NWRI Contribution 90-135

# ASSESSING THE PERFORMANCE OF LABORATORIES IN LARGE EXTERNAL QUALITY ASSURANCE PROGRAMS

by

K.I. Aspila and A.S.Y. Chau

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

September 1990

## MANAGEMENT PERSPECTIVE

A successful strategy in a quality management process is to include an external quality assurance program with effective and informative techniques to rapidly analyze data in order to define the performance of a group laboratories. Information on poor performance, indicated by very inaccurate or very imprecise laboratory measurement systems, must be swiftly conveyed to the laboratory or program manager in order to implement corrective action and internal review. This paper describes such procedures to assess performance in laboratory measurement systems.

Dr. J. Lawrence
Director
Research and Applications Branch

## PERSPECTIVE GESTION

Il est bon d'incorporer au processus de gestion de la qualité un programme externe d'assurance de la qualité dont les techniques, informatives et efficaces, permettent d'analyser rapidement des données pour évaluer le rendement d'un groupe de laboratoires. Dans les cas où le rendement n'est pas satisfaisant, ce qui se voit par la grande inexactitude ou la grande imprécision des systèmes de mesure, il faut promptement signaler la chose au gestionnaire du laboratoire ou du programme, afin que les mesures correctives et la révision interne nécessaires soient mises en oeuvre. Dans cet article, il est question des méthodes employées pour évaluer le rendement des systèmes de mesure des laboratoires.

J. Lawrence
Directeur
Direction de la recherche et des applications

## ABSTRACT

Interlaboratory quality assurance studies (external QA) are effective techniques to monitor the performance of laboratories analyzing environmental constituents. Environment Canada through the National Water Research Institute conducts intercomparison studies on waters, sediments and fish for inorganic and toxic organic constituents. Over 400 laboratories in Canada and the United States are involved in eight different ongoing external QA programs. Most studies incude over 50 laboratories that analyze 15 to 30 different constituents in ten different samples. Data sets are addressed by non-parametric statistics (ranking) in order to discern laboratory measurement bias (a systematic error high or low). Individual sample results that deviate significantly from interlaboratory medians are flagged either very high, high, low, or very low. In large studies (10 samples, 50 labs, 20 constituents) excellent performance is recognized when a laboratory has a very low frequency of flags and bias. Poor performance is identified when laboratories have a high frequency of bias and flags. Performance over time is analyzed through a data base management system (System 2000) and evidence clearly indicates many laboratories improving.

Rendement dans les grands programmes d'assurance de la qualité

## RÉSUMÉ

Les études d'assurance de la qualité interlaboratoires (AQ externe) sont un moyen efficace de surveiller le rendement des laboratoires où se fait l'analyse de composants environnementaux. Par l'entremise de l'Institut national de recherche sur les eaux, Environnement Canada mène des études comparatives sur l'analyse des composants inorganiques et des toxiques organiques des eaux, des sédiments et des poissons. Au Canada et aux États-Unis, plus de 400 laboratoires participent à huit programmes externes d'AQ permanents. La plupart des études portent sur plus de 50 laboratoires où 10 échantillons sont traités pour l'analyse de 15 à 30 composants différents. Les ensembles de données sont traités par des méthodes statistiques non paramétriques (classement), ce qui permet de déceler les cas où les mesures sont biaisées (erreur systématique, grande ou petite). Les résultats de l'analyse d'un échantillon qui s'écartent dans une mesure significative des médianes interlaboratoires sont classés très élevés, élevés, peu élevés ou très peu élevés. Dans les études de grande envergure (10 échantillons, 50 laboratoires, 20 composants), les laboratoires où les biais et les classements sont de très faible fréquence sont considérés comme ayant un excellent rendement. On parle de mauvais rendement lorsque les biais et les classements sont de fréquence élevée. On analyse l'évolution du rendement au moyen d'un système de gestion de base de données (Système 2000) : les résultats indiquent une nette amélioration dans de nombreux laboratoires.

## KEY WORDS

Quality Assurance
Environment
Performance
Interlaboratory Studies

Performance in Large QA Programs

## BACKGROUND

For over 15 years the Quality Assurance Group, at the National Water Research Institute has provided external quality assurance support to federal, provincial and international environmental programs. This support has included interlaboratory studies and the associated performance appraisals. Some external QA programs are listed in Table 1. Interlaboratory studies have included softwaters, rainwaters, sediments, standard solutions and fish. The constituents have included pesticides, various toxic organics, trace metals, major ions, nutrients and physical parameters. How such studies are administered is documented in a quality assurance manual(1).

Not all external QA studies are classified as large. The large studies are those in the LRTAP (Long Range Transport of Airborne Pollutants Program), IJC (International Joint Commission studies on the Great Lakes), some National Studies (within Canada) and the Eulerian Model Program. These studies normally involve up to 50 to 100 laboratories which are provided 10 to 15 test samples for analysis of up to 15 to 30 constituents. Small QA studies include few samples, few laboratories and a small number of constituents. Such studies can be addressed by manual processing whereas the large studies need more elaborate computer processing. Details on the data handling are described elsewhere(1).

## KEY ISSUES

When one compares the results reported in an intercomparison studies to true value targets or the consensus value (interlaboratory medians) one can quickly organize individual performance into various groups. These groups are presented graphically in Figures 1a and 1b. These figures are illustrations of extreme performance. If external studies are frequent and provide rapid performance appraisals then corrective action can dramatically improve the quality of laboratory measurements.

## DATA BASE MANAGEMENT SYSTEMS

To administer effective interlaboratory QA programs that create a vast amount of data has required an electronic processing system. A data base system was developed in the late 1970's and is now routinely used to handle almost all interlaboratory data. The structure of this data base system (System 2000) is given figuratively in Figure 2. The system archives virtually all pertinent information on the laboratory, including all results reported, the history of all test samples and all laboratory appraisals generated for each participant. The merits of this system are recognized by its ability to provide a rapid track record on laboratory performance and on the quality of the QA test samples. Additional information is described in the QA Manual (1).

## DATA ASSESSMENT

Two distinctly different techniques are employed to assess data from large studies. The first is the discernment of bias using a non-parametric technique and the second is a simple assignment of a flag to a result when it that result deviates significantly from the interlaboratory median. These two procedures are described below.

## RANKING TO DISCERN BIAS

The Youden bias assessment technique (2,3,4) 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) is ranked such that the lowest result has a rank 1 assigned, the second lowest has a rank 2 and so on. The highest result has a rank of 50 if there are 50 laboratories. When laboratories report "equal values" then the rank assigned is an average. Examples are provided by Youden (2-5).

Performance in Large QA Programs

The next step in the ranking process is to review the total laboratory rank (sum of the ranks) or the average rank. The immediate impact is recognition that some laboratories have an overall reach which is very high or very low. The question to resolve is whether these anomolously 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 of declaring a laboratory biased when it is not, is one chance in 20 (5%).

A description of this process is given in Table 2. This example is derived from a LRTAP Study. Youden's original work(2,3) describes total ranks for which a matrix of critical ranks were calculated manually. The probability calculations described in this paper were developed by Clark(4,5) and are parallel to those of Youden. Both methods provide very informative statements when appraising interlaboratory results(1) for systematic errors in the laboratory measurement system.

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. Laboratories with

Performance in Large QA Programs

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. 1).

In other situations, where results by a laboratory are so erratic, the statistical inference is "no bias", simply because some results are extremely high and some are extremely low. A severe laboratory problem nevertheless exists. To address this issue a flagging formula was adopted to address "poor" intra-laboratory precision.

The Youden bias assessment in many large studies can successfully address and discern the presence of inaccuracy in the laboratory measurement process. The rigor with which this method identifies inaccuracies is clouded when serious blank issues occur or if the entire group of laboratories is 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).

## FLAGGING RESULTS

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 bias cannot be easily determined since there are on average, very high and very low results. Fig. 1 provides examples.

## Performance in Large QA Programs

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 increases almost linearly fom low concentration to high concentration (see Fig. 3).

The relationship between interlaboratory precision and concentrations allows for the 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. 3) is referred to as the concentration error increment (CEI). These three variables (CEI, LLBAE and BAE) are given in the schematic (Fig. 4).

The relationship between the observed precision function and the flagging formula is quite clear. The principle issues to be resolved 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. Flagging criteria chosen can be adjusted so that about 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 results are those that deviate from median values by 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.

Performance in Large QA Programs

## PERFORMANCE WITHIN A STUDY

A data base output computer program referred to as LABCOMP provides information on the relative performance of laboratories within a single study. An example output is given in Table 3. Its primary purpose is to provide each laboratory with a precise statement on its relative performance with respect to its peer group within a study. It is particularly useful for large studies involving many laboratories that analyze many different constituents.

The computer program can isolate and also accommodate any group of parameters. A very wide choice of outputs are available. The example output (Table 3) 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 one particular laboratory to other 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 indicates poor performance.

The output table created by the computer program LABCOMP illustrates relative laboratory performance. 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

Performance in Large QA Programs

are biased, then the score will be 60%. Some labs are both biased and flagged and can have very high scores and are declared poor within the study. Corrective action is required.

Experience in analyzing many studies has created performance guidelines (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 good 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%) is cause for immediate internal review. To this end this output program (LABCOMP) has merit and evidence now accumulating suggests the impact for many laboratories has been constructive.

**PERFORMANCE OF A GROUP OF LABORATORIES BASED ON FREQUENCY OF BIAS AND FLAGS (FLGTBL)**

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 LRTAP QA studies involve about 50 laboratories who in general analyze the same constituents. With this resource (almost 20 major ion studies) it is now possible to compare the frequency in which laboratories have their data assessed as biased or flagged and it is possible to provide a track record on the performance of each laboratory over time.

Performance in Large QA Programs

Within any study, a laboratory that gets most of its parameters declared biased and has most of its results flagged, is considered as having a very poor performance while 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, helps in this effort. An example output is given in Table 5. This output is the integrated results of LABCOMP and when transferred from the mainframe to a personal computer, performance can be graphically displayed. Because the output is created from a data base management system, it is possible to create an 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. 5. The performance index in this figure (and Table 4) are the same as used in LABCOMP. They are arbitrary and may be modified when all evidence has been reviewed.

## AUTOMATED APPRAISALS

When the original bias assessment techniques 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 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. When the program "Apprais" is initiated it retrieves 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 5. It is this appraisal that is attached to a covering letter accompanied by all support data (LABCOMP, FLGTBL and YOUDS2K) This critical support information is essential and is provided to each participant when a study is formally completed.

## SUMMARY

This paper has described the basic elements used in evaluating large interlaboratory studies. The techniques have been applied in over 38 studies over the past seven years. Work is currently under way to utilize the information base to create accurate precision functions to define the criteria for flagging and to employ the statistical tests to estimate the magnitude of laboratory measurement bias. Feedback from client laboratories has been very positive and preliminary assessments from graphics illustrate that many laboratories are improving precision and reducing their bias. In brief, the external QA program has had a positive impact on the quality of data produced by the participating laboratories.

## ACKNOWLEDGMENTS

The authors extend a sincere thanks to Bob White and Dr. John Clarke of the International Joint Commission Regional Office, (Windsor, Ontario) for their development of the computerized technique to address bias. A sincere thanks is also extended to Don King of the Ontario Ministry of Environment (Rexdale, Ontario) for his insight into demonstrating the utility of using bias and flag frequencies to identify laboratory performance. Mrs. Karon Miles is acknowledged for the patient development of the data base computer programs that have made the external quality programs a rewarding experience for all involved.

Performance in Large QA Programs

## REFERENCES

1. K.I. Aspila. A Manual for Effective Interlaboratory Quality Assurance. NWRI Contribution No. 89-99, National Water Research Institute, P.O. Box 5050, Burlington, Ontario, L7R 4A6, February, 1989.

2. 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.

3. 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).

4. 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, Alexandria, Virginia, April 8-10, 1981.

5. 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).

## TABLE 1.

## EXTERNAL QUALITY ASSURANCE PROGRAMS

| PROGRAM | NUMBER OF LABS | CLIENTS |
|---|---|---|
| LRTAP | 102 | US-Canada Acid Rain Labs |
| IJC | 140 | Great Lakes Surveillance |
| UGLCCS | 16 | Bi-National (connecting channels) |
| National | 230 | Canada (national program) |
| FP and PPWB | 23 | Federal-Provincial program |
| CAPCO (FICP) | 110 | Pesticide Labs |
| Eulerian | 8 | US-Canada (Acid Rain) |
| National Dioxin QA | 20 | Commercial and Federal |

Special Studies:
- PCB fire, St. Basile-le-Grand
- Tainted fuels program
- Groundwater QA Studies
- Pulp and Paper Mills (Dioxin issue)

1) UGLCCS: Upper Great Lakes Connecting Channels Study
2) FP: Federal-Provincial Quality Assurance Program
3) PPWB: Provincial Prairie Water Board
4) FICP: Federal Interdepartmental Committee on Pesticides
5) CAPCO: Canadian Association of Pesticide Control Officials

Table 2: YOUDN21 Output

PARAMETER: 16001 SULFATE NON IC METHODS mg/L

LRTAP 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 2: 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 MTB |
| L085 | 44.00 | 4.400 | 10 | | | | ICPS |
| L086 | 51.50 | 5.150 | 10 | | | | AA-MTB |
| 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-MTB |
| L006 | 68.00 | 6.800 | 10 | | | TECHNICON |
| L067 | 74.50 | 7.450 | 10 | | | AUTO MTB |
| L004 | 78.50 | 7.850 | 10 | EHEH | BIASED HIGH | 16306 |
| OVERALL AVERAGE RANK IS | | 5.361 | | | | |

SULFATE NON IC METHODS

Performance in Large QA Programs

Table 3.: 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

Table 4 : 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 | | |
| 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)

Performance in Large QA Programs

Table 4: 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 |
|---|---|---|---|---|---|---|---|---|
| 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.),

Table 5 : 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<br>RANKING INDICATES RESULTS ARE BIASED LOW |
| NITRATE + NITRITE | SATISFACTORY |
| AMMONIA | SATISFACTORY |
| TOTAL KJELDAHL NITROGEN | NO RESULTS REPORTED. |
| SODIUM | FLAGGED EXTREMELY HIGH ON SAMPLE 9<br>THIS EXTREMELY HIGH RESULT SUGGESTS<br>THE MEASUREMENT PROCESS IS OUT OF CONTROL |
| MAGNESIUM | FLAGGED HIGH ON SAMPLE 4 10 |
| REACTIVE SILICA | FLAGGED HIGH ON SAMPLE 6 7<br>FLAGGED VERY HIGH ON SAMPLE 5 |
| SULFATE,IC METHOD | ALTHOUGH NO RESULTS ARE FLAGGED<br>RANKING INDICATES A SLIGHT BIAS HIGH |
| SULFATE NON IC METHODS | FLAGGED EXTREMELY LOW ON SAMPLE 1<br>THIS EXTREMELY LOW RESULT SUGGESTS<br>THE MEASUREMENT PROCESS IS OUT OF CONTROL |
| CHLORIDE IC | SATISFACTORY |
| CHLORIDE NON IC METHODS | SATISFACTORY |
| POTASSIUM | ALTHOUGH NO RESULTS ARE FLAGGED<br>RANKING INDICATES A SLIGHT BIAS HIGH |
| CALCIUM | SATISFACTORY |

Performance in Large QA Programs

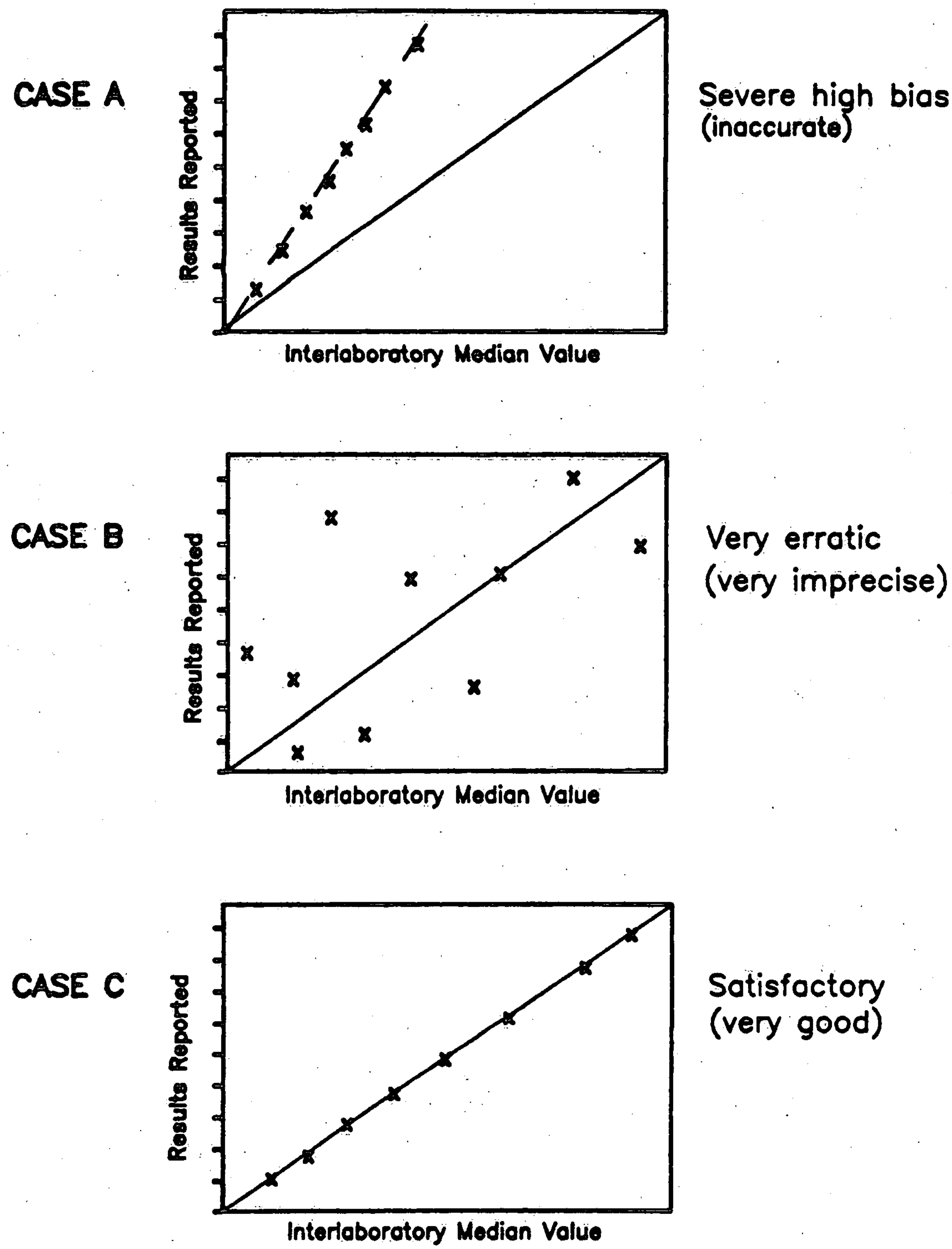



**Fig. 1a Some typical types of Laboratory Performance revealed by External QA Studies**

Performance in Large QA Programs

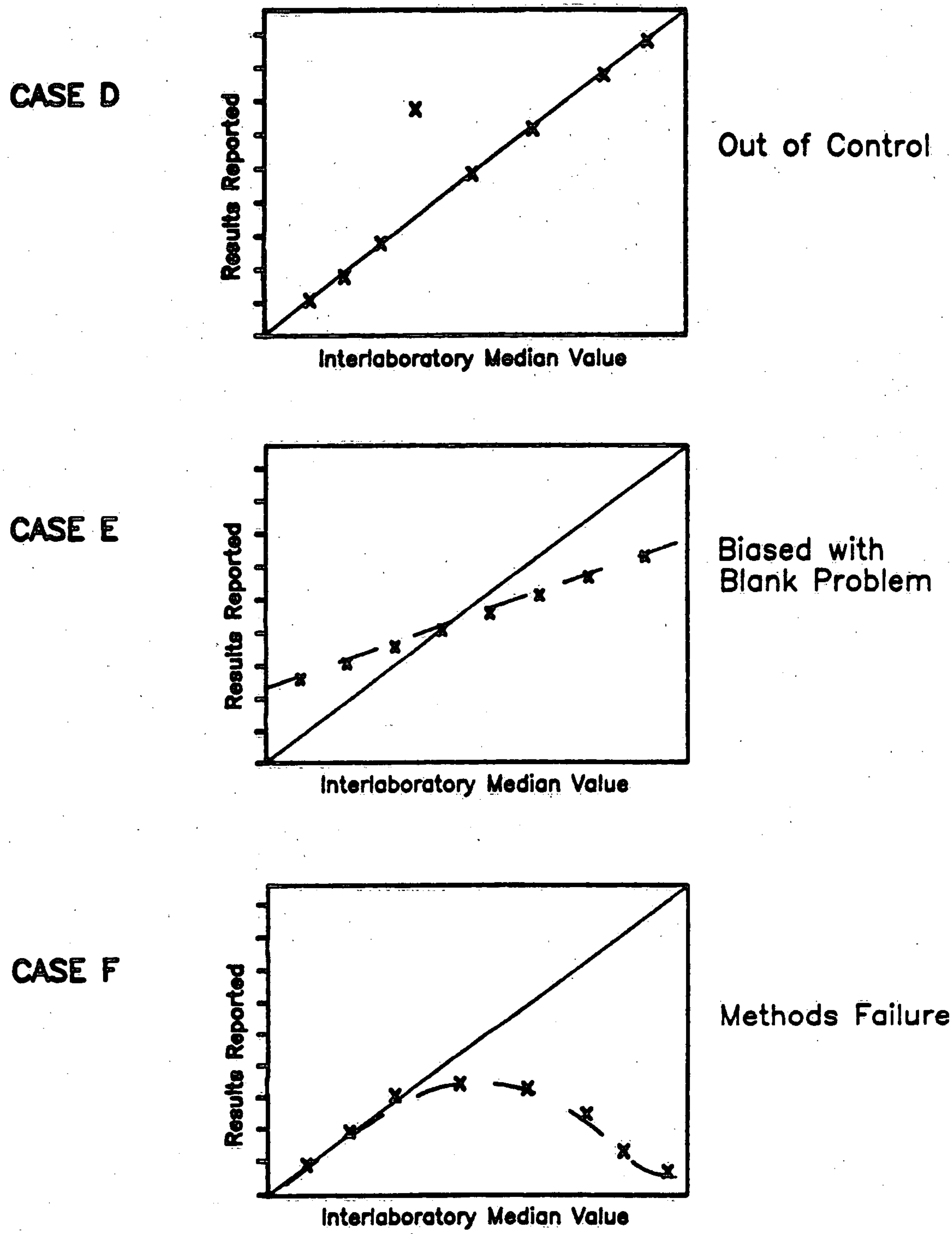



**Fig. 1b Some typical types of Laboratory Performance revealed by External QA Studies**

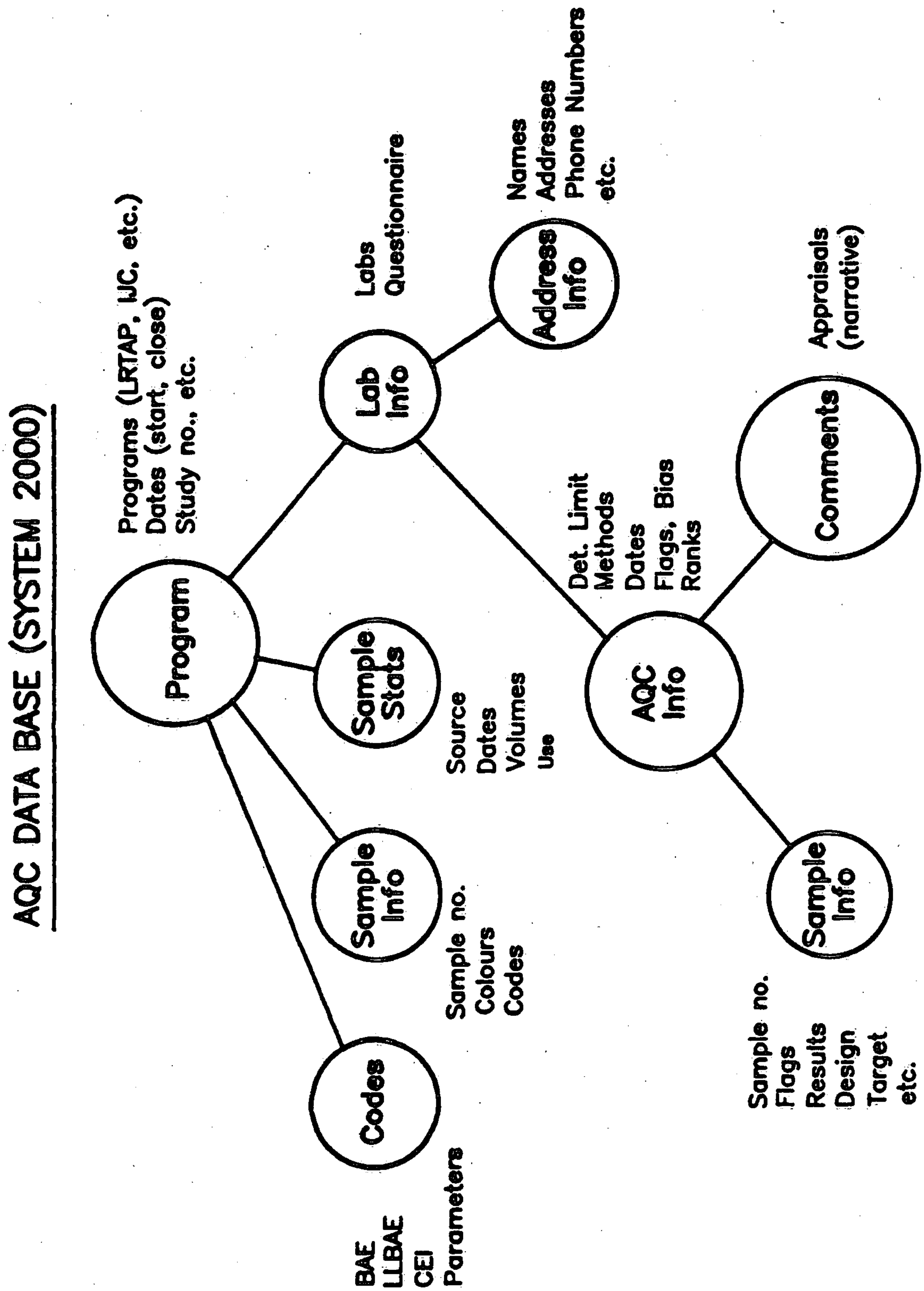
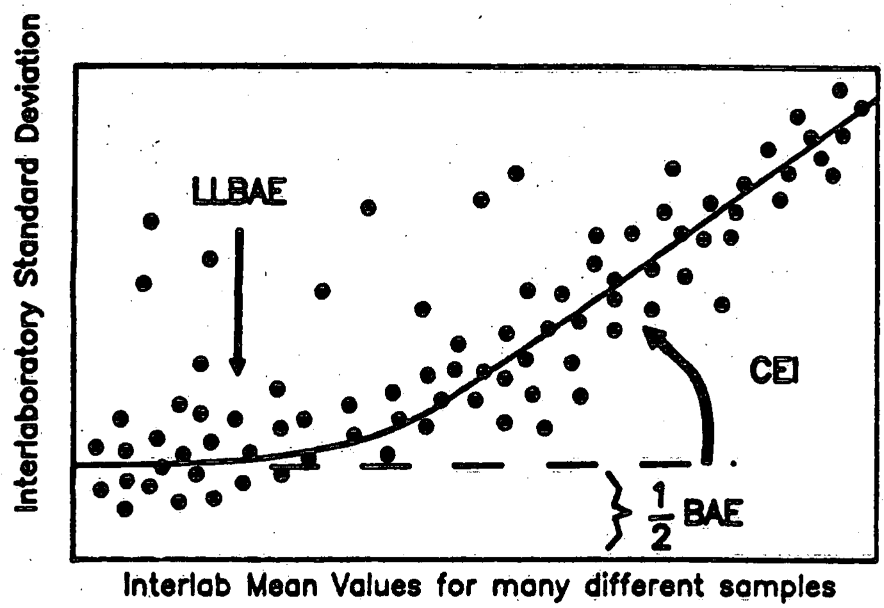


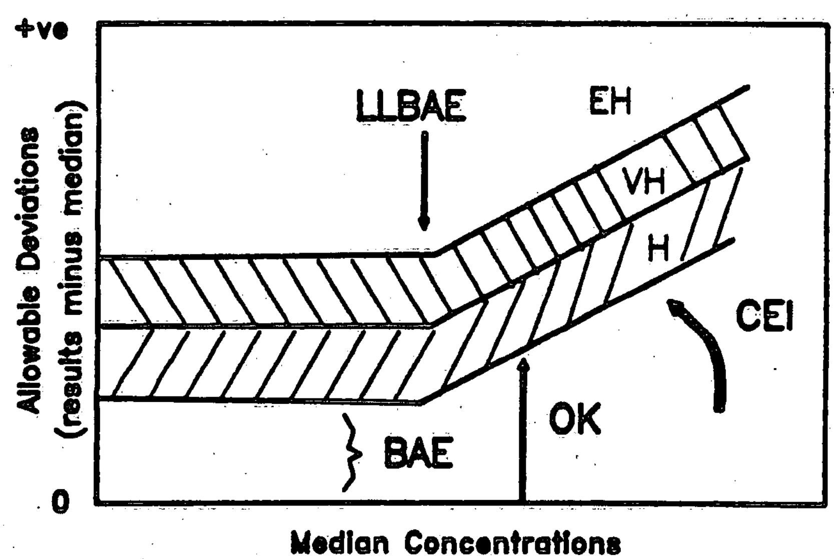

Fig. 2 AQC Data Base System

Performance in Large QA Programs




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

**Fig. 3 A Typical Precision Function**




NOTE: Flags assigned are EH,VH,H

**Fig. 4 An Example of Flagging Criteria**

Performance in Large QA Programs

Fig. 5