaged goods or for special follow up by laboratories requesting assistance to verify problem areas that may be corrected. The development, maintenance and inventory of extra sets of samples are wise and precautionary measures.

Chapter 11 addresses a data base management system that allows for a routine administration of inventory of stock materials as well as analytical data.

7.6 References

1. Amendment to Canada Labour Code. Statutes of Canada. 1987, Chapter 30, Part II, Canadian Government Publishing Centre, Supply and Services Canada, Ottawa K1A 0S9.
2. Control Products Regulations. SOR/88-64, Canada Gazette, Part II.
3. Hazardous Products Act. Statutes of Canada 1987, Chapter 30, Part I.
4. International Civil Aviation Organization. Technical Instructions for the Safe Transport of Dangerous Goods by Air, 1983 edition. Published in Canada by ICAD, P.O. Box 400, Place de l'aviation internationale, 1000 Sherbrooke Street West, Montreal, Quebec, H3A 2R2.

8.0 THE NATIONAL WATER RESEARCH INSTITUTE AQC DATA BASE MANAGEMENT SYSTEM

Over the course of 15 years of external QA development (IJC, National, and LRTAP), K.I. Aspila of the Quality Assurance group at NWRI, with the programming support of Karon Miles (Computing and Programming Services Section, NWRI), developed a series of software programs to address QA data and issues on the mainframe computer. This software has proved suitable for addressing large studies (40 labs, 10 samples and 20 constituents). The highlight program (Youdn21) was made[1,2,3] possible through the development work of Dr. John Clark and Mr. R. White of the International Joint Commission's Great Lakes Regional Office. Through their insight, the nonparametric techniques of Youden[2,5] for bias assessment, were developed and computerized to allow the assessment of bias in large data arrays (100 labs, 100 samples).

As the QA programs at NWRI increased, it became apparent that information between programs and different studies would be required and indeed were essential. The original software produced single outputs (flat files) for studies in isolation. To improve the retrieval of study-to-study data and make long term projections more efficient, Karon Miles developed the essential data base structure on System 2000 (a commercial database software package). The following chapter describes many evaluation procedures now recognized essential for large external QA studies.

8.1 AQCPROC (QA Procedure File)

AQCPROC is the acronym for "analytical quality control procedures". In brief, it contains the Network Operating System (NOS) commands that are needed to execute many different software packages maintained on the NWRI mainframe computer (Cyber 180/310A). Several of the essential computer programs used for handling larger studies for the IJC, Ocean Dumping, Eulerian, LRTAP terrestrial and LRTAP aquatic QA programs are described in the following sections. Example outputs are presented.

In the early phase of software development, the AQCPROC had flat file outputs. The output results of computer programs such as RAWDAT1, MEANPRT and YOUDN21 were produced in isolation. They have since been extended with System 2000 retrievals to produce information across different studies using programs such as Median2, and FLGTBL. These programs have proved very useful are highlighted in the following sections.

None of these programs are possible without a dictionary to translate parameter codes into memories. A small portion of the dictionary is given in Table 5. It is structured to have elements listed in ascending atomic numbers.

8.2 RAWDAT 1 (Raw Data Summary)

This program is the first and essential program that is initiated after the results have been entered. An example of output is shown in Table 6. The primary role of RAWDAT 1 is to search for errors in the data files. No output is achieved until all laboratory and parameter codes, method summary and data values are correctly in place. Once verified, the output (Table 6) is produced. This output can then be proof-read for transcription errors. A nominal study of 60 laboratories yields 60 pages of data (one page for each laboratory).

Future plans to effect data entry with the use of electronic scanners or electronic mail systems may reduce input errors. Scanners which are presently available will be able to read pages of data directly into electronic files located in personal computers and those data can then be downloaded into the mainframe data base. Other options available to to streamlining data entry make use of floppy diskettes created by each laboratory for reading date directly into a personal computer and then into the mainframe computer. These methods are currently under review.

Table 5: A Portion of the AQCPROC Dictionary

CODE	PARAMETER	SHORTFORM	UNITS
07092	NITRATE + NITRITE	$NO_3 + NO_2$	mg N/L
07102	ORGANIC NITROGEN	ORGAN N	% N
07192	AMMONIA	NH3	mg N/L
07202	NITRATE	NO_3	ug N/G
07292	TOTAL NITROGEN UV	TOT N UV	mg N/L
07392	TOTAL KJELDAHL NITROGEN	TKN	mg N/L
07492	NITRATE-IC	NO_3	mg N/L
07592	NITRATE-NON IC	NO_3	mg N/L
08002	OXYGEN	O	% O
09002	FLORINE	F	ug F/G
09092	FLUORIDE	F	mg/L
1100P	SODIUM IN PLANTS	Na(ODW)	ug/G(ODW)
11001	SODIUM	Na	% Na
11091	SODIUM	Na	mg/L
11092	SODIUM IN SEDIMENTS	Na	ug/G
1200P	MAGNESIUM IN PLANTS	Mg(ODW)	mg Mg/G(ODW)
12001	MAGNESIUM	Mg	% Mg
12091	MAGNESIUM	Mg	mg/L
12092	MAGNESIUM IN SEDIMENTS	Mg	ug/G
1300P	ALUMINIUM IN PLANTS	Al(ODW)	ug/G(ODW)
13001	ALUMINIUM	Al	% Al

Table 6: RAWDAT1 - Raw Data Summary (for a laboratory file)

STUDY: INTERLABORATORY STUDY NO. 20 MAJOR IONS IN WATER
PRINTOUT PREPARED: 89/04/12.
LAB: L004

PARAMETER CODE		METHOD	SAMPLE NUMBER 1	2	3	4	5	6	7	8	9	10
COLOUR	HAZEN UNIT	02022	3	31	3	31	0	0	0	60	3	6
SPEC COND	UMHO/CM	02041	10.4	17.9	60.0	67.2	12.9	10.4	8.9	22.4	34.2	52.8
PH UNITS	PH UNITS	10301	5.2	6.7	4.3	6.4	5.0	5.5	5.1	5.8	6.4	7.2
DOC	MG C/L	06104	0.4	3.0	1.1	3.6	0.2	0.3	0.2	4.7	1.8	2.4
ALKA-FE. PH4.5	MG/L CACO3	10101	T	3.0	T	2.2	T	T	T	0.7	2.6	12.4
GRAN ALK	MG/LCAC03	10110	0.08	3.35	-2.59	2.48	-0.48	0.08	-0.28	0.82	2.88	12.89
DIC	MG C/L	06180	T	0.4	T	0.2	T	T	T	0.3	0.4	2.2
NO3 + NO2	MG N/L	07112	0.12	0.13	1.75	0.27	0.08	0.26	0.18	0.04	0.15	0.11
NH3	MG N/L	07555	0.073	0.007	0.250	0.004	0.050	0.208	0.029	0.057	0.018	0.029
NA	MG/L	11105	0.67	1.08	0.16	5.7	0.79	0.13	T	0.49	0.94	0.91
MG	MG/L	12102	0.07	0.31	0.92	0.71	0.05	0.07	0.15	0.45	0.69	1.56
REAC SI	MG SI/L	14105	0.05	1.94	0.22	2.54	0.02	0.03	0.02	0.51	0.77	1.04
SO4 (NON-IC)	MG/L	16306	0.9	3.8	7.0	9.2	1.4	2.2	1.9	7.8	8.2	91
CL (NON-IC)	MG/L	17203	1.22	1.34	4.30	11.1	1.39	0.27	0.09	0.64	1.34	1.03
K	MG/L	19102	0.08	0.14	0.14	0.73	T	T	T	0.28	0.48	0.67
CA	MG/L	20103	0.18	1.64	3.99	4.48	0.43	0.72	0.52	2.12	3.17	5.8

Table 7: MEANPRT – Data Summary for one Parameter

LRTAP STUDY NO. 20 MAJOR IONS IN WATER
PRINTOUT PREPARED: 89/03/14.
PARAMETER: CHLORIDE NON IC METHODS MG/L

	SAMPLE RESULTS									
	1	2	3	4	5	6	7	8	9	10
LAB										
L002	1.2	1.4	4.4	10.9	1.4	.3			1.3	1.0
L004	1.22	1.34	4.30	11.1	1.39	.27	.09	.64	1.34	1.03
L006	1.26	1.34	4.34	11.30	1.41	.316	.088	.762	1.35	1.06
L010	1.1	1.2	3.8	10.7	1.2 T	.5 T	.5	.7	1.3	1.0
L023	1.20	1.40	4.30	10.	1.40	.35 W	.10	.80	1.30	1.10
L024	1.17	1.06	4.22	11.00	1.35	.26	.08	.33	1.21	.89
L025	1.0	1.0	3.80	11.0	1.3 T	.2 W	.1 T	.6	1.2 T	.9
L045	1.19	1.22	4.18	10.9	1.48	.30	.10	.64	1.32	.95
L049	1.2	1.2	4.0	11.1	1.7	.5	.2	.5	1.2	.9
L063	1.10	1.20	4.20	11.3	1.27	.28	.13	.67	1.20	.90
L067	1.0	1.1	3.9	11.0	1.2	.3 T	.1	.6	1.1	.8
L085	1.09	1.02	4.01	11.5	1.27	.27	.07	.37	1.18	.89
L086	1.0	1.2	3.8	10.5	1.6	.4 W	.2	.6	1.2	.9
L089	1.11	1.16	4.30	10.7	1.30	.30	.06	.33	1.17	.91
TOTAL LABS REPORTING	14	14	14	14	14	14	14	14	14	14
TOTAL LABS USED	14	14	14	14	14	14	10	13	14	14
MEAN	1.13143	1.20286	4.11071	10.92857	1.37643	.32471	.14180	.58015	1.24071	.94500
STD DEV	.08743	.13100	.21925	.37505	.14281	.08683	.13190	.15490	.07580	.08178
MEDIAN	1.14000	1.20000	4.19000	11.00000	1.37000	.30000	.09500	.60000	1.20500	.90500

8.3 MEANPRT (Summary of Data)

The MEANPRT program provides the type of output shown in Table 7. Its primary use is to overview raw data in a convenient summary format during the course of a study, and perhaps more importantly, to scan the raw data for blunders (wrong units, ppb vs ppm) due either to input errors or transcription errors by the analyst. The combination of output from RAWDAT1 and MEANPRT facilitates for a smooth search for errors prior to finalizing a report to analysts and updating results in the database.

8.4 YOUDN21 (Summary of Bias and Flags)

This procedure program though simple to initiate as a batch job is by far the most sophisticated in construction. An example of the output is given in Table 8. For a 20 parameter study involving 50 laboratories, the output can approach 80 to 100 pages. The example in Table 8 is for but one parameter for a small group of laboratories.

The key elements in the YOUDN21 output are - (a) discerning whether or not a laboratory data set is biased, and (b) whether an individual result reported by a laboratory is sufficiently different from the median to warrant the flagging of tt individual result. The following is a brief overview on bias and flagging.

8.4.1 Ranking to Discern Bias

The Youden bias assessment technique is a non-parametric process in which a matrix of results (for example, 10 samples - 50 laboratories) are converted into a matrix of ranks. Each sample (with say 50 results) are ranked such that the lowest result has a rank 1 assigned, the second lowest is a rank 2 and so on. The highest results is rank 50 if there are 50 laboratories. When laboratories report equal values then the rank assigned is an average. Examples are provided by Youden (Refer to suggested reading - Chapter 15 and Tables 8, 9 and 10.

Table 8: YOUDN21 Output

PARAMETER: 16001 SULFATE NON IC METHODS mg/L

LRIAP STUDY NO. 20 MAJOR IONS IN WATER

LOWER LIMIT FOR USE OF BASIC ACCEPTABLE ERROR= 2.00 BASIC ACCEPTABLE ERROR= .50 CONCENTRATION ERROR INCREMENT= .2000
LABORATORIES YET TO REPORT: L014C,L009,L022,L052,L098
LABORATORY RESULTS OMITTED ARE NONE

SAMPLE	1		2		3		4		5		6	
LAB NO	REPORTED VALUE	RANK	REPORTED VALUE	RANK	REPORTED VALUE	RANK	REPORTED VALUE	RANK	REPORTED VALUE	RANK	REPORTED VALUE	RANK
L002	.9	3.50		.00		.00		.00	1.8	9.50	1.8	2.50
L003	1.0	9.00	2.1	4.00	6.5	3.50	7.7	3.00	1.6	7.00	1.95	6.00
L004	.9	3.50	3.8 EH	9.00	7.0	7.00	9.2	8.00	1.4	4.00	2.2	10.00
L006	.984	7.00	2.22	6.00	6.84	6.00	7.72	4.00	1.62	8.00	1.93	5.00
L023	.95	5.00	2.85 H	8.00	6.45	2.00	7.55	2.00	1.50	5.00	2.00	8.00
L049	1.0	9.00	1.5 L	1.00	4.8 L	1.00	13.5 EH	9.00	1.3	2.00	1.5	1.00
L063	.983	6.00	1.920	3.00	6.647	5.00	8.380	6.00	1.323	3.00	1.830	4.00
L067	1.0	9.00	2.3	7.00	7.2	8.50	7.3	1.00	1.8	9.50	2.0	8.00
L085	.852	2.00	1.68	2.00	7.2	8.50	7.74	5.00	1.59	6.00	1.80	2.50
L086	.8	1.00	2.2	5.00	6.5	3.50	8.5	7.00	1.2	1.00	2.0	8.00
MEDIAN CONC.	.967		2.200		6.647		7.740		1.545		1.940	

SAMPLE	7		8		9		10	
LAB NO	REPORTED VALUE	RANK	REPORTED VALUE	RANK	REPORTED VALUE	RANK	REPORTED VALUE	RANK
L002	1.7	3.50	6.6	7.00	7.9	5.00	8.5	3.50
L003	1.7	3.50	6.1	4.00	8.05	6.00	8.6	5.50
L004	1.9	8.50	7.8	10.00	8.2	8.50	91. EH	10.00
L006	1.73	5.00	6.76	8.00	8.30	10.00	8.94	9.00
L023	1.80	7.00	6.45	6.00	7.80	4.00	8.50	3.50
L049	1.5	1.00	4.0 VL	1.00	5.2 VL	1.00	7.0	1.00
L063	1.733	6.00	6.050	3.00	7.600	2.00	8.320	2.00
L067	2.0	10.00	6.4	5.00	8.2	8.50	8.8	8.00
L085	1.62	2.00	5.94	2.00	8.13	7.00	8.64	7.00
L086	1.9	8.50	6.8	9.00	7.7	3.00	8.6	5.50
MEDIAN CONC.	1.732		6.425		7.975		8.600	

(Continued on next Page)

Continued
Table 8: YOUDN21 Output

(IN ORDER OF LAB CODE)

LAB NO.	TOTAL RANK	AVERAGE RANK	NO.OF SAMPLES RANKED	SUMMARY OF FLAGGING			METHOD CODING
L002	34.50	4.929	7				
L003	51.50	5.150	10				16304
L004	78.50	7.850	10	EH	EH	BIASED HIGH	16306
L006	68.00	6.800	10				TECHNICON
L023	50.50	5.050	10	H			METHYLTHYMOL
L049	27.00	2.700	10	L L EH	VLVL	BIASED LOW	COLORMETRIC
L063	40.00	4.000	10				TURBIDOMETRIC
L067	74.50	7.450	10				AUTO MIB
L085	44.00	4.400	10				ICPS
L086	51.50	5.150	10				AA-MIB
OVERALL AVERAGE RANK IS		5.361					

(IN ORDER OF TOTAL RANK)

LAB NO.	TOTAL RANK	AVERAGE RANK	NO.OF SAMPLES RANKED	SUMMARY OF FLAGGING			METHOD CODING
L049	27.00	2.700	10	LLEHVLVL		BIASED LOW	COLORMETRIC
L063	40.00	4.000	10				TURBIDOMETRIC
L085	44.00	4.400	10				ICPS
L002	34.50	4.929	7				
L023	50.50	5.050	10	H			METHYLTHYMOL
L003	51.50	5.150	10				16304
L086	51.50	5.150	10				AA-MIB
L006	68.00	6.800	10				TECHNICON
L067	74.50	7.450	10				AUTO MIB
L004	78.50	7.850	10	EHEH		BIASED HIGH	16306
OVERALL AVERAGE RANK IS		5.361					

SULFATE NON IC METHODS

Table 9. A Typical Interlab Study Design

Lab Code	SAMPLE NUMBER 1 2 3 . . . . . 9 10 . . . . n
A B C D . . K . m	e.g. 10 samples 56 laboratories 23 different constituents concentrations range from 1 unit to 100 units

Table 10. An example of Ranking Results to Discern Bias

Lab Code	Rank on Sample Results 1 2 3 . . . . 8 9 10	Total Rank	Average Rank
A	12 16 41 20 26 28	250	25
B	43 58 49 45 57 59	550	55
C	3 2 1 5 4 2	20	2
.	. . . . . .	.	.
.	. . . . . .	.	.
M	23 22 20 21 24 25	240	24

Note: Lowest result (on a sample) = rank 1
Second lowest result = rank 2
Highest result is highest rank = no. of labs

The next issue in the ranking process is to review the total laboratory rank (sum the ranks) or the average rank. The immediate impact is recognition that some laboratories have a unique ability to rank very high or very low. The question to resolve is whether these anomolous high or low ranks are rare events (less than 5% chance of occurring). To evaluate if bias exists, one needs to use a traditional hypothesis test. First it is assumed that no bias exists. The next step is to calculate the probability of total ranks from the matrix that is composed of ranks (e.g., 10 by 50). This calculation (found in gambling handbooks) is synonomous to calculating the probability of scores when 10 dice (samples) are thrown and each dice has 50 sides (50 labs). The probabilities of interest are the very high and very low scores. When extreme scores (very high or low ranks) are found in the matrix of ranks with occurrence probabilities of less than 5% of the time, then the null hypothesis is rejected and the laboratory data set is declared as biased. The risk in declaring a laboratory biased when it is not, is one chance in 20 (5%).

A description of this process is given in Tables 8 to 10. The example is derived from a LRTAP Study (Chapter 14, see List of Studies). Youden's original work[3] describes total ranks for which a matrix of critical ranks were calculated manually and found in the literature (Chapter 15). The probability calculations described above and developed by Clark are parallel to those of Youden. Both methods provide very informative statements when appraising interlaboratory results.

Non-parametric techniques are powerful procedures for discerning small systematic errors in calibrations. In some cases the decision is valid but is so slight that some laboratories are unable to react and adjust their calibration to remove the slight difference between their standards and the error implied from the interlaboratory study evaluation. Some laboratories which achieve considerable precision and a statistical control in determinations associated with blanks and secondary blanks have had the ability to adjust calibration 2 or 3% using backup verification from standard refrence materials (e.g.,

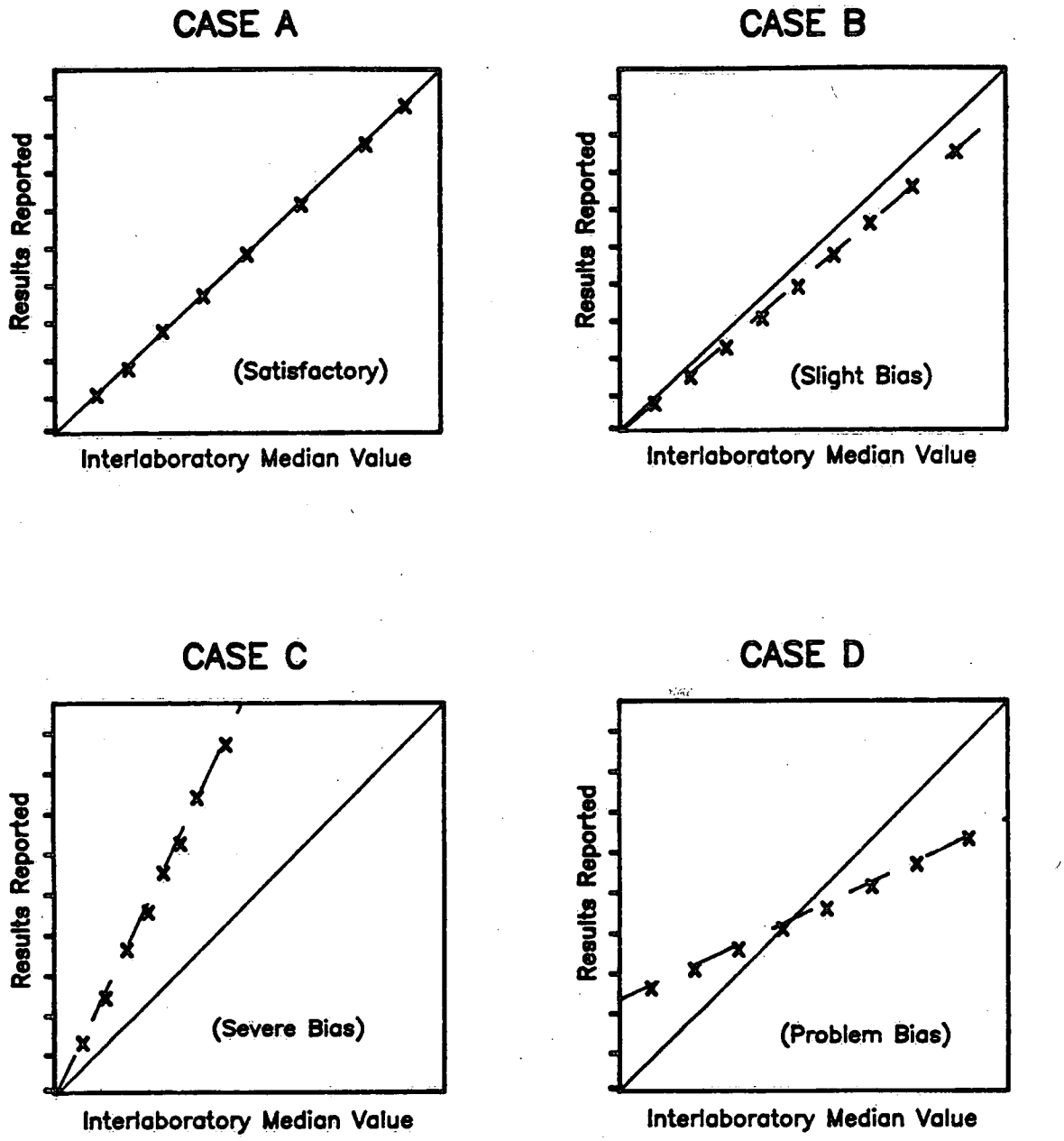


Fig. 4 Examples of Biased Data

for NO_3 and SO_4). Laboratories with gross bias (10 to 30% error) are sometimes so severe that bias assessment by Youden's ranking method need not be applied since a graphical format or simple review of the matrix of results is visually adequate (see Fig. 4).

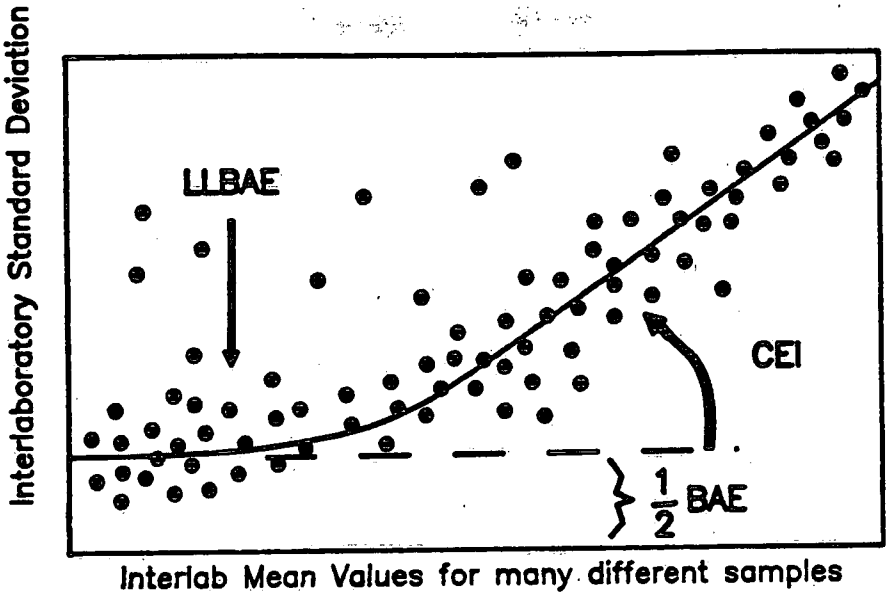
Assessment of bias by the Youden method is not always informative. A typical set of results in Fig. 4 serves as an example for discussion.The data set in this example has a blank problem and the calibration is also biased. To correct this shortcoming, the performance of a laboratory needs to be complemented by a flagging system or by an inference using graphics as a visual aid.

8.4.2 Flagging Results

The Youden bias assessment in many large studies can successfully address and discern the presence of inaccuracy in the laboratory measurement process. The rigour in which this method identifies inaccuracies is of course clouded when serious blank issues or if the entire group of laboratories are all in error. The entire group of laboratories being wrong is itself a rare event (for large studies) but vigilance and review must be maintained when difficult substrates and constituents are under review (e.g., toxic organics in fish or sediments).

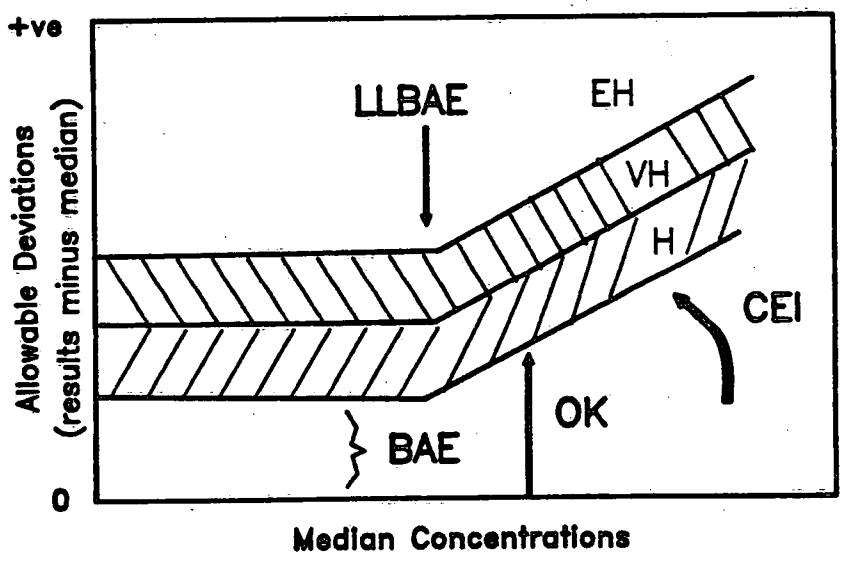
To complement bias assessment, large or small studies can use a flagging procedure that identifies a laboratory result as very high or very low. The flagging process and the bias assessment are two different and separate evaluation procedures. Flagging is critical since some laboratories are imprecise and as such the degree of biasness cannot be easily determined since there are on average very high and very low results. Fig. 4 provides as examples.

A formula to flag individual results on a sample within a study has been developed for many traditional constituents. Experience has shown that within any study covering a concentration range of 1 or 2 orders of magnitude, the interlaboratory standard deviation varies and



BAE = Basic Acceptable Error
LLBAE = Lower Limit for use of BAE
CEI = Concentration Error Increment

Fig. 5 A Typical Precision Function



NOTE: Flags assigned are EH,VH,H

Fig. 6 An Example of Flagging Criteria

increases almost linearly fom low concentration to high concentration (see Fig. 5).

The above figure (see Fig. 5) allows for construction of a simple formula for flagging. Three variables are required to decide if a result reported deviates sufficiently from an interlab median to warrant a flag (high or low). The first is the basic acceptable error (BAE) and this is the allowable deviation fixed over all concentrations. The second is the lower limit for use of basic acceptable error (LLBAE). This lower limit is the concentration at which the acceptable deviation (result reported minus the median) begins to increase. The rate of increase,similar to the slope of the precision function (Fig. 5) is referred to as the concentration error increment (CEI). These three variables are given in the schematic (Fig. 6).

The relationship between the observed precision function and the flagging formula is quite close. The principle issue to resolve are the values assigned to the BAE, LLBAE and CEI. Some trial and error may be required if the information on the correct precision function is unknown. The median is chosen as a target since medians are more robust than the average values. The average or mean values are often influenced by extreme results. Criteria chosen can be adjusted so that some (10 to 30%) of all results reported are flagged either H (high) or L (low). When results are very different they can be flagged VH (very high) or VL (very low). These extremes can be arbitrarily recognized when the results reported are more than 1-1/2 times the acceptable deviation. A third flag (EL or EH) extremely low or high is assigned if the deviation is more than two times the acceptable deviation.

8.5 System 2000 - Data Base Management System

The three procedure programs (RAWDAT1, MEANPRT and YOUDN21) serve well in providing a rapid data assessment for large or small studies. They have been and will continue to be utilized for large NWRI studies (error search). This flat file processing is, however, unable

AQC DATA BASE (SYSTEM 2000)

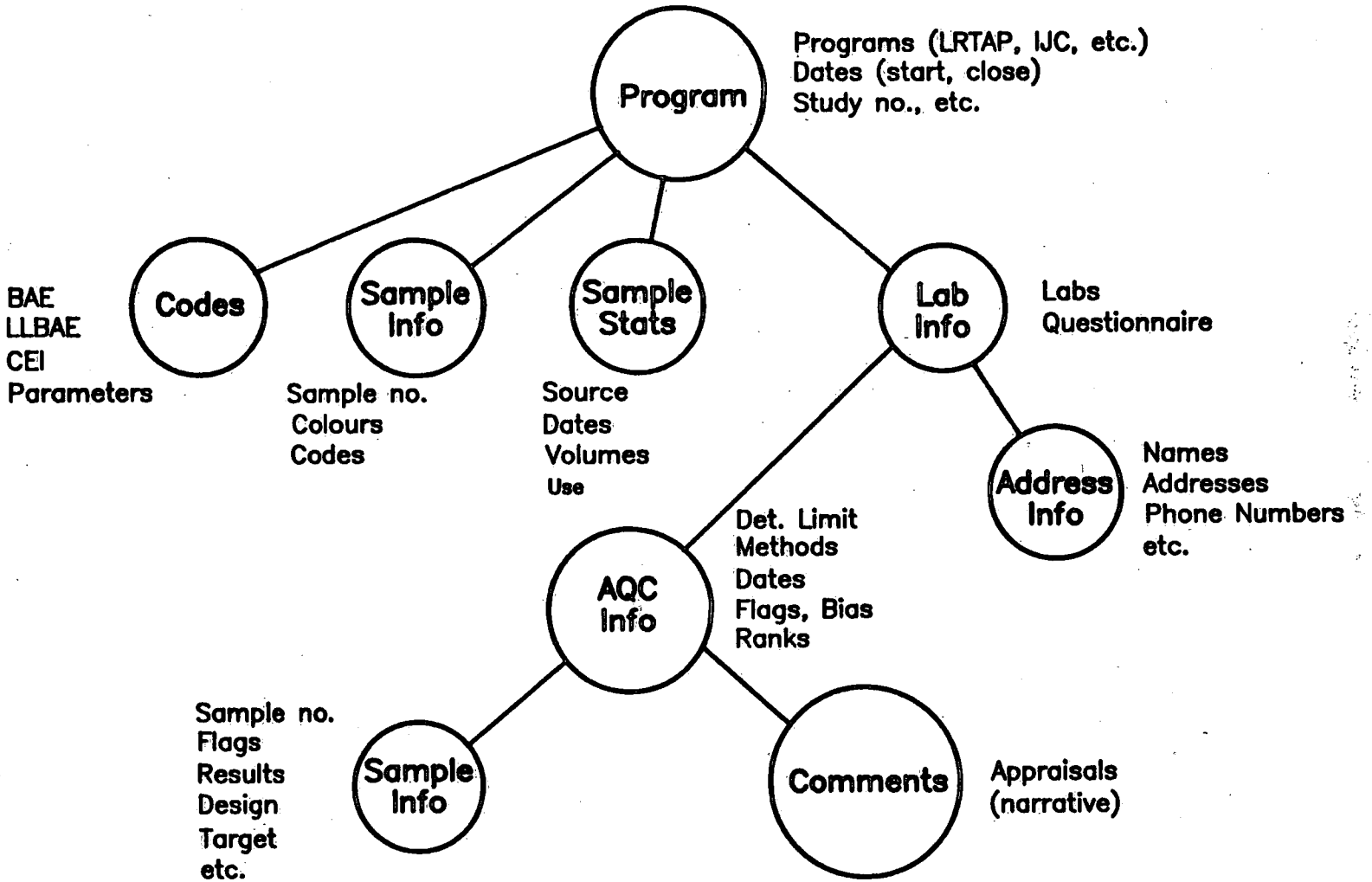


Fig. 7 AQC Data Base System

Table 11: Data Base Definition (System 2000)

```
1*  PROGRAM (CHAR X(10));
2*  STUDY (NON-KEY CHAR X(40));
3*  STUDY CODE (CHAR XXXX);
4*  STARTDATE (NON-KEY DATE);
5*  STOPDATE (NON-KEY DATE);
100*  CODE INFO (RECORD);
101*  CODEA (NON-KEY CHAR X(5) IN 100);
102*  BAE (NON-KEY DECIMAL NUMBER 999.999 IN 100);
103*  CEI (NON-KEY DECIMAL NUMBER 999.999 IN 100);
104*  LLBAE (NON-KEY DECIMAL NUMBER 999.999 IN 100);
105*  AVER RANK1 (NON-KEY DECIMAL NUMBER 9999.999 IN 100);
200*  SAMPLE INVEN (RECORD);
201*  SAMPLE NO. (NON-KEY INTEGER NUMBER 99 IN 200);
202*  QAMS CODE (NON-KEY CHAR X(12) IN 200);
203*  SAMPLE COLOUR ID (NON-KEY CHAR X(10) IN 200);
250*  SAMPLE STATS (RECORD);
251*  QCODE (NON-KEY CHAR X(12) IN 250);
252*  SOURCE (NON-KEY CHAR X(50) IN 250);
253*  DATE OF PREPARATION (NON-KEY DATE IN 250);
254*  INITIAL VOLUME (NON-KEY DECIMAL NUMBER 9999.99 IN 250);
255*  VOLUME ON HAND (NON-KEY DECIMAL NUMBER 9999.99 IN 250);
256*  UNITS (NON-KEY CHAR XXX IN 250);
300*  LAB INFO (RECORD);
301*  LAB CODE (CHAR X(5) IN 300);
302*  DATE QUESTIONNAIRE (NON-KEY DATE IN 300);
303*  TAKING PART (NON-KEY CHAR X IN 300);
400*  AQC INFO (RECORD IN 300);
401*  PARM CODE (CHAR X(5) IN 400);
402*  DETECTION STATED (NON-KEY CHAR X(10) IN 400);
403*  METHOD (NON-KEY CHAR X(18) IN 400);
404*  DATE OF ENTRY (NON-KEY DATE IN 400);
405*  OTHER INFO (NON-KEY CHAR X(18) IN 400);
406*  BIAS FLAG (NON-KEY TEXT X(42) IN 400);
407*  BIAS STATEMENT (CHAR XXXX IN 400);
408*  TOTAL RANK (NON-KEY DECIMAL NUMBER 999.99 IN 400);
409*  NO. OF SAMPLES RANK (NON-KEY INTEGER NUMBER 99 IN 400);
500*  SAMPLE INFO (RECORD IN 400);
501*  SAMPLE NUMBER (NON-KEY INTEGER NUMBER 99 IN 500);
502*  FLAG (NON-KEY CHAR X IN 500);
503*  VALUE (NON-KEY CHAR X(8) IN 500);
504*  NO. OF DECIMALS (NON-KEY INTEGER NUMBER 9 IN 500);
505*  REPLICATE SAMPLE FLAG (NON-KEY INTEGER NUMBER 99 IN 500);
506*  DESIGNED VALUE (NON-KEY CHAR X(7) IN 500);
507*  MEDIAN (NON-KEY CHAR X(9) IN 500);
600*  COMMENTS (RECORD IN 400);
601*  LINE NUMBER (NON-KEY CHAR XX IN 600);
602*  NARRATIVE (NON-KEY CHAR X(58) IN 600);
700*  ADDRESS INFO (RECORD IN 300);
701*  ADDRESS NUMBER (CHAR XX IN 700);
702*  ADDRESS TYPE (CHAR XXXX IN 700);
703*  NAME (NON-KEY CHAR X(36) IN 700);
704*  TITLE (NON-KEY CHAR X(36) IN 700);
705*  AFFILIATION1 (NON-KEY CHAR X(36) IN 700);
706*  AFFILIATION2 (NON-KEY CHAR X(36) IN 700);
707*  ADDRESS (NON-KEY CHAR X(36) IN 700);
708*  CITY PROVINCE POST CODE (NON-KEY CHAR X(36) IN 700);
709*  COUNTRY (NON-KEY CHAR X(20) IN 700);
710*  PHONE NO. (NON-KEY CHAR X(14) IN 700);
711*  AFFILKEY (CHAR X(10) IN 700);
800*  SUBCON (RECORD IN 700);
801*  SUBSTRATE CONSTITUENT (NON-KEY CHAR X(10) IN 800);
```

to provide rapid electronic study to study information. As studies developed and the QA programs at NWRI escalated, Karon Miles (Computing and Programming Services Section, NWRI) provided the initiative in developing the S2K data base management system. This system was designed to allow a search for almost every conceivable piece of information entered as data and much of the calculated information generated by the very successful Youdn21 program. It was recognized early in the systems development that one could create files to (a) confirm swiftly the stability of reference samples (RMs and CRMs), (b) seek specific information to identify laboratory performance (bias and flags) over all studies, (c) obtain hard evidence very swiftly on improvements or change in laboratory performance or (d) make accessible all data or calculated data entered for any lab on any study over all programs.

The structure of the database is graphically given in Fig. 7 and in narrative format in Table 11. The components on lab specific information (names, addresses and phone numbers, etc.) are included to create a fully automated report writing system where covering and informational letters (sometimes as many as 200 letters) need to be swiftly generated and in which appraisals and a vast amount of support data (hundreds of pages per lab) are handled. This part of the system relating to merging, letters, names, and appraisals and suppoort data is still under development.

The AQC data management is now a very useful system with extensive use in the NWRI QA programs (LRTAP, LRTAPP, Eulerian, Ocean Dumping, and for some elements of the National and and International Joint Commission QA programs). The data base is populated when RAWDAT 1 creates the data format, a program PLSUPDA makes the program, study and laboratory information available and the program DATAUP updates the data on to the S2K data base. All pertinent data can then be accessed and retrieved. Various programs are outlined in the following sections to illustrate some key outputs.

The inherent value of the AQC data base is realized in administratively controlling the vast amount of data generated, be it on

large RMs that are transient (2 to 5 studies) or in large RMs set for certification (10 to 100 studies). The data base currently has about 1/4 of a million QA results and within a few years should contain several million individual laboratory data.

8.6 MEDIAN1 (S2K)

This program is simple and is performed after a study is completed or whenever a summary copy of an earlier study is required. The program simply abstracts interlaboratory median values for all parameters on all samples for a specified study. A typical output is given in Table 12. Although partly administrative, the program serves a value in providing a capsule summary of data for a large study and internally a quick means of selecting previous samples when preparing for a new study. This type of output is provided in study reports.

8.7 MEDIAN 2 (S2K) Track Record of an RM/CRM

This program has exceptional value in its ability to administer a continually updated track record on the characteristics of a CRM or RM. The National Water Research Institute has many QA programs that involve or have involved many hundreds of reference materials (water, rain, sediments, vegetation, and fish). Maintaining a swift electronic data summary of how each sample has performed over time is essential. A typical output is given in Table 13. It is now available on command within minutes.

The usefulness of this tracking program is apparent when one has a data base containing over a million results for several external QA programs employing many samples. Besides being administratively useful, the program can provide a summary output that quickly confirms the stability of data over time. Knowledge and hard evidence on the stability of constituent concentration is critical if any external QA program is to be successful.

Table 12: MEDIAN1: Summary of Interlaboratory Median Values

PARAMETER		SAMPLE NUMBER						
		LR-PRV-13 SAMPLE 1	LR-SSW-20 SAMPLE 2	LR-PRC-04 SAMPLE 3	LR-SSW-16 SAMPLE 4	EU-07 SAMPLE 5	EU-08 SAMPLE 6	EU-09 SAMPLE 7
COLOUR	HAZEN UNIT	3.000	34.000	3.000	32.000	1.750	1.000	1.000
SPECIFIC CONDUCTANCE	UMHO/CM	10.500	17.550	62.320	68.000	13.330	10.600	9.010
GRAN ACIDITY	MG/LCACO3	1.370	1.860	4.220	1.590	1.760	1.060	1.700
ACIDITY TO PH 8.3	MG/LCACO3	2.325	2.678	5.800	2.765	2.367	2.118	2.310
PH	PH UNITS	5.173	6.600	4.293	6.440	4.985	5.505	5.165
DISOLVED ORG CARBON	MG C/L	.500	3.920	1.400	4.720	.200	.224	.271
ALKALINITY-FIXED ENDPT. PH4.5	MG/L CACO3	1.400	4.500	.000	4.000	1.100	1.700	1.400
ALKALINITY-GRAN,INFLEC,EXTRAP	MG/L CACO3	.000	3.200	.000	2.460	.000	.030	.000
GRAN TITRA ALK	MG/LCACO3	-.240	3.245	-2.575	2.570	-.490	.060	-.265
DISSOLVED INORG CARBON	MG C/L	.220	.983	.285	.740	.296	.328	.241
NITRATE + NITRITE	MG N/L	.110	.124	1.539	.249	.070	.250	.160
AMMONIA	MG N/L	.079	.008	.290	.010	.051	.232	.022
TOTAL KJELDAHL NITROGEN	MG N/L	.154	.121	.470	.236	.100	.300	.135
SODIUM	MG/L	.710	1.183	.185	6.026	.860	.150	.020
MAGNESIUM	MG/L	.082	.310	.954	.730	.055	.086	.162
REACTIVE SILICA	MG SI/L	.053	1.940	.230	2.550	.022	.039	.027
SULFATE,IC METHOD	MG/L	.954	1.510	6.910	7.421	1.630	1.920	1.743
SULFATE NON IC METHODS	MG/L	.967	2.200	6.647	7.740	1.545	1.940	1.732
CHLORIDE IC	MG/L	1.220	1.140	4.100	11.000	1.390	.280	.062
CHLORIDE NON IC METHODS	MG/L	1.140	1.200	4.190	11.000	1.370	.300	.095
POTASSIUM	MG/L	.110	.173	.178	.791	.012	.030	.022
CALCIUM	MG/L	.196	1.600	4.000	4.447	.430	.717	.540

PARAMETER		SAMPLE NUMBER		
		AUD-04 SAMPLE 8	AUD-05 SAMPLE 9	AUD-06 SAMPLE 10
COLOUR	HAZEN UNIT	56.500	5.000	8.000
SPECIFIC CONDUCTANCE	UMHO/CM	22.250	34.250	53.100
GRAN ACIDITY	MG/LCACO3	1.870	1.190	1.040
ACIDITY TO PH 8.3	MG/LCACO3	3.310	2.018	2.230
PH	PH UNITS	5.830	6.580	7.189
DISOLVED ORG CARBON	MG C/L	6.180	2.370	3.400
ALKALINITY-FIXED ENDPT. PH4.5	MG/L CACO3	2.350	4.200	13.800
ALKALINITY-GRAN,INFLEC,EXTRAP	MG/L CACO3	.870	2.790	12.880
GRAN TITRA ALK	MG/LCACO3	.940	2.880	12.880
DISSOLVED INORG CARBON	MG C/L	.446	.956	3.295
NITRATE + NITRITE	MG N/L	.027	.120	.083
AMMONIA	MG N/L	.040	.019	.022
TOTAL KJELDAHL NITROGEN	MG N/L	.390	.192	.232
SODIUM	MG/L	.543	1.031	.990
MAGNESIUM	MG/L	.476	.750	1.613
REACTIVE SILICA	MG SI/L	.520	.770	1.050
SULFATE,IC METHOD	MG/L	5.600	7.975	8.420
SULFATE NON IC METHODS	MG/L	6.425	7.975	8.600
CHLORIDE IC	MG/L	.356	1.260	.940
CHLORIDE NON IC METHODS	MG/L	.600	1.205	.905
POTASSIUM	MG/L	.320	.550	.730
CALCIUM	MG/L	2.085	3.155	5.770

Table 13: MEDIAN2: Comparision of Interlaboratory Median Values Between Studies

PARAMETER		PREVIOUS DATA (STUDY NO./SAMPLE NUMBER)					
		LR-PRD-10					
	STUDY NO. / SAMPLE NO.	0012 / 8	0010 / 9	0016 / 5	0019 / 2	AVG. MED.	STD. DEV.
COLOUR	HAZEN UNIT	30.000	35.000	30.000	30.000	31.250	2.500
SPECIFIC CONDUCTANCE	UMHO/CM	35.300	35.600	36.000	35.800	35.675	.299
ACIDITY TO PH 8.3	MG/LCACO3	7.000	6.875	6.740	5.610	6.556	.640
ACIDITY(ALL METHODS)	MG/LCACO3	6.530	6.770	-	-	6.650	.170
PH	PH UNITS	4.420	4.400	4.420	4.410	4.413	.010
NITRATE + NITRITE	MG N/L	.010	.007	.009	.006	.008	.002
AMMONIA	MG N/L	.010	.010	.007	.006	.008	.002
TOTAL KJELDAHL NITROGEN	MG N/L	.151	.178	.160	.140	.157	.016
SODIUM	MG/L	.531	.500	.535	.537	.526	.017
MAGNESIUM	MG/L	.410	.400	.403	.400	.403	.005
REACTIVE SILICA	MG SI/L	2.389	2.326	2.390	2.397	2.376	.033
SULFATE NON IC METHODS	MG/L	8.800	8.600	8.700	8.350	8.613	.193
SULFATE (ALL METHODS)	MG/L	8.120	8.155	-	-	8.138	.025
CHLORIDE NON IC METHODS	MG/L	.400	.400	.355	.415	.393	.026
CHLORIDE (ALL METHODS)	MG/L	.200	.210	-	-	.205	.007
POTASSIUM	MG/L	.141	.140	.140	.150	.143	.005
CALCIUM	MG/L	1.800	1.780	1.780	1.753	1.778	.019
SULFATE,IC METHOD	MG/L	8.080	8.040	8.220	8.096	8.109	.078
CHLORIDE IC	MG/L	.200	.200	.200	.210	.203	.005
ALKALINITY-GRAN,INFLEC,EXTRAP	MG/L CACO3	.000	.000	.000	.000	.000	.000
GRAN TITRA ALK	MG/LCACO3	-2.130	-	-2.000	-2.150	-2.093	.081
ALKALINITY-FIXED ENDPT. PH4.5	MG/L CACO3	.100	.000	.000	.000	.025	.050
GRAN ACIDITY	MG/LCACO3	5.340	6.590	5.545	4.409	5.471	.895
DISOLVED ORG CARBON	MG C/L	-	-	5.699	5.550	5.625	.105
DISSOLVED INORG CARBON	MG C/L	-	-	.275	.325	.300	.035

DATES: 0012(86/04/07.), 0010(85/10/29.), 0016(87/08/24.), 0019(88/08/24.)

Table 14: Comparison of Laboratory Performance (within a study)

	BIAS			FLAGS			
LAB CODE	NO. OF PARAMETERS ANALYZED	NO. OF PARAMETERS BIASED	PERCENTAGE OF PARAMETERS BIASED (%)	NO. OF RESULTS RANKED	NO. OF FLAGS ASSIGNED	PERCENTAGE OF RESULTS FLAGGED (%)	SUM OF % BIAS AND % FLAGS SCORE
L002	14	1	7.14	125	1	.80	7.94
L002C	9	0	.00	90	0	.00	.00
L003	12	0	.00	100	6	6.00	6.00
L004	15	8	53.33	136	11	8.09	61.42
L005	7	0	.00	66	0	.00	.00
L006	16	6	37.50	160	2	1.25	38.75
L007	10	5	50.00	100	15	15.00	65.00
L008	6	3	50.00	60	19	31.67	81.67
L010	10	4	40.00	100	21	21.00	61.00
L011	11	2	18.18	105	1	.95	19.13
L013	6	0	.00	54	0	.00	.00
L014	16	6	37.50	156	22	14.10	51.60
L019	10	4	40.00	96	9	9.38	49.38
L021	11	2	18.18	104	13	12.50	30.68
L023	13	3	23.08	117	13	11.11	34.19
L024	15	3	20.00	142	6	4.23	24.23
L025	20	5	25.00	188	1	.53	25.53
L029	16	1	6.25	157	8	5.10	11.35
L030	9	0	.00	86	0	.00	.00
L032	8	3	37.50	79	11	13.92	51.42
L033	8	3	37.50	71	3	4.23	41.73
L034	10	1	10.00	96	5	5.21	15.21
L035	6	1	16.67	60	9	15.00	31.67
L041	5	1	20.00	50	8	16.00	36.00
L043	4	1	25.00	40	0	.00	25.00
L045	13	4	30.77	126	4	3.17	33.94
L047	11	6	54.55	92	53	57.61	112.15
L048	12	0	.00	114	5	4.39	4.39
L049	16	10	62.50	155	31	20.00	82.50
L049B	2	0	.00	20	0	.00	.00
L053	11	3	27.27	99	0	.00	27.27
L058	13	4	30.77	130	3	2.31	33.08
L061	10	0	.00	99	0	.00	.00
L063	17	2	11.76	161	1	.62	12.39
L064	9	5	55.56	87	19	21.84	77.39
L066	10	3	30.00	100	10	10.00	40.00
L067	12	2	16.67	119	23	19.33	35.99
L069	16	2	12.50	156	4	2.56	15.06
L073	10	1	10.00	97	1	1.03	11.03
L078	2	0	.00	20	0	.00	.00
L081	11	1	9.09	101	5	4.95	14.04
L082	12	5	41.67	120	8	6.67	48.33
L085	14	2	14.29	124	7	5.65	19.93
L086	13	5	38.46	128	16	12.50	50.96
L087	10	0	.00	100	5	5.00	5.00
L088	6	1	16.67	60	0	.00	16.67
L089	12	1	8.33	120	5	4.17	12.50
L089B	1	0	.00	10	0	.00	.00
L089C	10	0	.00	97	0	.00	.00
L090	11	1	9.09	106	2	1.89	10.98
L090B	1	0	.00	10	0	.00	.00
L091	14	2	14.29	140	20	14.29	28.57
L092	3	0	.00	27	4	14.81	14.81
L093	13	1	7.69	129	15	11.63	19.32
L094	14	1	7.14	135	9	6.67	13.81
L095	10	1	10.00	100	6	6.00	16.00
L096	7	4	57.14	70	27	38.57	95.71
L097	11	0	.00	98	3	3.06	3.06

THE FOLLOWING CODES WERE USED IN THE ANALYSIS
00392 01090 01092 06002 06194 07092 07192 11091 12091 14092
16001 17001 19091 20091 16000 17000 06193 06282 06592 07392
01089

THE FOLLOWING CODES WERE EXCLUDED
00292

8.8 LABCOMP: Performance Within a Study

The output from this program is derived by searching the S2K data base. An example output is given in Table 14. Its primary purpose is to provide each laboratory with a precise statement on its relative performance with its peer group within a study. It is particularly useful for large studies involving many laboratories that analyze many different constituents.

The program can isolate and also respond to all or any group parameters. A very wide choice of outputs are available. The example output (Table 14) is for a LRTAP study and the footnotes indicates those parameters used and those which were excluded when this table was created. This program option is particularly useful when a request is made to compare a particular contract lab to existing program laboratories for a specific series of constituents.

The program LABCOMP ranks laboratory performance and provides a score. This score is the summation of the percentage of parameters biased and percentage of results flagged. A very low score is indicative of superior performance whereas a very high score is poor performance.

Performance is accepted in LABCOMP as quite relative. It includes bias (which reflects accuracy) and precision (indicated by many flags). Laboratories, that are severely imprecise will, if the flagging process is correctly established, have as many as half their results flagged (any flag H,L,VL,VH) is counted). If half the data are flagged their score will be 50%. On the other hand, if a lab is precisely inaccurate (no flags) it may be frequently discerned as biased by the Youden technique. If six out of ten parameters are biased, then the score will be 60%. Some labs are both biased and flagged and can have very high scores and is declared poor within the study. Corrective action is suggested.

Experience in analyzing many studies has created guidelines on performance (as viewed through the Youden bias and flagging process).

Scores of over 60% are poor (maximum score is 200%, all data flagged and all parameters biased). Scores of less than 25% are satisfactory, scores of less than 10% are satisfactory well done and those results between 25% and 60% are moderate.

The visual impact to a laboratory which in LABCOMP is graded with a high score is informative. To have a very low score (satisfactory) creates satisfaction. A very high score (over 60%) can be very surprising when judged as poor. It is certainly stimulating and cause for immediate internal view. To this end this output program (LABCOMP) has merit and evidence given in Section 8.8 would suggest the impact for some laboratories has been constructive.

8.9 FLGTBL: Performance of a Group of Laboratories Based on Frequency of Bias and Flags

Some studies, such as for the Federal/Provincial LRTAP intercomparison program are (a) frequent (three per year), (b) involve laboratories of equivalent capability, (c) use the same types of water (soft) and (d) have criteria for flagging that have remained constant over several years. These studies also involve about 50 laboratories who in general analyze the same constituents. With this resource (almost 20 major ion studies) it is possible to compare the frequency in which laboratories have their data assessed as biased or flagged. In fact, it is possible to provide a track record on the performance of each laboratory over time.

Within any study, a laboratory that gets most of its parameters declared biased and has most of its results all flagged, is considered as giving a very poor performance. On the other hand, a laboratory with no bias and no flags may be considered satisfactory and an excellent performer. Between these two extremes lie average or moderate performances. When studies are frequent, it is possible to examine trends in the frequency of biased and flagged results (e.g., improvements over time may be observed). The program, called FLGTBL,

Table 15: FLGTBL - Comparison of Laboratory Performance over Several Studies (LRTAP)

LAB CODE	%BIAS AND %FLAGS ON STUDIES 0015	0016	0017	0018	0019	0020	MEDIAN SCORE	COMMENTS
L002	27.0	.8	7.4	12.6	10.7	7.9	9.3	SATISFACTORY, WELL DONE
L002C	24.6	11.1	13.5	.0	2.2	.0	6.7	SATISFACTORY, WELL DONE
L003	10.5	.0	2.5	18.8	28.8	6.0	8.3	SATISFACTORY, WELL DONE
L004	23.3	67.2	14.5	4.4	37.6	61.4	30.4	MODERATE
L005	.0	17.5	14.3	41.5	33.3	.0	15.9	SATISFACTORY
L006	.6	13.1	23.2	18.4	30.7	38.8	20.8	SATISFACTORY
L007	27.5	59.1	13.0	17.1	38.8	65.0	33.2	MODERATE
L008	41.7	-	126.7	80.0	96.0	81.7	81.7	POOR
L010	51.5	32.8	43.1	58.1	46.3	61.0	48.9	MODERATE
L011	23.9	.9	46.7	14.5	-	19.1	19.1	SATISFACTORY
L013	6.5	16.7	31.2	.0	1.8	.0	4.1	SATISFACTORY, WELL DONE
L014	12.5	54.0	20.6	24.0	28.8	51.6	26.4	MODERATE
L014C	47.0	60.8	8.1	24.0	32.5	-	32.5	MODERATE
L017	-	-	-	45.3	-	-	-	-
L019	-	2.0	24.2	28.3	.0	49.4	24.2	SATISFACTORY
L020	5.3	-	25.9	-	-	-	15.6	SATISFACTORY
L020C	5.2	-	1.3	-	-	-	3.2	SATISFACTORY, WELL DONE
L021	1.3	10.1	4.9	32.0	39.2	30.7	20.4	SATISFACTORY
L022	39.0	-	14.4	-	37.2	-	37.2	MODERATE
L023	29.1	6.1	43.5	35.3	27.1	34.2	31.7	MODERATE
L024	10.0	15.1	26.3	16.2	37.2	24.2	20.2	SATISFACTORY
L025	31.8	39.4	26.5	15.1	21.2	25.5	26.0	MODERATE
L027	56.0	51.8	1.5	-	-	-	51.8	MODERATE
L029	35.5	25.1	12.1	-	13.7	11.3	13.7	SATISFACTORY
L030	4.1	1.4	1.3	24.0	.0	.0	1.3	SATISFACTORY, WELL DONE
L031	7.9	14.4	16.0	37.5	10.1	-	14.4	SATISFACTORY
L032	63.3	37.0	53.1	61.5	60.0	51.4	56.5	MODERATE
L033	2.9	38.8	.0	55.0	20.1	41.7	29.4	MODERATE
L034	33.9	.0	18.7	2.8	15.1	15.2	15.2	SATISFACTORY
L035	30.0	.0	.0	14.3	30.9	31.7	22.1	SATISFACTORY
L041	-	-	15.0	10.0	-	36.0	15.0	SATISFACTORY
L043	.0	-	25.0	.0	-	25.0	12.5	SATISFACTORY
L045	23.6	-	8.5	11.8	25.7	33.9	23.6	SATISFACTORY
L047	-	-	74.8	75.6	98.5	112.2	87.0	POOR
L048	14.5	-	34.1	7.7	32.2	4.4	14.5	SATISFACTORY
L049	62.7	-	26.9	79.7	60.1	82.5	62.7	POOR
L052	9.5	38.4	40.5	33.9	29.2	-	33.9	MODERATE
L053	18.2	-	9.1	-	-	27.3	18.2	SATISFACTORY
L054	50.0	-	-	106.0	-	-	78.0	POOR
L056	101.5	-	-	-	-	-	-	-
L057	51.4	54.2	50.0	16.3	-	-	50.7	MODERATE
L058	12.1	.9	43.6	93.3	-	33.1	33.1	MODERATE
L059	-	-	-	96.7	48.2	-	72.4	POOR
L060	45.7	-	-	-	-	-	-	-
L061	2.9	-	11.0	-	10.0	.0	6.4	SATISFACTORY, WELL DONE
L063	13.6	30.8	27.1	25.0	23.2	12.4	24.1	SATISFACTORY
L064	73.3	16.0	35.0	34.0	20.0	77.4	34.5	MODERATE
L066	-	13.7	13.4	42.5	23.0	40.0	23.0	SATISFACTORY
L067	21.7	53.3	67.7	51.2	-	36.0	51.2	MODERATE
L069	-	-	32.9	18.3	40.8	15.1	25.6	MODERATE
L073	21.3	6.8	71.4	.0	10.0	11.0	10.5	SATISFACTORY
L074	51.3	44.7	61.4	63.1	27.8	-	51.3	MODERATE
L078	50.0	.0	.0	.0	5.0	.0	.0	SATISFACTORY, WELL DONE
L081	77.4	3.6	15.2	22.6	30.4	14.0	18.9	SATISFACTORY
L082	-	44.4	-	48.4	-	48.3	48.3	MODERATE
L083	37.6	34.9	-	-	-	-	36.2	MODERATE
L084	-	31.0	-	-	-	-	-	-
L085	-	-	35.7	34.6	-	19.9	34.6	MODERATE
L086	-	65.7	54.2	53.7	60.2	51.0	54.2	MODERATE
L087	-	-	-	35.3	19.8	5.0	19.8	SATISFACTORY
L088	-	-	-	10.0	20.0	16.7	16.7	SATISFACTORY
L089	-	-	-	78.8	46.7	12.5	46.7	MODERATE
L089C	-	-	-	.0	23.3	.0	.0	SATISFACTORY, WELL DONE
L090	-	-	-	18.0	10.4	11.0	11.0	SATISFACTORY

(Continued on next page)

Table 15: FLGTBL - Comparison of Laboratory Performance over Several Studies (LRTAP)

LAB CODE	%BIAS AND %FLAGS ON STUDIES						MEDIAN SCORE	COMMENTS
	0015	0016	0017	0018	0019	0020		
L091	-	-	-	31.4	25.7	28.6	28.6	MODERATE
L092	-	-	-	62.4	31.3	14.8	31.3	MODERATE
L093	-	-	-	65.0	42.0	19.3	42.0	MODERATE
L094	-	-	-	57.3	27.4	13.8	27.4	MODERATE
L095	-	-	-	75.0	-	16.0	45.5	MODERATE
L096	-	-	-	-	-	95.7	-	-
L097	-	-	-	-	-	3.1	-	-

STUDY DATES: 0015(87/04/06.);0016(87/08/24.);0017(87/12/01.);
0018(88/05/02.);0019(88/08/24.);0020(89/01/10.);

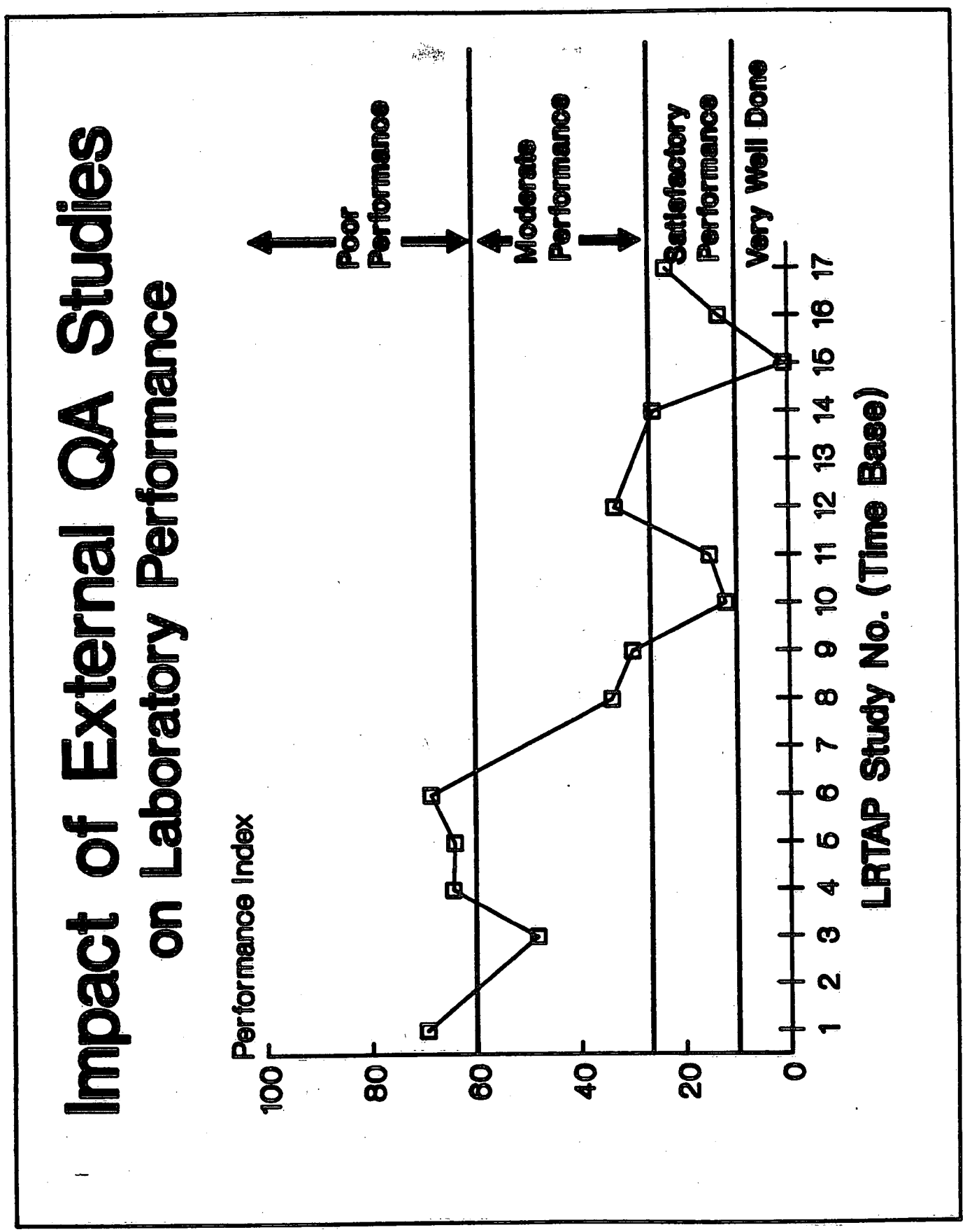
(Continued on next Page)

(Continued)

Table 15: FLGTBL - Comparison of Laboratory Performance over Several Studies (LRTAP)

LAB CODE	MEDIAN SCORE (%)	NUMBER OF STUDIES	LAB CODE	MEDIAN SCORE (%)	NUMBER OF STUDIES
L001	-	0	L005	15.9	6
L009	-	0	L088	16.7	3
L012	-	0	L053	18.2	3
L013A	-	0	L081	18.9	6
L016	-	0	L011	19.1	5
L017	-	1	L087	19.8	3
L026	-	0	L024	20.2	6
L028	-	0	L021	20.4	6
L031C	-	1	L006	20.8	6
L037	-	0	L035	22.1	6
L038	-	0	L066	23.0	5
L039	-	0	L045	23.6	5
L040	-	0	L063	24.1	6
L042	-	0	L019	24.2	5
L044	-	0	L069	25.6	4
L046	-	0	L025	26.0	6
L050	-	0	L014	26.4	6
L051	-	0	L094	27.4	3
L056	-	1	L091	28.6	3
L060	-	1	L033	29.4	6
L068	-	0	L004	30.4	6
L070	-	0	L092	31.3	3
L071	-	0	L023	31.7	6
L077	-	0	L014C	32.5	5
L080	-	0	L058	33.1	5
L084	-	1	L007	33.2	6
L096	-	1	L052	33.9	5
L097	-	1	L064	34.5	6
L098	-	0	L085	34.6	3
L089C	.0	3	L083	36.2	2
L078	.0	6	L022	37.2	3
L030	1.3	6	L093	42.0	3
L020C	3.2	2	L095	45.5	2
L013	4.1	6	L089	46.7	3
L061	6.4	4	L082	48.3	3
L002C	6.7	6	L010	48.9	6
L003	8.3	6	L057	50.7	4
L002	9.3	6	L067	51.2	5
L073	10.5	6	L074	51.3	5
L090	11.0	3	L027	51.8	3
L043	12.5	4	L086	54.2	5
L029	13.7	5	L032	56.5	6
L031	14.4	5	L049	62.7	5
L048	14.5	5	L059	72.4	2
L041	15.0	3	L054	78.0	2
L034	15.2	6	L008	81.7	5
L020	15.6	2	L047	87.0	4

Fig. 8


Impact of External QA Studies
on Laboratory Performance
Performance Index
100
80
60
40
20
0
Poor Performance
Moderate Performance
Satisfactory Performance
Very Well Done
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
LRTAP Study No. (Time Base)

helps in this effort. Output is given in Table 15. This output is the integrated results of LABCOMP and when transferred from the mainframe to a PC, performance can be graphically displayed. Because the output is created from S2K, it is possible to create sn output for any group of laboratories, studies or parameters.

A typical graphic display of performance for one laboratory abstracted from FLGTBL is given in Fig. 8. The performance index in this figure (and Table 15) are the same as used in LABCOMP. They are arbitrary and may be modified when all evidence has been reviewed.

8.10 APPRAIS - Automated Appraisals

When the original YOUDN21 flat file programs were applied to large 50 lab, 10 sample, 20 parameter studies, a great deal of manual effort was required to prepare narrative comments on each lab for every parameter and each sample result. Not only was it tedious but it was subject to human and transcription error.

With the development of the S2K data base the preparation of an appraisal became extremely rapid since sufficient space was built into the data base structure to store the calculated outputs. Table 16 gives the criteria developed so that when the program Apprais is initiated it would retrieve from the data base the necessary information to formulate a written narrative. A typical narrative is defined as a "laboratory specific appraisal" and is given in Table 17. It is this appraisal that is attached to a covering letter accompanied by all support data (MEDIAN 1,2, LABCOMP, FLGTBL and YOUDS2K) This critical support information is essential and is provided to each participant when a study is formally completed.

8.11 AQC Programs in Development

The above AQC Data Base Management system has proven very effective and efficient in addressing large external QA studies. The

Table 16:

CRITERIA USED TO PREPARE STATEMENTS FOR THE AUTOMATED APPRAISALS

Note: Please refer to the "Glossary of Terms" and associated references for an explanation of A) How the non-parametric process of ranking is able to discern bias in a laboratory data set and B) The calculation and conditions that warrents a reported result to be flagged L, H, VL or VH.

Status of Data	Statement Produced in Appraisal
1(A) No flags, no bias in the data.	Satisfactory.
(B) No bias, only 1 minor flag (H or L).	Satisfactory except for low/high flag on sample _.
2. No data or results reported by laboratory.	No results reported.
3. Data reported on less than half of the samples, no results flagged.	Insufficient data to assess bias.
4. Same as item 3, but some results are flagged.	Flagged __ on sample __,__ and flagged __ on sample __. Insufficient data to assess bias.
5. No results are flagged but data set identified as biased high or low.	Although no results are flagged, ranking indicates results are biased high/low.
6. Some results are flagged, the data set is discerned as biased.	Flagged __ on sample __,__; Flagged __ on sample __,__. Ranking indicates results are biased.
7. Some results are flagged.	Flagged __ on samples __,__ and flagged __ on samples __.
8. No bias statement but two or three results are flagged, one is very high, the other is very low.	Flagged very high on sample __ and very low on sample __. These results are slightly erratic.
9. No bias statement but two or more results are flagged very high and and two or more results are flagged very low.	Flagged very high on samples __,__ and flagged very low on samples __,__,__. These results are erratic.
10. Results are ranked, the data set is not biased but one result is flagged very high or very low.	Flagged __ on sample __. This extreme result suggests the measurement process is out of control.

Table 17: Laboratory Appraisal (an example)

SPECIFIC CONDUCTANCE	SATISFACTORY
GRAN ACIDITY	NO RESULTS REPORTED.
ACIDITY TO PH 8.3	SATISFACTORY
PH	SATISFACTORY EXCEPT FOR LOW ON SAMPLE 4
DISOLVED ORG CARBON	SATISFACTORY
ALKALINITY-FIXED ENDPT. PH4.5	NO RESULTS REPORTED.
ALKALINITY-GRAN,INFLEC,EXTRAP	INSUFFICIENT DATA TO ASSESS BIAS
GRAN TITRA ALK	NO RESULTS REPORTED.
DISSOLVED INORG CARBON	FLAGGED LOW ON SAMPLE 10 RANKING INDICATES RESULTS ARE BIASED LOW
NITRATE + NITRITE	SATISFACTORY
AMMONIA	SATISFACTORY
TOTAL KJELDAHL NITROGEN	NO RESULTS REPORTED.
SODIUM	FLAGGED EXTREMELY HIGH ON SAMPLE 9 THIS EXTREMELY HIGH RESULT SUGGESTS THE MEASUREMENT PROCESS IS OUT OF CONTROL
MAGNESIUM	FLAGGED HIGH ON SAMPLE 4 10
REACTIVE SILICA	FLAGGED HIGH ON SAMPLE 6 7 FLAGGED VERY HIGH ON SAMPLE 5
SULFATE,IC METHOD	ALTHOUGH NO RESULTS ARE FLAGGED RANKING INDICATES A SLIGHT BIAS HIGH
SULFATE NON IC METHODS	FLAGGED EXTREMELY LOW ON SAMPLE 1 THIS EXTREMELY LOW RESULT SUGGESTS THE MEASUREMENT PROCESS IS OUT OF CONTROL
CHLORIDE IC	SATISFACTORY
CHLORIDE NON IC METHODS	SATISFACTORY
POTASSIUM	ALTHOUGH NO RESULTS ARE FLAGGED RANKING INDICATES A SLIGHT BIAS HIGH
CALCIUM	SATISFACTORY

external QA studies provided to the LRTAP, Eulerian, IJC, LRTAP terrestrial and the National QA programs have created a data base of some significance. This resource is now being analyzed to improve the overall QA assessment program and to allow more ready access of information to analysts, program managers and the users of data.

8.11.1 Precision functions

The main strength of the YOUDN21 (or YOUDS2K) program has been the ability to discern bias (systematic error) and provide analysts information on precision (flags). The formula for flagging has been a series of educated guesses so that the criteria (BAE, LLBAE and CEI) would provide reasonable distribution of flags.

The heavily populated data base system can now be used to analyze interlaboratory means and standard deviations to create more meaningful distributions of precision (the precision functions). This possibility can be realized because of existence of a larger body of information on many different samples (20 to 200) contained in such programs as the Eulerian and LRTAP. A typical output that is anticipated is given in Fig 9.

The selective nature of the S2K data base can allow the graphics to be created to isolate such functions on groups of labs (surface water, rainwater, government, contract, etc.) for selected samples or for all samples. The more serious benefit will be in the development of functions clearly stating the level of performance that satisfactory laboratories have demonstrated. All other laboratories (e.g., contract laboratories) will need to either achieve this performance or be one which excel beyond this minimum standard. Future applications of the data system have considerable potential to serving environmental programs and the quality management issues inherent.

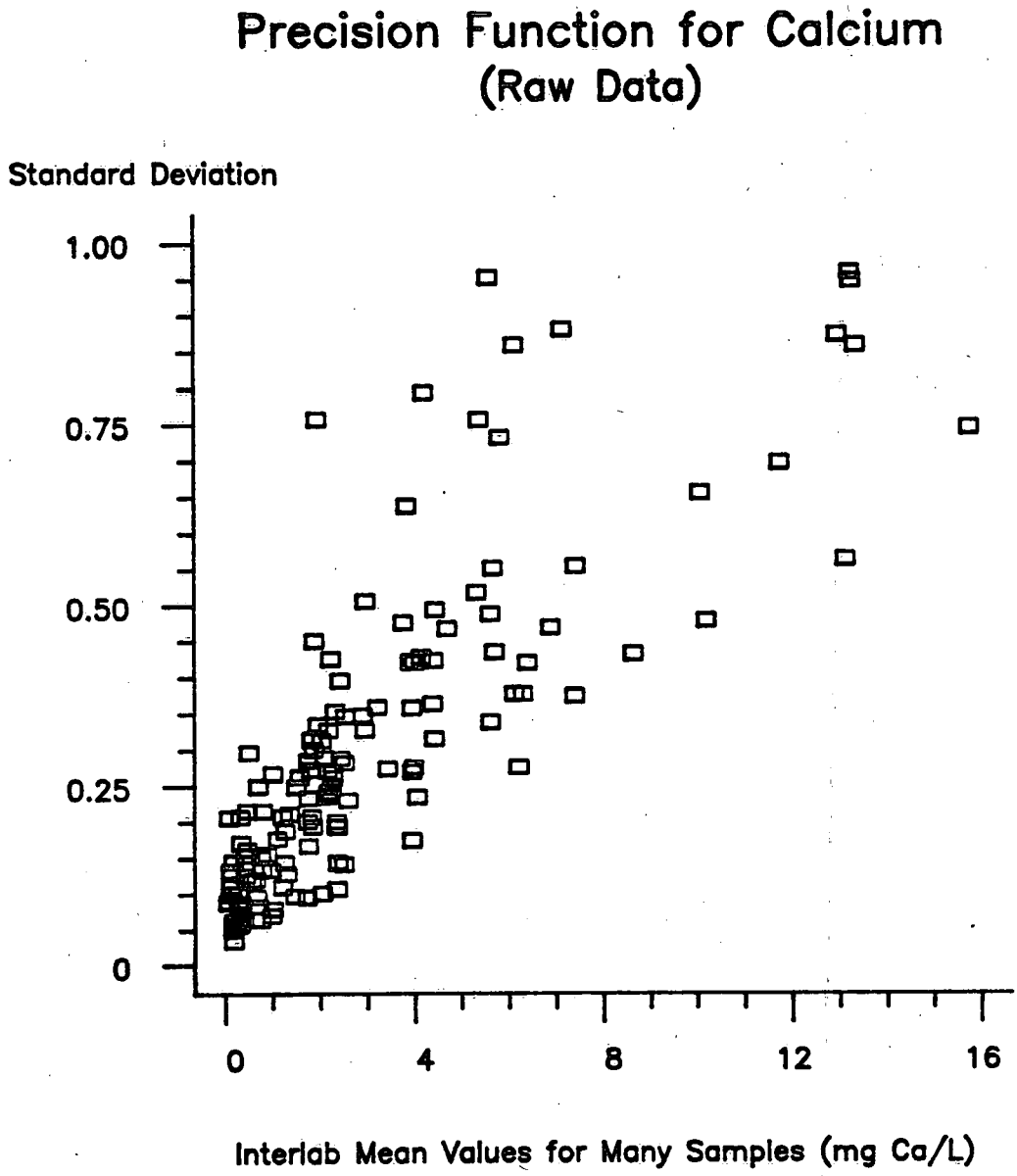


Fig. 9a Precision Function for Calcium – Raw Data (LRTAP QA Program)

Precision Function for Calcium
(After Data Rejection)

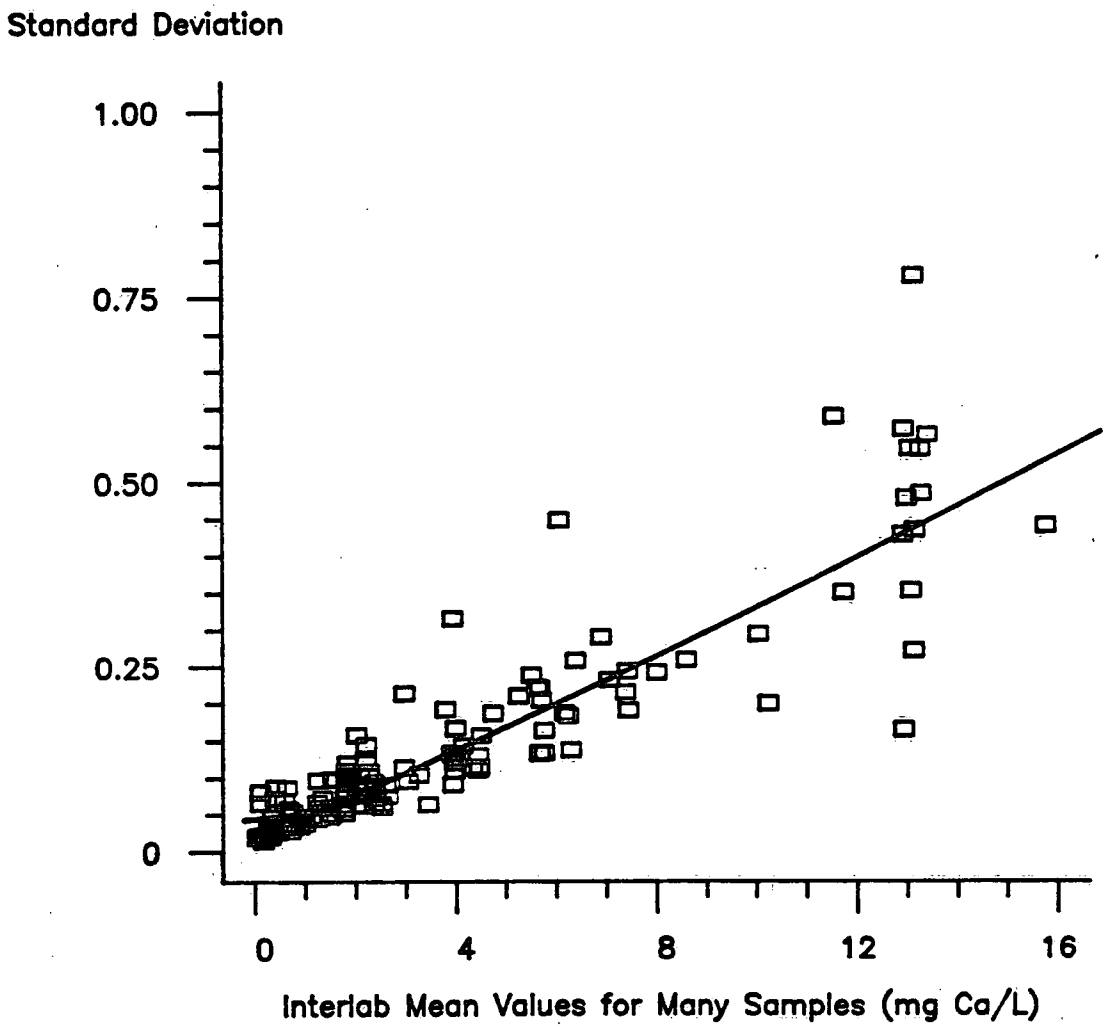


Fig. 9b Precision Function for Calcium
– After removing some outliers

8.11.2 Charge balance and conductance ratios

One specific intralab quality control procedure that all large water quality laboratories often utilize to check results is the calculation of charge balance (anions versus cations) and for soft waters, the ratio of measured to calculated conductance.

As with precision functions, the AQC data base is sufficiently large for water studies to analyze the overall results to (a) confirm that interlab medians indeed have integrity and (b) create % error or precision functions on charge balance and conductance calculations. The ability to create a distribution of percent error or uncertainties as a function of ionic strength will have merit for users of data and will give information on the criteria required for the selection of contract laboratories. This work is in progress.

8.12 References

1. J.L. Clark. Evaluation of Performance of laboratories Determining Water Quality Constituents Through Natural Water Samples Whose True Values are Unknown. In Summary of Conference Presentations. Envirometrics 81, p. 54-55, 1981, Alexandria, Virginia, April 8-10, 1981.
2. (ASTM-D19), D4210-83. Intralaboratory Quality Control Procedures and a Discussion Reporting Low-Level Data, A Standard Practice. ASTM, 1916, Race St., Philadelphia, PA 19103.
3. K.I. Aspila, R.E. White, J.L. Clark. Quality Assurance Aspects of the International Joint Commission Great Lakes Monitoring Program. In ASTM Special Technical Publication 867 (1985), a symposium on Quality Assurance of Environmental Measurements, Aug. 8-12, 1983, Boulder, Colorado (published by ASTM, 1916 Race St., Philadelphia, PA 19103).

4. W.J. Youden. Ranking Laboratories by Round-Robin Tests. In Precision Measurement and Calibration, Harry H. Ku, Editor, NBS Special Publication 300, Vol. 1, p. 165-9 to p. 169-13, US Government Printing Office, Washington, DC, 1969.
5. W.J. Youden and E.H. Steiner. Statistical Manual of the Association of Official Analytical Chemists. AOAC, P.O. Box 540, Benjamin Franklin Station, Washington, DC, 20044 (1975).

9.0 THE FEDERAL-PROVINCIAL AND PRAIRIE PROVINCE WATERBOARD QA PROGRAMS

The preceding section has described how large interlaboratory studies are presented and processed by various computer assessment techniques. These have served well for LRTAP-aquatics, LRTAPP-terrestrial, recent national and the Eulerian QA studies. The QA programs that use alternative methods for assessing laboratory results are the Federal-Provincial and Prairie Province Water Board QA programs. The following is an overview of these two programs.

9.1 General

Under terms of the federal-provincial agreements and the agreements under the Prairie Province Water Board (PPWB)[1], quality assurance programs have been implemented to assess and improve the comparability[2] of water quality data. These programs provide ongoing bimonthly studies for some 40 inorganic constituents in surface waters. Participants include eight federal and eight provincial laboratories. A laboratory from the Department of Indian and Northern Affairs also participates in these programs. The essential activities to implement the FP and PPWB QA programs are highlited in Figure 10.

The objectives for the FP and PPWB QA Program are briefly expressed as:

- to detect laboratory measurement anomolies and report them quickly to laboratory managers to allow for swift remedial action;
- to ensure and define comparability and reliability of data that are eventually stored in the national data base system (NAQUADAT) in order to assist users of data;
- to provide evidence on the effectiveness of intralaboratory and interlaboratory quality control procedures;
- to provide regular reports which summarize results for laboratory and project managers.

9.2 Study Design

A single bimonthly study[3] consists of four or five standard reference samples of known values. Half of these samples are for trace metal analysis at two concentration levels. For the other half of the samples, the laboratories report on 25 major ions and on nutrient and physical parameters. Altogether, 100 analysis methodologies and individual results are tabulated in the data summaries. In 1988 the FP QA Program was expanded to address toxic organics using injection ready ampul standards.

A standard reporting form showing various inorganic parameters is given in Table 18.

9.3 Data Evaluation

Analyses for four or five test samples are made by each laboratory during a two month period and results reported to the NWRI QA chemist in charge. Data are entered into the Cyber 180/830A mainframe computer and programs are executed to format data and prepare printouts. A sample data report showing various laboratory methodologies and parameter statistics is given in Table 19. An overview is given in Table 20.

Since it is difficult to have all participants analyze all samples simultaneously, several preliminary reports are often provided during each bimonthly study. This service allows laboratories the opportunity of rapid review and corrective action if their data are indicated as inadequate. A formal report is distributed at the completion of each study. The following describes how flags are assigned to deviating results.

In the FP and PPWB QA program individual results are flagged if evidence suggests that they deviate significantly from design or target values. The concentration range in the FP and PPWB programs often covers two or three orders of magnitude (e.g., 0.010 ppm to 10 ppm). For this reason, the evaluation of data[4,5] has required two

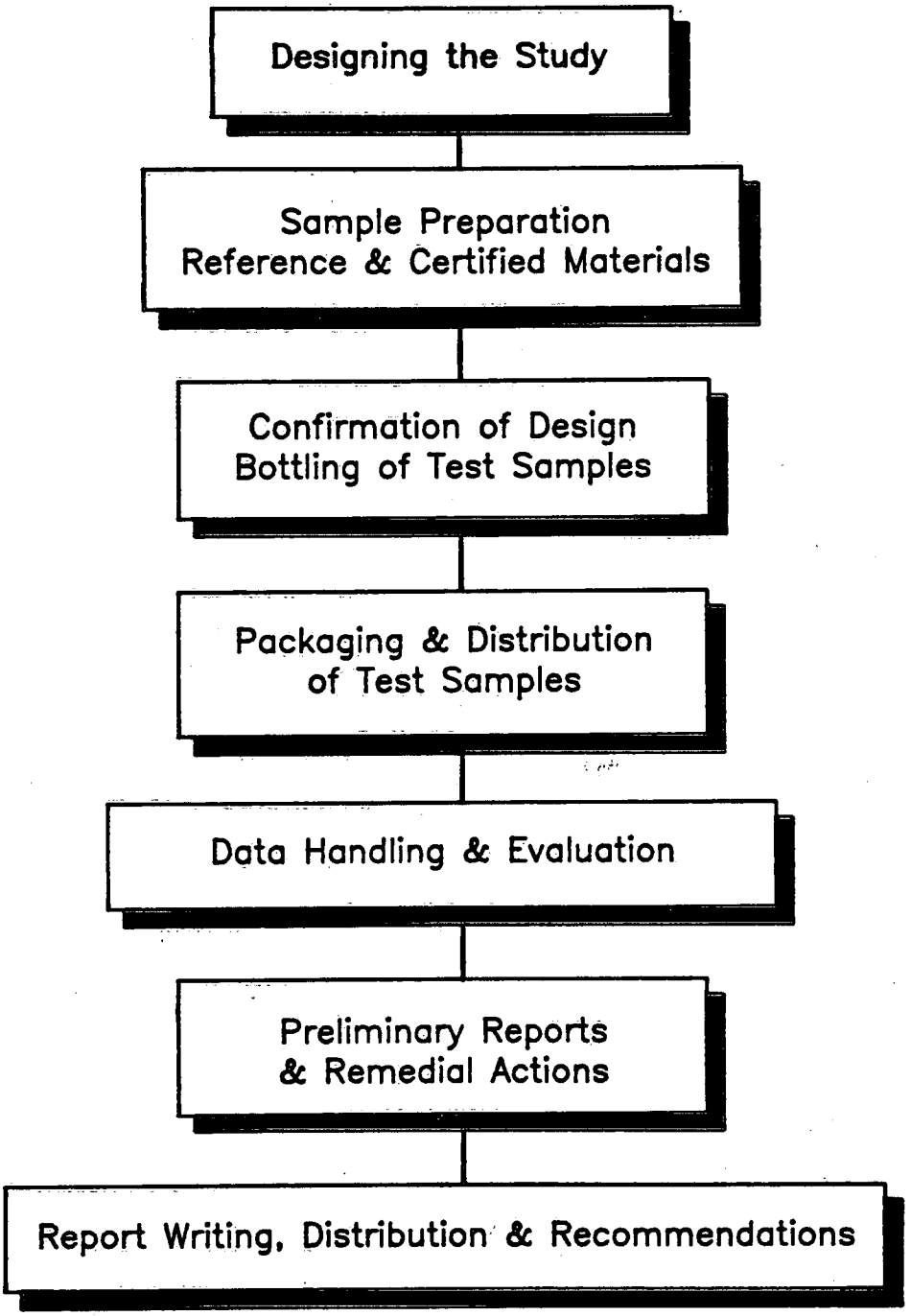


Fig. 10 Flow Chart of Activities to Implement the FP & PPWB QA Program

DATA REPORT SHEET

FEDERAL-PROVINCIAL & PRAIRIE PROVINCES QUALITY ASSURANCE PROGRAMS

LAB # _____ STUDY # _____ DATE _____

Major Ions, Nutrients	NAQUADAT CODE	SAMPLE RESULTS, ppm
Color (Units)	020	
Specific Conductance (µΩ)	020	
Turbidity (JTU/NTU)	020	
Boron	051	
DOC	06	
DIC	06	
TKN	07	
NO_3 + NO_2	071	
Ammonia	075	
Total Nitrogen	076	
Fluoride	091	
Total Alkalinity	101	
Acidity	102	
pH (Units)	103	
Total Hardness	106	
Sodium	11	
Magnesium	12	
Silica (as SiO_2)	14	
Total P (as P)	15	
Sulfate (as SO_4)	16	
Chloride	17	
Potassium	19	
Calcium	20	

METALS	NAQUADAT CODE	SAMPLE RESULTS, ppm
Aluminum	13	
Vanadium	23	
Chromium	24	
Manganese	25	
Iron	26	
Cobalt	27	
Nickel	28	
Copper	29	
Zinc	30	
Strontium	38	
Molybdenum	42	
Cadmium	48	
Barium	56	
Lead	82	

MI CALCULATED PARAMETERS

Sum of Cations meq/l	00120	
Sum of Anions meq/l	00125	
% Difference	00110	

PLEASE REPORT ALL DATA IN PPM, UNLESS OTHER UNITS ARE INDICATED

Table 19: DATA SUMMARY

FEDERAL-PROVINCIAL & PRAIRIE PROVINCES QUALITY ASSURANCE PROGRAMS

STUDY NO. FP 35 PP 75 DATE: 01/11/88 PAGE 1

Sample 1 Trace Metals D/A. (in 3.0% HNO_3)

LAB	13009 Al Tot 5X ICP	13111 Al Dis ICP DA	13302 Al Ext AAS DA	13306 Al UF AAS OX	13321 Al Ext ICP DA	13322 Al Ext DCP DA	13999 Aluminum COMMON
1	-	-	-	1.014	-	-	1.014
2	-	-	1.1	-	-	-	1.1
3	-	-	0.980	-	1.11	-	0.980
6	1.0	-	-	-	-	-	1.0
8	-	-	1.2	-	-	-	1.2 *
9	-	1.05	-	-	-	-	1.05
10	-	0.97	-	-	-	-	0.97
15	-	-	-	-	1.02	-	1.02
16	-	-	-	-	-	1.03	1.03
MEAN	1.0000	1.0100	1.0933	1.0140	1.1100	1.0300	1.0404
STD DEV	-	.0566	.1102	-	-	-	.0712
REL STD	-	5.6	10.1	-	-	-	6.8
DES VAL	-	-	-	-	-	-	1.048

LAB	25004 Mn Tot AAS GF	25011 Mn Tot 5X ICP	25012 Mn Tot 5X DCP	25302 Mn Ext AAS DA	25311 Mn Ext ICP DA	25321 Mn Ext ICP DA	25999 Manganese COMMON
1	-	-	-	-	0.101	-	0.101
3	-	-	-	-	-	0.099	0.099
6	-	0.095	-	-	-	-	0.095
8	0.097	-	-	-	-	-	0.097
9	-	-	-	-	-	-	0.099
10	-	-	-	-	-	-	0.094
11	-	-	-	0.093	-	-	0.093
13	-	-	-	-	0.09	-	0.09
14	-	-	-	-	-	-	-
15	-	-	-	-	0.098	-	0.098
16	-	-	0.112	-	-	-	0.112 R
MEAN	.0970	.0950	.1120	.0930	.0940	.0990	.0978
STD DEV	-	-	-	-	.0057	-	.0060
REL STD	-	-	-	-	6.0	-	6.1
DES VAL	-	-	-	-	-	-	.0972

NOTE:

All concentration units are expressed in MG/L of each element, the exceptions being: Colour in relative units, Conductivity in USIE/Cm, Turbidity in JTU or NTU, Nitrogen analysis in "N", Alkalinity & Hardness in $CACO_3$, Silica in SiO_2, and Sulfate in SO_4.

Table 20:
SUMMARY of FLAGGED RESULTS for the FEDERAL-PROVINCIAL QA PROGRAM
for STUDIES FP 31-32

LAB 2	FLAGS :	NITRATE	-16%		PTASSIUM	-91%	R	ZINC	25%	
		NITRATE	-62%	L						
	HDL :	T N DIS								
LAB 3	FLAGS :	NITRATE	-81%	L	T N DIS	-89%	L			
LAB 4	FLAGS :	NONE								
LAB 5	FLAGS :	D O C	73%		D I C	16%		AMMONIA	-75%	L
		PH	-14%		D O C	337%	R			
	HDL :	SILICA								
LAB 7	FLAGS :	NITRATE	-26%		TOT P	106%				
LAB 9	FLAGS :	COPPER	-23%		MOLYBNUM	-23%		PTASSIUM	-21%	
	HDL :	ALKLINTY								
LAB 10	FLAGS :	COPPER	39%		SILICA	-12%		ALUMINUM	50%	
		COBALT	-36%		LEAD	40%		T N DIS	55%	
	HDL :	AMMONIA			TOT P			AMMONIA		
LAB 11	FLAGS :	NICKEL	16%		D O C	368%	R	D I C	-22%	
		TOT P	106%		IRON	38%		AMMONIA	126%	
		HARDNESS	-19%	R	PTASSIUM	13%		CALCIUM	-25%	R
	HDL :	D O C			NITRATE					
LAB 12	FLAGS :	TKN	22%		ALUMINUM	25%				
	HDL :	D O C			D I C			SILICA		
LAB 13	FLAGS :	IRON	-20%		CONDUCT	-14%		ALKLINTY	19%	R
		CONDUCT	-12%		SULFATE	13%		PTASSIUM	31%	
	HDL :	CHROMIUM			AMMONIA					
LAB 14	FLAGS :	MANGNESE	54%	R	COPPER	180%	R	CADMIUM	144%	R
		LEAD	48%	R	SODIUM	-35%	R	MGNESIUM	-15%	R
		ALUMINUM	23%		IRON	-31%		ZINC	-38%	
		SODIUM	-38%	R						
LAB 15	FLAGS :	MANGNESE	-15%		ZINC	-16%		NITRATE	41%	
		FLUORIDE	25%	R	ALUMINUM	358%	R	VANADIUM	82%	R
		MANGNESE	-50%		IRON	-83%	R	COBALT	55%	
		ZINC	-63%	R	MOLYBNUM	-23%	L	CADMIUM	-55%	R
		LEAD	-95%	L	SULFATE	-11%				
	HDL :	NICKEL			MOLYBNUM					
LAB 16	FLAGS :	MANGNESE	30%		IRON	19%		STRNTIUM	35%	R
		LEAD	17%		TURBIDTY	203%	R	D I C	25%	
		TKN	143%	R	SODIUM	-23%		TOT P	3075%	R
		SULFATE	55%	R	CHLORIDE	18%	R	CALCIUM	-12%	R
		ALUMINUM	85%	R	VANADIUM	118%	R	MANGNESE	67%	
		IRON	24%		COPPER	38%		STRNTIUM	74%	R
		MOLYBNUM	-62%	L	LEAD	50%		ALKLINTY	552%	R
		SODIUM	-21%	R	TOT P	2281%	R	SULFATE	67%	R
		CALCIUM	-11%							
	HDL :	CADMIUM			AMMONIA			FLUORIDE		
		MOLYBNUM			NITRATE			AMMONIA		

NOTE: A VERY HIGH FREQUENCY OF FLAGGED RESULTS (OR A HIGH %) IS INDICATIVE OF POOR PERFORMANCE. ON THE OTHER HAND, LABS WITH FEW IF ANY FLAGS ARE JUDGED TO HAVE VERY GOOD PERFORMANCE.

ALSO, AN "R" FLAG INDICATES A NON COMPARABLE RESULT, THAT IS, ONE PRODUCED WITH NON RANDOM FACTORS. AN "L" FLAG INDICATES A 'LESS THAN' RESULT LOWER THAN THE COMPARATOR.

AND, "HDL" MEANS THAT THE THE METHOD USED HAS A HIGH DETECTION LIMIT, AND CONSEQUENTLY, DATA IN THIS RANGE MAY NOT BE APPROPRIATE RELATIVE TO THE OTHER LABORATORIES WHICH USE MORE SENSITIVE METHODS.

general techniques to determine whether or not a result will be flagged. One approach is the statistical outlier test of Grubbs[6,7] and the other approach is the 10% or one standard deviation rule. These two approaches are described below.

9.3.1 The 10% or One Standard Deviation Rule

The general approach in discerning a flag by the 10% process is given in Fig. 11. The concept of having a 10% rule was introduced and accepted by the FP and PPWB programs since for hard surface waters at high concentrations, adequate precision (10%) was very achievable for most constituents (metals, nutrients and physical parameters). Unfortunately, at low concentrations 10% of the design values (e.g., 10% of 0.010 µg Al/L) would be a small percentage of the interlaboratory standard deviation. For this reason, the criteria to flag at low concentrations was assigned in reference to the interlaboratory standard deviation. A result at low concentrations is assigned a flag "*" if the result reported deviates from the target (or design value) by more than one standard deviation. The interlaboratory standard deviation used is that calculated for each sample within each study. Results for manganese in Table 19 illustrates the above flagging process.

9.3.2 Grubb's outliers (FP and PWGB QA Program)

The second method for flagging data in the FP and PPWB QA program is the method of Grubbs[6,7,8]. This statistical test discerns an outlier when a result deviates from the population mean by more than two times the standard deviation. Mathematically this requires calculating the Grubb's statistic for the suspect values (highest or lowest) and if this statistic exceeds a critical value the result reported is declared an outlier. Details of the procedure are found in the literature.

An example of a Grubb's outlier is given for manganese in Table 19. The outlier is noted by the letter "R". Refer to Fig. 12 for a graphic example of an outlier.

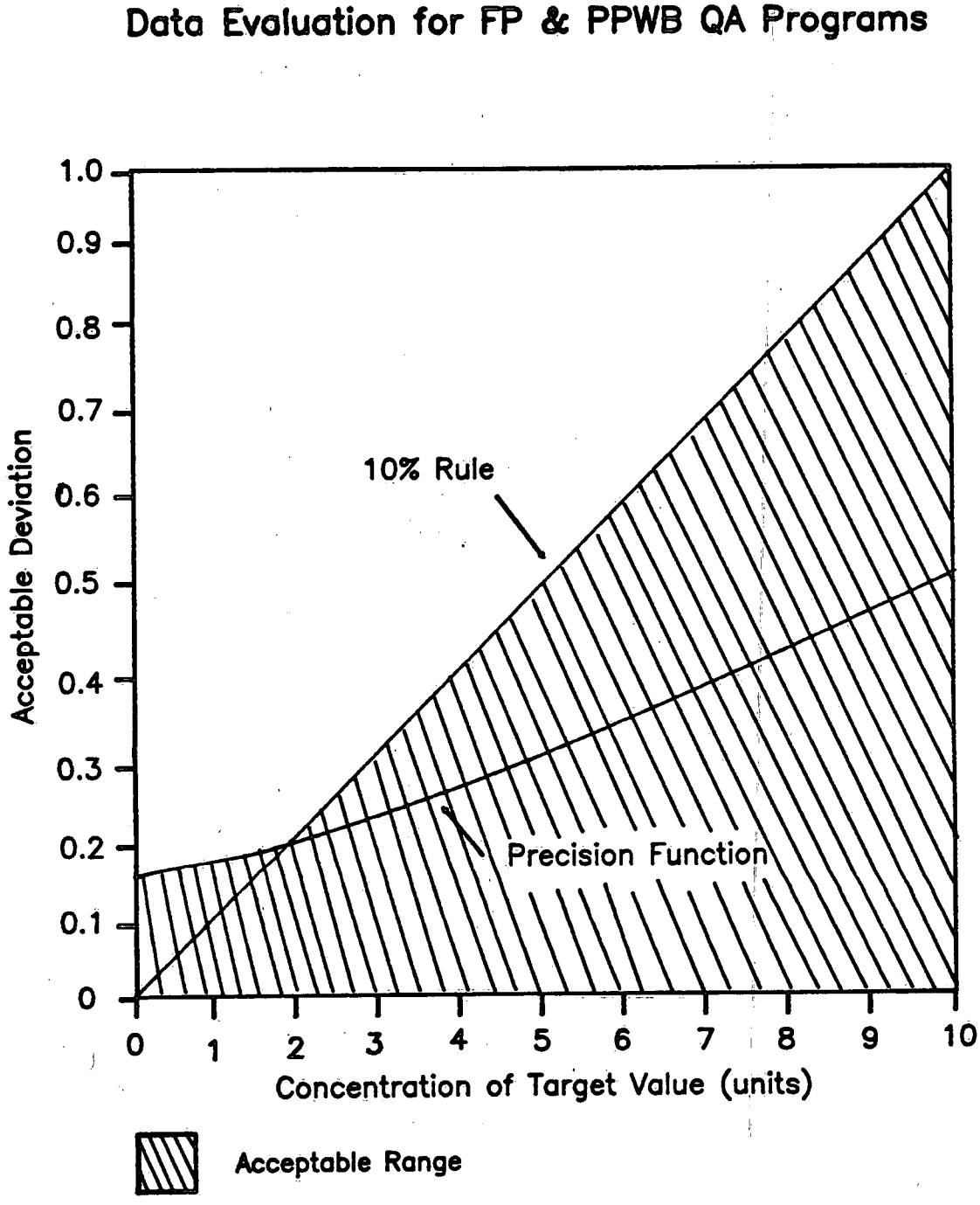


Fig. 11 Evaluation of Data based on 10% & 1 Std. Dev. Rule

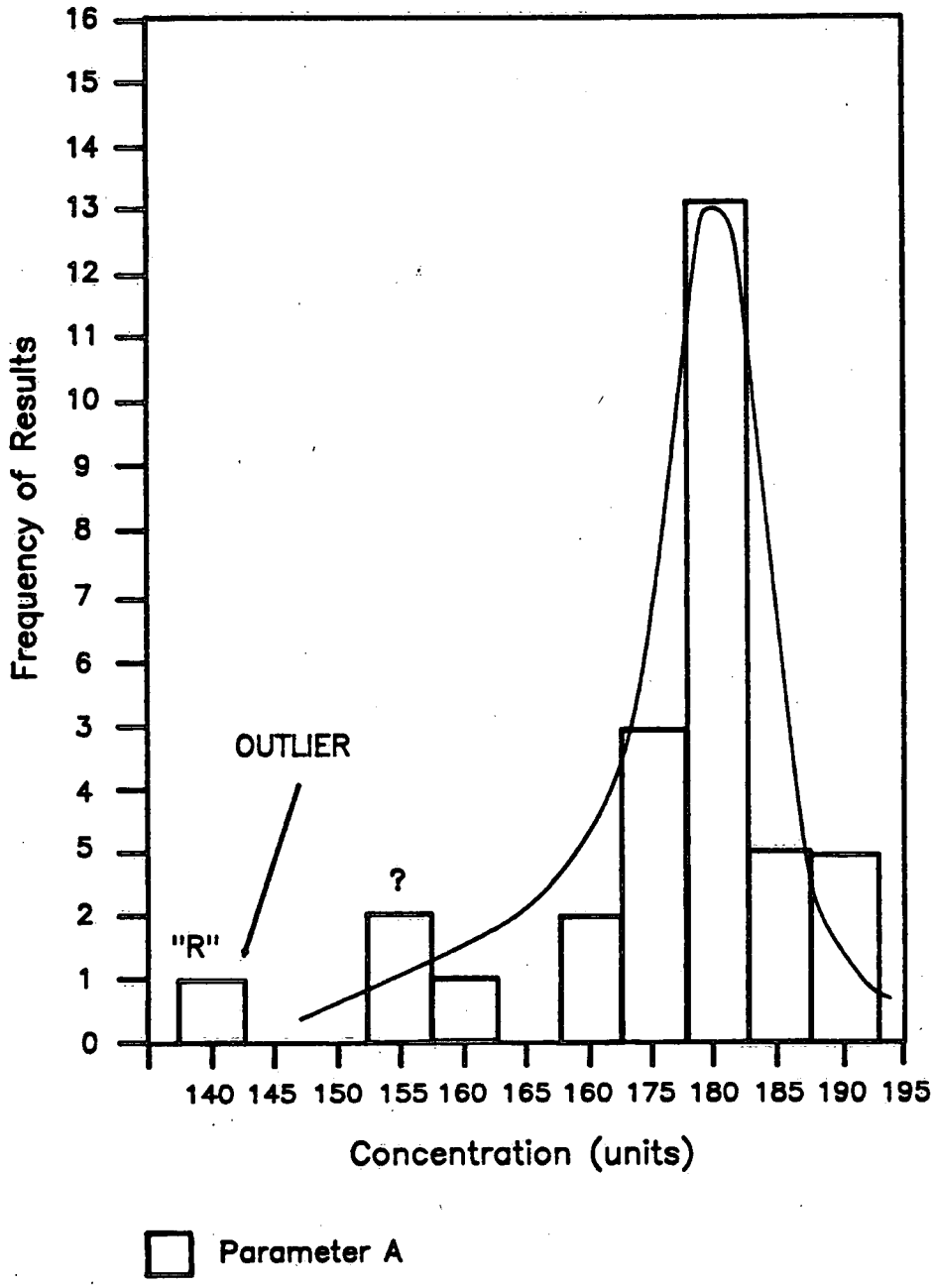


Fig. 12 An Example of a Statistical Outlier (Grubbs)

9.4 Laboratory Performance

For each bimonthly study, laboratory heads and their managers are shown their laboratory's comparability of data and the laboratory's performance in terms of the number of flagged results. A sample listing of flagged data in a typical study is given in Table 20 (Table of Flagged Data).

To enhance the value of this table, computer programs are currently being written, to (a) reveal the percentage of flags; (b) generate appraisals via a computer program and (c) have graphics reveal how performance varies over time.

9.5 Impact of the FP and PPWB QA Studies

One of the key features of the original Federal-Provincial QA studies (referred to as the IRQC - the inter-regional quality control) was the swift comparison of information on data and methods for a long list of parameters. The data and information program implemented in the early 1970's is currently very active and continues to provide rapid service. A typical report form is given in Table 19. When each laboratory reports data, it includes a NAQUADAT code for each methodology. A new methodology receives a particular code assigned to the method after application to and approval from NAQUADAT officials (Environment Canada - Ottawa). The assigned codes are essential in tabulating data by methods. Such tabulations of data by method are useful to laboratory heads when they are evaluating the performance of their own laboratories. A typical summary given to each laboratory is given in Table 20. This summary is circulated to all participants, laboratory managers and program managers. On the issue of inadequate methodology the QA chemist often refers a laboratory manager to another laboratory that has more suitable methods and/or his satisfactory performance. This is viewed as a constructive transfer of information.

The FP and PPWB QA programs' most important feature is the quick response evaluation. Reports indicating results that are suspect are often returned within four weeks, and by noting that results are flagged, the laboratory manager can take corrective action to discover the possible error sources within the measurement systems. Anomalous results may relate to a random error, a blunder, an incorrect calibration, poor precision or unsuitable method or simply inadequate QC procedures within the laboratory. Laboratory managers agree that the overall effort within the quick response evaluation is constructive.

9.6 References

1. Task Two Report, April 1981. Water Quality Analytical Methods Task Force, Prairie Provinces Water Board, Regina, Saskatchewan.
2. James P. Dux. Handbook of Quality Assurance for the Analytical Chemistry Laboratory. Van Nostrand Reinhold Co., NY, 1986.
3. V. Cheam and A.S.Y. Chau. Manual for the Bi-Monthly Inter-Regional Quality Control Studies. Manuscript No. 48-AMD-82-VC, 1982, National Water Research Institute, P.O. Box 5050, Burlington, Ontario.
4. D.T.E. Hunt and A.L. Wilson. Chemical Analysis of Water. Second edition, 1988, The Royal Society of Chemistry, London.
5. R. Caulcutt and R. Boddy. Statistics for Analytical Chemistry. 1983, Chapman and Hall.
6. F.E. Grubbs. Procedures for Detecting Outlying Observations in Samples. Technometrics Vol. II, p. 1-21, 1969.
7. M.J. Mazu. Detecting Outlying Observations. In Transactions of the 35th Annual ASQC Quality Congress and Exposition, San Francisco, May 26-29, 1981.
8. J.K. Taylor. Quality Assurance of Chemical Measurements. Lewis Publishers Inc., 121 Main Street, P.O. Drawer 519, Chelsea Michigan, 48118.

10.0 OTHER EVALUATION TECHNIQUES

Chapters 9 and 10 have described how interlaboratory data are evaluated for the major external QA studies which are conducted routinely by the Quality Assurance Group in the National Water Research Institute. These methods represent the more usual evaluation techniques but are often supplemented by other techniques as described below.

10.1 Outliers

No measurement system is absolutely free from error or uncertainty. Hence data from interlaboratory studies will fall into some form of distribution pattern (e.g., normal, skewed, etc.). Some results may be extreme and may not fall within a normal and expected range. Such results are referred to as outliers[1,2]. An example of such an outlier is given in Fig 12.

Inclusion of outliers into simple statistics such as calculated means (or averages) or variances will taint or bias the calculated estimates. If confidence intervals are required the inclusion of outliers means broader intervals that may cause analysts to make erroneous decisions about the performance of a method of a laboratory.

How one addresses outliers can be summarized into several categories. One can arbitrarily throw out any results (high or low) that appear suspicious. One can be firm and use all results and exclude nothing. On the other hand one can be cautious, analyze all methods and results seeking out all possible reasons for a deviant result and then accept or reject the result or results. A fourth and more cautious approach is to use traditional statistical techniques to identify outlying values such as described by Grubbs, Dixon, and Ferguson. Many techniques exist and literature is found in the list of references. The Grubbs technique is now routinely applied in the FP and PPWB programs outlined in Chapter 9.

Whichever outlier detection method is adopted it is very critical to be cautious in rejecting or removing of data. This is

especially so for large interlaboratory data files since the distribution of results may be bimodel or multi modal, a situation which is influenced by the diffrences in methods. Sometimes, it may be necessary to separate data and analyze the individual data groups.

The two principle external programs described earlier (Chapters 8 and 9) differ in their approach to outliers. In Chapter 8, the concept of an outlier is not even considered since no data are rejected. All extreme results (high or low) are ranked and used to estimate bias for the whole data set. Very extreme results merely accentuate the bias. If many results are extreme the decision is either severely biased or simply erratic. Although not used in Chapter 8, outliers and their detection certainly is an issue implied when flagging data. The flags assigned are not ligitimate outliers but simply a mild warning to each lab that their results are deviating. A warning is implicit that some internal investigation should be considered. An interesting and perhaps ligitimate outlier appears in Chapter 8 where the program APPRAIS yields an out of control statement. This decision occurs when the laboratory is very competent on all but one of 10 measurements (refer to Fig. 2, Chapter 1). In this case (out of control), the isolated result is severely different.

10.2 The Youden Paired Sample Approach

Data are also graphically evaluated by the paired-sample plot technique originated by Youden[3]. The technique requires that the two samples be of similar composition and analyte concentration. It has been successfully used in several national interlaboratory studies. The two common merits that the plot offers are the visual display of (a) data quality[4] of each laboratory; and (b) methods performance[5]. For example, Fig. (13) with the circles as acceptability limits, reveals the data quality of arsenic determinations by the various laboratories, whereas Fig. (14) effectively compares the performance of various methodologies used for analyzing SO_4^{-2} in colored waters.

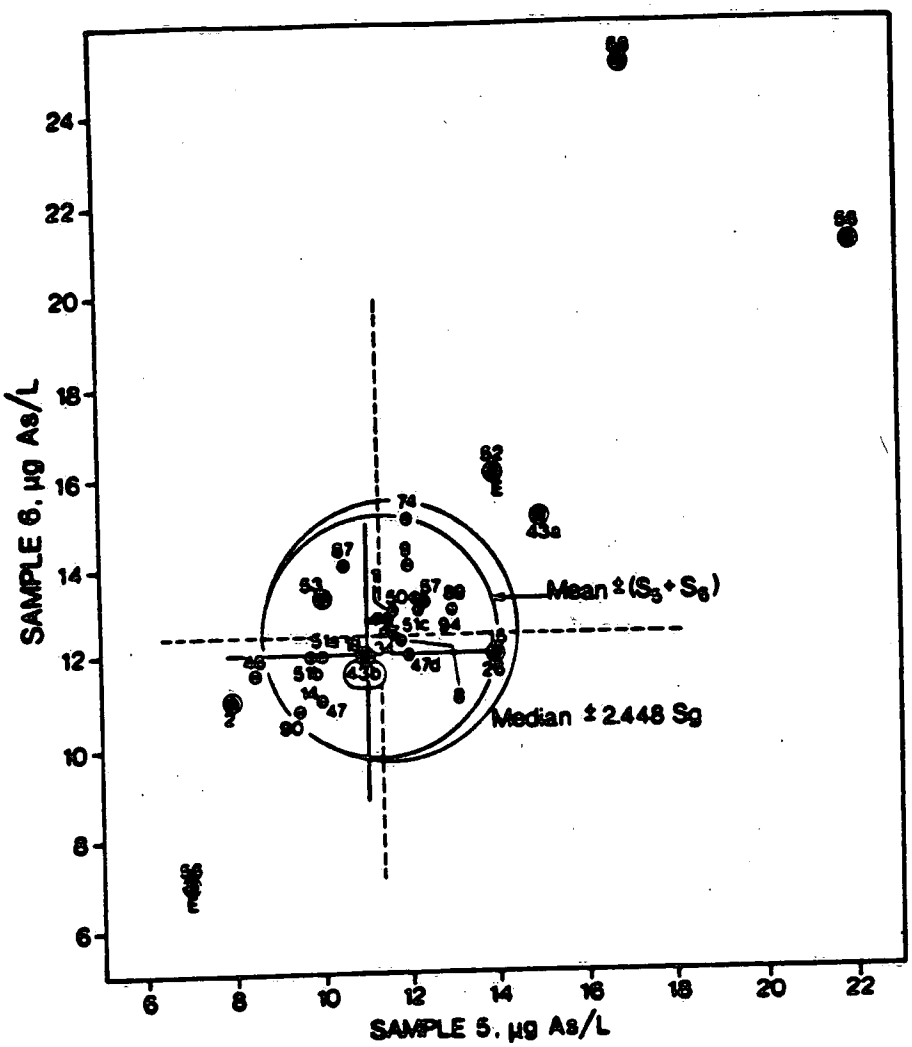


Figure 13 Arsenic paired sample plot for samples 5 and 6.

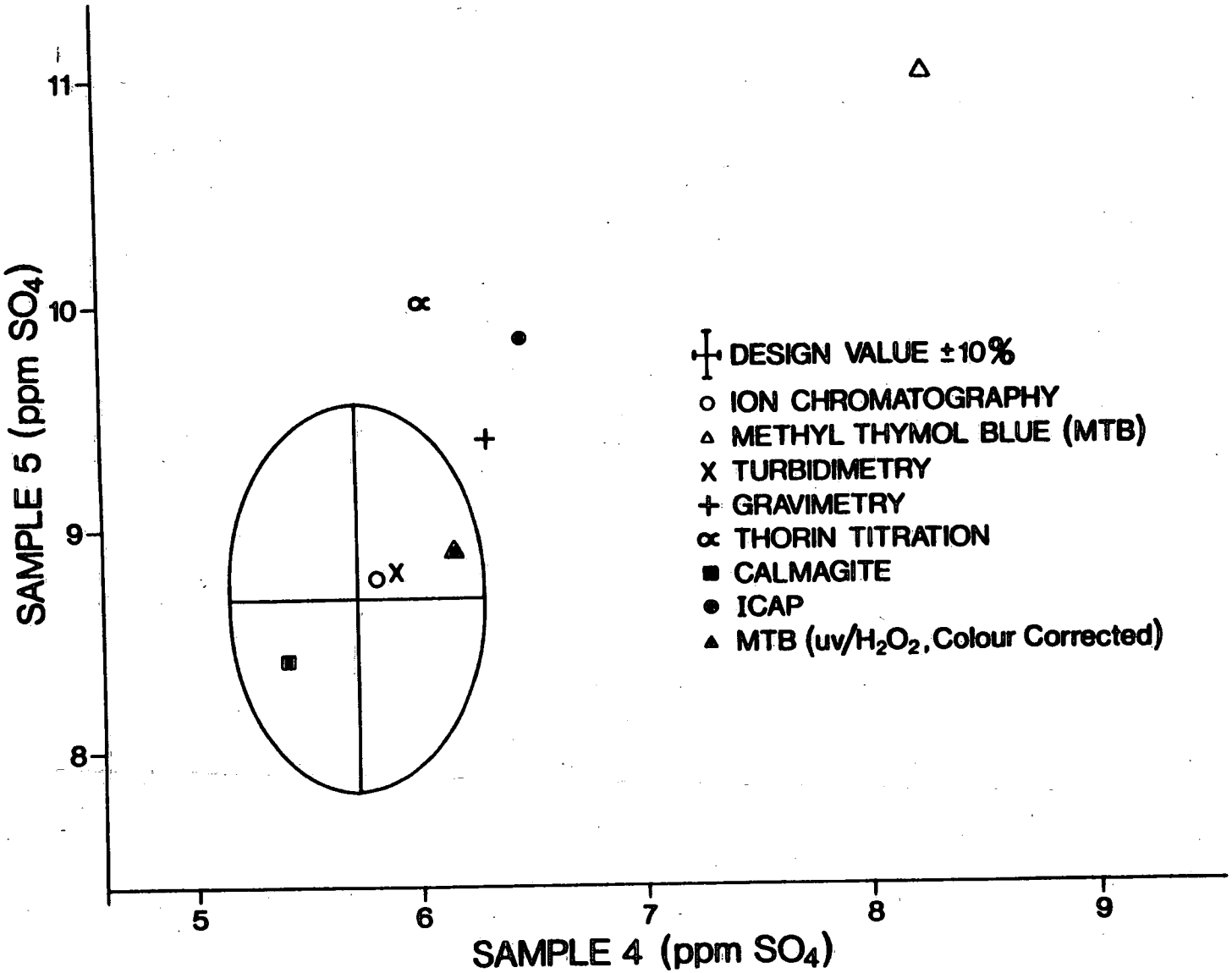


FIGURE 14. COMPARISON OF THE DESIGN VALUE WITH MEDIAN VALUES IN PAIRED SAMPLES 4 AND 5 DETERMINED BY VARIOUS METHODOLOGIES

10.3 References

1. F.E. Grubbs. Procedures for Detecting Outlying Observations in Samples. Technometrics Vol. 11, p. 1-21,. 1969.
2. M.J. Mazu. Detecting Outyling Observations. In Transactions of the 35th Annual ASQC Quality Congress and Exposition, San Francisco, May 26-29, 1981.
3. W.J. Youden and E.H. Steiner. 1975. Statistical Manual of the AOAC, Association of Official Analytical Chemists, Washington, D.C.
4. V. Cheam and K.I. Aspila. 1980. Interlaboratory Quality Control Study No. 26 - Arsenic and Selenium in Water. IWD Report Series No. 68, Burlington, Ontario.
5. V. Cheam and A.S.Y. Chau. 1987. National Interlaboratory Quality Control Study No. 23 - Sulfate in Colored Waters. IWD Report Series No. 75, Burlington, Ontario.

11.0 THE REPORTING PROCESS (Feedback)

When a large effective interlaboratory study is nearing completion, it is necessary to prepare two reports in quick succession.

The first report to the analyst is often a preliminary computer report highlighting only the specific data that is specific to that laboratory. This simple report allows the analyst to verify that the transcription of raw data into the computer data base system was correct and clear of transcription errors.

The second report, whether for a large or a small study, is the more important. It is a report to the analyst, laboratory manager, and any other program person connected to the laboratory (e.g., QC person or program manager). This report is a one or two page letter and includes a laboratory specific performance appraisal that is pertinent to the laboratory. Attached to the letter are all relevant supporting data such as (a) a description of test samples, (b) previous history of test sample, (c) summary of historical data on test samples, (d) a list of all participants, (e) comparison of the performance of all laboratories, and (f) a copy of all reported results.

The primary intent of this second report, normally mailed 6 to 8 weeks after a participant has first received the test samples, is to inform him/her of any problems in the measurement system of the participating laboratory.

The third report is a formality and consists of a final document published and circulated for general reference. Effort should be taken to have this completed within the year of the study.

The above reporting process relates to how the initiator of an interlaboratory study should relay information to the client laboratories. The specific protocols can be quite flexible depending on the particular program. It is important that there be rapid feedback on any problems perceived, since many measurement problems left unaddressed, can, for some large high production laboratories, quickly taint an important environmental data base.

The reporting process within a laboratory is also critical. In most cases, the quality control specialist responsible for the laboratory should be quickly notified in addition to the analyst. The process to initiate change is referred to as remedial action and requires management involvement. This issue is discussed in the next chapter.

12.0 REMEDIAL ACTION TO QA REPORT

Interlaboratory comparison studies are normally produced not for the academic benefits of the producer or sponsor but for the mutual benefits of the client laboratories and the environmental program. As such, there should always be some reaction to the final results product of interlaboratory comparison studies. The reaction will vary differently at the various levels of the management structure.

12.1 Intralaboratory Action

The bench analyst and laboratory manager are on the front-lines and are the individuals who should always receive first notice of preliminary reports to each study. If the study had been designed well and the supporting data are clear, concise and informative, then the bench analyst and the manager involved should be able to use the evidence and react to the appraisal constructively. The usual and most common problem is the calibration (e.g., bias) although in other cases it may be lack of precision (e.g., erratic data).

For large laboratories that have an active interchange of information between analysts, managers and quality control officers, the first line reaction should be a review of intralaboratory control data obtained at the date when the interlaboratory study test samples were analyzed. This review should be constructive with analysts comparing their data against that of their peers, examining method issues (if pertinent), the possibility of a calibration failure (e.g., standards), and perhaps simply poor application of the method (erratic recovery or poor precision). If the study was elaborate in design, then the study report and associated laboratory specific statements might cover some of these quality control issues. The responsibility for much of the data interpretation remains with the analyst and the pertinent manager. Feedback to the originator of the study, although not essential is sometimes of value.

12.2 Management Action

Large facilities very often have project or program managers who have a responsibility to manage several environmental projects for which analyses are carried out in one or more laboratories. These individuals should be informed on the performance of their laboratories. It is a line of authority that is often overlooked. It is an important area especially where performance in the laboratory has been judged as poorand requires internal review. How management reacts is an internal matter but react it must, if the data base it has, or will produce, is to be protected and retained as credible for program requirements.

The nature and process of reaction to a study can vary. The appraisal which states satisfactory, well done, or average performance, may be met with satisfaction. An average appraisal may yield to an internal review or audit by laboratory staff. A poor performance, with severe bias, out of control statements, or the use of the term erratic may or impel management and program managers to request the laboratory to cease data production until proof of adequate performance can be provided. This extreme reaction would be an internal management decision reached by management after careful review of the study report and the internal data quality objectives of the program.

On matters of performance it remains the responsibility of the interlaboratory program and the originator of studies to retain spare test samples on hand to support the follow up needs of the client laboratory if assistance is required to investigate identified problems. As a summary, the remedial actions to any study should be always viewed as a constructive process and indeed one with all parties involved.

12.3 Users of Data

Very often the users of environmental data are the least informed on issues of quality. Many agencies tacitly assume that data are acceptable since they have been continually told that the laboratory producing the data has internal quality control. This can become an increasingly dangerous position if the agency data files originate from several laboratories. Experience has shown that different databases can be disjointed. Data users should maintain files defining internal and external performance. This performance must match or exceed the objectives of their data needs. If they do not, acceptance limits must be defined.

13.0 GLOSSARY OF TERMS

13.1 NWRI External QA Programs

LRTAP: Long Range Transport of Airborne Pollutants. This QA program involves almost 100 laboratories. Each laboratory is provided ten different unpreserved "soft waters" three times per year and are requested to analyze up to 23 constituents (major ions nutrients and physical parameters). Twenty studies have been completed.

LRTAPP: The same program as the LRTAP but LRTAPP refers to the studies involving plant materials for nutrients and metals. These studies are provided through the Great Lakes Forestry Centre, Sault Ste. Marie, Ontario (Dr. Ian Morrison).

IJC: International Joint Commission. This QA program is in reference to the Canada-US Lakes Water Quality Agreement, Great Lakes International Surveillance Program (GLISP). The external QA studies are provided two to four times per year to about 30 to 100 laboratories. Studies include a) phosphorus in water or sewage plant effluents; b) toxic organics and inorganics in fish homogenates and sediments; and c) major ions, nutrients, physical properties and trace metals in water. Thirty studies have been completed (1976 to 1988).

Eulerian: This external QA program supports the "Eulerian Model Field Study" and involves external monitoring of the four primary laboratories that report precipitation data to Environment Canada, Ontario Ministry of Environment, US Environmental Protection Agency and the US-based Electric Power Research Institute. Twenty-four studies (one per month) involving eight laboratories are in progress. The program parallels LRTAP but is specific to rainwater.

FP: The Federal-Provincial QA program. This QA program was called the Inter-Regional Quality Control Program (IRQC) when implemented in 1974. The studies are monthly and involve trace metals and major ions. Over 158 studies have been completed.

PPWB: Prairie Provinces Water Board. This QA program involves laboratories in Alberta, Saskatchewan and Manitoba. The program runs concurrent with the FP and is similar in design (see chapter 8). Thirty-six studies have been completed.

UGLCCS: Upper Great Lakes Connecting Channel Studies. This QA program, recently completed was a bi-national program (Canada-US).

National: This is the NWRI QA program presented to all federal, provincial, university and private sector laboratories. This large program initiated in 1970 involves a diverse series of substrate and constituents. Thirty-seven studies have been completed.

FICP: Federal Interdepartmental Committee on Pesticides. The QA program is interdepartmental and involves toxic organics in a wide variety of substrates. The NWRI involvement (one study per year) involves aqueous and sediments for a variety of toxic organics.

CAPCO: Canadian Association of Pesticide Control Officials.

13.2 Specific Terms

Accuracy: Of a test method: the degree of agreement between the true value of the property being tested (or an accepted standard value) and the average of many observations made according to the test method, preferably by many observers (see also Bias and Precision).

Between-Laboratory Precision: The multi-laboratory, single-sample, single-operator-apparatus-day (within-laboratory) precision of a method; the precision of a set of statistically independent test results, all of which are obtained by testing the same sample of material and each of which is obtained in a different laboratory by one operator using one apparatus to obtain the same number of observations by testing randomly drawn specimens over the shortest practical time interval.

Bias: A constant or systematic error in test results. Bias can exist between the true value and a test result obtained from one method; between test results from two methods; or between two test results obtained form a single method, for example, between operators or between laboratories.

Certified Reference Materials (CRMs): Are stable homogeneous and well-characterized reference materials prepared in quantity having essentially identical or very similar matrices to the field program materials in order to eliminate or minimize the matrix effect between reference and test samples.

Control Charts: A charting of the variability of a procedure such that, when some limit in variability is exceeded, the method is deemed to be out of control.

Criterion of Detection: The minimum quantity (analytical result) which must be observed before it can be stated that a substance has been discerned with an acceptable probability that the statement is true. Expression of the criterion of detection must always be accompanied by the stated probability.

Erratic: The term used in evaluating laboratory performance when results are very scattered (high and low) relative to the peer group.

Grubbs: The notation used for the FP and PPWB program to discern an outlier (refer to Chapter 9, Ref. 6).

Limit of Detection: A concentration twice the criterion of detection when it has been decided that the risk of making a Type II error is to be equal to that of a Type I error.

Non-parametric: In reference to an ordering process and a statistical approach that requires no knowledge of the distribution of data.

Median: For a series of results the median is the middle value.

Parametric: In reference to the statistical approach that implies knowledge or asumptions of the distribution of data.

Out of control: The term used in evaluating laboratory performance. When all data for a laboratory are satisfactory except for one result which is significanty different.

Precision: In general, the degree of agreement within a set of observations or test results. Various measures are in use. The measure, the set of samples (and concentration range) used to calculate it, and the extent of the sampling/analytical system to which it applies, must be stated with the numeric value of the measure. The measures used are usually inverse measures of precision, such as the standard deviation or relative standard deviation.

Quality Assurance: Activities that define the way in which tasks are to be performed to ensure a final product that meets pre-defined data quality goals. Quality assurance ensures that operations and procedures requiring control are identified, and that effective control protocols are defined and implemented.

Quality Control: Operational level activities to determine and verify the suitability of data generation procedures, equipment and materials, in relation to acievement of predefined data quality goals and to identify and eliminate measurements that do not meet these goals. Data quality goals must be reduced to quantitative control limits for this purpose.

Quality Management: Management activities undertaken to ensure that staff are informed of their responsibilities to establish, maintain and document a defined level of data quality, and are held accountable for achieving these goals. Quality management includes documentation of the management structure, and explicit endorsement of data quality goals, audits and procedures.

Quality Planning: An exercise which ensures that resources are used wisely by defining data qualtiy needs in advance, in an explicit manner that permits objective assessment of whether these needs have been achieved.

Reference Materials (RMs): Are similar to CRMs except they are less rigorously characterized.

Relative Difference (%): As an accuracy metric, the difference between the mean measurement of a sample and a reference value, divided by the reference value, multiplied by 100.

Relative Standard Deviation (%): As a precision metric, the standard deviation of replicate measurements of a sample, divided by the mean of those measurements, multiplied by 100.