NBSIR 73-163

A Proficiency Test Assessment of Clinical Laboratory Capability in the United States

Peter W. Finkel, Ted R. Miller

Technical Analysis Division
Institute for Applied Technology
National Bureau of Standards
Washington, D. C. 20234

May 1973

Final Report

Available NTIS - #COM74 10542 (163page)
see IR73-197

Supersedes NTIS Nos. COM 73-11190 and COM 73-11193

Prepared for
Division of Health Evaluation
Office of the Assistant Secretary for Planning and Evaluation
Department of Health, Education, and Welfare
Washington, D. C. 20201

NBSIR 73-163

A PROFICIENCY TEST ASSESSMENT OF CLINICAL LABORATORY CAPABILITY IN THE UNITED STATES

Peter W. Finkel, Ted R. Miller

Technical Analysis Division
Institute for Applied Technology
National Bureau of Standards
Washington, D. C. 20234

May 1973

Final Report

Prepared for
Division of Health Evaluation
Office of the Assistant Secretary for Planning and Evaluation
Department of Health, Education, and Welfare
Washington, D. C. 20201

U. S. DEPARTMENT OF COMMERCE, Frederick B. Dent, Secretary
NATIONAL BUREAU OF STANDARDS, Richard W. Roberts, Director

ACKNOWLEDGMENTS

The authors gratefully acknowledge the participation of the following persons during the conduct of this study:

Dr. Margaret M. Gurney, Statistical Research Division, Bureau of the Census, for consultation in the sampling of the various laboratory groups;

Dr. Joan R. Rosenblatt, Statistical Engineering Section, Applied Mathematics Division, NBS for technical review and critique of the many drafts and the final version;

Drs. Roy N. Barnett, Joseph H. Boutwell, Harvey Gralnick, James D. MacLowry, W. Wayne Meinke, M. Keith Weikel and Donald S. Young—members of the Scientific Advisory Committee without whose contributions the study could not have been completed;

Ms. Patricia Cullen, Mr. Eugene Balzer, Mr. Donald Kronish, Mr. William Young, Mr. Charles Sloane and so many other dedicated people at Warner-Lambert, for unstinting, sincere and tireless support and technical consultation;

Mrs. Ruth Warnock and Miss June Watkins for arranging and preparing the many drafts and final version;

Dr. John Rowen, Mr. John Flannagan, Mrs. Mary Friend, Mr. Robert Lemieux, Mr. Eli Mantel, Mr. Rolfe MacCullough, Ms. Marilyn Westfall, Mr. Barry Nocks and other members of the staff of the Technical Analysis Division for their work in support of the project.

EXECUTIVE SUMMARY

A. BACKGROUND

There are estimated to be at least 15,000 clinical laboratories in the United States, ranging in size from small laboratories operated in conjunction with a physician's practice to extremely large laboratories which process millions of specimens each year on a mail order basis. Various accreditation and licensing programs periodically monitor the performance of some clinical laboratories through inspection and/or proficiency testing. Other laboratories are subject to little or no control.

The effect of regulatory activities on the performance of clinical laboratories has been investigated in prior evaluations. In general, these studies and surveys have been of limited scope and directed toward specific groups of laboratories, such as those within a state, a subset of hospital laboratories, or laboratories engaged in interstate commerce.

Unfortunately, it is not possible to "pool" data from these experiments in order to establish relative performance measures for various classes of laboratories. Analysis of the relevant literature shows a lack of compatibility: variations in temporal aspects, sample sizes, constituents analyzed, and a host of other discrepancies which make valid data aggregation or extrapolation impractical.

In order to obtain comparable data on laboratory performance, the Division of Health Evaluation, Department of Health, Education, and Welfare, contracted with the Technical Analysis Division of the National Bureau of Standards in January 1971 to develop and implement a comprehensive national survey of clinical laboratories.

The major objective of the study was to obtain measures of capability for several different types or groups of clinical laboratories and to determine if there are basic differences in analytical accuracy among these groups which would warrant remedial action by public agencies or the private sector. Analytical capability was judged by the performance of participating laboratories in analyzing proficiency test specimens in the areas of clinical chemistry, hematology, and microbiology. A secondary objective was to compare the accuracy and precision of the various analytical methods used by the participating laboratories.

The procedure utilized was to: (1) establish a Scientific Advisory Committee of governmental and health industry representatives to assist in the selection of proficiency test components and monitor the progress of the study from a clinical viewpoint, (2) develop a survey design, (3) contact professional and regulatory groups to solicit laboratory participation, (4) procure and distribute two sets of laboratory samples in clinical chemistry, hematology and bacteriology to laboratories participating in the study, and (5) statistically analyze the results reported by the laboratories.

B. PARTICIPANT LABORATORIES

Six types or groups of clinical laboratories participated in the study:

1. Interstate—laboratories engaged in interstate commerce and licensed by the Center for Disease Control (CDC), Health Services and Mental Health Administration, under the Clinical Laboratory Improvement Act of 1967 (CLIA '67) in one or more of the specialties under consideration in this study. This Act requires licensees to participate in either the CDC or the College of American Pathologists proficiency testing program.
2. American Academy of Family Physicians (AAFP)—private physician laboratories, generally small, which are currently affiliated with AAFP. These laboratories are presently exempt from any Federal licensure program.
3. American Society of Internal Medicine—laboratories operated in conjunction with a private physician's practice which, like AAFP affiliated laboratories, are presently exempt from any Federal licensure program.

4. Joint Commission on Accreditation of Hospitals (JCAH)—laboratories within hospitals which are accredited under JCAH. The JCAH inspects these laboratories, but does not require them to participate in proficiency testing.
5. Medicare Certified Hospitals—laboratories in hospitals which are Medicare providers under Title XVIII, Health Insurance for the Aged. These laboratories are also Medicaid providers. Hospitals and laboratories accredited under JCAH were excluded from this category. As a result, the laboratories in this category tend to be the smallest laboratories which are not operated in conjunction with a private physician's practice.
6. Medicare Certified Independent—private and commercial laboratories which are reimbursed for certain laboratory procedures under Medicare and Medicaid, but which are not normally licensed under CLIA '67 or accredited by JCAH. The law requires that these laboratories participate in a proficiency testing program administered or approved by the regulatory program in the state in which they operate.

Approximately 1,000 laboratories participated in the study. The number within a category ranged from 43 (AAFP) to 231 (Interstate).* A seventh category of 18 reference laboratories served as a control group.

C. METHODOLOGY

Separate proficiency test specimens were prepared for clinical chemistry, hematology and microbiology. Criteria for selection of the clinical chemistry and hematology constituents to be analyzed were that laboratory analysis should be routine, and that fairly well developed analytical procedures exist.

Each clinical chemistry shipment included one normal and one abnormal sample. The laboratories were asked to determine the concentration of eight constituents: glucose, urea nitrogen, calcium, total bilirubin, cholesterol, uric acid, sodium, and total protein. Each hematology shipment required analysis of red blood count, white blood count, hemoglobin, hematocrit, and mean corpuscular volume levels in both normal and abnormal specimens. Five pure cultures of ordinary and easily identifiable bacteria were used as sample cultures for identification in the microbiological portion of the study.

Two shipments of specimens and/or cultures were sent to each participating laboratory. All analytic results were reported by the laboratory on forms shipped with the test specimens.

D. RESULTS

For analysis, the AAFP and ASIM data were pooled together into a Doctor's Office group. All groups were then sufficiently large to allow extrapolation of the study results to the unsampled laboratories with 95% confidence. However, since participation was on a voluntary basis, the sample is a selected rather than a random sample, and it would be invalid to extrapolate to the unsampled population.

Clinical Chemistry

There were no significant differences (at the 95% confidence level) among the average clinical chemistry laboratory results obtained by the groups participating in the study. The interlaboratory consistency (interlaboratory precision) of the laboratory groups can be exhibited as follows, where groups joined by the same line did not exhibit significantly different precision at the 90% confidence level.

---

*Accurate information is not available on the number of laboratories which are members of each group. Even if such information were available, it would be improper to use the relative numbers of laboratories as measures of group activity because the average number of analyses performed by a laboratory in a year is much higher for some groups than for others.

Rank Order	Laboratory Group
Most Precise	Medicare Independent
	Interstate
	JCAH
	AAFP/ASIM
Least Precise	Medicare Hospital

The line diagram indicates that the Medicare Independent Laboratory analyses were significantly more precise than those of all other groups except the Interstate group. No other differences in group precision were statistically significant.

The techniques used had a considerable effect on the accuracy and precision of reported analyses. Table 1 lists the techniques which were most satisfactorily applied, and the percentage of the participating laboratories which applied each technique. As the table implies, a large percentage of the participating laboratories are using outmoded clinical chemistry techniques. In most instances, automated methods were applied with equal or better average accuracy and considerably better precision than the corresponding manual methods. Results reported by laboratories using diagnostic kits were consistently less precise than other determinations.

Medical usefulness of the clinical chemistry analyses was assessed using the criterion that analyses should be sufficiently precise to permit the interlaboratory monitoring over time of the variation in an individual patient's constituent concentrations. This criterion was not applied to total bilirubin because data on individual bilirubin variation was not available. Of the remaining seven constituents, only cholesterol was analyzed by the study participants with sufficient precision to satisfy the criterion. In contrast to the performance of the study participants, reference laboratory analyses of cholesterol, uric acid, urea nitrogen, sodium and total protein were all sufficiently precise to permit interlaboratory monitoring of individual variation. Those participating laboratories using the best applied techniques also achieved acceptable interlaboratory precision in analyses of these five constituents.

Hematology

As with clinical chemistry, the average hematology results obtained by the participating groups did not differ significantly at the 95% confidence level. The interlaboratory precision of the laboratory groups can be exhibited as follows where groups joined by the same line did not exhibit significant differences at the 90% confidence level.

Rank Order	Laboratory Group
Most Precise	Medicare Independent
	JCAH
	Interstate
	Medicare Hospital
Least Precise	AAFP/ASIM

Table 2 shows the best applied techniques and the percentages of participating laboratories using these techniques.

Table 1. Clinical Chemistry - Best Applied Techniques

Constituent	Best Applied Techniques	% Applying
Glucose	Ferricyanide AutoAnalyzer	12.9
Urea Nitrogen	Diacetyl Monoxime Automated	34.0
Calcium	Atomic Absorption Cresolphthalein Complexone Automated	4.0 33.4
Bilirubin	Diazo-Other Coupling (J & G) Automated	24.9
Cholesterol	$FeCl_3$-H_2SO_4 with Prior Extraction Manual or Automated	5.9
Uric Acid	Uricase Phosphotungstate Automated	2.2 24.2
Sodium	Flame Photometer Automated	12.6
Total Protein	Biuret Automated Refractometer	34.5 15.7

Table 2. Hematology - Best Applied Techniques

Constituent	Best Applied Techniques	% Applying
Red Cell Count	All Coulter Models Kits	43.3
White Cell Count	All Coulter Models Kits	45.7
Hematocrit	Microhematocrit	75.9
Hemoglobin	All techniques applied equally well	100.0
Mean Corpuscular Volume	Impossible to judge	---

Microbiology

The performance in microbiology of the Interstate group was significantly better than the performance of the other groups at the 95% confidence level; 7.6% of the Interstate laboratory determinations were incorrect while 19.9% of all other determinations were incorrect. However, even a 7.6% misidentification rate is not satisfactory because the organisms selected are common and should be easily identifiable. Most worrisome are such misidentifications as Neisseria NOS, N. gonorrhoeae or N. meningitidis for the pure culture of Streptococcus faecalis.

The relative performance of the laboratory groups can be portrayed as shown, where the lines join groups whose performance did not differ significantly at the 95% confidence level.

Rank Order	Laboratory Group
Best Performance	Interstate
	JCAH
	Medicare Independent
	AAFP/ASIM
Worst Performance	Medicare Hospital

E. CONCLUSIONS

The data indicate that high volume laboratories may be more proficient than smaller laboratories, such as those which serve Doctors' Offices and Medicare Certified Hospitals. In microbiology, 7.6% of the Interstate laboratory determinations were incorrect, while 16.5% of the determinations by other large laboratories (JCAH and Medicare Independents) were incorrect. Thus, it would appear that the CDC proficiency testing program has considerably improved the microbiology performance of the enrolled laboratories (although further improvement is still needed). Conversely, clinical chemistry and hematology analyses by the Interstate laboratories were no better than comparable analyses by other large laboratories, many of whom do not engage in routine proficiency testing programs. This seems to indicate that the CDC proficiency testing programs in clinical chemistry and hematology have had relatively little effect upon the performance of laboratories participating in the program. This conclusion is further substantiated in a companion report.* It is particularly important to improve the effectiveness of these programs because the interlaboratory consistency of study participants with respect to clinical chemistry and hematology was too often insufficient to support monitoring of an individual's constituent concentrations over time. It appears that poor selection of techniques is an important factor in the low rate of acceptability of laboratory determinations.

F. LIMITATIONS

It must be clearly understood that the results of this survey are limited by four important considerations:

1. Because all of the laboratories participated on a purely voluntary basis, no straightforward extrapolation can be made to the larger universe of unsampled clinical laboratories.
2. It is probable that the results of this study do not represent routine laboratory performance for two reasons: (a) a laboratory probably would not volunteer if its

*Clinical Laboratory Performance Analysis Using Proficiency Test Statistics, NBS Report, NBSIR 73-197, June, 1973.

management felt that to do so would be disadvantageous, and (b) the sample materials probably received special attention in many of the smaller laboratories which were unfamiliar with analyzing proficiency test samples.

3. The clinical chemistry test specimens were prepared by a dialyzation process which removes naturally occurring substances. As a result, the accuracy of some methods, as applied to the test specimens, might differ from their accuracy in analyses of human serum.
4. The true constituent concentrations of cholesterol and the hematology constituents could not be exactly determined. For these constituents, accuracy was assessed relative to the mean reference laboratory assays.

G. RECOMMENDATIONS

The following recommendations are offered as logical outgrowths of the conclusions noted earlier in E:

1. Satisfactory performance in a microbiology proficiency testing program conducted under the auspices of either Federal or other approved authorities should be a requirement for all clinical laboratories analyzing microbiological specimens. It is unclear whether this recommendation should be implemented through new legislation or a reinterpretation of existing legislative authority.
2. A Technical Advisory Committee consisting of government and professional society representatives should be established to identify the most accurate and precise analytical methods available, and encourage their use by the largest possible number of clinical laboratories. Zones of acceptable performance for proficiency testing should be constructed in a manner which reflects the variability associated with the more accurate and precise methods and systems. In this way, failure to accept the recommended procedures would increase the risk of unacceptable performance ratings.
3. An experimental study should be undertaken to determine a better design for proficiency testing programs in clinical chemistry and hematology. An empirical description of the causes of inadequate laboratory work should be used in defining alternative testing strategies for consideration. This study should deal with such questions as frequency of sampling; feedback to participants; number of levels at which to test; long-term monitoring of intralaboratory variability; follow-up procedures on outlier values; and the criteria for scoring, ranking or rating laboratory performance and its medical usefulness. Until the results of this recommended study become available, it does not appear justified (or warranted) to alter the frequency of CDC proficiency testing in clinical chemistry and hematology.

CONTENTS

CONTENTS (continued)

CONTENTS (continued)

A PROFICIENCY TEST ASSESSMENT OF CLINICAL LABORATORY CAPABILITY IN THE UNITED STATES

Peter W. Finkel
Ted R. Miller

The proficiency of a representative sample of physician, hospital and independent laboratories was assessed with respect to their ability to analyze clinical chemistry and hematology samples and to identify microbiological organisms. For the assessment of clinical chemistry and hematology proficiency, the laboratories were grouped, and determinations of group accuracy and group precision were made. Further analyses were performed to determine relative accuracy and precision of the techniques presently applied by these groups. There was no significant difference at the 95% confidence level in the accuracy achieved by the various laboratory groups involved in clinical chemistry and hematology analysis. In clinical chemistry, the Medicare Certified Independent laboratories, CDC Tested laboratories and JCAH Members generally proved more precise than Physicians' Office and Medicare Certified Hospital laboratories. However, none of the laboratory groups were sufficiently accurate to permit the monitoring over time of variation in an individual patient's constituent concentrations. It would appear that poor selection of techniques was an important contributor to this low performance level. In hematology the Physicians' Office laboratories proved to be the least precise of the groups. There was no noticeable difference in precision between participants in the CDC proficiency testing program and nonparticipants. With respect to microbiology, 7.6% of the identifications by laboratories participating in the CDC testing program were incorrect, while 19.4% of all other identifications were incorrect.

Key words: Accuracy; clinical chemistry; hematology; medical usefulness; microbiology; proficiency testing.

1. INTRODUCTION

There are estimated to be at least 15,000 clinical laboratories in the United States, ranging from small laboratories operated in conjunction with a physician's practice to extremely large laboratories which process millions of specimens each year on a mail order basis. A major group of clinical laboratories are involved in interstate commerce. These laboratories are normally licensed in one or more "procedure" or "category of procedure" (microbiology, serology, etc.) under Public Law 90-174 (Partnership for Health Amendments 1967). Section 5 of this law includes provisions relating to the application for and issuance of licenses, quality control, personnel standard setting, and proficiency testing. The Center for Disease Control (CDC), Department of Health, Education, and Welfare (DHEW), serves as the Federal regulatory agency for this legislation, which applies to between 5 to 10 percent of all clinical laboratories in the U. S.

Recently, many states have passed new legislation or strengthened existing laws in an effort to insure high quality laboratory results. Such legislation is primarily directed to hospitals and independent laboratories not involved in interstate commerce. Regulatory activities, including inspection, staff accreditation and proficiency testing, are dependent on the strength of the legislation involved. Twenty-six states are currently conducting some form of regulatory activity. However, in all instances, laboratories operated in conjunction with a physician's practice are excluded from legislative actions as long as the laboratory is operated solely for the benefit of the physician's patients.

In addition to Federal and state regulatory activities, three other accreditation programs are worthy of mention. In 1962 the College of American Pathologists (CAP) initiated its own program of laboratory inspection and accreditation for member laboratories. This program involves on-site inspections, proficiency testing and the use of certification criteria pertaining to laboratory facilities, services, procedures, and staff qualifications. A second laboratory accreditation program, which does not presently include proficiency testing, is administered by the Joint Commission on Accreditation of Hospitals (JCAH). Currently, more than 2,000 hospitals throughout the U. S. are surveyed annually by the

Joint Commission. The third accreditation program, also lacking proficiency tests, is associated with the Medicare program; state agencies accredit hospitals and independent commercial firms as being eligible for payment of Medicare funds.

Thus, various accreditation and licensing programs periodically monitor clinical laboratory performance through inspection and/or proficiency testing. Most independent and hospital laboratories are accredited or licensed under one or more of these programs. However, almost all laboratories operated exclusively by and for the private physician in conjunction with his practice are not similarly covered although such laboratories perform approximately 25% of all laboratory tests each year.

The effect of regulatory activities on the performance of clinical laboratories has been the subject of prior evaluations. A review of the literature indicates that a sizable amount of laboratory performance data has been collected since the late 1940's. In general, these studies and surveys have been of limited scope and were directed to specific groups of laboratories, such as those within a state, a subset of hospital laboratories, or laboratories engaged in interstate commerce. Unfortunately, it is not possible to "pool" data from these experiments in order to establish relative performance measures for various classes of laboratories. Analysis of the relevant literature shows lack of compatibility variations in temporal aspects, sample sizes, constituents analyzed (where proficiency testing was involved), and a host of other differences which make valid data aggregation or extrapolation impractical.

In order to obtain comparable data on laboratory performance, the Division of Health Evaluation, DHEW, contracted with the Technical Analysis Division (TAD) of the National Bureau of Standards (NBS) to develop and implement a comprehensive national survey of clinical laboratories. Laboratories participating in this study were to include laboratories involved in various licensure/accreditation programs (CLIA '67, JCAH, Medicare), as well as laboratories which do not participate in such programs. Since disguising samples for study purposes is impractical, proficiency tests were to be administered to each laboratory openly. Work commenced in January, 1971.

Concurrent with this study program, a complementary analysis was performed of quantitative proficiency test results compiled by the CDC. These data relate to the performance of roughly 270 clinical laboratories licensed under the Clinical Laboratory Improvement Act of 1967 (CLIA '67) and approximately 400 additional laboratories that participate in this program on a voluntary basis. The results of that study are described in a companion (TAD/NBS) report entitled "Clinical Laboratory Performance Analysis Using Proficiency Test Statistics." Some of the results derived from this work affected the data collection and analysis phase of the national survey reported herein. An additional report which deals with CLIA '67 was prepared by CDC and is entitled "Evaluation Report on the Effectiveness of the Clinical Laboratories Improvement Program" (July 1, 1972).

Two general caveats on the interpretation of the results of the current study should be emphasized. The absence of any legal requirement for participation in the study implied a self-selection aspect that negates extrapolating results to the unsampled universe of clinical laboratories. In addition, the relationship between laboratory performance in open proficiency testing and routine analysis is unknown. However, a laboratory's routine level of performance is unlikely to be superior to the effort it expends in the analysis of known proficiency test samples.

The major objective of the study was to obtain measures of performance from several different types or groups of clinical laboratories, and to determine if there are basic differences in the capabilities of these laboratory groups from the standpoint of analytical accuracy. Secondary objectives of the study were to compare the precision of clinical chemistry and hematology determinations by these groups and to compare the accuracy and precision of the various analytical methods used. Analytical capability of participating laboratories was measured through the use of proficiency test samples in the areas of bacteriology, clinical chemistry, and hematology.

2. METHODOLOGY

The specific tasks performed to carry out this study included:

(a) Establishment of a Scientific Advisory Committee to act as an advisory body to the study
(b) Contact with professional groups and their member laboratories to insure widespread participation in the study
(c) Development of a sample survey design
(d) Production of laboratory specimens through contractual agreement
(e) Design of reporting forms and a computer card layout to facilitate analysis
(f) Mailing of specimens to the laboratories and coordination of the return of laboratory results
(g) Statistical analysis of the data and preparation of a report summarizing the study results.

2.1. Scientific Advisory Committee

A Scientific Advisory Committee of recognized authorities was formed to assist in the selection of those proficiency test components which could be regarded as representative of laboratory analytical capability, and to monitor the progress of the study from a clinical viewpoint. The members of the Committee are listed in Appendix A.

2.1.1. Proficiency Test Components

There was general agreement among the Committee members that the study should concentrate on routine laboratory procedures which are commonly performed and for which fairly well developed analytical procedures exist. It was assumed that a laboratory which is unable to adequately accomplish routine analyses is unlikely to adequately perform less common or more complicated procedures. Based on this rationale, the following quantitative procedures were selected for inclusion in the study:

Clinical Chemistry - glucose, urea nitrogen, calcium, total bilirubin, cholesterol, uric acid, sodium, total protein
Hematology - red cell count, white cell count, hemoglobin, hematocrit, mean corpuscular volume.

In the area of microbiology the samples consisted of pure cultures of five rather ordinary bacteria: Diplococcus pneumoniae, Salmonella oranienberg, Klebsiella pneumoniae, Pseudomonas aeruginosa, and Streptococcus faecalis. Each participating laboratory was asked to analyze two shipments of proficiency test samples.

2.2. Types of Clinical Laboratories Represented

The study encompassed voluntary participation by a sample of the most common types of clinical laboratories, as evidenced by the following list:

Doctors' Offices - physician members of the American Academy of Family Practice (AAFP) and the American Society of Internal Medicine (ASIM)
Interstate - laboratories licensed by the Center for Disease Control under the Clinical Laboratory Improvement Act of 1967
Large Hospitals - laboratories in hospitals which are accredited by the Joint Commission on Accreditation of Hospitals
Smaller Hospitals - laboratories in hospitals which are accredited for the Medicare program by state agencies
Smaller Independents - commercial laboratories which are accredited for the Medicare program by state agencies.

The next few paragraphs briefly discuss each of the participating laboratory groups. The anonymity of all participants was assured through the use of code numbers known only to the

individual laboratory and the National Bureau of Standards. The manner in which laboratories were solicited for the study apparently played an important part in the decision for or against participation. Since each laboratory was contacted by the NBS study team, Warner-Lambert, and at least one other group (JCAH, ASIM, etc.), some laboratories were uncertain as to whether they were being asked to cooperate in one, two or three studies.

2.2.1. Interstate Laboratories

One of the major objectives of the study was to compare the performance capability of laboratories covered by CLIA '67 with laboratories exempt from the Act.

The assistance of the Laboratory Division, CDC, was secured in order to solicit the voluntary participation of those clinical laboratories which are licensed under CLIA '67 in one or more of the specialties included in this study. Of the 257 laboratories originally contacted by CDC for the study team, there were 9 declinations and 248 acceptances.

2.2.2. American Academy of Family Physicians

The American Academy of Family Physicians (AAFP) was one of two professional societies contacted in an effort to bring private physicians' laboratories into the study. Through the cooperation of Dr. William E. Loterhos, then President of AAFP, the approval of the Board of Directors, and the efforts of Mr. Arthur T. Smith (Secretary of the Committee on Clinical Investigation, AAFP), a list of 300 research-minded Academy members was made available.[1] Dr. Roger A. Peatee, Chairman of the Committee on Clinical Investigation, sent letters to the 300 physicians informing them that their names had been given to the study group and advising them that they might be asked to participate in the testing program.

Subsequently, written contact was made with each physician, asking for an expression of interest in the study and an indication of the areas of testing (bacteriology, chemistry, and hematology) in which the physician was currently active. Negative replies were received from 61 doctors—39 indicated that they did little or no laboratory testing in conjunction with their practice, and 22 said that they preferred not to participate. There were 150 physicians who expressed an interest in participating, and who also indicated which tests they were capable of performing. No replies were received from the remaining 89 doctors. Table 2.2.1 below summarizes this information.

Table 2.2.1. Physician Participation, AAFP

	Number	Percent
Total Number of Physicians Sent the AAFP Letter Dated 8/18/71 and the TAD/NBS Letter Dated 8/23/71	300	100
Total Number of Physicians Declining to Participate (but were capable)	22	7.3
Total Number of Physicians Unable to Participate (no lab facilities, no longer in private practice, retired, ill, etc.)	39	13.0
Total Number of Physicians Not Responding	89	29.7
Total Number of Physicians Volunteering to Participate	150	50.0

[1]The Academy had previously polled its members in order to ascertain the extent of their interest in participating in research studies. The list given to NBS contained the names of all physicians who had indicated a positive response to the 1969 query.

2.2.3. American Society of Internal Medicine

The second professional group contacted, the American Society of Internal Medicine, has a membership of approximately 12,000 physician specialists (internists) many of whom operate a clinical laboratory in conjunction with their practice. Because of their significant role in the health care delivery system, and the fact that their laboratories are generally exempt from the Federal licensure program, their participation in this study was considered to be essential.

On July 20, 1971, Dr. Otto C. Page, then President of the ASIM, wrote to each of 496 physicians who had previously expressed an interest in having their laboratory evaluated. Dr. Page's letter requested an expression of continued interest. Replies were received from 306—19 stated that they did not want to be in the study, an additional 23 reported that they did not have a laboratory in their practice, and 264 agreed to participate. All communication was between the Society and the physician members. The TAD study group did not contact any of the ASIM participants except in connection with the data reporting or to respond to questions. Table 2.2.2 below summarizes the numbers of contacts, responses, etc.

Table 2.2.2. Physician Participation, ASIM

	Number	Percent
Total Number of Internists Written to by Dr. Page on 7/20/71	496	100
Total Number of Internists Declining (but were capable)	19	3.8
Total Number of Internists Unable to Participate (no lab, etc.)	23	4.6
Total Number of Internists Not Responding	190	38.2
Total Number of Internists Volunteering to Participate	264	53.4

2.2.4. Joint Commission on Accreditation of Hospitals

The JCAH was created in 1952 through a joint effort of the American College of Surgeons, the American College of Physicians, the American Hospital Association, the American Medical Association, and the Canadian Medical Association. (The latter group discontinued participation in 1959, however, upon the creation of the Canadian Council on Hospital Accreditation.) Basically the JCAH exists to: (1) establish standards for the operation of hospitals and other health care facilities and services; and (2) conduct survey and accreditation programs that will encourage members of the health professions, hospitals and other health care facilities and services to voluntarily:

(a) Promote high quality of care in all aspects in order to give patients the optimal benefits that medical science has to offer;
(b) Apply certain basic principles of physical plant safety and maintenance, and of organization and administration of function for efficient care of the patient; and
(c) Maintain the essential services in the facilities through coordinated effort of the organized staffs and the governing bodies of the facilities.

Through the support of the JCAH, the study was able to include more than 200 hospital clinical laboratories selected strictly randomly from all JCAH hospitals. This was accomplished through the use of a table of random numbers, after assigning a unique number to each accredited hospital. After 222 hospitals were selected, each was sent a letter from the Commission, advising them that participation in the study was considered mandatory as part of the accreditation program. Seven hospitals received JCAH approval to drop out of the study for various reasons, leaving 215 hospital laboratories in this group.

2.2.5. Medicare Certified Hospitals and Independent Laboratories

Most of the Nation's hospitals and independent clinical laboratories are providers of Medicare services under Title XVIII, Health Insurance for the Aged. These same laboratories are approved providers of Medicaid services. It was considered extremely important that the study include a sample of these laboratories in order to permit comparisons between laboratory groups involved in different Federal certification programs.

Personnel in the Community Health Service (Dr. Sheriden Weinstein and Mrs. Barbara Nagel) and the Division of State Operations (DSO), Bureau of Health Insurance (BHI)/Social Security Administration (SSA) (Messrs. Brown, Burk and Byers) were instrumental in securing the participation of several hundred Medicare certified hospitals and independent laboratories. Through the combined efforts of these individuals and the TAD study team, a letter was issued on July 23, 1971, over the signature of the Director, BHI, to every BHI/SSA regional office. The regional offices contacted the state health departments who, in turn, contacted the individual hospitals and independent laboratories. The manner in which the study was described to the laboratory by the state agency apparently played an important role in the laboratory's decision on whether to participate.

The names and addresses of 300 hospitals and 200 independent laboratories (randomly selected from those who were not JCAH members), broken down by state, were supplied to DSO/BHI/SSA for transmittal to the regional offices. Of the 300 hospitals, positive replies were received from 236; 153 of the 200 independent laboratories contacted said that they would participate in the study.

2.2.6. Reference Laboratories

To obtain a check on the accuracy of information supplied by the manufacturer of the test samples, the participation of 19 reference laboratories was solicited. These laboratories were selected so as to be representative of the "best" possible state-of-the-art. Eighteen of the laboratories (see Appendix B) elected to participate in the study. Of these, 16 laboratories normally performed clinical chemistry determinations (one of these labs did not do cholesterol), 15 performed hematology, and 16 performed microbiology.

2.3. Selection of Contractor for Sample Materials

At the very early discussion stages of the study, consideration was given to the possibility of having the Center for Disease Control prepare the sample materials. The major reason for not pursuing this course of action was that CDC was heavily committed to preparing samples for another program and could not redirect their resources toward this effort during the established time frame.

Accordingly a contract was awarded to the General Diagnostic Division of Warner-Lambert (through competitive bid) to prepare and mail all samples, specimens and cultures used in the testing program.

2.4. Specifications for Sample Materials

2.4.1. Establishing Target Values for Clinical Chemistry

It was possible to stipulate desired levels for seven of the eight constituents in clinical chemistry[2] because of Warner-Lambert's unique manufacturing process. In this

---

[2]A desired level of cholesterol could not be specified in advance because it is functionally related to other blood constituents, some of which are removed in the dialysis and ion exchange procedure.

process, many of the normally occurring constituents are removed from the blood through an ion exchange and dialysis procedure until measurable amounts can no longer be recovered. Known amounts of the desired constituents are then weighed and remixed ("weighed back") in the blood serum to arrive at specified levels.[3] Unfortunately, the process also removes naturally occurring reducing substances which are not replaced. Consequently, comparisons of laboratory methods based on performance in analyzing these study samples may not accurately reflect differences in performance in analyzing normal patient samples.

The first column in Table 2.4.1 shows the constituent values requested by the Advisory Committee, and the second column shows the actual amounts weighed back by Warner-Lambert. These latter values are alleged to be exact measures of the contents of the test sample material (nonhomogeneity in the material results in some vial-to-vial variability, but this variability never exceeds .5%). The close agreement between requested and actual values is evident. The third column, headed MRA, is the Mean Reference Assay—the arithmetic mean assay reported by the reference laboratories. Thirty reference laboratory assays are included in these means. One of the reference laboratory determinations for normal glucose and one for abnormal uric acid were reported as more than seven standard deviations from their respective means and were therefore discarded as outliers. Only 4 of the 14 Warner-Lambert weigh-back determinations (both levels of calcium and bilirubin) differed from the MRAs by as much as 2%, and no determinations differed by more than 10%. Warner-Lambert's average cholesterol assay at the normal level seriously differs from the mean reference assay. A reference technique for determining exact constituent concentration was available only for calcium. Because calcium sometimes binds to glass and the calcium MRAs and weigh back values differed fairly widely, the mean reference assay by the reference technique (MRART) was examined. The MRART values were 9.0 and 11.7 mg/100 ml.

Table 2.4.1. Specification of Constituent Levels, Chemistry

	Normal			Abnormal		
Constituent	Requested	Actual	MRA	Requested	Actual	MRA
Glucose (mg/100 ml)	84	82.2	83.4	307	304	301
Sodium (mEq/l)	143	141	141	118	117	116
Urea Nitrogen (mg/100 ml)	19	18.6	18.6	66	65.4	65.7
Bilirubin (mg/100 ml)	0.9	0.9	1.0	4.7	4.7	4.5
Uric Acid (mg/100 ml)	6.3	6.2	6.2	11.1	11.0	11.0
Calcium (mg/100 ml)	9.2	9.2	8.7	12.2	12.2	11.5
Total Protein (g/100 ml)	6.8	6.8	6.9	3.3	3.3	3.4
Cholesterol (mg/100 ml)[4]		160	133		70	74

In this study, laboratory performance was judged by the ability to reproduce the actual specimen contents as displayed in Table 2.4.1. These actual values were defined as the target values for clinical chemistry. Analysis presented in Section 3.7 was used to justify the use of mean reference assays as target values for cholesterol. In analyses where minor differences in the mean might prove important, calcium was assessed in terms of both the weigh-back and MRART values.

---

[3]A more complete description of the process is available upon request to the authors.

[4]A desired level of cholesterol could not be specified in advance because it is functionally related to other blood constituents, some of which are removed in the dialysis and ion exchange procedure. The "actual" cholesterol values shown are the average Warner-Lambert assays since weigh-back values do not exist.

In the first shipment, dialyzed specimens were sent to all 900 laboratories participating in the clinical chemistry portion of the study. In the second shipment, 200 laboratories, randomly selected from the 900, were sent specimens prepared from serum pool batches (validate specimens), while the other 700 were again sent dialyzed sample material. This substitution was designed to provide insight into the method comparison problem. The assay values for these specimens, as experimentally determined by Warner-Lambert, and the mean of the 200 laboratory analyses are shown in Table 2.4.2.

Table 2.4.2. Specification of Constituent Levels, Validate Chemistry

	Normal		Abnormal	
	W-L Assay	Mean Analysis	W-L Assay	Mean Analysis
Glucose	88.0	80.6	250	241
Sodium	134	135	130	129
Urea Nitrogen	10.6	10.6	51.2	49.7
Bilirubin	.5	.5	5.3	4.8
Uric Acid	5.5	5.4	9.9	9.2
Calcium	8.4	8.3	12.1	12.3
Total Protein	6.3	6.1	5.3	5.2
Cholesterol	174	172	183	167

It is apparent that two of the Warner-Lambert average assays (normal glucose, abnormal cholesterol) differ substantially from the mean assays. Consequently, a choice of proper serum pool "target values" is prerequisite to an evaluation of differences in performance on dialyzed and serum pool specimens. This topic is discussed at length in Appendix C.

2.4.2. Desired Levels of Hematology Samples

The hematology samples, prepared by Charles Pfizer & Co., Inc. under contract to Warner-Lambert, were the most critical materials with regard to shelf life. According to Warner-Lambert, the lyophilized chemistry and bacteriology samples were stable for at least two years; in contrast, Pfizer recommends a shelf life of only 21 days for their commercially prepared blood standards. As a result, it was necessary to prepare separate hematology batches for the two shipments.

One of the reference laboratories agreed to monitor the stability of the hematology samples used in the study. Manual and automated determinations were made on each working day for a 30-day period starting on the day that each batch of samples was mailed. During this period, each batch remained stable. Previous studies[5] have reported that mailing does not affect the stability of hematology samples, and it seemed unnecessary to reverify this result.

Pfizer did not make special hematology batches for this study; the samples which they supplied to Warner-Lambert were part of their routine commercially prepared batches. As a result, the Advisory Committee could not control the choice of normal and abnormal levels used. This was particularly unfortunate in the case of white cell count since the Committee would have preferred an abnormally low value to the abnormally high value in the prepared batches.

---

[5]See, e.g., I. Schoen, G. Thomas, and S. Lange, "The Quality of Performance in Physician's Office Laboratories," American Journal of Clinical Pathology, Vol. 55, No. 2, 2/71, pp. 163-170.

Table 2.4.3 contains values for each constituent in each hematology batch, as reported by Pfizer and the reference laboratories (MRA). Within each cell the top line refers to the Coulter Model S assay and the bottom line is for all other methods. At least five assays were used in computing each MRA.

Pfizer had reported separate Coulter S targets because of the occasional presence of substances in their artificially prepared samples which do not occur naturally and have different effects on Coulter S measurements than on others. The Advisory Committee suggested that it would be preferable to use a single set of targets if such problems were uncommon in the serum batches. The frequency of bias therefore had to be checked before proceeding with analysis of the study results. The Pfizer recommendations and the reference laboratory results provided sufficient data for resolution of this question. Because of the small reference laboratory sample sizes, a non-parametric test was constructed. First, the difference between each pair of Coulter S and non-Coulter S values was determined for the Pfizer data and the MRAs. Under the hypothesis that Coulter S determinations are unbiased, for any constituent in any batch at any level, the probability is one-half that the sign of the difference between Pfizer's assays using Coulter S and other methods will match the sign of the difference in MRA assays. In actuality 10 of the 19 differences agree in sign.[6] It is thus possible to state that bias occurred rarely, if at all, for specimen analyses performed in this study, and it is not necessary to use separate targets for the Coulter S determinations. However, it is important to note any analytic results which arise strictly as a result of using a single set of hematology target values.

2.4.3. Selection of Organisms for Microbiology

The bacteria were selected jointly by the Advisory Committee and the Microbiology Department of Warner-Lambert. This was done so that the samples would be representative of organisms routinely encountered in clinical laboratories and to insure that Warner-Lambert had prior experience in isolating and growing pure cultures of the chosen bacterium. Another very real concern on the part of Warner-Lambert was that the organisms should not be too pathogenic in case a vial accidently burst or was dropped.[7]

A very brief description of the "source" of each vial of freeze-dried organisms was included in the package insert. Table 2.4.4 below shows the organism and its vial number and simulated source in the two shipments. The permutation of the organisms was intended to identify any laboratories which, upon noting the same "source" for the same vial on the second shipment, concluded that the organism contained therein was the same as they had previously identified. Care was taken to assure that each organism was commonly found in both of the sources ascribed to it. There were only two laboratories which misidentified all five vials in the second shipment by reporting them as they were in the first shipment.

Table 2.4.4. Simulated Source of Organisms

Vial Number	Source	Organism Shipment 1	Organism Shipment 2
1	Sputum, 46 yr. old male	Diplococcus pneumoniae	Klebsiella pneumoniae
2	Stool, 9 yr. old girl	Salmonella oranienberg	Streptococcus faecalis
3	Sputum, 60 yr. old female	Klebsiella pneumoniae	Pseudomonas aeruginosa
4	Blood, 22 yr. old male	Pseudomonas aeruginosa	Diplococcus pneumoniae
5	Urine, 25 yr. old female	Streptococcus faecalis	Salmonella oranienberg

[6]One data point was discarded because the difference between MRA values was unsigned (0).

[7]A description of the manufacturing procedure for the sample cultures is available upon request to the authors.

Table 2.4.3. Constituent Levels for Hematology

CONSTITUENT		NORMAL				ABNORMAL			
		Batch 1		Batch 2		Batch 3		Batch 4	
		Pfizer	MRA	Pfizer	MRA	Pfizer	MRA	Pfizer	MRA
Hemoglobin (grams/100 ml)	Coulter Model S	12.6	12.5	13.0	12.8	6.3	6.2	6.6	6.5
	All Other	13.1	12.6	13.2	13.1	6.6	5.7	7.0	6.4
Hematocrit (%)	Coulter Model S	37.6	37.2	38.6	38.0	18.0	18.2	18.6	19.8
	All Other	43.0	40.8	43.5	41.5	20.0	17.7	22.0	20.5
Red Blood Count (million/mm^3)	Coulter Model S	4.09	4.21	4.29	4.35	1.71	1.77	2.14	2.19
	All Other	4.1	4.14	4.24	4.17	1.76	1.73	2.19	2.08
White Blood Count (thous./mm^3)	Coulter Model S	7.2	7.57	8.3	8.3	19.2	19.1	19.3	19.1
	All Other	7.35	7.26	8.25	8.3	20.3	19.8	19.6	18.8
Mean Corpuscular Volume (cubic microns)	Coulter Model S	92	88	90	88	105	101	87	89
	All Other	105	98	103	99	114	102	100	99

2.5. Execution of the Proficiency Test Program

2.5.1. Shipment of Samples

Approximately 7% of the laboratories that originally agreed to participate in the study did not carry through with their earlier declaration. Samples were therefore mailed to 1,195 laboratories, 87 less than the number originally agreeing to be in the study. Table 2.5.1 summarizes these dropouts by laboratory type. The majority of the dropouts were AAFP members.

Table 2.5.1. Number and Percent of Dropouts

Lab Type	Number Originally Agreeing	Number Dropped Out		Shipments Mailed to	
		n	%	n	%
AAFP	150	49	32.7	101	67.3
ASIM	264	13	4.9	251	95.1
Interstate	257	9	3.5	248	96.5
JCAH	222	7	3.2	215	96.8
Medicare Hospitals	236	4	1.7	232	98.3
Medicare Independent	153	5	3.3	148	96.7
Total	1282	87	6.8	1195	93.2

2.5.2. Response

Of the 1,195 laboratories to which samples were sent, some did not respond. Table 2.5.2 contains the number of laboratories in each group which received shipments and returned analytical results for neither, one, or both of the shipments (as of 5/25/72).

Table 2.5.2. Partial and Total Participation, by Laboratory Group

Lab Type	Shipments Mailed to	No Response		One (First) Shipment Analyzed		Both Shipments Analyzed		
		n	%	n	%	n	%	
AAFP	101	58	57.4	17	16.8	26	25.7	(8.7)[8]
ASIM	251	44	17.5	43	17.1	164	65.3	(33.2)
Interstate	248	17	6.8	23	9.3	208	83.9	(80.9)

[8] ( ) figures are percent of original list.

Table 2.5.2. Partial and Total Participation, by Laboratory Group (continued)

Lab Type	Shipments Mailed to	No Response		One (First) Shipment Analyzed		Both Shipments Analyzed	
		n	%	n	%	n	%
JCAH	215	14	6.5	24	11.2	177	82.3 (79.7)
Medicare Hospitals	232	31	13.4	30	12.9	171	73.7 (57.0)
Medicare Independent	148	18	12.2	24	16.2	106	71.6 (53.0)
Total	1195	182	15.2	161	13.5	852	71.3 (48.0)

Fifteen percent of the laboratories did not return results for either shipment. Approximately another 14% of the laboratories analyzed and returned the report forms for only one of the two shipments. Consequently, only 71% (852 out of 1195) of the laboratories analyzed and completed the report forms for both shipments. Sample sizes were sufficient to permit statistically significant method comparisons in clinical chemistry and hematology for all groups except the AAFP. Fortunately, statistical tests showed that the AAFP and ASIM pooled method data did not differ significantly, and it was possible to retain all data by creating a pooled Doctor's Office group. Not every laboratory had the capability to perform all of the assays routinely; consequently, the number of laboratory results reported will change from one constituent to another. Table 2.5.3 shows the number of usable analysis reports returned.

Table 2.5.3. Number of Analyses by Constituent

	Dialyzed Chemistry	Serum Pool Chemistry
Glucose	1292	174
Sodium	887	144
Urea Nitrogen	1256	172
Bilirubin	1163	176
Uric Acid	1433	199
Calcium	969	164
Total Protein	1136	175
Cholesterol	1284	180
	Hematology Shipment 1	Hematology Shipment 2
Hemoglobin	864	784
Hematocrit	847	770
RBC	790	717
WBC	866	789
MCV	731	653

2.5.3. Follow-Up to Non-Respondents

On April 27, 1972, a follow-up letter was sent from the National Bureau of Standards to all 475 laboratories which had not returned one or both of the two report forms. Because of the large number of laboratories from whom little or nothing had been heard it was interesting to tally their responses to our follow-up letter. As of May 31, 1972, more than four weeks after the letter was mailed, no reply had been received from 275 of the laboratories to whom it had been sent. The majority of the responses which were received fell largely into the following five categories:

(a) Failed to receive either one or both shipments - 44;
(b) Samples were broken upon arrival - 11;
(c) Time and personnel not available - 64;
(d) Insufficient knowledge about study - 37;
(e) All laboratory work sent out - 24.

Some other problems which arose were:

(a) Hospital administrators agreed to participate but did not tell the laboratory personnel—consequently when the unexpected samples arrived they were discarded;
(b) Physicians in private practice as well as hospital personnel received a letter from Warner-Lambert two weeks prior to receiving the first shipment, but never opened the envelope because it looked like "junk mail";
(c) A consulting pathologist told a hospital not to participate because the laboratory was not very good;
(d) Laboratories discarded the second shipment because they did not remember that there would be two shipments.

It is reasonable to expect that direct contact with the laboratories would have minimized these problems.

2.5.4. Other Problems

A great deal of time was spent reviewing the incoming laboratory report forms. Almost all of the problems were associated with laboratories which have little or no routine contact with data reporting forms (e.g., private physicians offices and small, rural hospitals which are not normally engaged in proficiency testing programs). Quite a few laboratories did not honor the decimal point when entering their results on the form; hence, order of magnitude errors were not uncommon.

Other problems which arose were:

(a) Laboratories whose submitted report forms were incomplete, and whose telephone numbers were not listed in their local telephone directory, making it impossible to complete the form and include their data;
(b) Laboratories returned the report form intended for their files (because it did not contain their name and address on it) and kept the copy they were supposed to return (which had the only means of identifying the source of the data);
(c) Physicians and laboratories not sending the report forms back to NBS in the return envelope provided—and addressing their return envelopes to CDC, JCAH, state health departments, etc.;
(d) Physicians and laboratories not observing the deadlines for sample analyses and submission of reports.

2.5.5. Treatment of Outliers

Every possible effort was expended toward identifying the rational basis for roughly 500 reported analyses which, upon first glance, appeared to be totally unrealistic. Through a series of letters and telephone calls almost 200 of these aberrant values were resolved—dilution errors, inversion of normal and abnormal results, arithmetical mistakes, etc. Table 2.5.5.1 shows the distribution by constituent of the 315 determinations which were categorized as obviously deviant and were eliminated from the data base.

Table 2.5.5.1. Reductions Due to Removal of Obviously Deviant Values[8]

Constituent	AAFP	ASIM	Interstate	JCAH	Medicare		Totals	
					Hospital	Indep.		
Glucose	0/67	3/513	3/697	2/610	6/634	3/426	17/2947	0.58
Urea Nitrogen	0/58	12/463	1/704	2/611	12/630	5/417	32/2883	1.11
Calcium	0/28	1/129	3/719	5/633	3/456	3/313	15/2278	0.66
Bilirubin	1/45	2/247	5/735	5/676	2/602	3/387	18/2692	1.67
Cholesterol	3/61	7/460	7/750	12/689	10/603	7/404	46/2967	1.55
Uric Acid	2/90	7/614	1/780	3/718	7/646	2/434	22/3282	0.67
Sodium	0/28	0/54	1/610	3/646	5/488	1/244	10/2070	0.48
Total Protein	0/44	0/201	2/769	6/713	2/557	0/346	10/2630	0.38
Red Blood Count	1/70	3/529	6/596	5/705	3/684	2/436	21/3020	0.70
White Blood Count	2/100	5/717	4/603	5/719	12/726	3/456	31/3321	0.93
Hemoglobin	2/99	6/709	3/604	2/716	2/717	1/455	16/3300	0.48
Hematocrit	2/97	8/646	4/612	4/720	1/705	5/461	24/3237	0.74
Mean Corpuscular Volume	2/59	9/404	18/596	11/702	11/639	7/416	58/2816	2.06
Total Number of Assays Deleted:	15/846	63/5686	58/8775	66/8858	76/8087	42/5195	320/37446	
Percent of Assays Submitted:	1.77	1.11	0.67	0.75	0.94	0.81	0.85	

[8] Numerator is number of assays deleted and denominator is total number of assays reported.

After removing the obvious deviants, it was necessary to determine whether the data base still contained any outliers; i.e., values so extreme that they should not be included in further data analysis. Data was excluded which did not meet the criteria of plus and minus four (4) standard deviations from the mean of the determinations by the given analytical technique. Applying the rule of $\pm$ 4 standard deviations eliminated data from the various laboratory groups as summarized in the Tables 2.5.5.2 through 2.5.5.4. The tables are interpreted as follows: each table entry gives the number of results which were more than $\pm$ 4 standard deviations away from the target value for the indicated constituent and the total number of results reported for the constituent. In all there were 273 analyses that were classified as outliers and not included in any subsequent treatment of the data.

Table 2.5.5.5 summarizes the overall percentage reductions by constituent and laboratory group. In all, 1.58% of the data were discarded. It was immediately obvious that, on a laboratory group basis, the largest frequency of outliers belonged to the private physicians' laboratories. On a percentage basis, the federally-licensed laboratories had the smallest number of extreme values rejected; the Medicare Independents were next, followed by the Joint Commission accredited laboratories and smaller hospitals, in that order.

On a constituent by constituent basis, the largest percentage of deletions were associated with cholesterol, mean corpuscular volume and hematocrit. The constituents with the smallest percentage of deletions were calcium, hemoglobin, uric acid and total bilirubin.

2.6. Limitations

It must be clearly understood that any results derived from this survey will be limited by four important considerations:

(a) Because all of the laboratories participated on a purely voluntary basis, no straightforward extrapolation can be made to the larger universe of unsampled clinical laboratories.
(b) It is probable that the results of this study do not represent routine laboratory performance for two reasons: (1) a laboratory probably would not volunteer if its management felt that to do so would be disadvantageous, and (2) the sample materials probably received special attention in many of the smaller laboratories which were unfamiliar with analyzing proficiency test samples.
(c) The clinical chemistry test specimens were prepared by a dialyzation process which removes naturally occuring substances. As a result, the accuracy of some methods, as applied to the test specimens, might differ from their accuracy in analyses of human serum.
(d) The true constituent concentrations of cholesterol and the hematology constituents could not be exactly determined. For these constituents accuracy was assessed relative to the MRAs.

3. CLINICAL CHEMISTRY

3.1. Definition

3.1.1. Accuracy

The accuracy of a laboratory determination is a measure of its distance from the true value. Unfortunately, there is no known method for determining true values of most clinical chemistry and hematology specimens. Therefore, accuracy could only be assessed relative to measured values which were considered to be close to the true values. Initially, the decision was made to assess relative accuracy as either: (a) the relative percent difference between the observed analytical result and the target value established during the weigh-back and dialysis procedure[9] for all chemistry constituents except cholesterol; or (b) the

[9] [(Obs. Value - Target) ÷ Target] x 100 = Relative Accuracy for Chemistry (except cholesterol).

Table 2.5.5.2. Reductions in Dialyzed Chemistry Sample Analyses Due to Elimination of Outliers[10]

Constituent	AAFP	ASIM	Interstate	JCAH	Medicare		Totals	
					Hospital	Indep.		
Glucose	1/57	2/438	7/545	2/539	4/582	2/421	18/2582	0.70%
Urea Nitrogen	0/48	1/383	2/551	5/544	1/572	1/410	10/2508	0.40
Calcium	0/24	1/84	1/558	3/545	1/417	1/308	7/1936	0.36
Bilirubin	1/34	0/191	6/567	1/596	5/556	1/382	14/2326	0.60
Cholesterol	0/50	5/381	6/579	9/603	8/554	7/395	35/2562	1.37
Uric Acid	0/76	6/525	2/609	0/633	5/591	0/430	13/2864	0.45
Sodium	1/24	0/36	6/465	4/564	9/443	0/241	20/1773	1.13
Total Protein	1/36	3/156	5/601	6/623	5/511	1/344	21/2278	0.92
Total Number of Assays Deleted:	4/349	18/2194	35/4475	30/4647	38/4226	13/2931	138/18822	0.73
Percent of Assays Submitted:	1.15	0.82	0.78	0.65	0.90	0.44	0.73	

[10] Numerator is number of assays deleted and denominator is total number of assays reported.

Table 2.5.5.3. Reductions in Validate Chemistry Sample Analyses Due to Elimination of Outliers[11]

Constituent	AAFP	ASIM	Interstate	JCAH	Medicare		Totals	
					Hospital	Indep.		
Glucose	0/10	1/72	1/149	0/69	0/46	0/2	2/348	% 0.57
Urea Nitrogen	0/10	0/68	1/152	0/65	0/46	0/2	1/343	0.29
Calcium	0/4	0/44	0/158	0/83	0/36	0/2	0/327	0.00
Bilirubin	0/10	0/54	0/163	0/75	0/44	0/2	0/348	0.00
Cholesterol	0/8	2/72	0/164	2/74	1/39	0/2	5/359	1.39
Uric Acid	1/12	2/82	0/170	0/82	0/48	0/2	3/396	0.76
Sodium	0/4	1/18	0/144	0/79	1/40	0/2	2/287	0.70
Total Protein	0/8	0/45	1/166	1/84	0/44	0/2	2/349	0.57
Total Number of Assays Deleted:	1/66	6/455	3/1266	3/611	2/343	0/16	15/2757	0.54
Percent of Assays Submitted:	1.52	1.32	0.24	0.49	0.58	0.00	0.54	

[11] Numerator is number of assays deleted and denominator is total number of assays reported.

Table 2.5.5.4. Reduction in Hematology Analyses Due to Elimination of Outliers[12]

	Constituent	AAFP	ASIM	Interstate	JCAH	Medicare Hospital	Medicare Indep.	Total	%
Shipment 1									
	Red Blood Count	1/41	3/260	1/327	3/369	2/354	0/226	10/1577	0.63
	White Blood Count	2/55	1/353	0/333	4/377	5/369	0/238	12/1725	0.70
	Hemoglobin	1/56	3/350	0/334	2/375	2/374	2/239	10/1728	0.58
	Hematocrit	2/55	8/314	1/339	2/376	4/368	1/239	18/1691	1.06
	Mean Corp. Vol.	0/35	1/194	3/320	1/367	4/332	1/214	10/1462	0.68
	Total Number of Assays Deleted:	6/242	16/1471	5/1653	12/1864	17/1797	4/1156	60/8183	0.73
	Percent of Assays Submitted:	2.48	1.09	0.30	0.64	0.95	0.35	0.73	
Shipment 2									
	Red Blood Count	3/28	2/266	2/263	4/330	0/327	2/208	13/1422	0.91
	White Blood Count	1/43	1/359	2/266	2/337	1/345	1/215	8/1565	0.51
	Hemoglobin	0/41	3/353	0/267	4/339	2/341	3/215	12/1556	0.79
	Hematocrit	2/40	5/324	1/269	3/340	2/336	4/217	17/1526	1.11
	Mean Corp. Vol.	0/22	2/201	4/258	2/324	2/296	0/195	10/1296	0.77
	Total Number of Assays Deleted:	6/174	13/1503	9/1323	15/1670	7/1645	10/1050	60/7365	0.81
	Percent of Assays Submitted:	3.45	0.86	0.68	0.90	0.43	0.95	0.81	

[12] Numerator is number of assays deleted and denominator is total number of assays reported.

Table 2.5.5.5. Total Reduction in the Data Base[13]

Constituent	Number	%
Glucose	37/2947	1.26
Urea Nitrogen	43/2883	1.49
Calcium	22/2278	0.97
Bilirubin	32/2692	1.19
Cholesterol	86/2967	2.90
Uric Acid	38/3282	1.16
Sodium	32/2070	1.55
Total Protein	33/2630	1.25
Red Blood Count	44/3020	1.46
White Blood Count	51/3321	1.54
Hemoglobin	38/3300	1.15
Hematocrit	59/3237	1.82
Mean Corpuscular Volume	78/2811	2.77

Lab Group	Number	%
AAFP	32/846	3.78
ASIM	116/5686	2.04
Interstate	110/8775	1.25
JCAH	126/8858	1.42
Medicare Hospital	140/8087	1.73
Medicare Lab	69/5195	1.33
All	593/37446	1.58

[13]Numerator is number of assays deleted and denominator is total number of assays reported.

relative percent difference between the observed analytical result and the manufacturer's mean assay[15] for cholesterol and the hematology samples (manufacturer-relative accuracy). However, the Advisory Committee felt that this definition was not wholly satisfactory, particularly with regard to hematology and cholesterol, because it relied on determinations by a single laboratory. Therefore, the decision was made to also assess accuracy relative to: (c) the mean reference assay (MRA) (reference-relative accuracy); and (d) the mean of all non-outlier laboratory analyses. These further assessments, reported in Section 3.7, showed that reference-relative accuracy was a more reliable measure than manufacturer-relative accuracy.

Regardless of which definition of relative accuracy is considered, the clinical chemistry results obtained will be subject to question because the weigh-back and dialysis procedure, as used by the manufacturer in preparing the samples for this study, removes naturally occurring substances. Consequently, some analytic methods will not perform on these samples in the same manner as on normal patient samples. This becomes particularly problematic when comparisons of performance by different methods are attempted.

A laboratory technique is statistically accurate if its mean does not differ significantly, at the 95% confidence level, from the target value.

3.1.2. Precision

The precision of laboratory determinations is a measure of reproducibility or consistency of results. It would be desirable to assess precision within individual laboratories, but it was not possible to obtain the necessary replicate analyses. Therefore, only inter-laboratory precision was assessed in this study.

3.1.3. Acceptability of Laboratory Determinations

The definition of criteria for measuring the acceptability of laboratory results is a topic which has received considerable attention. In 1968, the Subcommittee on Criteria of Medical Usefulness of the College of American Pathologists developed a set of basic subjective guidelines for evaluation of the medical usefulness of laboratory determinations.[16] These included:

(a) Desirable limits for accuracy and precision must be defined at each level of medical usefulness. Maximal accuracy and precision are necessary at levels where decisions are made regarding diagnosis or treatment.
(b) Accuracy and precision of a degree greater than is clinically useful should not be required if extra time or expense is thereby necessitated.
(c) Desirable precision should be such that errors induced by the measurement process do not significantly widen the range of values for the normal population.
(d) Desirable accuracy (of a technique) should be such that the method will create no substantial divergence from generally accepted values for normal and disease states.
(e) Ability to distinguish normal from abnormal values is often more important than the determination of absolute values.
(f) A less precise analytic technique free of large error may be preferable to a more precise method subject to erratic performance.

Unfortunately, the Subcommittee did not develop quantitative techniques for setting allowable limits for medical significance; however, other investigators have, and a review of such developments is appropriate.

---

[15] [(Obs. Value - Mfg. Mean) ÷ Mfg. Mean] x 100 = Relative Accuracy for hematology and cholesterol.

[16] R. Barnett, Clinical Laboratory Statistics, Little Brown, 1971, pp. 132-137.

Early empirical attempts[17] to define allowable limits of error in accuracy as a percentage deviation from the mean relied on the formula: Allowable limits of error (in %) = $\pm \frac{(1/4 \text{ of the normal range})}{(\text{mean of the normal range})} \times 100\%$.

> "If the normal range for sodium is taken to be 135-145 mEq/L, then the allowable limits of error calculated by this formula are ± 1.8%. The maximum limits for any determination, however, were set at ± 10%, even though in some cases those calculated by the above formula exceeded this figure."

The normal range, although intended as a measure of human deviation, is compounded with analytical variability. As a result, allowable limits calculated by this formula may not have medical significance when analytical variability is large relative to the human variations.

CLIA '67 defines satisfactory laboratory performance as the ability to achieve values which fall between the lowest lower limit and the highest upper limit of three superimposed subsets of empirically-derived limits for the sample under consideration.

These subsets of limits were obtained from data accumulated from (a) Volunteer Laboratories, (b) Reference Laboratories, and (c) Clinical Requirements.

Subset No. 1 (Volunteer or Licensed Laboratory Data)

This subset of limits encompasses the central 95% of all volunteer laboratory results. Obviously deviant results are not used in establishing the limits.

Subset No. 2 (Reference Laboratory Data)

This subset of limits encompasses all reference laboratory results and therefore consists of the lowest and highest results. Obviously deviant results are not used in establishing the limits. The median reference laboratory value (MRLV) for each sample is used as an estimate of the true concentration value for the desired constituent.

Subset No. 3 (Clinical Requirements)

This subset of limits is centered on the median reference laboratory value. For sample values in the normal range, the limits encompass an interval equal to one-half of the normal range of values given in several literature references. The limits for sample values outside the normal range encompass an interval equal to one-half of the value calculated by dividing the normal range of values by the midpoint of this normal range and multiplying this value by the median reference laboratory results.[18]

Two problems exist with limits based on these criteria: (a) they are difficult to interpret; and (b) the use of the volunteer laboratory data subset forces these to be limits for interlaboratory comparisons of ability rather than for measurement of acceptable quality. By criteria Subset 1, at most 5% of the volunteer laboratories can be performing in an unsatisfactory manner.

Recent research directed by the Clinical Pathology Department at the National Institutes of Health has yielded a theoretical basis for judging the medical acceptability of laboratory

---

[17]See, e.g., D. Tonks, "A Study of the Accuracy and Precision of Clinical Chemistry Determinations in 170 Canadian Laboratories," Clinical Chemistry, Vol. 9, No. 2, 10/63, pp. 217-233.

[18]CDC, Proficiency Testing, Clinical Chemistry Summary Analysis Survey I, 2/2/72.

precision and for appraising the limits of good laboratory performance as measured against what can be achieved within the state-of-the-art in clinical chemistry.[19] The basic hypothesis is that analytic variability in measurement within the normal range should not exceed one-half of the composite biological variation (personal and group) of the normal population. In cases where the abnormal target is larger than the normal target, Dr. Young suggests that the acceptable abnormal variability be computed by multiplying the acceptable normal variability by the ratio of the abnormal target value to the normal target value. Suggested maximum allowable variations for normal and abnormal clinical chemistry are labeled S_{BN} and S_{BA} respectively in Table 3.1.3.1. Replicate analyses were run over 30-day periods in the reference laboratory at NIH in an attempt to compare the state-of-the-art analytic capability with these requirements for medical acceptability of results. Analytically achieved standard deviations are displayed in Table 3.1.3.1 as S_A. As can be seen by comparing S_A and S_B, the current state-of-the-art does not provide the necessary capability to perform medically acceptable determinations for sodium and calcium. This problem is intensified in a proficiency testing situation where the test specimens are not completely homogeneous. In the current study, the vial-to-vial coefficient of variation was .5% of the target values for the normal specimens and .45% for the abnormal. The columns labeled S_{AN} and S_{AA} display the best normal and abnormal test specimen standard deviations achievable within the state-of-the-art. These deviations were calculated from the vial-to-vial deviations and the achievable analytic standard deviations.

Table 3.1.3.1.

	S_A	S_{BN}	S_{BA}	S_{AN}	S_{AA}
Glucose	1.5	3.3	12.2	1.6	2.0
Sodium	1.4	1.3	1.3	1.6	1.5
BUN	.52	1.7	6.0	.53	.60
Bilirubin	Not Available				
Uric Acid	.10	.39	.69	.10	.11
Calcium	.085	.046	.061	.096	.101
Total Protein	.13	.13	.13	.13	.13
Cholesterol	9.9	15.	15.	9.9	9.9

For the purposes of this study, acceptable precision of clinical chemistry results is defined as the larger of medically required precision (S_{BN} or S_{BA}) or 1.5 times the best achievable precision for the specimen (S_{AN} or S_{AA}) as displayed in Table 3.1.3.1. Acceptable relative accuracy of clinical chemistry results is defined by the range of acceptable precision around the target value for the test specimen.

3.2. Topics Addressed

Examination of the clinical chemistry data indicated that the following seven topics could be addressed:

---

[19] D. Young, E. Cotlove, E. Harris, et al., "Biological and Analytic Components of Variation in Long Term Studies of Serum Constituents in Normal Subjects," Clinical Chemistry, Parts I through III, Vol. 16, No. 12, 1970, pp. 1016-1032, Part IV, Vol. 17, No. 5, 1971, pp. 403-410.

(a) Differences in both precision and relative accuracy among laboratory groups;
(b) Differences in both precision and relative accuracy among laboratory techniques;[20]
(c) The acceptability of clinical laboratory determinations;
(d) The accuracy of manufacturer's average assays which are frequently used as target values in routine laboratory quality control programs;
(e) The feasibility of using the mean reference assays, the mean laboratory assays, or the mean assays performed with the techniques used by the manufacturer as target values;
(f) The likelihood that results of a proficiency test using dialyzed clinical chemistry specimens would differ significantly from results of a test using serum pool specimens;
(g) The performance of laboratories in this study in comparison to performance in other published proficiency tests.

Since analysis of topic (f) proved inconclusive, the relevant discussion has been consigned to Appendix C.

3.3. Analysis Plan

Both precision and accuracy can vary depending on the type of laboratory performing the analysis and on the laboratory technique used. Consequently, comparison of techniques must control for differences in laboratory group, and comparison of groups must control for differences in laboratory technique used. In the clinical chemistry analysis, the two physician groups will be pooled.

To simplify the ensuing discussion, the technique mean has been defined as the mean value of determinations by a single technique pooled over all groups. The group mean has been defined as the mean value of determinations by a single laboratory group pooled over all techniques. The technique variance and group variance have been defined as the variances analogous to the technique mean and group mean. The analysis discussed below has been carried out for each constituent at both the normal and abnormal levels. A plan for evaluation of the differences in techniques is discussed in subsections 3.3.1 and 3.3.2; the same plan will be applied in evaluating the differences in groups.

3.3.1. Relative Accuracy

To analyze whether a group applied a technique in a statistically accurate manner, a t-test was used to compare the group's technique mean with the appropriate target value. Some techniques were applied much more precisely or by many more laboratories than others. The t-test results may, therefore, be counter-intuitive. For example, a very imprecise technique might be judged statistically accurate, while a more precise technique which has a mean considerably closer to the target value might be judged statistically inaccurate. A second measure of relative accuracy, the average percentage difference of the group's normal and abnormal technique means from the target values, was therefore introduced. Comparative judgments on the relative accuracy of techniques applied by a group were based mainly on average percentage differences.

A second question addressed was whether the technique means differed significantly from one another. This question was analyzed using a one-way analysis of variance (ANOVA). ANOVA allows evaluation of the significance of differences in technique means based on the distribution of determinations by each group.

---

[20]Laboratory technique, in this report, is considered to be a unique combination of a particular analytical method and the system employed; i.e., the glucose oxidase/SMA-12 is considered to be different from the glucose oxidase/AutoAnalyzer.

3.3.2. Precision

Differences in precision of analyses can be assessed by a comparison of standard deviations using Bartlett's test for the homogeneity of variances. In cases where Bartlett's test indicates that technique variances cannot be considered homogeneous at the 95% confidence level, the precision of the techniques has been ranked, with the highest ranking (1) assigned to the most precise determination and subsequent rankings assigned in order of decreasing precision. A two-stage procedure was used to formulate the rankings. First, two-way analysis of variance results were examined to determine whether individual differences in technique precision were significant. A count was then made of the number of times (over applications by different groups and applications to normal and abnormal specimens) that each technique was significantly more precise than other techniques. This relative performance count was used as the basis for the rankings.

3.3.3. Medical Acceptability of Results

The acceptability of the results was determined by using χ^2-tests to evaluate whether the achieved group standard deviations significantly exceeded the medically acceptable precision as defined in subsection 3.1.3. In addition, for each constituent, 25% samples of the determinations by each group were selected randomly and plotted. The medically acceptable accuracy range was then plotted. The fraction of the determinations which lie outside the acceptable range provides a reasonable estimate of the likelihood that a laboratory report will not be medically useful.

3.3.4. Target Value Surrogates

The accuracy of manufacturer's average assays as target values can be judged by comparisons (using t-tests) with the weigh-back target values. Similar comparisons can be used to assess other possible targets.

3.3.5. Dialyzed Versus Pooled Serum Specimens

Because the laboratories receiving the validate specimens were randomly selected, pooling of the data over groups is permissible for the analysis of dialyzed versus pooled serum specimens. If the use of dialyzed specimens did affect the performance of a laboratory technique, then a t-test would confirm whether the percentage bias of the mean with respect to the target value for the dialyzed specimens differed significantly from the percentage bias with respect to the target value for the pooled serum specimens.

3.3.6. Comparison with Other Proficiency Tests

Since different studies use proficiency test samples with different assay values, it would be impossible to use means and variances as a basis for comparison. However, the coefficient of variation (the ratio of the standard deviation to the mean) provides a basis for comparison and has been used in this study.

3.4. Reporting of Results

In this study very large quantities of data were collected. Analysis of the data by laboratory group, sample batch, and analytical technique resulted in over 1,000 calculations of different means and standard deviations. These values are likely to be useful to many researchers and proficiency test program administrators, and consequently it seemed desirable to report this information. If all of the test statistics and intermediate statistical calculations performed in this study were also reported, there would be an overwhelming quantity of data. Therefore, the results of the statistical analysis are reported, but details such as the exact values of test statistics are omitted. All results are significant at the 95% confidence level unless otherwise stated.

3.5. Precision and Accuracy

One part of each table in this section displays information on both levels (normal and abnormal) of a single constituent; the information is broken down by laboratory group and technique. Results were entered only for those laboratory group/technique combinations which include 15 or more determinations. For each line item, the sample size (n), mean ($\bar{x}$), and standard deviation (s) are shown. For ease of reference, the table includes technique-dependent results pooled over all groups and group-dependent results pooled over all techniques. The weigh-back values for the normal and abnormal determinations (T) and the manufacturer's average assays (A) are indicated in parentheses in the table headings. At the bottom of each column, the relative ranking of the precision of each of the groups is indicated.

A discussion of differences in the group means has been omitted because the ANOVA analyses revealed that, at the 95% confidence level, there is no evidence that the technique means obtained by any laboratory group differed significantly from the technique means of any other group; thereby indicating that relative accuracy among groups does not differ significantly at this level. Because means do not differ among groups, it was possible to pool the data on all groups to analyze the difference in technique means. Differences in technique means could thus be discussed for any technique used for 15 or more laboratory determinations.

For each constituent, the discussion following the table contains a relative ranking of the precision of the techniques and an analysis of the relative accuracy of the techniques. For each constituent an associated table contains the pooled data used to make this analysis. This table displays, by technique, the means, standard deviations, sample sizes and percentage biases from the target values in cases where targets exist or from mean reference assays in cases where no targets exist. The percentage difference (% $\bar{d}$) is defined as 100 times the average bias from the target value divided by the target value. A star appears after those technique means which a t-test indicates are significantly different from the target values.

Bartlett's test revealed that precision differed by both group and technique for all constituents. Because it was demonstrated that the groups did not differ significantly in their overall accuracy, it was possible to judge them based on their relative precision. However, since laboratory techniques differed in both accuracy and precision, comparisons among them must consider both descriptors.

3.5.1. Glucose

The data are displayed in Tables 3.5.1.1 and 3.5.1.2. The precision of techniques[21] is ranked:

1	Ferricyanide (AutoAnalyzer)
1	Copper Neocuproine (SMA 12-60)
2	Glucose Oxidase
2	Folin-Wu
3	o-Toluidine
3	Kits.

Table 3.5.1.2 reveals that only three techniques (Somogyi Nelson, Ferricyanide (Auto-Analyzer) and Folin-Wu) were statistically accurate at both the normal and abnormal levels. Two techniques (Glucose Oxidase and Copper Neocuproine (SMA 12-60)) were not accurately applied at either level.

Of the glucose techniques used in this study, Ferricyanide (AutoAnalyzer) appears to be the most successfully applied. It is interesting to note that this is not one of the techniques which, when correctly applied, are considered to be the "best" glucose techniques.

---

[21]Throughout Section 3.5, techniques are manual or partially automated unless otherwise stated.

Table 3.5.1.1. Glucose Normal Dialyzed Specimens (T = 82.2, A = 80.7)

Technique	Doctors' Offices			Interstate			JCAH			Medicare Hospitals			Medicare Labs			All Groups		
	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
Glucose Oxidase	31	71.8	9.3										18	75.9	5.4	66	73.7	8.9
o-Toluidine	115	82.0	7.6	48	83.3	7.4	107	82.9	7.9	172	81.1	7.3	107	82.7	7.6	549	82.1	7.6
Folin-Wu	32	85.1	11.8										17	86.9	5.9	67	84.3	11.4
Ferricyanide (AutoAnalyzer)				45	80.2	4.8	68	82.3	5.2	26	81.3	7.0	22	80.9	3.2	165	81.4	5.3
Copper Neocuproine (SMA 12-60)				116	89.1	8.4	38	89.8	5.5				17	89.5	6.9	179	89.2	5.3
Kits	17	78.4	13.5							20	83.2	6.0				50	82.2	10.2
All Methods	245	80.8	10.3	270	85.8	7.1	268	84.4	7.8	289	81.5	8.2	210	83.0	7.4	1282	83.1	8.3
	Glucose Abnormal Dialyzed Specimens (T = 304, A = 294)																	
Glucose Oxidase	30	291.	24.4										18	283.	30.8	65	287.	26.2
o-Toluidine	116	293.	28.4	48	294.	25.6	107	289.	26.9	172	290.	26.7	106	288.	20.6	549	291.	25.9
Folin-Wu	33	297.	31.5										17	312.	24.2	68	300.	29.5
Ferricyanide (AutoAnalyzer)				43	298.	26.2	69	301.	18.0	26	298.	14.2	22	303.	12.8	165	300.	19.4
Copper Neocuproine (SMA 12-60)				117	311.	19.5	38	320.	18.5				17	309.	13.4	180	312.	18.8
Kits	18	289.	35.0							20	289.	30.5				51	289.	28.8
All Methods	248	296.	31.5	272	301.	30.9	269	299.	25.7	291	291.	28.8	210	294.	24.5	1282	297.	26.2
Ranking of Group Precision		5			4			3			2			1				

Table 3.5.1.2. Glucose (mg/100 ml)

Technique	Normal Level (T = 82.2)				Abnormal Level (T = 304)			
	n	$\bar{x}$	$\%\bar{d}$	s	n	$\bar{x}$	$\%\bar{d}$	s
Somogyi Nelson	31	81.5	-.9	7.0	32	298.5	-1.8	37.4
Glucose Oxidase	66	73.6*	-10.4	8.9	65	286.6*	-6.1	26.2
Ferricyanide (AutoAnalyzer)	165	81.4	-.9	5.3	164	300.2	-1.8	19.4
o-Toluidine	553	82.3	0.1	8.5	549	290.6*	-4.4	25.9
Hexokinase	19	89.3*	8.6	11.2	19	296.2	-2.6	25.7
Copper Neocuproine (AutoAnalyzer)	25	89.1*	8.4	5.2	25	310.6	2.2	19.8
Copper Neocuproine (SMA 12-30)	33	90.8*	10.5	6.4	32	299.5	-1.5	22.1
Copper Neocuproine (SMA 12-60)	181	89.9*	8.6	8.7	180	312.5*	2.8	18.8
Hycel Carbohydrate	39	83.4	1.5	7.2	38	292.1*	-3.9	30.6
Folin-Wu	67	84.3	2.6	11.4	68	300.3	-3.5	29.5
Kits	51	81.2	-1.2	12.5	51	289.0*	-4.9	28.8
All Methods	1282	$\bar{\bar{x}}$ = 83.1		8.4	1282	$\bar{\bar{x}}$ = 296.5		26.2

$$\%\bar{d} = \left(\frac{\bar{x} - T}{T}\right) 100$$

* Significantly different from the target value at the 95% confidence level.

3.5.2. Urea Nitrogen

Table 3.5.2.1 contains the data stratified by group and technique. Precision of techniques is ranked:

1	Diacetyl monoxime (SMA 12-60)
2	Diacetyl monoxime (AutoAnalyzer)
3	Berthelot Reaction
3	Chromatographic
3	Urease Nessler's
4	Diacetyl monoxime (Manual)
4	Kits.

The data displayed in Table 3.5.2.2 indicates that none of the techniques, as applied, accurately reproduced both the normal and abnormal target values. Two techniques were statistically accurate, however, at the normal level: Urease Nessler's and Kits. The Diacetyl monoxime determinations performed with the SMA 12-30 and SMA 12-60 were statistically accurate in the abnormal range. Only the automated Diacetyl monoxime techniques had average percentage differences (normal and abnormal combined) of less than 5%.

The automated Diacetyl monoxime techniques appear to outperform other techniques as they are currently being used.

3.5.3. Calcium

The relevant data are displayed in Tables 3.5.3.1 and 3.5.3.2. The Doctors' Offices were excluded from the ranking of groups because of insufficient sample size. Precision of techniques is ranked:

1	Cresolphthalein Complexone (SMA 12-60)
2	Cresolphthalein Complexone (AutoAnalyzer)
3	Calcein Fluorimetric
3	Chloranilate precipitate
4	Emission flame photometer
5	Kits.

Table 3.5.3.2 shows that, on average, all of the techniques underestimate the calcium weigh-back value at the abnormal level. Nine of the 11 techniques also underestimate the mean reference assay by the reference technique (MRART). Four techniques yielded statistically accurate normal determinations: Oxalate precipitate, Chloranilate precipitate, Emission flame photometry and Cresolphthalein Complexone (AutoAnalyzer). Averaging the errors at the normal and abnormal levels yields a mean error of less than 5% for Oxalate precipitate, Chloranilate precipitate, Atomic Absorption Spectrometry and all three automated Cresolphthalein techniques.

Sample sizes were insufficient to support statistically valid conclusions about the precision of the Oxalate precipitate and Atomic Absorption techniques. However, it would appear that Atomic Absorption was applied with a precision roughly equal to that of the automated Cresolphthalein techniques. Similarly, the precision achieved with the Oxalate precipitate technique appears to be roughly equivalent to the precision achieved with the Chloranilate precipitate technique. Thus, when both accuracy and precision are considered, it would appear that calcium determinations performed with either Atomic Absorption Spectrometry or the automated Cresolphthalein Complexone techniques were the most satisfactory.

3.5.4. Total Bilirubin

Table 3.5.4.1 presents the data for total bilirubin stratified by technique and group. Precision of techniques is ranked:

Table 3.5.2.1. Urea Nitrogen Normal Dialyzed Specimens (T = 18.6, A = 18.7)

Technique	Doctors' Offices			Interstate			JCAH			Medicare Hospitals			Medicare Labs			All Groups		
	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
Urease-Nessler's	15	17.2	3.9										21	18.7	1.8	64	18.0	3.1
Diacetyl monoxime (manual)	84	18.1	2.2	37	18.1	2.4	72	18.0	2.1	103	18.1	2.7	69	18.4	2.5	365	18.1	2.4
Diacetyl monoxime (AutoAnalyzer)				47	19.4	1.3	83	19.3	1.4	33	19.5	1.0	20	19.9	1.1	189	19.4	1.3
Diacetyl monoxime (SMA 12-60)				125	19.0	.8	38	19.1	.9				23	19.0	.8	195	19.0	.9
Berthelot Reaction	42	15.4	2.9	17	16.7	2.5	37	14.9	2.2	33	15.0	3.4	48	15.9	2.7	177	15.5	2.8
Chromotographic	29	15.8	2.5							51	17.4	3.5				99	16.7	3.1
Kits	26	16.8	4.7							46	18.8	3.3				105	17.9	3.7
All Methods	217	17.2	3.3	275	18.6	1.8	269	18.0	2.5	286	18.0	3.2	205	18.0	2.5	1252	18.0	2.7
Urea Nitrogen Abnormal Dialyzed Specimens (T = 65.4, A = 65.4)																		
Urease-Nessler's	15	52.3	16.1										21	63.0	14.0	64	58.2	15.9
Diacetyl monoxime (manual)	82	57.8	10.4	36	57.5	11.4	74	63.3	9.6	103	60.8	11.1	68	61.7	10.8	363	60.5	10.8
Diacetyl monoxime (AutoAnalyzer)				47	67.3	3.3	83	66.4	4.5	33	66.0	2.6	20	67.8	2.6	188	66.6	3.9
Diacetyl monoxime (SMA 12-60)				125	65.0	2.2	38	65.8	2.5				23	65.9	1.5	195	65.2	2.4
Berthelot Reaction	42	57.0	10.1	17	52.6	6.1	36	54.5	5.1	33	55.1	9.0	48	58.3	6.6	176	56.0	7.9
Chromotographic	23	53.0	12.0							50	58.9	9.7				97	57.5	11.5
Kits	26	57.1	12.7							46	61.4	15.1				105	59.5	13.1
All Methods	213	57.3	11.7	274	63.3	7.2	270	62.8	8.3	285	60.4	11.7	204	62.1	9.4	1246	61.3	10.0
Ranking of Group Precision		4			2			2			3			1				

Table 3.5.2.2. Urea Nitrogen (mg/100 ml)

Technique	Normal Level (T = 18.6)				Abnormal Level (T = 65.4)			
	n	$\bar{x}$	%$\bar{d}$	s	n	$\bar{x}$	%$\bar{d}$	s
Urease reaction-Nessler's	64	18.0	-3.3	3.1	64	*58.2	-11.0	15.9
Diacetyl monoxime (Manual or partially automated)	365	*18.1	-2.7	2.4	363	*60.5	-7.5	10.8
Diacetyl monoxime (AutoAnalyzer)	189	*19.4	4.3	1.3	188	*66.6	1.8	3.9
Diacetyl monoxime (SMA 12-30)	42	*19.1	2.7	1.0	42	66.0	.9	2.3
Diacetyl monoxime (SMA 12-60)	195	*19.0	2.2	.9	195	65.2	-.3	2.4
Berthelot Reaction	177	*15.5	-16.7	2.8	176	*56.0	-14.4	7.9
Chromotographic (Urograph)	99	*16.7	-10.2	3.1	97	*57.5	-12.1	11.5
Kit	105	17.9	-3.7	3.7	105	*59.5	-9.0	13.1
All Methods	1252	$\bar{\bar{x}}$ = 18.0		2.7	1246	$\bar{\bar{x}}$ = 61.3		10.0

* Significantly different from the target value at the 95% confidence level.

Table 3.5.3.1. Calcium Normal Dialyzed Specimens (T = 9.2, A = 9.2)

Technique	Doctors' Offices			Interstate			JCAH			Medicare Hospitals			Medicare Labs			All Groups		
	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
Chloranilate precip.							32	9.4	1.2	27	9.2	.8	27	9.2	1.3	107	9.3	1.1
Emission Flame Photom.							33	9.2	1.3	47	8.9	1.8	15	8.9	1.2	108	9.1	1.5
Calcein-Fluorimetric				19	7.7	1.0	38	7.7	1.0	32	7.7	.9				98	7.7	1.0
Cresolphthalein Complex. (AutoAnalyzer)				21	9.1	.8	16	9.5	.4							50	9.2	.9
Cresolphthalein Complex. (SMA 12-60)				136	8.9	.3	54	8.8	.2				30	9.0	.3	228	8.9	.3
Kits				26	7.6	1.3	48	8.0	1.6	66	8.0	2.2	28	8.5	1.2	181	8.0	1.7
All Methods	54	8.7	1.3	279	8.6	.8	272	8.7	1.3	208	8.4	1.7	153	8.8	1.0	966	8.7	1.2
Calcium Abnormal Dialyzed Specimens (T = 12.2, A = 11.9)																		
Chloranilate precip.							31	11.8	1.4	27	11.4	1.6	27	11.4	1.2	106	11.6	1.3
Emission Flame Photom.							33	10.7	1.5	46	10.5	1.8	15	10.5	.9	107	10.6	1.5
Calcein-Fluorimetric				19	10.5	1.8	38	10.3	1.6	33	9.9	1.4				99	10.3	1.6
Cresolphthalein Complex. (AutoAnalyzer)				21	11.6	.6	15	11.3								50	11.3	1.1
Cresolphthalein Complex. (SMA 12-60)				136	11.4	.4	54	11.3	.3				30	11.4	.4	228	11.4	.4
Kits				25	9.8	1.8	48	10.3	2.2	66	9.7	2.9	28	10.8	1.8	180	10.1	2.4
All Methods	53	11.0	1.3	278	11.3	.8	270	11.0	1.3	208	10.4	1.7	154	11.1	1.0	963	10.9	1.6
Ranking of Group Precision					2			2			3			1				

Table 3.5.3.2. Calcium (mg/100 ml)

Technique	Normal Level (T = 9.2, MRART = 9.0)				Abnormal Level (T = 12.2, MRART = 11.7)			
	n	$\bar{x}$	%$\bar{d}$	s	n	$\bar{x}$	%$\bar{d}$	s
Oxalate precipitate	26	9.0	-1.3	1.1	25	*11.7	-4.1	.9
Chloranilate precipitate	107	9.3	1.1	1.1	106	*11.6	-4.9	1.3
Emission Flame photometry	108	9.1	-1.1	1.5	107	*10.6	-13.1	1.5
Atomic Absorption Spectrometry	39	*9.0	-2.2	0.3	39	*11.8	-3.3	0.5
Calcein-Fluorimetric	98	*7.7	-16.3	1.0	99	*10.3	-15.6	1.6
Calcein-non-fluorimetric	35	*8.4	-8.7	1.4	35	*10.7	-12.3	1.9
Cresolphthalein Complexone (Manual or partially automated)	17	*8.0	-13.0	1.2	17	*10.6	-13.1	1.9
Cresolphthalein Complexone (AutoAnalyzer)	50	9.2	0.0	.9	50	*11.3	-7.4	1.1
Cresolphthalein Complexone (SMA 12-30)	45	*8.9	-3.3	0.3	45	*11.5	-5.7	0.5
Cresolphthalein Complexone (SMA 12-60)	228	*8.9	-3.3	0.3	228	*11.4	-6.6	0.4
Kit	181	*8.0	-13.0	1.7	180	*10.1	-17.2	2.4
All Methods	966	$\bar{\bar{x}}$ = 8.7		1.2	963	$\bar{\bar{x}}$ = 10.9		1.6

* Significantly different from the target value at the 95% confidence level.

MRART: Mean reference assay by the reference techniques.

Table 3.5.4.1 Total Bilirubin Normal Dialyzed Specimens (T = .9, A = .98)

Technique	Doctors' Offices			Interstate			JCAH			Medicare Hospitals			Medicare Labs			All Groups		
	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
Diazo-Alcohol	57	1.05	.32	68	.97	.29	145	1.05	.22	149	1.07	.33	119	1.08	.28	537	1.05	.28
Diazo-Other (Manual)	29	1.26	.42	34	1.13	.37	62	1.33	.38	69	1.22	.36	23	1.05	.32	217	1.22	.38
Diazo-Other (AutoAnalyzer)				15	1.15	.28	15	1.41	.28							42	1.30	.29
Diazo-Other (SMA 12-60)				122	1.20	.16	51	1.19	.15				30	1.14	.16	212	1.19	.16
All Methods	113	1.19	.40	282	1.15	.27	297	1.18	.29	276	1.16	.37	190	1.09	.24	1158	1.15	.31
	Total Bilirubin Abnormal Dialyzed Specimens (T = 4.7, A = 4.5)																	
Diazo-Alcohol	56	4.8	.8	66	4.7	.6	145	4.8	.7	148	4.8	1.0	119	4.8	.7	538	4.8	.9
Diazo-Other (manual)	29	5.0	1.0	34	4.8	.7	63	5.0	.9	70	4.6	.9	23	4.6	.7	219	4.8	.9
Diazo-Other (AutoAnalyzer)				15	4.8	.4	15	5.0	.3							42	5.0	.6
Diazo-Other (SMA 12-60)				119	5.3	.6	51	5.3	.7				30	5.1	.5	212	5.2	.7
All Methods	111	4.9	1.0	279	5.0	.7	298	5.0	.8	275	4.8	1.1	191	4.9	.7	1154	4.9	.9
Ranking of Group Precision		3			2			2			3			1				

1 Diazo-Other (AutoAnalyzer)
1 Diazo-Other (SMA 12-60)
2 Diazo-Alcohol
3 Diazo-Other (Manual)

Table 3.5.4.2 indicates that none of the techniques, as applied, were statistically accurate at the normal level. Furthermore, all techniques overestimated the normal bilirubin content. Only three techniques (Diazo-Other Coupling (Manual), Spectrophotometric, and Kits) were statistically accurate at the abnormal level.

The data indicates that, as presently applied, there are no techniques which are both relatively precise and statistically accurate. Therefore, a choice between techniques will require a decision on whether an accurate technique with low precision or a biased, but precise, technique is preferable. A feasible basis for this choice is that if a reasonably precise but biased technique is used to redefine the normal range, then determination of the normality of specimens tested with that technique will be reasonably consistent. Based on this reasoning Diazo-Other (AutoAnalyzer) and Diazo-Other (SMA 12-60) are the techniques that are preferred as they are presently utilized.

3.5.5. Cholesterol

Tables 3.5.5.1 and 3.5.5.2 present the data for cholesterol. The precision of the techniques is ranked:

1 Lieberman-Burchard w/o Extraction (SMA 12)
2 Lieberman-Burchard w/o Extraction (Manual)
2 $FeCl_3$-H_2SO_4 with Extraction (AutoAnalyzer)
3 Kit with Extraction
4 Kit w/o Extraction
4 Paratoluene-Sulfonic Acid w/o Extraction.

Cholesterol was unique among the chemistries in that there was no weigh-back value associated with the samples. Consequently, to determine the relative accuracy of the various techniques it was necessary to use a proxy for this value. In Section 3.7, the mean reference assay is explained to be the most reasonable target value. Table 3.5.5.2 shows that the $FeCl_3$-H_2SO_4 technique, with prior extraction (Manual) was statistically accurate at both normal and abnormal levels. The $FeCl_3$-H_2SO_4 (AutoAnalyzer) technique, with prior extraction, also proved statistically accurate at the abnormal level and had an average error of less than 5%. Among the many techniques used to analyze cholesterol samples in this study, these two appear to be the most successfully applied.

3.5.6. Uric Acid

The data for uric acid are displayed in Tables 3.5.6.1 and 3.5.6.2. The precision of the techniques is ranked as:

1 Phosphotungstate (AutoAnalyzer)
1 Phosphotungstate (SMA 12)
2 Phosphotungstate (Manual)
2 Phosphotungstate (Kit)
3 Other Kit.

The data in Table 3.5.6.2 indicates that participants in this study who analyzed uric acid with the Uricase technique achieved statistical accuracy. However, the small sample size made it impractical to include this method in the rankings of precision. Of the other techniques used, Kits were statistically accurate at the normal level but were more than 5% inaccurate at the abnormal level; while Phosphotungstate (SMA 12) determinations were statistically accurate at the abnormal level and less than 5% inaccurate at the normal level. In addition, the Phosphotungstate (AutoAnalyzer) determinations were less than 5% inaccurate at both levels.

Table 3.5.4.2. Total Bilirubin (mg/100 ml)

Technique	Normal Level (T = 0.9)				Abnormal Level (T = 4.7)			
	n	$\bar{x}$	$\%\bar{d}$	s	n	$\bar{x}$	$\%\bar{d}$	s
Diazo-Alcohol Coupling (Manual or partially automated)	537	*1.0	11.1	0.28	534	*4.8	2.1	0.81
Diazo-Alcohol Coupling (SMA 12-60)	17	*1.2	33.3	0.19	17	*5.3	12.8	0.86
Diazo-Other Coupling (Manual or partially automated)	217	*1.2	33.3	0.38	219	4.8	2.1	0.89
Diazo-Other Coupling (AutoAnalyzer)	42	*1.3	44.4	0.29	41	*4.9	4.3	0.39
Diazo-Other Coupling (SMA 12-30)	34	*1.3	44.4	0.14	36	4.5	-4.3	0.57
Diazo-Other Coupling (SMA 12-60)	212	*1.2	33.3	0.16	208	*5.3	12.8	0.61
Spectrophotometric	16	*1.3	44.4	0.57	16	4.4	-6.4	1.34
Kits	65	*1.4	55.5	0.40	65	4.9	4.3	1.33
All Methods	1158	$\bar{\bar{x}}$ = 1.2		0.31	1154	$\bar{\bar{x}}$ = 4.9		0.86

* Significantly different from the target value at the 95% confidence level.

Table 3.5.5.1. Cholesterol Normal Dialyzed Specimen (T' = 134, A = 160)

Technique	Doctors' Offices			Interstate			JCAH			Medicare Hospitals			Medicare Labs			All Groups		
	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
Liebermann-Burchard w/o Extraction (Manual)	95	149.	14.1	55	151.	11.3	144	150.	10.9	172	152.	15.6	97	149.	12.1	563	150.	13.3
Liebermann-Burchard w/o Extraction (SMA 12)				137	149.	8.9	53	149.	10.7				33	150.	8.7	232	149.	9.4
Paratoluene Sulfonic w/o Extraction	18	159.	17.3										20	150.	18.6	67	153.	17.4
$FeCl_3$-H_2SO_4 w/Extr. (AutoAnalyzer)				16	127.	18.9	18	124.	11.0							41	126.	14.2
Kit w/o Extraction	32	148.	18.0							30	149.	26.3				86	148.	20.0
Kit with Extraction							17	149.	20.6	19	149.	8.1				58	150.	15.3
All Methods	214	150.	16.1	286	147.	12.4	300	146.	15.5	273	150.	18.8	194	147.	13.3	1267	148.	15.5
				Cholesterol Abnormal Dialyzed Specimen (T' = 74, A = 70)														
Liebermann-Burchard w/o Extraction (Manual)	93	97.	12.8	54	101.	13.0	141	99.	10.1	172	99.	11.7	96	95.	9.2	556	98.	11.4
Liebermann-Burchard w/o Extraction (SMA 12)				140	95.	8.9	51	95.	12.8				34	96.	7.6	234	95.	9.7
Paratoluene Sulfonic w/o Extraction	18	100.	18.8										20	97.	19.4	67	99.	18.7
$FeCl_3$-H_2SO_4 w/Extr. (AutoAnalyzer)				15	72.	8.1	18	71.	9.7							40	72.	8.3
Kit w/o Extraction	31	91.	15.8							30	93.	18.6				85	94.	15.7
Kit with Extraction							17	90.	15.3	19	97.	13.2				59	94.	18.4
All Methods	212	95.	16.8	287	93.	14.0	294	93.	15.2	273	97.	14.6	194	93.	12.1	1260	94.	14.7
Ranking of Group Precision		2			2			3			3			1				

Table 3.5.5.2. Cholesterol (mg/100 ml)

Technique	Normal Level (T* = 134.2) n	$\bar{x}$	%d'	s	Abnormal Level (T* = 74.1) n	$\bar{x}$	%d'	s
w/o Prior Extraction								
Liebermann-Burchard (Manual)	563	*150.5	12.1	13.3	556	*98.3	32.7	11.4
Liebermann-Burchard (AutoAnalyzer)	35	*148.5	10.7	11.8	34	*97.7	31.8	15.4
Liebermann-Burchard (SMA 12)	232	*148.9	11.0	9.7	234	*95.3	28.6	9.4
Paratoluene Sulfonic Acid	67	*152.9	13.9	17.4	67	*98.9	33.5	18.7
$FeCl_3$-H_2SO_4 (Manual)	67	*143.8	7.2	17.8	67	*82.8	11.7	16.1
Kit	86	*148.4	10.6	20.0	85	*93.5	26.2	15.7
w/Prior Extraction								
Liebermann-Burchard (Manual)	20	139.9	4.2	20.3	20	*92.5	24.8	15.4
Liebermann-Burchard (SMA 12)	15	*142.9	6.5	12.0	15	*92.1	24.3	10.4
$FeCl_3$-H_2SO_4 (Manual)	34	127.7	-4.8	20.2	34	71.6	3.4	11.2
$FeCl_3$-H_2SO_4 (AutoAnalyzer)	41	*125.5	-6.5	14.2	40	71.8	3.1	8.3
Kit	58	*150.2	11.9	15.3	59	*93.7	26.5	18.4
All Methods	1267	$\bar{\bar{x}}$ = 147.8		15.5	1260	$\bar{\bar{x}}$ = 94.3		14.7

T* = mean reference assay

1/ %d' = $[(\bar{x} - \bar{\bar{x}})/\bar{\bar{x}}] \times 100$ where $\bar{x}$ and $\bar{\bar{x}}$ are shown in the table.

* Significantly different from the target value at the 95% confidence level.

Table 3.5.6.1. Uric Acid Normal Dialyzed Specimen (T = 6.2, A = 6.3)

Technique	Doctors' Offices			Interstate			JCAH			Medicare Hospitals			Medicare Labs			All Groups		
	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
Phosphotungstate (Manual)	149	6.0	1.1	72	6.0	.7	176	6.1	.7	201	6.1	.9	131	6.3	.8	729	6.1	.9
Phosphotungstate (AutoAnalyzer)				41	6.3	.3	44	6.3	.3				17	6.4	.4	122	6.3	.4
Phosphotungstate (SMA 12)				131	6.4	.3	51	6.4	.2				31	6.3	.3	223	6.4	.3
Phosphotungstate Kit	44	6.5	.9	20	6.8	.9	18	7.1	.9	16	6.9	1.7				104	6.7	1.1
Other Kit	41	6.4	1.5							37	6.0	1.2	18	6.0	.8	117	6.1	1.2
All Methods	300	6.2	1.2	303	6.3	.6	317	6.3	.7	292	6.1	1.0	215	6.3	.7	1427	6.2	.9
Uric Acid Abnormal Dialyzed Specimen (T = 11.0, A = 12.1)																		
Phosphotungstate (Manual)	145	10.2	1.5	71	10.3	1.4	176	10.4	1.2	203	10.3	1.3	131	10.6	1.3	726	10.3	1.3
Phosphotungstate (AutoAnalyzer)				41	11.4	.4	44	11.3	.6				17	11.6	.4	122	11.3	.6
Phosphotungstate (SMA 12)				133	10.9	.6	50	11.0	.5				31	11.0	.3	223	10.9	.5
Phosphotungstate Kit	44	10.7	1.0	20	11.1	1.0	18	11.7	1.2	15	10.9	1.7				103	11.0	1.2
Other Kit	40	10.5	1.5							38	9.8	1.7	18	9.9	1.5	117	10.2	1.6
All Methods	295	10.4	1.5	304	10.9	1.0	316	10.7	1.1	294	10.3	1.4	215	10.7	1.2	1424	10.6	1.3
Rankings of Group Precision		2			1			1			3			1				

Table 3.5.6.2. Uric Acid (mg/100 ml)

Technique	Normal Level (T = 6.2)				Abnormal Level (T = 11.0)			
	n	$\bar{x}$	$\%\bar{d}$	s	n	$\bar{x}$	$\%\bar{d}$	s
Phosphotungstate (Manual)	729	*6.1	-1.6	0.87	726	*10.3	-6.4	1.33
Phosphotungstate (AutoAnalyzer)	122	*6.3	1.6	0.37	123	*11.3	2.7	0.63
Phosphotungstate (SMA 12)	223	*6.4	3.2	0.29	223	10.94	-.5	0.54
Uricase/manual	31	6.2	0.0	0.66	31	10.9	-.9	1.30
Kit	117	6.2	0.0	1.20	117	*10.2	-7.3	1.61
All Methods	1427	$\bar{\bar{x}}$ = 6.2		0.88	1424	$\bar{\bar{x}}$ = 10.6		1.27

* Significantly different from the target value at the 95% confidence level.

Therefore, it would appear that study participants achieved adequate results using the Uricase, Phosphotungstate (SMA 12) and Phosphotungstate (AutoAnalyzer) techniques.

3.5.7. Sodium

The stratified data is displayed in Table 3.5.7.1. Precision of techniques is ranked:

1 Flame photometer (SMA 6)
2 Flame photometer (Manual).

Table 3.5.7.2 indicates that all of the techniques except Flame photometer (Manual) yielded statistically accurate determinations at both the normal and abnormal levels. Although the sample sizes are too small to make operationally significant statements about the relative precision of three of the four acceptably accurate techniques, it would appear that the Flame photometer (AutoAnalyzer) and Flame photometer (SMA 12) techniques were applied with a precision roughly equivalent to that of the Flame photometer (SMA 6), while the kits were applied considerably less precisely. Thus, it appears that the automated Flame photometer techniques were applied most successfully by participants in this study.

3.5.8. Total Protein

Based upon the data displayed in Table 3.5.8.1, the precision of the techniques is ranked:

1 Biuret (AutoAnalyzer)
1 Biuret (SMA 12)
1 Refractometer
2 Biuret (Manual).

Examination of Table 3.5.8.2 reveals that Biuret (Manual) and two types of kits were applied with statistically accurate results at the normal level. None of the techniques were statistically accurate at the abnormal level. Consideration of the average percentage inaccuracy reveals that all techniques, except the kits, achieved less than a 5% average inaccuracy.

It appears that the Refractometer, Biuret (AutoAnalyzer), and Biuret (SMA 12) techniques were all adequately applied in analyzing total protein.

3.5.9. Summary of Group and Technique Performance

There were no significant differences in the relative accuracy of constituent analyses by the different laboratory groups. Table 3.5.9.1 summarizes the ranked precision of the groups. In order to facilitate analysis, tied ranks have been assigned an average ranking; for example, if two analyses were tied as the most precise, they would both be ranked 1.5, the average of 1 and 2. Application of Friedman's Two-Way Analysis of Variance by Ranks[22] revealed that the groups differ significantly in precision (at the 99% confidence level).

---

[22]For further details on this test, see R. G. Miller, Jr., Simultaneous Statistical Inference, McGraw-Hill, 1966, pp. 171-173.

Table 3.5.7.1. Sodium Normal Dialyzed Specimen (T = 141, A = 141)

Technique	Doctors' Offices			Interstate			JCAH			Medicare Hospitals			Medicare Labs			All Groups		
	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
Flame photometer (Manual)	27	140.	3.1	187	140.	2.6	230	140.	3.3	196	140.	3.5	99	141.	3.1	739	140.	3.2
Flame photometer (SMA 6)				18	140.	2.5	26	141.	1.9							56	140.	2.0
All Methods	30	140.	3.6	230	140.	2.6	281	140.	3.1	217	140.	3.6	121	140.	3.0	879	140.	3.1
	Sodium Abnormal Dialyzed Specimen (T = 117, A - 116)																	
Flame photometer (Manual)	26	120.	5.7	187	117.	3.7	228	118.	5.2	196	120.	6.1	98	118.	3.8	735	118.	5.1
Flame photometer (SMA 6)				17	118.	3.5	26	116.	2.5							55	117.	2.8
All Methods	29	121.	6.8	229	117.	4.2	282	118.	4.9	217	120.	8.1	120	118.	3.7	874	118.	5.8
Ranking of Group Precision		3			2			2			4			1				

Table 3.5.7.2. Sodium (mEq/L)

Technique	Normal Level (T = 140.5)				Abnormal Level (T = 116.3)			
	n	$\bar{x}$	$\%\bar{d}$	s	n	$\bar{x}$	$\%\bar{d}$	s
Flame photometer (Manual)	739	*140.1	-.3	3.2	735	*118.3	1.7	5.1
Flame photometer (AutoAnalyzer)	33	140.1	-.3	2.9	33	115.9	-0.3	3.3
Flame photometer (SMA 12)	22	140.4	-.1	2.0	22	117.1	0.7	2.7
Flame photometer (SMA 6)	56	140.5	0.0	2.0	55	116.8	0.4	2.8
Kit	26	139.1	-1.0	5.0	26	119.3	2.6	18.7
All Methods	879	$\bar{\bar{x}}$ = 140.1		3.1	874	$\bar{\bar{x}}$ = 118.1		5.8

* Significantly different from the target value at the 95% confidence level.

Table 3.5.8.1. Total Protein Normal Dialyzed Specimen (T = 6.8, A = 6.8)

Technique	Doctors' Offices			Interstate			JCAH			Medicare Hospitals			Medicare Labs			All Groups		
	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
Biuret (Manual)	42	6.7	.3	67	6.8	.4	133	6.8	.3	158	6.8	.4	95	6.8	.3	495	6.8	.4
Biuret (AutoAnalyzer)				86	6.9	.2	48	7.0	.3				21	6.9	.2	167	6.9	.3
Biuret (SMA 12)				112	6.9	.2	50	6.8	.3	17	6.9	.3	30	6.8	.2	221	6.9	.2
Refractometer				28	7.0	.2	69	6.9	.3	50	6.9	.2	20	7.0	.1	177	6.9	.2
All Methods	93	6.8	.4	299	6.9	.3	309	6.9	.3	253	6.8	.4	172	6.8	.3	1126	6.8	.3
Total Protein Abnormal Dialyzed Specimen (T = 3.3, A = 3.3)																		
Biuret (Manual)	42	3.5	.4	66	3.6	.3	132	3.5	.3	159	3.5	.3	94	3.5	.4	493	3.5	.4
Biuret (AutoAnalyzer)				86	3.5	.2	47	3.6	.1				21	3.5	.2	166	3.5	.2
Biuret (SMA 12)				111	3.6	.1	50	3.5	.2	16	3.5	.2	30	3.5	.1	221	3.5	.2
Refractometer				28	3.3	.2	70	3.1	.2	50	3.1	.2	20	3.2	.2	178	3.2	.2
All Methods	95	3.5	.4	297	3.5	.2	308	3.4	.3	253	3.4	.4	171	3.5	.3	1124	3.5	.3
Ranking of Group Precision		2			1			2			2			1				

Table 3.5.8.2. Total Protein (gm/100 ml)

Technique	Normal Level (T = 6.8)				Abnormal Level (T = 3.3)			
	n	$\bar{x}$	%$\bar{d}$	s	n	$\bar{x}$	%$\bar{d}$	s
Biuret (Manual)	495	6.8	0.0	0.36	493	*3.5	6.1	0.35
Biuret (AutoAnalyzer)	167	*6.9	1.5	0.25	166	*3.5	6.1	0.18
Biuret (SMA 12)	221	*6.9	1.5	0.23	221	*3.5	6.1	0.18
Refractometer	177	*6.9	1.5	0.22	178	*3.2	-3.0	0.20
Kit, Type 1	27	6.8	0.0	0.78	27	*3.7	12.1	0.48
Kit, Type 2	32	7.0	2.9	0.50	32	*3.8	15.2	0.57
All Methods	1126	$\bar{\bar{x}}$= 6.8		0.33	1124	$\bar{\bar{x}}$ = 3.5		0.33

* Significantly different from the target value at the 95% confidence level

Table 3.5.9.1. Ranking of Precision of Groups by Constituent

	Doctors' Offices	Interstate	JCAH	Medicare Hospitals	Medicare Labs.
Glucose	5	4	3	2	1
Urea Nitrogen	5	2.5	2.5	4	1
Calcium		2.5	2.5	4	1
Bilirubin	4.5	2.5	2.5	4.5	1
Uric Acid	2.5	2.5	4.5	4.5	1
Cholesterol	4	2	2	5	2
Sodium	4	2.5	2.5	5	1
Total Protein	4	1.5	4	4	1.5

The simultaneous confidence interval for multiple comparisons based on the Friedman test (as given by Miller) was adjusted using the correction factor given by Gibbons.[23] This adjustment was necessary because ties are present in the data. Applications of the adjusted multiple comparison test to the mean ranks when calcium was and was not included in the analysis yielded identical conclusions, all of which are significant at the 90% confidence level:

(a) The precision of the Independent Medicare laboratories was significantly better than the precision of any other group except the Interstate Hospital group.
(b) The JCAH, AAFP-ASIM, Interstate and Medicare Hospital groups did not differ significantly in precision.

In summary, the precision of the different groups of laboratories can be exhibited as follows, where groups joined by the same line are not significantly different from each other, at the 90% confidence level.

Rank Order	Laboratory Group
Most Precise	Medicare Laboratories
	Interstate
	JCAH
	AAFP/ASIM
Least Precise	Medicare Hospitals

Table 3.5.9.2 displays the referee analytical methods suggested by a member of the Advisory Committee, the percentage of study participants who employed these methods, the techniques which the study participants appeared to have applied most successfully, and the percentage of the study participants who used these techniques. In examining this table, it is important to recall that these conclusions are based on performance in analyzing dialyzed serum samples and may not be directly transferable to human serum analyses.

Examination of Table 3.5.9.2 indicates that, for six of the eight constituents, at least one of the best applied techniques was based on the suggested referee method; very few

[23] J. Gibbons, Non-Parametric Statistical Inference, McGraw-Hill, 1971, pp. 256-257.

Table 3.5.9.2. Clinical Chemistry - Referee Methods and Best Applied Techniques

Constituent	Referee Method	% Using	Best Applied Techniques	% Using
Glucose	Hexokinase	1.5	Ferricyanide AutoAnalyzer	12.9
Urea Nitrogen	Diacetyl monoxime	63.2	Diacetyl monoxime Automated	34.0
Calcium	Atomic Absorption	4.0	Atomic Absorption Cresolphthalein Complexone Automated	4.0 33.4
Bilirubin	Diazo-Other Coupling (Jendrassik & Grof)	43.6	Diazo-Other Coupling (J & G) Automated	24.9
Cholesterol	Abell Kendall	0.6	$FeCl_3$-H_2SO_4 with Prior Extraction Manual or Automated	5.9
Uric Acid	Uricase	2.2	Uricase Phosphotungstate Automated	2.2 24.2
Sodium	Flame Photometer	96.7	Flame Photometer Automated	12.6
Total Protein	Biuret	78.4	Biuret Automated Refractometer	34.5 15.7

laboratories used the suggested referee methods in analyzing the other two constituents. For three of the constituents (urea nitrogen, calcium and total protein) more than 30% of the analyses were performed with the best applied techniques. Conversely, less than 15% of the cholesterol, glucose, and sodium analyses were performed with the techniques which were most effectively applied.

Assessment of the results reported in Subsections 3.5.1 through 3.5.8 reveals that automated techniques, as applied by the laboratories in this study, yielded equal or better accuracy and considerably superior precision than manual techniques employing the same analytical methods. In contrast, the various techniques commonly referred to as diagnostic kits were consistently the most imprecisely applied.

3.6. Medical Acceptability and Usefulness of Results

In Section 3.5, the precision and accuracy of the reported analyses were assessed. In this section, the relationship between the reported analyses and medical utility will be treated. For each constituent, discussion will focus on the acceptability of the group precision and the likelihood that determinations made by members of the group will be medically useful. The criterion chosen for measuring medical utility is that interlaboratory precision and accuracy should be within the limits necessary to support the monitoring over time of the variation in an individual's constituent concentrations. Such monitoring is important in therapeutic medicine as a means of checking on the progress of treatment and in preventive medicine as a means of early detection of medically alarming changes in a person's physical and functional condition. An intralaboratory monitoring capability is insufficient because people move from one area to another area, physicians sometimes switch laboratories or split their work between laboratories, and patients sometimes switch physicians.

3.6.1. Acceptability of Laboratory Group Precision

For each of the constituents except total bilirubin, Table 3.6.1.1 displays the acceptable precision (S_B) as defined in Subsection 3.1.3[24] and the sample size and standard deviation for:

the reference laboratory determinations,
each laboratory group, and
the best applied technique with the smallest standard deviation.

Total bilirubin has been omitted from this analysis because acceptable limits for this constituent were not available. Standard deviations which were not significantly greater than the acceptable precision values at the 99% confidence level have been starred in the table.

Examination of the table indicates the following:

(a) Precision of the reference laboratory analyses was within the medically acceptable range for five of the seven constituents—urea nitrogen, cholesterol, uric acid, sodium, and total protein.
(b) Precision of the cholesterol determinations was, in general, medically acceptable for all of the laboratory groups.
(c) The Interstate laboratory group achieved medically acceptable precision for urea nitrogen and sodium at the normal level, and total protein at the abnormal level.
(d) All other group precisions exceeded the medically acceptable precision values.
(e) Those laboratories using the best applied technique achieved acceptable precision for urea nitrogen, cholesterol, uric acid, sodium, and total protein. Thus, it is possible to state that a better selection of techniques would probably increase the acceptability of interlaboratory precision.

---

[24]Some of the acceptable precision values contained more significant digits than were desired for the analysis; these values were rounded upwards.

Table 3.6.1.1 Acceptability of Group Precision

Constituent		S_B	Deviations													
			Reference Labs		Doctors' Offices		Interstate		JCAH		Medicare Hospitals		Medicare Labs		Best Applied Technique	
			n	s	n	s	n	s	n	s	n	s	n	s	n	s
Glucose	- Normal	3.3	29	6.6	245	10.3	270	7.1	268	7.8	289	8.2	210	7.4	165	5.3
	Abnormal	12.2	30	19.7	248	31.5	272	30.9	269	25.7	291	28.8	210	24.5	165	19.4
Urea Nitrogen	- Normal	1.7	30	*2.2	217	3.3	275	*1.8	269	2.5	286	3.2	205	2.5	195	*0.9
	Abnormal	6.0	30	*2.05	213	11.7	274	7.2	270	8.3	285	11.7	204	9.4	195	*2.4
Calcium	- Normal	0.2	30	0.6	54	1.3	279	0.8	272	1.3	208	1.7	153	1.0	228	0.3
	Abnormal	0.2	30	0.6	53	1.3	278	0.8	270	1.3	208	1.7	154	1.0	228	0.4
Cholesterol	- Normal	15.	29	*14.0	214	*16.1	286	*12.4	300	*15.5	273	18.8	194	*13.3	41	*14.2
	Abnormal	15.	29	*12.6	212	16.8	287	*14.0	294	*15.2	273	*14.6	194	*12.1	40	*8.3
Uric Acid	- Normal	0.4	31	*0.3	300	1.2	303	0.6	317	0.7	292	1.0	215	0.7	223	*0.3
	Abnormal	0.7	30	*0.4	295	1.5	304	1.0	316	1.1	294	1.4	215	1.2	223	*0.5
Sodium	- Normal	2.4	31	*1.8	30	3.6	230	*2.6	281	3.1	217	3.6	121	3.0	56	*2.0
	Abnormal	2.3	31	*1.7	29	6.8	229	4.2	282	4.9	217	8.1	120	3.7	55	*2.8
Total Protein	- Normal	0.2	31	*0.2	93	0.4	299	0.3	309	0.3	253	0.4	172	0.3	398	*0.2
	Abnormal	0.2	31	*0.2	95	0.4	297	*0.2	308	0.3	253	0.4	171	0.3	399	*0.2

Starred values are not significantly greater than S_B at the 99% confidence level.

The conclusions just cited are open to question on two grounds. First, they are based on precision values calculated about the group means. It would be possible to argue that the precision values were inflated by the procedure of pooling over techniques. It is important to investigate whether the conclusions would have changed if techniques had been differentiated in the precision calculations. Secondly, the conclusions are based on performance in analyzing dialyzed specimens. It would be desirable to compare the precision values in this study with values from a survey which employed serum pool specimens. These further assessments should yield insight into the validity and transferability of the conclusions.

Table 3.6.1.2 displays data which can be used to make these further assessments. These data come from two sources, the present study (labeled NBS in the table) and a document entitled The 1971 Chemistry Survey Program—An Analysis of the Data by Dr. Roger Gilbert which was published by the College of American Pathologists and is based on their 1971 survey data (labeled CAP in the table). It should be noted that the CAP program is the only proficiency program which the CDC certifies as equivalent to its own. The CAP report gives coefficients of variation which, when multiplied by the average of the normal and abnormal target values used in the NBS study, yield projected approximations of the precision CAP participants would have achieved in analyzing the NBS test samples. This procedure is of questionable validity in the case of cholesterol where CAP supplied abnormally high specimens while the NBS abnormal specimen was abnormally low. The reported NBS data are an average of the normal and abnormal precision values.

The first two columns in Table 3.6.1.2 show the constituents and the medically acceptable precision (S_B) which should be used when the normal and abnormal precisions are averaged. The third column, headed NBS All-Lab Pooled Technique Precision, displays, for each constituent, the standard deviation of all reported non-outlier analyses. The fourth column, headed NBS Median Technique Precision, displays the precision of that technique which had the median technique precision (as defined in Section 3.3) among technique precisions calculated for the constituent. Comparison of these data should permit assessment of whether the conclusions on medical acceptability were biased by using data pooled over techniques. Since these two sets of data are roughly consistent in magnitude, it is unlikely that pooling affected the conclusions.

The last six columns in Table 3.6.1.2 provide the data needed for intersurvey comparisons. The three sets of precisions selected for comparison were:

(a) The NBS Interstate and CAP data on the standard deviations of all reported non-outlier analyses (Pooled Technique Precision).
(b) The NBS Interstate and CAP data on the median standard deviations among the deviations of the different techniques used in analyzing a constituent (Median Technique Precision).
(c) The NBS All-Lab and CAP data on the standard deviations of those techniques which were applied with the least variability by study participants (Minimum Variance Technique Precision).

For the most part, the data prove to be comparable. Thus, the precision achieved by an analysis of dialyzed specimens, as reported in this study, is not atypical of precision achieved in analyzing serum pool specimens, and the conclusions on medical acceptability are probably robust ones.

3.6.2. Likelihood of a Medically Useful Laboratory Analysis

To estimate the probability that a correct laboratory assay of one of the seven clinical chemistry constituents (total bilirubin is again excluded) could be acceptably reproduced by laboratories in a given group, a multi-step procedure was utilized. First, a 25% random sample of the shipment 1 assay reports (including outliers) returned by each group was selected, and separate plots of the seven constituents were prepared. In each case, normal determinations reported by the group were plotted on the horizontal axis and abnormal determinations were plotted on the vertical axis. The medically acceptable region was then plotted about the target value. (For calcium, the region about the MRART was also plotted.) The percentage of points lying in the acceptable region gives the desired estimate

Table 3.6.1.2. Acceptability of Technique Precision and of Precision in the 1971 CAP

Constituent	Average S_B	NBS All-Lab Pooled Technique Precision	NBS All-Lab Median Technique Precision	NBS Interstate Pooled Technique Precision	CAP Pooled Technique Precision	NBS Interstate Median Technique Precision	CAP Median Technique Precision	NBS Minimum Variance Technique Precision	CAP Minimum Variance Technique Precision
Glucose	7.8	17.3	17.1	19.0	13.3	15.5	13.6	12.0	9.7
Urea Nitrogen	3.8	6.3	6.6	4.5	5.1	6.6	4.7	*1.65	*2.8
Calcium	0.2	1.4	1.2	0.8	0.6	2.1	0.6	0.35	0.3
Cholesterol	15.	*15.1	*14.4	*13.2	*8.7	*12.1	*8.3	*9.5	*6.5
Uric Acid	0.6	1.1	1.1	0.8	0.85	1.4	0.8	*0.4	*0.5
Sodium	2.4	4.4	3.3	3.4	2.6	3.1	*2.3	*2.4	*2.3
Total Protein	0.2	0.3	*0.2	0.25	*0.2	*0.2	*0.2	*0.2	*0.15

Starred values are not significantly greater than S_B at the 99% confidence level.

of the percent of medically useful results. The plots are displayed in Appendix D. Table 3.6.2.1 summarizes the estimates. Reference laboratory capability assessment was based on all reported analyses.

Examination of Table 3.6.2.1 reveals that more than half of the reference laboratory determinations for all constituents except glucose and calcium were within medically acceptable precision limits from the target values. In only four instances were more than half of the analyses of a constituent by a laboratory group medically useful—the Interstate and Medicare Independent uric acid analyses and the Interstate and JCAH urea nitrogen analyses. Overall, the average likelihood of a correct assay being reproduced within medically acceptable limits was 28%. Such a low likelihood of a medically acceptable determination is particularly noteworthy in view of the frequency with which these laboratory tests are performed and the cost of their administration.

Further examination of the plots in Appendix D reveals that the medically unacceptable values tend to cluster about a 45° line. Youden has demonstrated that such clustering results from systematic errors.[25] Known causes of systematic error include miscalibration of equipment, inadequate calibration specimens, incorrect application of procedures, and inherent bias in procedures.

3.7. Target Value Surrogates

Exact determination of the composition of pooled serum test specimens is impossible within the current state-of-the-art of clinical chemistry. Consequently, manufacturer's average assays may be used as the target values for pooled serum calibration and proficiency test specimens. In this proficiency test, the availability of actual weigh-back values makes it possible to assess whether the manufacturer's average assay is a good choice for the target value or whether another choice might be preferable in proficiency testing situations. To address these questions three possible target surrogates were considered in addition to the manufacturer's average assay. These were the mean reference assays, the mean laboratory assays, and the mean assay of laboratories using the technique that the manufacturer used. Table 3.7.1 contains all of the data used in this analysis. An asterisk next to the mean value indicates that this value differs significantly from the weigh-back value at the 95% confidence level.

3.7.1. Adequacy of the Manufacturer's Average Assay as a Target Value

In all cases, the manufacturer's determinations were performed with either the recommended referee method or the best applied technique identified in Section 3.5.9. A series of t-tests run on the manufacturer's average assays indicates that 13 of 14 differ significantly from the weigh-back values. However, an examination of the percentage difference of the manufacturer's average assays from the weigh-back values reveals that 10 of the 14 assays differ from the weigh-back values by less than 3%. The statistical significance of the differences between these assays and the weigh-back values may have resulted from the use of the precision values of the manufacturer's assays in performing the t-tests. These values were used because the precision values of the techniques were not available. Precision was extremely high for the manufacturer's assays. This may be partially attributed to the fact that all 25 replicate determinations were run on the same day on the same piece of equipment by the same technician. Thus, the precision of the manufacturer's assays is largely a measure of the vial-to-vial variation, and the significance of the t-tests for the manufacturer's assays may be artificial.

Examination of the reference laboratory data facilitates an evaluation of whether the apparent inaccuracy of the manufacturer's assays results from unavailability of data on the precision of the manufacturer's technique. Reference laboratory sample sizes and average

---

[25]W. J. Youden, Statistical Techniques for Collaborative Tests, Association of Official Analytical Chemists, 1967.

Table 3.6.2.1 Probability that a Correct Laboratory Assay Is Acceptably Reproduced

Constituent	Reference Labs	Doctors' Offices	Interstate	JCAH	Medicare Hospitals	Medicare Labs	All Groups
Glucose	.35	.14	.26	.28	.15	.30	.21
Urea Nitrogen	.76	.13	.64	.50	.29	.32	.36
Calcium *	.35	.04	.20	.14	.10	.13	.12
Cholesterol	.63	.19	.32	.09	.18	.28	.19
Uric Acid	.81	.35	.59	.39	.26	.56	.43
Sodium	.82	.35	.37	.37	.23	.33	.32
Total Protein	.59	.19	.32	.35	.29	.39	.31
Average	.62	.20	.39	.30	.21	.33	.28

* Probabilities based on MRART range, Probabilities based on the target range were lower.

Table 3.7.1. Analysis of Potential Target Surrogates

Test		Manufacturer's Technique	Manufacturer's Average Assay				All Determinations—All Techniques				All Determinations—Same Technique				Reference Assays				Target
			n	$\bar{x}$	s	$\lvert\%\bar{d}\rvert$	n	$\bar{x}$	s	$\lvert\%\bar{d}\rvert$	n	$\bar{x}$	s	$\lvert\%\bar{d}\rvert$	n	$\bar{x}$	s	$\lvert\%\bar{d}\rvert$	
Glucose	- Normal	Ferricyanide	25	*80.4	0.87	2.2	1282	*83.10	8.42	1.1	165	81.40	5.30	1.0	29	83.45	6.64	1.5	82.2
	Abnormal	(Auto. Syst. 1)	25	*294.6	7.6	3.1	1282	*296.5	26.2	2.5	164	*300.2	19.4	1.3	30	300.5	19.73	1.2	304
Urea Nitrogen	- Normal	Diacetyl Monoxime	25	*18.7	0.21	0.5	1252	*18.01	2.72	3.2	189	*19.41	1.34	4.4	30	18.61	2.21	0.1	18.6
	Abnormal	(Auto. Syst. 1)	25	65.36	0.85	0.1	1246	*61.30	9.98	6.3	188	*66.65	3.93	1.9	30	65.73	2.06	0.5	65.4
Calcium	- Normal	Atomic Absorp.	25	*9.18	0.013	0.2	966	*8.65	1.24	6.0	39	*8.99	.26	2.3	30	*8.71	0.62	5.3	9.2
	Abnormal	Spectroscopy	25	*11.9	0.14	2.5	963	*10.94	1.59	10.3	39	*11.81	.47	3.2	30	*11.46	0.56	6.1	12.2
Bilirubin	- Normal	Diazo-Other	25	*0.98	0.009	8.9	1158	*1.15	0.31	27.8	42	*1.30	0.29	44.4	30	*1.02	0.18	13.3	0.9
	Abnormal	(Auto. Syst. 1)	25	*4.5	0.027	4.3	1154	*4.90	0.86	4.3	41	*4.88	0.39	3.8	30	4.51	0.56	4.0	4.7
Cholesterol	- Normal	Abell-Kendall	25	159.9	8.0	-	1267	147.8	15.5	-	8	133.8	13.6	-	29	134.2	14.0	-	-
	Abnormal		25	70	3.7	-	1260	94.32	14.7	-	8	73.62	11.7	-	29	74.13	12.6	-	-
Uric Acid	- Normal	Phosphotungstate	25	*6.31	0.07	1.8	1427	6.25	0.88	0.8	122	6.28	0.37	1.3	31	6.2	.32	0.0	6.2
	Abnormal	(Auto. Syst. 1)	25	*12.08	0.23	9.8	1424	*10.59	1.27	3.7	123	*11.30	0.64	2.7	30	11.0	.37	0.0	11.0
Sodium	- Normal	Flame Photometer	25	*140.5	0.08	0.4	879	*140.1	3.13	0.6	739	*140.1	3.16	0.6	31	140.9	1.77	0.1	141
	Abnormal	(Manual)	25	*116.3	0.47	0.6	874	*118.1	5.76	0.9	735	*118.3	5.07	1.1	31	116.3	1.75	0.6	117
Total Protein	- Normal	Biuret	25	*6.77	0.053	0.4	1126	*6.85	0.33	0.7	167	*6.91	0.25	1.6	31	6.88	0.20	1.2	6.8
	Abnormal	(Auto. Syst. 1)	25	*3.27	0.04	0.9	1124	*3.47	0.33	5.2	166	*3.54	0.18	6.8	31	*3.43	0.19	3.9	3.3
Average % Difference						2.6				5.2				5.5				2.7	

Star indicates significant difference from target value.

percent differences from the weigh-back values (Table 3.7.1) are comparable to those in the manufacturer's determinations, but the deviations are much larger because of interlaboratory variability. None of the nine reference assays which differ from the weigh-back values by less than 3% are statistically different from the weigh-back values. In contrast, 9 of the 10 manufacturer's assays which differed from the weigh-back values by less than 3% are statistically different from the weigh-back values. Thus, it would appear that the accuracy of manufacturer's average assays may, in general, be acceptable, but these assays are performed in a manner which does not protect against occasional human and calibration errors.

3.7.2. Other Possible Targets for Proficiency Testing

The first surrogate considered as a replacement for the manufacturer's mean assay, the mean value of all reported determinations, differed significantly from the weigh-back value in 13 of 14 instances and differed from the weigh-back value by more than 3% in 8 of 14 instances. This level of performance is far below that achieved by the manufacturer, and this possible target was rejected. The mean value of all determinations performed by the technique employed by the manufacturer was the second target considered. This measure differed significantly from the weigh-back value in 12 of 14 cases and differed from the weigh-back value by more than 3% in 5 of 14 cases. This level of performance is similar to the level of performance achieved by the manufacturer. However, the average percentage difference of manufacturer's means from the weigh-back values is only 2.6% while this group's average is 5.2%. Therefore, this possible target was also rejected.

The final target considered was the mean reference assay. Only 4 of the 14 reference assays differed significantly from the weigh-back values. Five of the 14 assays were more than 3% from the weigh-back values and the average percent difference was 2.7%. Thus, the accuracy of the mean reference assay is comparable to the accuracy of the manufacturer's average assay. Furthermore, the mean reference assay is determined from heterogeneous data sources and forms a more reliable base than the replicate analyses of a single laboratory. When available, mean reference assays (MRAs) appear to be a better choice of target values than manufacturer's average assays. This finding lends credence to the choice of MRAs as the target values for cholesterol and hematology in this study.

3.8. Comparison with Other Proficiency Surveys

The coefficients of variation (CV) provide a basis for comparing the proficiency with which individual laboratory methods were applied. Unfortunately, although information identified by method is regularly reported by some groups administering open proficiency tests, many studies that are one-time assessments of laboratory capability report only CV's grouped over methods. In order to make comparisons with these surveys, the CV's from this study were combined over all groups, methods, and levels (normal and abnormal).[26] Although this procedure increases the number of surveys available for comparison, the concomitant masking of information decreases the reliability of the comparisons.

Table 3.8.1 displays the CV's observed in this study and 11 other studies. Bibliographic references for these studies are given in Appendix E.

The relative variability observed in 4 of the 12 other studies, the initial (1949) CAP survey, a Canadian study in 1963, an assessment of Italian laboratory proficiency (1968), and, surprisingly, the initial CDC survey in 1968, was greater than in this study. Variability displayed in the other seven surveys seems to be consistent with the relative variability in this study.

---

[26]Although they are not shown, the ungrouped CV's can be calculated from the data in Section 3.5.

Table 3.8.1. Comparison of Coefficients of Variation (%) for Clinical Chemistry

Constituent	Glucose	BUN	Calcium	Bilirubin	Cholesterol	Uric Acid	Sodium	Total Protein
NBS Study - Dialyzed	9.0%	15.0%	13.8%	27.0%	10.5%	14.5%	2.2%	4.4%
NBS Study - Pooled Serum	8.2	11.6	8.2	39.0	7.9	9.0	2.8	5.8
CDC - Initial Analyses - 1968	25.8	32.4	15.9		14.0	45.9	4.3	7.4
CDC - After 21 Mos. in Pgm. - 1970	5.5	7.1	6.1		12.4	8.2	2.9	5.8
CAP - 1949 (Aqueous Samples)	16.2	63.5	27.6		23.7			
CAP - 1969	8.0	13.3	11.1		18.5			
CAP - 1972	7.2	13.4	6.3	13.5	8.9	9.8	2.4	4.0
AAB - 3 Quarters of 1970 & 71	7.5	13.9	7.2	20.2	8.2	11.2	3.5	4.7
Canada - 1963	10.4	27.3			20.1		4.3	8.9
Scandinavia - 1969	11.3	13.6	7.0		13.4		3.1	
New Zealand - 1964	7.1	17.4	6.0		13.2	15.9	1.9	
Australia - 1964	17.3	18.7	6.1		15.6		2.8	
LA Physicians	9.9	19.1			11.0	16.0		3.6
Italy - 1968	16.8	40.4			26.2			12.0

4. HEMATOLOGY

The topics addressed in analyzing hematology include:

(a) Differences in precision and relative accuracy among laboratory groups;
(b) Differences in precision and relative accuracy among laboratory techniques;
(c) The performance of laboratories in this study in comparison to the performance of laboratories in other published proficiency tests.

The analysis follows the same plan which was outlined for clinical chemistry in Section 3.3, and the discussion employs the definitions presented in Section 3.1. The mean reference assays were used as the target values in the hematology analysis.

Because the two shipments of hematology specimens differed in composition, it was necessary to analyze each separately. The rankings of precision in these two separate analyses were virtually identical. Therefore, it was deemed sufficient to present the data on precision for the first shipment only. There was no significant difference in relative accuracy among the laboratory groups; consequently, it was possible to pool data over groups for the analysis of relative accuracy of the techniques.

4.1. Precision and Relative Accuracy

The table formats used in the previous chapter will be adopted here. The ranking of group performance again appears at the bottom of the first table in each set.

4.1.1. Red Cell Count

The red cell count data for shipment 1, stratified by group and laboratory technique, appear in Table 4.1.1.1. Precision of techniques is ranked:

1	Coulter Model S
2	Coulter Model A, B, D2
2	Kits, Coulter F
3	Hemocytometer.

Examination of Table 4.1.1.2 reveals that only the Fischer Autocytometer technique was applied in a statistically accurate manner for all four determinations. However, the average percentage difference from the target value did not exceed 3% for any of the techniques, and only the Hemocytometer technique had a percentage difference which exceeded 5%. Although the sample size was insufficient to make operationally significant statements about the precision of the Fischer Autocytometer, it appears that this technique was applied with a precision comparable to that achieved with the Hemocytometer.

Based on criteria of precision and accuracy, it would appear that Kits and the Coulter Models A, B, D2, F, and S were successfully applied techniques for red cell counting.

4.1.2. White Cell Count

Analysis of the white cell count data displayed in Table 4.1.2.1 reveals that the precision of techniques is ranked:

1	Coulter Model S
2	Coulter Model A, B, D2, F
2	Kits
3	Hemocytometer.

Table 4.1.2.2 indicates that statistically accurate performance was achieved most frequently with Kits and the Coulter Model S. Six techniques, Kits and the Coulter Models A, B, D2, F and S, were applied with average inaccuracies of less than 3%; and appear to be the most successfully applied techniques for white cell counts.

Table 4.1.1.1. Red Cell Count Normal Specimens

Technique	Doctors' Offices			Interstate			JCAH			Medicare Hospitals			Medicare Labs			All Groups		
	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
Hemocytometer	121	4.26	.346	26	4.33	.285	59	4.30	.322	112	4.32	.375	66	4.36	.352	384	4.31	.349
Coulter Model A, B, D2				18	4.16	.246	29	4.18	.236	17	4.25	.165	15	4.01	.353	85	4.15	.260
Coulter Model S				69	4.23	.115	42	4.25	.118	19	4.22	.183				149	4.24	.129
Kits, Coulter F	16	4.13	.401	25	4.17	.265	35	4.03	.213	20	4.17	.234				105	4.12	.263
All Methods	150	4.24	.346	164	4.24	.215	180	4.20	.266	176	4.28	.335	112	4.26	.333	782	4.24	.301
							Red Cell Count Abnormal Specimens											
Hemocytometer	119	1.84	.227	25	1.85	.184	60	1.80	.206	114	1.82	.233	66	1.88	.193	384	1.84	.218
Coulter Model A, B, D2				18	1.76	.181	29	1.79	.174	16	1.82	.146	15	1.75	.187	84	1.78	.167
Coulter Model S				68	1.84	.083	46	1.83	.090	17	1.80	.086	15	1.81	.055	151	1.83	.088
Kits, Coulter F	15	1.68	.252	25	1.78	.177	35	1.76	.177	21	1.77	.195				105	1.76	.191
All Methods	147	1.82	.229	162	1.82	.149	186	1.79	.170	176	1.81	.210	114	1.84	.175	785	1.81	.189
Ranking of Group Precision		2			1			1			1			1				

Table 4.1.1.2. Red Cell Count Normal

Technique	Shipment 1 (T = 4.18) n	$\bar{x}$	$\%\bar{d}$	s	Shipment 2 (T = 4.27) n	$\bar{x}$	$\%\bar{d}$	s
Hemocytometer	384	*4.31	3.1	.349	345	*4.33	1.4	.413
Coulter Model A, B. D2	85	4.15	-0.7	.260	79	*4.19	-1.9	.282
Coulter Model S	149	*4.24	1.4	.129	146	*4.30	0.7	.117
Fisher Autocytometer	20	4.16	-0.5	.292	16	4.28	0.2	.373
Kits, Coulter F	105	*4.12	-1.4	.263	105	*4.19	-1.9	.290
All Methods	782	$\bar{\bar{x}}$ = 4.24		.301	711	$\bar{\bar{x}}$ = 4.28		.340

Red Cell Count Abnormal

	Shipment 1 (T = 1.75)				Shipment 2 (T = 2.14)			
Hemocytometer	384	*1.84	5.1	.218	340	2.15	0.5	.271
Coulter Model A, B, D2	84	*1.79	2.3	.167	75	2.10	-1.9	.187
Coulter Model S	151	*1.83	4.6	.088	145	*2.19	2.3	.096
Fisher Autocytometer	21	1.74	-0.6	.188	16	2.21	3.3	.356
Kits, Coulter F	105	1.76	0.6	.191	104	*2.06	-3.7	.190
All Methods	785	$\bar{\bar{x}}$ = 1.81		.189	700	$\bar{\bar{x}}$ = 2.15		.231

* Significantly different from the mean reference assay at the 95% confidence level.

Table 4.1.2.1. White Cell Count Normal

Technique	Doctors' Offices			Interstate			JCAH			Medicare Hospitals			Medicare Labs			All Groups		
	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
Hemocytometer	175	6.8	.95	24	7.1	.98	55	7.2	.84	120	6.9	1.02	70	7.0	.95	444	6.9	.96
Coulter Model A, B, D2, F				32	7.3	.63	55	7.2	.52	24	7.3	.74	18	7.5	.72	137	7.3	.61
Coulter Model S				70	7.5	.56	42	7.3	.67	19	7.4	.64				149	7.4	.62
Kits	21	7.3	.71	23	7.5	.59	29	7.2	.50	20	7.1	.84				106	7.3	.66
All Methods	208	6.9	.94	168	7.4	.68	184	7.2	.66	184	7.0	.94	118	7.0	.86	862	7.1	.85
							White Cell Count Abnormal											
Hemocytometer	168	20.4	2.58	24	21.2	1.91	55	21.1	2.56	118	20.8	2.67	70	20.7	2.14	435	20.7	2.51
Coulter Model A, B, D2, F				31	19.5	1.30	56	19.9	1.18	23	19.7	1.01	18	19.8	.73	136	19.8	1.12
Coulter Model S				69	19.4	1.13	45	18.9	1.43	18	19.5	.96	15	19.0	.83	151	19.2	1.22
Kits	17	18.9	3.54	23	20.4	1.75	30	19.8	.95	20	19.5	2.12				104	19.7	2.02
All Methods	197	20.2	2.63	165	19.7	1.82	189	20.0	1.89	180	20.4	2.39	120	20.2	1.83	851	20.1	2.18
Ranking of Group Precision		3			3			2			3			1				

Table 4.1.2.2. White Cell Count Normal

Technique	Shipment 1 (T = 7.4) n	$\bar{x}$	$\%\bar{d}$	s	Shipment 2 (T = 8.3) n	$\bar{x}$	$\%\bar{d}$	s
Hemocytometer	444	*6.9	-6.8	.96	400	*7.4	-10.8	1.21
Coulter Model A, B, D2, F	137	7.3	-1.3	.61	119	*8.0	-3.6	.61
Coulter Model S	149	7.4	0.0	.62	148	8.2	-1.2	.68
AutoAnalyzer Model 4 or 7	16	*7.0	-5.4	.57	17	8.1	-2.4	1.99
Kits	106	7.3	-1.3	.66	91	*7.9	-4.8	.76
All Methods	862	$\bar{\bar{x}}$ = 7.1		.85	784	$\bar{\bar{x}}$ = 7.7		1.08

White Cell Count Abnormal

	Shipment 1 (T = 19.4)				Shipment 2 (T = 19.0)			
Hemocytometer	435	*20.7	6.7	2.51	397	*19.5	2.6	2.79
Coulter Model A, B, D2, F	136	*19.8	2.1	1.12	117	*19.3	1.6	1.25
Coulter Model S	151	19.2	-1.0	1.22	144	*18.8	-1.1	1.13
AutoAnalyzer Model 4 or 7	15	*17.9	-7.7	2.87	17	*17.4	-8.4	2.97
Kits	104	19.7	1.5	2.02	89	*19.5	2.6	2.16
All Methods	851	$\bar{\bar{x}}$ = 20.1		2.18	773	$\bar{\bar{x}}$ = 19.3		2.31

* Significantly different from the mean reference assay at the 95% confidence level.

4.1.3. Hematocrit

Based on the data displayed in Table 4.1.3.1, the precision of the hematocrit techniques was ranked:

1 Microhematocrit
1 Coulter Model S.

Examination of Table 4.1.3.2 shows that the abnormal determinations using Kits and one of the Coulter Model S abnormal determinations were the only statistically accurate determinations. It appears that use of a single normal target may have affected the results in this instance. The average inaccuracies of all techniques exceeded 3%, but only the Coulter Model S had an average inaccuracy of more than 4%.

Operationally significant statements about the relative precision of Kits cannot be made based on the small number of determinations reported in this study. However, these techniques appear to be more imprecise than the others used in hematocrit determinations. Among techniques applied by the study participants, the Microhematocrit technique appears to have been the most satisfactory.

4.1.4. Hemoglobin

Table 4.1.4.1 summarizes the hemoglobin analyses stratified by group and technique. Precision of the techniques is ranked:

1 Coulter Model S
2 Cyanmethemoglobin (Source Group A)
2 Cyanmethemoglobin (Source Group B).

Table 4.1.4.2 contains the hemoglobin data pooled over groups. The Hemoglobinometer (IL) was statistically accurate in all four determinations. None of the technique means was inaccurate by more than 5%, and the mean technique inaccuracy never exceeded 3%. The techniques where sample sizes were insufficient for reliable evaluation of relative precision did not appear to be less precise than the Cyanmethemoglobin techniques or the Kits.

It would appear that all of the hemoglobin techniques were equally well applied by the study participants.

4.1.5. Mean Corpuscular Volume

The mean corpuscular volume data is displayed in Table 4.1.5.1. Precision of techniques is ranked:

1 Coulter Model S
2 Best Chart
2 Hand Calculations.

Two of the Hand Calculation means and one Best Chart mean were statistically accurate. None of the Best Chart or Hand Calculation means were more than 5% inaccurate, and the average inaccuracy with these techniques was less than 3%. However, it should be noted that obviously deviant values were removed from these determinations at eight times the rate they were removed from the Coulter S determinations. As a result, it is virtually impossible to judge which techniques were applied best.

4.1.6. Summary

Relative accuracy did not differ significantly among the laboratory groups. However, application of Friedman's Two-Way Analysis of Variance by Ranks to the ranked precision data displayed in Table 4.1.6.1 showed that the precision of the groups differs significantly at the 99% confidence level.

Table 4.1.3.1. Hematocrit Normal

Technique	Doctors' Offices			Interstate			JCAH			Medicare Hospitals			Medicare Labs			All Groups		
	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
Microhematocrit	151	41.1	1.52	88	40.4	1.35	133	40.6	1.09	163	40.8	1.21	101	40.5	1.29	636	40.7	1.32
Coulter Model S				65	36.8	1.16	42	37.0	1.47	17	37.1	1.06				141	36.9	1.26
All Methods	180	41.2	2.19	171	39.1	2.22	186	39.6	2.04	184	40.4	1.73	117	40.0	1.84	838	40.1	2.15
							Hematocrit Abnormal											
Microhematocrit	150	18.6	1.17	87	18.3	1.08	132	18.0	.93	161	18.3	.86	104	18.2	.93	634	18.3	1.01
Coulter Model S				63	18.2	.73	45	17.9	1.02	15	17.8	.82				141	18.0	.87
All Methods	179	18.7	1.21	167	18.3	1.13	188	18.0	1.06	180	18.2	.86	121	18.2	1.00	835	18.3	1.09
Ranking of Group Precision		3			2			2			1			2				

Table 4.1.3.2. Hematocrit Normal

Technique	Shipment 1 (T = 39.0) n	$\bar{x}$	$\%\bar{d}$	s	Shipment 2 (T = 39.4) n	$\bar{x}$	$\%\bar{d}$	s
Microhematocrit	636	*40.7	4.4	1.32	561	*41.4	5.1	1.22
Coulter Model S	141	*36.9	-7.9	1.26	142	*37.4	-5.1	1.55
Kits	47	*40.7	4.4	4.00	45	*41.6	5.6	4.37
All Methods	838	$\bar{\bar{x}}$ = 40.1		2.15	761	40.7		2.28

Hematocrit Abnormal

	Shipment 1 (T = 18.0)				Shipment 2 (T = 20.1)			
Microhematocrit	634	*18.3	1.7	1.02	551	*19.9	-1.0	1.15
Coulter Model S	141	18.0	0.0	.87	138	*19.4	-3.5	1.08
Kits	46	18.4	2.2	1.67	46	20.1	0.0	1.99
All Methods	835	$\bar{\bar{x}}$ = 18.3		1.09	748	$\bar{\bar{x}}$ = 19.9		1.28

* Significantly different from the mean reference assay at the 95% confidence level.

Table 4.1.4.1. Hemoglobin Normal

Technique	Doctors' Offices			Interstate			JCAH			Medicare Hospitals			Medicare Labs			All Groups		
	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
Cyanmethemo. (Source A)	84	12.9	.51	42	12.8	.36	81	12.7	.45	103	12.9	.47	72	12.8	.37	382	12.8	.45
Cyanmethemo. (Source B)	45	12.9	.61	16	12.6	.57	21	12.8	.37	32	12.9	.59	18	12.6	.43	132	12.8	.55
Coulter Model S				69	12.7	.31	42	12.7	.35	18	12.7	.43				146	12.7	.34
Kits	47	12.7	.58	17	12.9	.35	27	12.6	.54	16	12.7	.56				117	12.7	.53
All Methods	204	12.8	.58	168	12.7	.38	185	12.7	.43	186	12.9	.51	117	12.7	.40	860	12.8	.48
							Hemoglobin Abnormal											
Cyanmethemo. (Source A)	81	6.1	.45	42	6.0	.47	82	5.9	.33	104	6.0	.43	74	5.9	.34	383	6.0	.41
Cyanmethemo. (Source B)	43	6.0	.42	16	6.0	.53	21	6.0	.24	32	6.0	.67	18	5.8	.36	130	6.0	.48
Coulter Model S				68	6.2	.29	45	6.2	.29	17	6.1	.29				148	6.1	.30
Kits	46	5.9	.49	16	6.1	.32	27	6.0	.38	16	5.9	.58				115	6.0	.45
All Methods	198	6.0	.48	166	6.1	.38	188	6.0	.34	186	6.0	.48	120	5.9	.34	858	6.0	.42
Ranking of Group Precision		3			2			2			3			1				

Table 4.1.4.2. Hemoglobin Normal

Technique	Shipment 1 (T = 12.6) n	$\bar{x}$	$\%\bar{d}$	s	Shipment 2 (T = 12.9) n	$\bar{x}$	$\%\bar{d}$	s
Cyanmethemoglobin	382	*12.8	1.6	.45	343	12.9	0.0	.48
Cyanmethemoglobin/other	132	*12.8	1.6	.55	110	12.9	0.0	.55
Oxyhemoglobin	30	12.6	0.0	.68	32	*12.6	-2.3	.66
Hemoglobinometer (IL)	30	12.7	0.8	.58	35	12.9	0.0	.54
SMA 4 or 7	16	*12.8	1.6	.25	17	12.9	0.0	.31
Coulter Model S	146	*12.7	0.8	.34	142	12.8	-0.8	.40
Kits	117	*12.7	0.8	.53	94	12.9	0.0	.66
All Methods	860	$\bar{\bar{x}}$ = 12.8		.48	779	$\bar{\bar{x}}$ = 12.9		.51

Hemoglobin Abnormal

	Shipment 1 (T = 6.0)				Shipment 2 (T = 6.5)			
Cyanmethemoglobin	383	6.0	0.0	.41	338	*6.4	-1.5	.40
Cyanmethemoglobin/other	130	6.0	0.0	.48	110	*6.2	-4.6	.49
Oxyhemoglobin	31	*5.8	-3.3	.45	32	*6.3	-3.1	.53
Hemoglobinometer (IL)	28	5.9	-1.7	.44	34	6.4	-1.5	.40
SMA 4 or 7	16	6.1	1.7	.31	17	6.4	-1.5	.32
Coulter Model S	148	*6.2	3.3	.29	137	6.5	0.0	.25
Kits	115	6.0	0.0	.45	91	*6.3	-3.1	.47
All Methods	858	$\bar{\bar{x}}$ = 6.0		.42	765	$\bar{\bar{x}}$ = 6.4		.42

* Significantly different from the near reference assay at the 95% confidence level.

Table 4.1.5.1. Mean Corpuscular Volume Normal

Technique	Doctors' Offices			Interstate			JCAH			Medicare Hospitals			Medicare Labs			All Groups		
	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
Hand Calculation	85	97.	8.1	70	96.	6.1	104	96.	7.3	114	95.	8.6	69	95.	8.9	442	96.	7.9
Test Chart	22	96.	9.5				29	96.	8.6	32	93	7.3	20	96.	6.5	114	96.	8.3
Coulter Model S				69	87.	2.0	46	87.	2.5	17	87.	2.2				153	87.	2.3
All Methods	116	96.	8.4	160	92.	7.0	180	94.	7.7	165	94.	8.3	106	94.	8.5	727	94.	8.0
							Mean Corpuscular Volume Abnormal											
Hand Calculation	83	103.	13.0	69	101.	12.5	107	101.	11.4	112	101.	13.1	70	99.	12.4	441	101.	12.5
Test Chart	20	104.	12.4				29	101.	9.2	32	101.	11.1	19	102.	8.7	111	102.	10.6
Coulter Model S				67	99.	5.3	49	97.	4.0	17	97.	2.9	15	97.	3.5	155	98.	5.6
All Methods	112	103.	12.4	157	100.	9.9	186	100.	9.8	163	100.	12.0	107	100.	10.9	725	101.	10.9
Ranking of Group Precision	2			1			2			2			2					

Table 4.1.5.2. Mean Corpuscular Volume Normal

Technique	Shipment 1 (T = 93) n	$\bar{x}$	$\%\bar{d}$	s	Shipment 2 (T = 93) n	$\bar{x}$	$\%\bar{d}$	s
Hand Calculation	442	*96.	3.2	7.9	363	*97.	4.3	8.4
Best Chart	114	*96.	3.2	8.3	116	*97.	4.3	9.0
Coulter Model S	153	*87.	-6.5	2.3	150	*86.	-8.1	2.7
All Methods	727	$\bar{\bar{x}}$ = 94.		8.0	648	$\bar{\bar{x}}$ = 94.		8.8

Mean Corpuscular Volume Abnormal

	Shipment 1 (T = 101)				Shipment 2 (T = 93)			
Hand Calculation	441	101.	0.0	12.5	359	94.	1.1	10.8
Best Chart	111	102.	1.0	10.6	115	*95.	2.2	9.2
Coulter Model S	155	*98.	-3.0	3.6	145	*88.	-5.4	4.2
All Methods	725	$\bar{\bar{x}}$ = 101.		10.9	638	$\bar{\bar{x}}$ = 93.		9.9

* Significantly different from the mean reference assay at the 95% confidence level.

Table 4.1.6.1. Ranking of Precision of Groups by Constituent and Shipment

Constituent	Shipment	Doctors' Offices	Interstate	JCAH	Medicare Hospitals	Medicare Labs.
Red Cell Count	1	5	2.5	2.5	2.5	2.5
	2	5	2.5	2.5	2.5	2.5
White Cell Count	1	4	4	2	4	1
	2	5	2	3.5	3.5	1
Hematocrit	1	5	3	3	1	3
	2	5	3.5	1.5	1.5	3.5
Hemoglobin	1	4.5	2.5	2.5	4.5	1
	2	4.5	3	1	4.5	2
Mean Corp. Vol.	1	3.5	1	3.5	3.5	3.5
	2	3.5	3.5	1.5	5	1.5

Further evaluation of the differences in precision was based on application of the Friedman multiple comparison procedure adjusted for ties. The precision of the groups of laboratories can be exhibited as follows, where groups joined by the same line are not significantly different from each other at the 90% confidence level.

Rank Order	Laboratory Group
Most Precise	Medicare Laboratories
	JCAH
	Interstate
	Medicare Hospitals
Least Precise	AAFP/ASIM

Table 4.1.6.2 indicates the techniques which the study participants appeared to have applied most successfully, and the percentage of study participants who used these techniques in shipment 1. For all constituents, more than 40% of the laboratories used one of the best applied techniques. In contrast, there was only one clinical chemistry constituent which more than 40% of the laboratories analyzed with one of the best applied techniques.

4.2. Comparison with Other Proficiency Surveys

The data displayed in Table 4.2.1 have been formatted in the same manner as the clinical chemistry information in Section 3.8. The coefficients of variation in all the listed surveys, with the exception of the initial CDC survey, appear to be comparable to those observed in this study.

4.3. Acceptability of Laboratory Group Precision

A literature search did not yield any data which could be used to assess the medical implications of the hematology analyses. However, one article indicated that Dr. Edward Burns of the Medical College of Ohio at Toledo had monitored the white cell count, hemoglobin, and hematocrit levels of a group of individuals from the Owen-Illinois Company.[27] Dr. Burns

[27]E. Burns, "Individual Health Pattern Profile Studies in Fifty-Six Healthy Adults," Journal of Occupational Medicine, Vol. 9, No. 10, October 1967, pp. 511-517.

Table 4.1.6.2. Hematology - Best Applied Techniques

Constituent	Best Applied Techniques	% Using
Red Cell Count	All Coulter Models Kits	43.3
White Cell Count	All Coulter Models Kits	45.7
Hematocrit	Microhematocrit	75.9
Hemoglobin	All techniques applied equally well	100.0
Mean Corpuscular Volume	Impossible to judge	---

Table 4.2.1. Comparison of Coefficients of Variation (%) for Hematology

Constituent	Red Count	White Count	Hemoglobin	Hematocrit	MCV
NBS Study	9.0	12.2	5.3	5.9	9.8
CDC - Initial Analyses - 1968	18.0		18.7	8.6	
CDC - After 21 Mos. in Pgm. - 1970	13.5		3.7	6.9	
CAP - 1965	10.2	10.4	5.0	5.5	
CAP - 1971	7.1	9.0	6.5	4.4	
AAB - 3 Quarters of 1970 & 71	7.3	12.6	4.1	4.5	
LA Physicians			5.5	3.7	
Am. Soc. Clin. Path. - 1961			3.9		
Am. Soc. Clin. Path. - 1964			5.1		

was contacted and agreed to supply data on the individual variability of thirty-nine male and two female subjects whose blood was tested 20 times under standardized conditions over a period of 60-80 days. These data on individual variability are, of course, confounded with the analytical variability of Dr. Burns' laboratory. Consequently, a delineation of the medically useful range based on Dr. Burns' data may be overly broad; and the estimates of acceptability of the precision of the hematology analyses may be overestimates.

The medically acceptable interlaboratory precision, as defined in Chapter 3, can be computed as one-half of the mean individual variability determined in Dr. Burns' study. These values are shown in the column headed S_B in Table 4.3.1. The table also displays the sample size and standard deviation for the reference laboratory determinations, and each laboratory group. Standard deviations which were not significantly greater than the acceptable precision values (defined and computed as above) at the 99% confidence level have been starred in the table.

Examination of this table indicates:

(a) The reference laboratories achieved medically acceptable precision for hematocrit at the abnormal level, and hemoglobin and white blood count at both levels of concentration.
(b) The precision of the JCAH-Accredited laboratory analyses of hemoglobin was medically acceptable at the abnormal constituent level.
(c) All other group precisions exceeded the medically acceptable precision values defined and computed in this report.

5. MICROBIOLOGY

5.1. Background

This phase of the evaluation study consisted of an analysis of two shipments of five vials, where each vial contained a lyophilized pure culture (see Subsection 2.4.3). The participating laboratory was asked to identify the organism in each vial, giving both genus and species wherever possible. According to information supplied by the Scientific Advisory Committee, more than one answer was acceptable for three out of the five organisms. The preferred and otherwise acceptable answers are displayed in Table 5.1.1.

Table 5.1.1. Preferred and Other Acceptable Identifications of Microbiological Specimens

Preferred	Other Acceptable	Vial & Shipment
Streptococcus pneumoniae Diplococcus pneumoniae Pneumococcus pneumoniae Pneumococcus	None	Vial 1, Shipment 1 and Vial 4, Shipment 2
Salmonella oranienberg Salmonella C_1 Salmonella montevideo Salmonella enteritidis Salmonella NOS	Salmonella arizona	Vial 2, Shipment 1 and Vial 5, Shipment 2
Klebsiella pneumoniae Klebsiella NOS	Klebsiella- Enterobacter group	Vial 3, Shipment 1 and Vial 1, Shipment 2
Pseudomonas aeruginosa	Pseudomonas fluorescens Pseudomonas stutzeri Pseudomonas NOS	Vial 4, Shipment 1 and Vial 3, Shipment 2

Table 4.3.1. Acceptability of Group Precision - Hematology

Constituent		S_B	Deviations											
			Reference Labs		Doctors' Offices		Interstate		JCAH		Medicare Hospitals		Medicare Labs	
			n	s	n	s	n	s	n	s	n	s	n	s
White Blood Count	- Normal	0.6	14	*0.4	208	0.9	168	0.7	184	0.7	184	0.9	118	0.9
	Abnormal	1.3	14	*1.1	197	2.6	165	1.8	180	1.9	180	2.4	120	1.8
Hematocrit	- Normal	0.7	14	1.9	180	2.2	171	2.2	186	2.0	184	1.7	117	1.8
	- Abnormal	0.7	14	*0.7	179	1.2	167	1.1	188	1.1	180	0.9	121	1.0
Hemoglobin	- Normal	0.3	14	*0.3	204	0.6	168	0.4	185	0.4	186	0.5	860	0.5
	- Abnormal	0.3	14	*0.3	198	0.5	166	0.4	188	*0.3	186	0.5	858	0.4

Starred values are not significantly greater than S_B at the 99% confidence level.

Table 5.1.1. Preferred and Other Acceptable Identifications of Microbiological Specimens (continued)

Preferred	Other Acceptable	Vial & Shipment
Streptococcus faecalis Enterococcus, group D Alpha Streptococcus, group D Beta hemolytic Strep., group D Gamma Strep., group D Enterococcus Streptococcus enterococcus Streptococcus zymogenes	None	Vial 5, Shipment 1 and Vial 2, Shipment 2

Upon the advice of the Scientific Advisory Committee, each instance of multiple identifications (a laboratory reporting that it had identified two or more organisms within the same vial) was treated separately. When one of the organisms identified was a pathogen which would not be expected to be found in the usual laboratory environment, the analysis of the vial was considered incorrect regardless of whether the correct organism was also one of those identified. If the correct organism was identified along with one of the innocuous bacteria which normally occur on the skin or in the air, the harmless contaminant was ignored, and the analysis of the vial was considered correct. There were several instances where the study team sought the advice of Dr. George Dillard, resident public health service physician at the National Bureau of Standards, as to whether a specific analytical report would have resulted in a medical diagnosis other than that which was intended.

5.2. Overall Performance

It is informative to look at the percent of incorrect answers by each laboratory group, recalling that participation in the study was completely voluntary, that these findings relate more closely to laboratory capability than they do to routine performance, and that the organisms were not unusual ones.

Although the overall frequency of incorrect identifications is 16.8%, it should be noted that deleting the Interstate group raises this figure to 19.4%—almost one out of every five.

Laboratory Group	Sample Size[28]	Incorrect Identifications	
		n	%
Interstate Licensure	1021	78	7.6
JCAH Hospitals	1639	259	15.8
Medicare Independent	778	139	17.9
Doctors' Offices	308	64	20.8
Medicare Hospitals	987	257	26.0
OVERALL	4733	797	16.8

[28]Sample Size = total number of identifications, not number of laboratories.

5.3. Comparison of Group Performance

Two statistical approaches—one parametric and one non-parametric—were considered for use in evaluating the relative performance of the groups. Each approach has some drawbacks. The non-parametric approach, based on a ranking procedure, ignores the quantitative nature of the data. Unnecessary reduction of the significance level of the performance rankings may result. The parametric approach relies upon assumptions concerning the normality of the distribution of the data and does not provide an adequate means for distinguishing "other acceptable" from "preferred" identifications. Since neither approach seemed totally satisfactory, both were tried. Some credence can be attached to the resulting performance rankings because they were consistent, although the significance levels differed. The higher parametric significance levels were used in reporting the results. Both analytic approaches are described in this section.

5.3.1. Non-parametric Approach

In order to examine relative performance non-parametrically, the submitted report forms were scored as follows: any of the Preferred identifications arbitrarily received a score of 20, identification of "Other Acceptable" received a score of 10, and all other identifications were scored as zero. It must be recognized that this scoring system assumes that the five organisms were equally difficult (or easy) to identify.

No attempt was made to analyze the laboratory scores on different vials parametrically because the assignment of equal scores to correct answers for all organisms was arbitrary, and parametric analysis would be invalid if the organisms were not equally easy to identify. Instead, the scores on each vial were ranked across laboratory groups. Although both shipments contained the same five cultures (with different vial numbers), their identifications were treated independently (e.g., 10 vials rather than 5) because of the variation in the number of laboratories reporting results of shipments 1 and 2. This variation is interesting to note because of the rather consistent decrease in number of participating laboratories between the first and second shipments. Doctors' Offices, the Interstate group, and Medicare Hospitals lost approximately 10 each, while the JCAH group was reduced by almost 20 laboratories in the second shipment.

The results were placed in a 10 x 6 matrix, each of the ten vials representing a row and the six laboratory groups representing the columns. Table 5.3.1 contains rankings by vial (or row); a rank of one indicates the lowest average score and the rank of six is given to the highest average score for that vial. For example consider S. pneumoniae in shipment 1. The lowest average score was received by the physician members of the AAFP, followed in order by the Medicare Certified Hospitals, the Medicare Certified Independent laboratories, the JCAH Hospitals, the ASIM physicians, and finally the CDC Interstate laboratories (the highest ranking group). In this and subsequent tables, n is the total number of analyses of the specimen reported by members of the group.

Table 5.3.1. Ranking of Group Performance, Microbiology

Specimen	Shipment	AAFP		ASIM		Interstate (CDC)		JCAH		Medic. Hosp.		Medic. Indep.	
		n	Rank	n	Rank	n	Rank	n	Rank	n	Rank	n	Rank
S. pneumoniae	1	9	1	27	5	107	6	174	4	108	2	83	3
S. pneumoniae	2	4	1	22	3	97	6	151	4	91	2	74	5
Salmonella NOS	1	9	1	23	2	107	6	170	4	105	3	80	5
Salmonella NOS	2	3	2	24	1	98	6	154	5	88	3	74	4
Klebsiella NOS	1	10	1	28	2	107	6	176	5	109	3	83	4

Table 5.3.1. Ranking of Group Performance, Microbiology (continued)

Specimen	Shipment	AAFP		ASIM		Interstate (CDC)		JCAH		Medic. Hosp.		Medic. Indep.	
		n	Rank	n	Rank	n	Rank	n	Rank	n	Rank	n	Rank
Klebsiella NOS	2	4	6	24	3	98	5	155	2	90	1	75	4
P. aeruginosa	1	10	2	27	4	105	6	175	5	108	1	80	3
P. aeruginosa	2	4	6	23	4	98	5	156	2	91	1	75	3
S. faecalis	1	8	4	28	3	107	6	173	5	108	1	81	2
S. faecalis	2	4	1	22	4	97	6	155	5	89	2	74	3
		$\bar{R} = 2.6$		$\bar{R} = 3.1$		$\bar{R} = 5.8$		$\bar{R} = 4.1$		$\bar{R} = 1.9$		$\bar{R} = 3.6$	

The statistical test used to determine whether the group rankings were significantly different was Friedman's Two-Way Analysis of Variance by Ranks.[29] If all of the groups had equal capability to identify the organisms, then the distribution of ranks in each column would be a matter of chance. Thus, one would expect ranks of 1, 2, ---6 to appear in each column with about equal frequency. The Friedman test determines whether the rank totals differ significantly.

The first application of the Friedman test involved all six laboratory groups and resulted in rejection of the hypothesis of equal capability at the .05 level of probability. In other words there is a very strong reason to believe that the six laboratory groups differ in ability. The multiple comparison portion of the test indicated that data from the two physician-office groups (the AAFP and the ASIM) could be pooled together. Analysis to justify pooling was not repeated in the parametric analysis.

After combining the AAFP and ASIM it was necessary to re-rank all of the groups, as shown in Table 5.3.2.

Table 5.3.2. Revised Ranking of Group Performance, Microbiology

Specimen	Shipment	Doctors' Offices		Interstate (CDC)		JCAH		Medic. Hosp.		Medic. Indep.	
		n	Rank	n	Rank	n	Rank	n	Rank	n	Rank
S. pneumoniae	1	36	2	107	5	174	4	108	1	83	3
S. pneumoniae	2	26	2	97	5	151	3	91	1	74	4
Salmonella NOS	1	32	1	107	5	170	3	105	2	80	4
Salmonella NOS	2	27	1	98	5	154	4	88	2	73	3
Klebsiella NOS	1	38	1	107	5	176	4	109	2	83	3
Klebsiella NOS	2	28	3	98	5	155	2	90	1	75	4
P. aeruginosa	1	37	2	105	5	175	4	108	1	80	3
P. aeruginosa	2	27	4	98	5	156	2	91	1	75	3
S. faecalis	1	36	3	107	5	173	4	108	1	81	2
S. faecalis	2	26	3	97	5	155	4	89	1	74	2
		$\bar{R} = 2.2$		$\bar{R} = 5.0$		$\bar{R} = 3.4$		$\bar{R} = 1.3$		$\bar{R} = 3.1$	

[29] R. Miller, op. cit.

Friedman's test was run again, and the differences among the laboratory groups were still significant at the 95% confidence level. The multiple comparison portion of the test was then applied at various confidence levels. The results of this analysis are discussed in Subsection 5.3.3.

5.3.2. Parametric Approach

A number of different additive analysis of variance (ANOVA) models were run. Scheffe's procedure was applied to the ANOVA results in order to rank the laboratory groups.[30] The models used were:

(a) Proportion of incorrect identifications by vial and shipment versus laboratory group;
(b) Same data as in (a) with a variance-stabilizing arcsine transformation;
(c) Proportion of incorrect identifications averaged over shipments versus laboratory group;
(d) Same data as in (c) with arcsine transformation.

An analysis weighted by sample sizes was also done for each of the four cases. None of the models fit the data as well as one would like; but the conclusions of all, significant at the 95% confidence level, were consistent.

5.3.3. Performance Ranking

The parametric ranking at the 95% confidence level is identical to the non-parametric ranking at the 75% confidence level. This ranking will be ascribed a confidence level of 95%. The performance (capability of the different groups of laboratories can then be portrayed as follows, where groups joined by the same vertical line are not significantly different from each other at the 95% confidence level.

Rank Order	Laboratory Group
High	Interstate (CDC)
	JCAH
	Medicare Independent
	AAFP/ASIM
Low	Medicare Hospitals

Several statements can be made based on this ranking:

(a) The demonstrated microbiological capability of the Interstate Licensure group (CDC) was significantly better than the capability of every other group.
(b) The JCAH Hospital and Medicare Independent laboratories were significantly better than the Medicare Hospital laboratories but were not significantly different from the Doctors' Office laboratories.
(c) The Medicare Hospital laboratories were significantly poorer than all other groups except the Doctors' Office laboratories.

5.4. Summary of Group Identification

It is interesting to review the various laboratory groups insofar as their identification capabilities are concerned. To facilitate such a review, two tables have been constructed for each particular group and for an aggregate of all groups. For each group the first of these tables contains information on the frequency of correct, marginal and incorrect identifications, while the second table lists these incorrect identifications. Of particular

---

[30] R. Miller, pp. 48-53.

interest is the heterogeneity of identifications within a vial. The medical implications of these incorrect identifications are beyond the scope of this report.

5.4.1. Overview of All Participants

The pooled data in Table 5.4.1.1 indicate that Streptococcus faecalis was much more difficult to identify properly than any of the other organisms. The easiest organism to identify properly was Pseudomonas aeruginosa, and the other three bacteria fell in between with very little difference in misidentification rate. The fact that S. faecalis was the most difficult organism to properly identify is a bit disconcerting because this organism is not that unusual, especially to a large hospital laboratory. From the list of misidentifications in Table 5.4.1.2, it will be evident that many of the organisms mistaken for Streptococcus faecalis were in the Streptococcus group—viridans, pyogenes, non hemolytic (gamma), etc. However, there were other instances where groups such as Staphylococcus, Pseudomonas, Neisseria, Salmonella and Enterobacter were named in place of S. faecalis.

Table 5.4.1.1. Overview of All Study Participants

Specimen	Shipment	Number of labs	Correct ID's		Marginal ID's		Incorrect ID's	
			n	%	n	%	n	%
S. pneumoniae	1	508	431	85	N/A	-	77	15
S. pneumoniae	2	438	364	83	N/A	-	74	17
Salmonella NOS	1	494	402	81	20	4	72	15
Salmonella NOS	2	439	345	78	26	6	68	16
Klebsiella NOS	1	513	387	76	32	6	94	18
Klebsiella NOS	2	445	353	79	25	6	67	15
P. aeruginosa	1	505	335	66	119	24	51	10
P. aeruginosa	2	446	301	67	112	25	33	8
S. faecalis	1	505	354	70	N/A	-	151	30
S. faecalis	2	440	330	75	N/A	-	110	25

Table 5.4.1.2. Misidentifications by All Study Participants

Shipment 1, S. pneumoniae: Yeast (1), Pseudomonas NOS (2), Salmonella NOS (2), Klebsiella NOS (1), K. pneumoniae (1), Klebsiella-Enterobacter group (1), H. influenzae (1), Staphylococcus NOS (2), S. aureus (4), S. epidermis (4), Neisseria NOS (3), M. polymorpha (2), Diplococcus (10), Streptococcus NOS (7), S. pyrogenes (4), S. viridans (22), S. faecalis (1), S. anaerobis (1), Non hemolytic strep (3), Mycobacterium NOS (1)

Salmonella NOS: Pseudomonas NOS (1), P. aeruginosa (2), E. coli (2), S. alkalescens (1), Edwardsiella (1), Shigella NOS (1), S. dysenteriae (1), S. typhi (13), S. paratyphi A (1), S. paratyphi B (8), S. paratyphi C (6), S. choleraesuis (7), S. typhimurium (2), Citrobacter group (7), Klebsiella-Enterobacter group (1), Enterobacter NOS (1), E. aerogenes (2), P. aerugenoides (2), Proteus NOS (4), P. vulgaris (1), P. mirabilis (2), S. aureus (1), S. epidermis (1), Neisseria NOS (1), S. viridans (1), S. faecalis (1), Bacillus NOS (1)

Table 5.4.1.2. Misidentifications by All Study Participants (continued)

Shipment 1, Klebsiella NOS: Pseudomonas NOS (1), E. coli (19), Salmonella NOS (1), Citrobacter group (2), K. ozaenae (1), Klebsiella-Enterobacter group (1), Enterobacter NOS (13), E. cloacae (16), E. aerogenes (16), E. liquefaciens (3), E. cloacae, atypical (1), P. aerugenoides (2), Proteus NOS (2), P. vulgaris (1), P. morganii (1), P. rettgeri (1), H. influenzae (2), Staphylococcus NOS (2), S. aureus (2), S. epidermis (1), S. pneumoniae (2), S. viridans (1), Corynebacterium (2), Bacillus NOS (1)

P. aeruginosa: P. pseudomallei (6), P. maltophilia (1), P. aeruginosa, aberrant strain (2), Alcaligenes (4), Achromobacter (1), E. coli (1), Edwardsiella (1), Salmonella NOS (3), Klebsiella NOS (4), Klebsiella-Enterobacter group (1), Enterobacter NOS (3), E. aerogenes (1), E. hafniae (2), Serratia (3), S. marcescens (2), Proteus NOS (4), P. vulgaris (1), P. mirabilis (1), H. influenzae (1), Mima (1), M. polymorpha (2), Herellea (1), Streptococcus NOS (1), S. faecalis (2), Bacillus NOS (1), B. subtilis (1)

S. faecalis: Pseudomonas NOS (3), P. aeruginosa (6), A. hydrophila (1), E. coli (4), S. alkalescens (2), Shigella NOS (1), S. sonnel (1), Klebsiella NOS (1), K. pneumoniae (1), Klebsiella-Enterobacter group (1), Enterobacter NOS (1), E. cloacae (1), E. liquefaciens (1), P. aerugenoides (1), Proteus NOS (1), Staphylococcus NOS (1), S. aureus (16), S. epidermis (17), Neisseria NOS (3), N. gonorrhoeae (1), N. meningitidis (1), Herellea (1), S. pneumoniae (1), Streptococcus NOS (17), S. pyogenes (9), S. viridans (19), Streptococcus MG, Beta hemolytic strep, Group B (4), Non hemolytic strep (Gamma) (19), Beta hemolytic strep (15), Salmonella NOS (1), Alcaligenes (1)

Shipment 2, Klebsiella NOS: Pseudomonas NOS (1), E. coli (10), S. typhi (1), K.ozaenae (1), Enterobacter NOS (9), E. cloacae (14), E. aerogenes (10), E. liquefaciens (1), E. hafniae (1), P. aerugenoides (1), Proteus NOS (2), P. vulgaris (1), P. mirabilis (2), P. rettgeri (1), Haemophilus NOS (1), H. influenzae (1), Neisseria NOS (1), M. polymorpha (1), S. pneumoniae (5), S. faecalis (1), Bacillus NOS (2)

S. faecalis: P. aeruginosa (1), A. hydrophila (1), E. coli (3), Shigella NOS (1), S. dysenteriae (1), Salmonella NOS (7), S. typhi (1), S. paratyphi C (1), Klebsiella NOS (2), E. cloacae (1), E. aerogenes (3), Proteus NOS (1), P. vulgaris (1), Staphylococcus NOS (4), S. aureus (19), S. epidermis (11), M. polymorpha (2), S. pneumoniae (1), Streptococcus NOS (10), S. pyogenes (4), S. viridans (12), Non hemolytic strep (Gamma) (11), Beta hemolytic strep (12)

P. aeruginosa: P. pseudomallei (1), P. maltophilia (1), P. aeruginosa, aberrant strain (1), Alcaligenes (3), A. species (1), E. coli (1), Klebsiella NOS (1), K. pneumoniae (2), Klebsiella-Enterobacter group (2), Enterobacter NOS (1), E. cloacae (2), E. aerogenes (3), Serratia (1), Proteus NOS (2), H. influenzae (2), Moraxella (1), Staphylococcus NOS (2), S. aureus (1), S. epidermis (2), S. pneumoniae (1), Streptococcus NOS (2)

S. pneumoniae: Pseudomonas NOS (5), P. aeruginosa (5), Achromobacter (1), E. coli (1), Salmonella NOS (1), Klebsiella NOS (2), K. pneumoniae (2), E. aerogenes (1), S. aureus (3), S. epidermis (2), Neisseria NOS (1), N. meningitidis (1), M. polymorpha (1), Streptococcus NOS (5), S. pyogenes (2), S. viridans (29), S. faecalis (6), Non hemolytic strep (Gamma) (1), Beta hemolytic strep (2), Diphtheroids NOS (1), L. monocytogenes (2)

Salmonella NOS: Pseudomonas NOS (2), P. aeruginosa (2), Alcaligenes (1), E. coli (1), S. alkalescens (2), Shigella NOS (1), S. typhi (6), S. paratyphi A (1), S. paratyphi B (5), S. choleraesuis (5), Citrobacter group (16), Enterobacter NOS (1), E. cloacae (2), Serratia (1), S. marcescens (1), Proteus NOS (2), P. vulgaris (2), P. mirabilis (3), Staphylococcus NOS (1), Mima (1), S. viridans (1), S. faecalis (8), Non hemolytic strep (Gamma) (1), Bacillus NOS (2)

5.4.2. Doctors' Offices

Table 5.4.2.1 indicates that the number of incorrect identifications of Klebsiella NOS, P. aeruginosa and S. faecalis significantly decreased from shipment 1 to shipment 2. Further investigation of the data disclosed this to be related to two causes. First, approximately half of the decrease is attributed to doctors' laboratories which misidentified these specimens in the first shipment and did not analyze the second. The remaining decrease could be attributed to laboratories which had improved their laboratory analyses. Dropouts accounted for proportionate decreases in correct and incorrect identifications. Thus, the change in rate is related to an improvement in performance. S. faecalis was the most frequently misidentified organism, while P. aeruginosa was least frequently misidentified. Table 5.4.2.2 lists misidentifications made by the Doctors' Offices.

Table 5.4.2.1. Doctors' Office Laboratories

Specimen	Shipment	Number of labs	Correct ID's		Marginal ID's		Incorrect ID's	
			n	%	n	%	n	%
S. pneumoniae	1	36	29	81	N/A	-	7	19
S. pneumoniae	2	25	19	76	N/A	-	6	24
Salmonella NOS	1	32	21	66	4	12	7	22
Salmonella NOS	2	26	16	61	2	8	8	31
Klebsiella NOS	1	38	24	63	4	10	10	26
Klebsiella NOS	2	27	22	81	3	11	2	7
P. aeruginosa	1	37	22	59	10	27	5	14
P. aeruginosa	2	26	19	73	7	27	0	0
S. faecalis	1	36	24	67	N/A	-	12	33
S. faecalis	2	25	18	72	N/A	-	7	28

Table 5.4.2.2. Misidentifications by Doctors' Office Laboratories

Shipment 1, S. pneumoniae: S. epidermis (1), Staphylococcus NOS (1), S. viridans (2), S. faecalis (1), Klebsiella (1), S. aureus (1)

Salmonella NOS: E. coli (1), S. paratyphi B (1), S. paratyphi C (1), Proteus NOS (2), Edwardsiella (1), P. aeruginosa (1)

Klebsiella NOS: E. coli (4), Pseudomonas NOS (1), Enterobacter NOS (1), E. cloacae (1), E. aerogenes (2), S. pneumoniae (1)

P. aeruginosa: Salmonella NOS (1), Enterobacter NOS (1), Proteus NOS (1), P. aeruginosa (1), Alcaligenes (1)

S. faecalis: S. aureus (2), S. epidermis (2), Streptococcus NOS (1), S. pyogenes (1), S. viridans (2), Non hemolytic strep (Gamma) (2), Beta hemolytic strep (1), E. coli (1)

Shipment 2, Klebsiella NOS: Enterobacter NOS (1), Proteus NOS (1)

S. faecalis: Proteus NOS (1), Staphylococcus NOS (1), S. aureus (3), Non hemolytic strep (Gamma) (1), Beta hemolytic strep (1)

P. aeruginosa: ----------

Table 5.4.2.2. Misidentifications by Doctors' Office Laboratories (continued)

Shipment 2, S. pneumoniae: Pseudomonas NOS (1), S. viridans (2), S. faecalis (2), S. aureus (1)

Salmonella NOS: Pseudomonas NOS (1), P. aeruginosa (2), S. paratyphi B (1), S. choleraesuis (1), Citrobacter group (1), Proteus NOS (1), Alcaligenes (1)

5.4.3. Interstate Licensure (CDC) Laboratories

The Interstate laboratories were found to have the highest percentage of correct identifications. As Table 5.4.3.1 indicates, the organism most frequently misidentified was S. faecalis, while that most frequently identified correctly was P. aeruginosa. Table 5.4.3.2 lists misidentifications made by the CDC members.

Table 5.4.3.1. Interstate Licensure (CDC) Laboratories

Specimen	Shipment	Number of labs	Correct ID's		Marginal ID's		Incorrect ID's	
			n	%	n	%	n	%
S. pneumoniae	1	107	102	95	N/A	-	5	5
S. pneumoniae	2	97	89	92	N/A	-	8	8
Salmonella NOS	1	107	97	91	0	-	10	9
Salmonella NOS	2	98	86	88	2	2	10	10
Klebsiella NOS	1	107	97	91	1	1	9	8
Klebsiella NOS	2	98	91	93	1	1	6	6
P. aeruginosa	1	105	83	79	20	19	2	2
P. aeruginosa	2	98	74	76	22	22	2	2
S. faecalis	1	107	95	89	N/A	-	12	11
S. faecalis	2	97	83	86	N/A	-	14	14

Table 5.4.3.2. Misidentifications by Interstate LIcensure Laboratories

Shipment 1, S. pneumoniae: M. polymorpha (1), Diplococcus (2), Streptococcus NOS (1), S. viridans (1)

Salmonella NOS: S. paratyphi A (1), S. paratyphi B (2), S. paratyphi C (3), S. choleraesuis (1), S. typhimurium (1), Citrobacter group (1), Proteus NOS (1)

Klebsiella NOS: E. coli (3), K. ozaenae (1), E. cloacae (1), E. aerogenes (2), E. liquefaciens (1), E. cloacae, atypical (1)

P. aeruginosa: P. pseudomallei (1), Achromobacter (1)

S. faecalis: A. hydrophila (1), E. liquefaciens (1), S. aureus (2), Neisseria NOS (1), Streptococcus NOS (3), S. viridans (2), Non hemolytic strep (Gamma) (1), Beta hemolytic strep (1)

Shipment 2, Klebsiella NOS: E. coli (1), Enterobacter NOS (1), E. liquefaciens (2), P. vulgaris (1), M. polymorpha (1)

Table 5.4.3.2. Misidentifications by Interstate Licensure Laboratories (continued)

Shipment 2, S. faecalis: A. hydrophila (1), E. coli (1), Salmonella NOS (2), E. cloacae (1), S. aureus (1), M. polymorpha (2), Streptococcus NOS (1), S. viridans (1), Non hemolytic strep (Gamma) (2), Beta hemolytic strep (2)

P. aeruginosa: E. cloacae (2)

S. pneumoniae: P. aeruginosa (1), K. pneumoniae (1), Neisseria NOS (1), Streptococcus NOS (3), S. viridans (1), Beta hemolytic strep (1)

Salmonella NOS: S. paratyphi A (1), S. paratyphi B (1), Citrobacter group (4), Serratia (1), Proteus NOS (1), S. faecalis (2)

5.4.4. JCAH Hospital Laboratories

Similar to the first two groups, the JCAH Hospital laboratories more frequently misidentified S. faecalis than any of the other four organisms, and correctly identified the P. aeruginosa organism most frequently. Table 5.4.4.1 shows the identification frequencies, and Table 5.4.4.2 lists the misidentifications.

Table 5.4.4.1. JCAH Hospital Laboratories

Specimen	Shipment	Number of labs	Correct ID's		Marginal ID's		Incorrect ID's	
			n	%	n	%	n	%
S. pneumoniae	1	174	146	84	N/A		28	16
S. pneumoniae	2	151	126	83	N/A		25	17
Salmonella NOS	1	170	142	84	6	4	22	13
Salmonella NOS	2	154	127	82	8	5	19	12
Klebsiella NOS	1	176	133	76	13	7	30	17
Klebsiella NOS	2	155	116	75	9	6	30	19
P. aeruginosa	1	175	124	71	42	24	9	5
P. aeruginosa	2	156	107	69	34	22	15	10
S. faecalis	1	173	124	72	N/A	-	49	28
S. faecalis	2	155	123	80	N/A		32	21

Table 5.4.4.2. Misidentifications by JCAH Hospital Laboratories

Shipment 1, S. pneumoniae: Yeast (1), K. pneumoniae (1), S. epidermis (2), Neisseria NOS (1), Diplococcus (5), Streptococcus NOS (4), S. pyogenes (2), S. viridans (7), Non hemolytic strep (Gamma) (2), Beta hemolytic strep (2), Pseudomonas NOS (1)

Salmonella NOS: S. alkalescens (1), S. typhi (6), S. paratyphi B (2), S. paratyphi C (2), S. choleraesuis (4), Citrobacter group (1), P. aerugenoides (1), P. vulgaris (1), P. mirabilis (1), S. epidermis (1), Neisseria NOS (1), Bacillus NOS (1)

Klebsiella NOS: E. coli (6), Salmonella NOS (1), Citrobacter group (1), Enterobacter NOS (5), E. cloacae (4), E. aerogenes (5), P. aerugenoides (2),

Table 5.4.4.2. Misidentifications by JCAH Hospital Laboratories (continued)

Shipment 1, Klebsiella NOS (continued): P. rettgeri (1), H. influenzae (1), S. aureus (1), S. epidermis (1), S. viridans (1), Bacillus NOS (1)

P. aeruginosa: P. pseudomallei (2), Alcaligenes (1), E. aerogenes (1), E. hafniae (1), S. marcescens (1), P. mirabilis (1), Mima (1), Bacillus NOS (1)

S. faecalis: Pseudomonas NOS (1), P. aeruginosa (4), E. coli (1), E. cloacae (1), P. aerugenoides (1), Staphylococcus NOS (1), S. aureus (4), S. epidermis (6), Herellea (1), Streptococcus NOS (5), S. pyogenes (3), S. viridans (8), Non hemolytic strep (Gamma) (5), Beta hemolytic strep (8)

Shipment 2, Klebsiella NOS: E. coli (5), S. typhi (1), K. ozaenae (1), Enterobacter NOS (1), E. cloacae (7), E. aerogenes (4), E. hafniae (1), P. aerugenoides (1), Proteus NOS (1), P. mirabilis (1), P. rettgeri (1), Haemophilus NOS (1), S. pneumoniae (3), S. faecalis (1), Bacillus NOS (1)

S. faecalis: Salmonella NOS (2), S. paratyphi C (1), Klebsiella NOS (1), E. aerogenes (3), P. vulgaris (1), Staphylococcus NOS (1), S. aureus (6), S. epidermis (4), Streptococcus NOS (5), S. pyogenes (2), S. viridans (1), Non hemolytic strep (Gamma) (2), Beta hemolytic strep (3)

P. aeruginosa: P. maltophilia (1), P. aeruginosa (1), Klebsiella NOS (1), K. pneumoniae (2), Klebsiella-Enterobacter group (1), E. aerogenes (3), Serratia (1), H. influenzae (1), S. aureus (1), S. epidermis (1), S. pneumoniae (1), Streptococcus NOS (1)

S. pneumoniae: Pseudomonas NOS (1), P. aeruginosa (4), Klebsiella NOS (2), E. aerogenes (1), S. aureus (1), S. epidermis (2), M. polymorpha (1), S. viridans (9), S. faecalis (3), Non hemolytic strep (1)

Salmonella NOS: Shigella NOS (1), S. typhi (3), S. paratyphi (1), S. choleraesuis (2), Citrobacter group (1), E. cloacae (1), P. vulgaris (2), P. mirabilis (2), Mima (1), S. viridans (1), S. faecalis (3), Bacillus NOS (1)

5.4.5. Medicare Hospital Laboratories

The Medicare Hospital laboratories most frequently misidentified the S. faecalis organism, and most frequently identified the P. aeruginosa organism, as did the other laboratory groups. Similar to the Doctors' Office data, the Medicare Hospital laboratory data indicate that a significant decrease in the number of misidentifications occurred from shipment 1 to shipment 2, especially with regard to S. faecalis where the percentage of misidentifications decreased from 45% to 27%. Further review of the data revealed that eight laboratories which misidentified this organism in the first shipment identified it correctly in the second shipment. The change in rate was largely a consequence of this improvement in performance. Tables 5.4.5.1 and 5.4.5.2 further detail the Medicare Hospital analyses.

Table 5.4.5.1. Medicare Hospital Laboratories

Specimen	Shipment	Number of labs	Correct ID's		Marginal ID's		Incorrect ID's	
			n	%	n	%	n	%
S. pneumoniae	1	108	84	78	N/A		24	22
S. pneumoniae	2	91	64	70	N/A		27	30
Salmonella NOS	1	105	75	71	6	6	24	23
Salmonella NOS	2	88	59	67	8	9	21	24

Table 5.4.5.1. Medicare Hospital Laboratories (continued)

Specimen	Shipment	Number of labs	Correct ID's		Marginal ID's		Incorrect ID's	
			n	%	n	%	n	%
Klebsiella NOS	1	109	74	68	9	8	26	24
Klebsiella NOS	2	90	63	70	8	9	19	21
P. aeruginosa	1	108	55	51	30	28	23	21
P. aeruginosa	2	91	50	55	29	32	12	13
S. faecalis	1	108	60	55	N/A		48	44
S. faecalis	2	89	56	63	N/A		33	37

Table 5.4.5.2. Misidentifications by Medicare Hospital Laboratories

Shipment 1, S. pneumoniae: Pseudomonas NOS (1), Klebsiella-Enterobacter group (1), H. influenzae (1), Staphylococcus NOS (1), S. aureus (1), S. epidermis (1), Neisseria NOS (2), M. polymorpha (1), Streptococcus NOS (1), S. pyogenes (1), S. viridans (9), Non hemolytic strep (Gamma) (2), Beta hemolytic strep (1), Mycobacterium (1)

Salmonella NOS: P. aeruginosa (1), E. coli (1), Shigella NOS (1), S. dysenteriae (1), S. typhi (5), S. paratyphi B (3), S. typhimurium (1), Citrobacter group (4), Klebsiella-Enterobacter group (1), Enterobacter NOS (1), E. aerogenes (2), P. aerugenoides (1), S. aureus (1), S. faecalis (1)

Klebsiella NOS: E. coli (3), Enterobacter NOS (4), E. cloacae (7), E. aerogenes (4), E. liquefaciens (1), Proteus NOS (2), P. vulgaris (1), H. influenzae, Staphylococcus NOS (2)

P. aeruginosa: P. aeruginosa; aberrant strain (1), Alcaligenes (1), E. coli (1), Edwardsiella (1), Salmonella NOS (1), Klebsiella NOS (4), Klebsiella-Enterobacter group (1), Enterobacter NOS (1), E. hafniae (1), Serratia (3), Proteus NOS (3), P. vulgaris (1), H. influenzae (1), Herellea (1), S. faecalis (1), B. subtilis (1)

S. faecalis: Pseudomonas NOS (1), P. aeruginosa (2), E. coli (1), S. alkalescens (2), Shigella NOS (1), Klebsiella NOS (1), Klebsiella-Enterobacter group (1), Enterobacter NOS (1), S. aureus (6), S. epidermis (7), Neisseria NOS (2), N. gonorrhoeae (1), N. meningitidis (1), S. pneumoniae (1), Streptococcus NOS (6), S. pyogenes (2), S. viridans (2), Streptococcus MG (1), Beta hemolytic strep, Group B (2), Non hemolytic strep (Gamma) (5), Beta hemolytic strep (2)

Shipment 2, Klebsiella NOS: Pseudomonas NOS (1), E. coli (2), Enterobacter NOS (4), E. cloacae (6), E. aerogenes (3), H. influenzae (1), Neisseria NOS (1), S. pneumoniae (1)

S. faecalis: E. coli (1), Shigella NOS (1), S. dysenteriae (1), Salmonella NOS (2), S. typhi (1), Klebsiella NOS (1), Staphylococcus NOS (2), S. aureus (6), S. epidermis (5), S. pneumoniae (1), Streptococcus NOS (4), S. pyogenes (2), S. viridans (3), Non hemolytic strep (Gamma) (2), Beta hemolytic strep (1)

P. aeruginosa: Alcaligenes (2), E. coli (1), Klebsiella-Enterobacter group (1), Proteus NOS (2), H. influenzae (1), Moraxella (1), Staphylococcus NOS (2), S. epidermis (1), Streptococcus NOS (1)

Table 5.4.5.2. Misidentifications by Medicare Hospital Laboratories (continued)

Shipment 2, S. pneumoniae: Pseudomonas NOS (3), Achromobacter (1), E. coli (1), Salmonella NOS (1), S. aureus (1), N. meningitidis (1), Streptococcus NOS (2), S. pyogenes (2), S. viridans (12), Diphteroids NOS (1), L. monocytogenes (2)

Salmonella NOS: Pseudomonas NOS (1), E. coli (1), S. alkalescens (2), S. typhi (2), S. paratyphi B (2), S. choleraesuis (1), Citrobacter group (8), S. marcescens (1), Staphylococcus NOS (1), S. faecalis (2)

5.4.6. Medicare Independent Laboratories

The data for the Medicare Independent laboratories, as shown in Tables 5.4.6.1 and 5.4.6.2, bear a very strong resemblance to the JCAH Hospital laboratory data. Once again, the most frequently misidentified organism was S. faecalis, while P. aeruginosa was the most frequently correctly identified organism.

Table 5.4.6.1. Medicare Independent Laboratories

Specimen	Shipment	Number of labs	Correct ID's		Marginal ID's		Incorrect ID's	
			n	%	n	%	n	%
S. pneumoniae	1	83	70	84	N/A		13	16
S. pneumoniae	2	74	66	89	N/A		8	11
Salmonella NOS	1	80	67	84	4	5	9	11
Salmonella NOS	2	73	57	78	6	8	10	14
Klebsiella NOS	1	83	59	71	5	6	19	23
Klebsiella NOS	2	75	61	82	4	5	10	13
P. aeruginosa	1	80	51	64	17	21	12	15
P. aeruginosa	2	75	51	68	20	27	4	5
S. faecalis	1	81	51	63	N/A		30	37
S. faecalis	2	74	50	67	N/A		24	32

Table 5.4.6.2. Misidentifications by Medicare Independent Laboratories

Shipment 1, S. pneumoniae: Salmonella NOS (2), S. aureus (2), Diplococcus (3), Streptococcus NOS (1), S. pyogenes (1), S. viridans (3), S. anaerobis (1)

Salmonella NOS: Pseudomonas NOS (1), S. typhi (2), S. choleraesuis (2), Citrobacter group (1), Proteus NOS (1), P. mirabilis (1), S. viridans (1)

Klebsiella NOS: E. coli (3), Citrobacter group (1), Enterobacter NOS (3), E. cloacae (3), E. aerogenes (3), E. liquefaciens (1), P. morganii (1), S. aureus (1), S. pneumoniae (1), Corynebacterium (2)

P. aeruginosa: P. pseudomallei (3), P. maltophilia (1), Alcaligenes (1), Salmonella NOS (1), Enterobacter NOS (1), S. marcescens (1), M. polymorpha (2), Streptococcus NOS (1), S. faecalis (1)

S. faecalis: Pseudomonas NOS (1), E. coli (1), S. sonnei (1), K. pneumoniae (1), Proteus NOS (1), S. aureus (2), S. epidermis (2), Streptococcus NOS (2), S. pyogenes (3), S. viridans (5), Beta hemolytic strep, Group B (2), Non hemolytic strep (Gamma) (6), Beta hemolytic strep (3)

Table 5.4.6.2. Misidentifications by Medicare Independent Laboratories (continued)

Shipment 2, Klebsiella NOS: E. coli (2), Enterobacter NOS (2), E. cloacae (1), E. aerogenes (1), E. liquefaciens (1), P. mirabilis (1), S. pneumoniae (1), Bacillus NOS (1)

S. faecalis: P. aeruginosa (1), E. coli (1), Salmonella NOS (1), S. aureus (3), S. epidermis (2), S. viridans (7), Non hemolytic strep (Gamma) (4), Beta hemolytic strep (5)

P. aeruginosa: P. pseudomallei (1), A. faecalis (1), A. species (1), Enterobacter NOS (1)

S. pneumoniae: K. pneumoniae (1), S. viridans (5), S. faecalis (1), Beta hemolytic strep (1)

Salmonella NOS: S. typhi (1), S. choleraesuis (1), Citrobacter group (2), Enterobacter NOS (1), E. cloacae (1), P. mirabilis (1), S. faecalis (1), Non hemolytic strep (Gamma) (1), Bacillus NOS (1)

6. SUMMARY OF RESULTS

This chapter summarizes the study results, presents some conclusions, and states the limitations of these conclusions.

6.1. Clinical Chemistry

There were no significant differences (at the 95% confidence level) among the average clinical chemistry results obtained by the groups participating in the study. The inter-laboratory consistency ("interlaboratory precision") of the laboratory groups can be exhibited as follows, where groups joined by the same line did not exhibit significantly different precision at the 90% confidence level.

Rank Order	Laboratory Group
Most Precise	Medicare Independent
	Interstate
	JCAH
	AAFP/ASIM
Least Precise	Medicare Hospital

The techniques used had a considerable effect on the accuracy and precision of reported clinical chemistry analyses. Table 3.5.9.2 on page 46 lists the techniques which were most satisfactorily applied and the percentage of the participating laboratories which applied each technique. As the table implies, a large percentage of the participating laboratories are using outmoded techniques. In most instances, automated methods were applied with equal or better average accuracy and considerably better precision than the corresponding manual methods. Results reported by laboratories using diagnostic kits were consistently less precise than other determinations.

Medical usefulness of the clinical chemistry analyses was assessed using the criterion that analyses should be sufficiently precise to permit the interlaboratory monitoring over time of the variation in an individual patient's constituent concentrations. This criterion was not applied to total bilirubin because data on individual bilirubin variation was not available. Of the remaining seven constituents, only cholesterol was analyzed by the study participants with sufficient precision to satisfy the criterion. In contrast to the performance of the study participants, reference laboratory analyses of cholesterol, uric acid, urea nitrogen, sodium and total protein were all sufficiently precise to permit interlaboratory monitoring of individual variation. Those participating laboratories using the best applied techniques also achieved acceptable interlaboratory precision in analyses of these five constituents.

6.2. Target Value Surrogates

In general, manufacturer's average assays appear to be acceptably accurate pooled serum target values, but these assays are overly reliant on human factors. As a result, when mean reference assays are available, they appear to be a better choice for target values.

6.3. Hematology

As with clinical chemistry, the average hematology results obtained by the participating groups did not differ significantly at the 95% confidence level. The interlaboratory precision of the laboratory groups can be exhibited as follows where groups joined by the same line did not exhibit significant differences at the 90% confidence level.

Rank Order	Laboratory Group
Most Precise	Medicare Independent
	JCAH
	Interstate
	Medicare Hospital
Least Precise	AAFP/ASIM

Table 4.1.6.2 on page 69 shows the best applied techniques and the percentages of participating laboratories using these techniques.

6.4. Microbiology

In microbiology, the performance of the Interstate group was significantly better than the performance of the other groups at the 95% confidence level; 7.6% of the Interstate laboratory determinations were incorrect while 19.9% of all other determinations were incorrect. However, even a 7.6% misidentification rate is not satisfactory because the organisms selected are common and should be easily identifiable. Most worrisome are such misidentifications as Neisseria NOS, N. gonorrhoeae or N. meningitidis for the pure culture of Streptococcus faecalis.

The relative performance of the laboratory groups can be portrayed as shown, where the lines join groups whose performance did not differ significantly at the 95% confidence level.

Rank Order	Laboratory Groups
Best Performance	Interstate
	JCAH
	Medicare Independent
	AAFP/ASIM
Worst Performance	Medicare Hospital

6.5. Conclusions

The data indicate that high volume laboratories may be more proficient than smaller laboratories, such as those which serve Doctors' Offices and Medicare Certified Hospitals. In microbiology, 7.6% of the Interstate laboratory determinations were incorrect, while 16.5% of the determinations by other large laboratories (JCAH and Medicare Independents) were incorrect. Thus, it would appear that the CDC proficiency testing program has considerably improved the microbiology performance of the enrolled laboratories (although further improvement is still needed). Conversely, clinical chemistry and hematology analyses by the Interstate laboratories were no better than comparable analyses by other large laboratories, many of whom do not engage in routine proficiency testing programs. This seems to indicate that the CDC proficiency testing programs in clinical chemistry and hematology have had relatively little effect upon the performance of laboratories participating in the program. This

conclusion is further substantiated in a companion report.[31] It is particularly important to improve the effectiveness of these programs because the interlaboratory consistency of study participants with respect to clinical chemistry and hematology was too often insufficient to support monitoring of an individual's constituent concentrations over time. It appears that poor selection of techniques is an important factor in the low rate of acceptability of laboratory determinations.

6.6. Limitations

It must be clearly understood that the results of this survey are limited by four important considerations:

(a) Because all of the laboratories participated on a purely voluntary basis, no straightforward extrapolation can be made to the larger universe of unsampled clinical laboratories.
(b) It is probable that the results of this study do not represent routine laboratory performance for two reasons: (1) a laboratory probably would not volunteer if its management felt that to do so would be disadvantageous, and (2) the sample materials probably received special attention in many of the smaller laboratories which were unfamiliar with analyzing proficiency test samples.
(c) The clinical chemistry test specimens were prepared by a dialyzation process which removes naturally occurring substances. As a result, the accuracy of some methods, as applied to the test specimens, might differ from their accuracy in analyses of human serum.
(d) The true constituent concentrations of cholesterol and the hematology constituents could not be exactly determined. For these constituents accuracy was assessed relative to the mean reference laboratory assays.

7. RECOMMENDATIONS AND THEIR IMPLEMENTATION

It is beyond the scope of this report to provide a complete plan to remedy existing problems with laboratory performance. However, it is possible to recommend several of the elements of such a plan, particularly with regard to proficiency testing.

7.1. Recommendations

(a) Satisfactory performance in a microbiology proficiency testing program conducted under the auspices of either Federal or other approved authorities should be a legislative requirement for all clinical laboratories analyzing microbiological specimens, regardless of their inter- or intrastate status. It is unclear whether this recommendation should be implemented through new legislation or a re-interpretation of existing legislative authority.
(b) An experimental study should be undertaken to determine a better design for proficiency testing programs in clinical chemistry and hematology. An empirical description of the causes of inadequate laboratory work should be used in defining alternative testing strategies for consideration. This study should deal with such questions as frequency of sampling; feedback to participants; number of levels at which to test; long-term monitoring of intralaboratory variability; follow-up procedures on outlier values; and the criteria for scoring, ranking or rating laboratory performance and its medical usefulness. Until the results of this recommended study become available, it does not appear justified (or warranted) to alter the frequency of CDC proficiency testing in clinical chemistry and hematology.

---

[31] Clinical Laboratory Performance Analysis Using Proficiency Test Statistics, NBS Report, NBSIR 73 197, 1973 (to be published).

(a) Through the cooperative efforts of the Federal Government, State and local agencies, professional societies and accreditation agencies, a concerted effort should be made to induce laboratories to use the best available analytical method/system combinations.

(d) In those instances where routine proficiency testing in one of the quantitative specialties is maintained, target values and zones of acceptable performance should be so constructed as to be consistent with generally accepted statistical concepts in quality control and sampling theory. This would include clearly defined rules for deletion of outliers and the use of appropriate standard deviations for the construction of confidence intervals or tolerance limits.

7.2. Implementation

Implementation of two of the recommendations enumerated in Section 7.1 may be facilitated if DHEW requests additional authority.

(a) The licensing authority of the Clinical Laboratory Improvement Act of 1967 could be extended to include those laboratories which are presently exempt but whose normal workload includes the identification of microbiological specimens. An alternate means of insuring the desired level of laboratory performance is to identify and accredit proficiency testing programs currently being operated by State or local governments or professional societies. Under this alternative, the Federal Government should maintain the authority to withhold or withdraw the license of any laboratory whose demonstrated performance in the identification of microbiological specimens does not at least meet existing minimum standards under CLIA '67.

(b) A Technical Advisory Committee on Clinical Laboratory Procedures consisting of government and professional society representatives should be established to identify the most accurate and precise analytical methods available and encourage their use by the largest possible number of clinical laboratories. Zones of acceptable performance for proficiency testing should be constructed in a manner which reflects the variability associated with the more accurate and precise methods and systems. In this way, failure to accept the recommended procedures would increase the risk of unacceptable performance ratings. It may be useful for this Committee to have legislated interfaces with Federal regulatory programs.

APPENDICES

Appendices A, B, and E contain report background information and references. Appendix C reports an inconclusive investigation of possible differences in clinical chemistry technique performance in analyses of serum pool and dialyzed specimens. Appendix D displays a series of graphs which were discussed in the body of the report.

Appendix A. Scientific Advisory Committee for Study of Clinical Laboratory Capability

Roy Barnett, M.D.
Chief, Pathology Department
Norwalk Hospital

Howard L. Bodily, Ph.D.
Assistant Director and Chief
of Laboratory Services
California State Department
of Health

Joseph H. Boutwell, Ph.D., M.D.
Chief, Licensure and Development Branch
Center for Disease Control

Harvey R. Gralnick, M.D.
Chief, Hematology Service
Clinical Center
National Institutes of Health

James D. MacLowry, M.D.
Chief, Clinical Pathology Division
Clinical Center
National Institutes of Health

W. Wayne Meinke, Ph.D.
Chief, Analytical Chemistry Division
Institute for Materials Research
National Bureau of Standards

Keith Weikel, Ph.D. (Chairman)
Office of the Assistant Secretary
for Planning and Evaluation
Department of Health, Education and
Welfare

Donald S. Young, Ph.D., M.D.
Chief, Clinical Chemistry Service
Clinical Center
National Institutes of Health

Appendix B. Reference Laboratories for Study of Clinical Laboratory Capability

Norwalk Hospital
ATTN: Dr. Roy N. Barnett
Norwalk, Connecticut 06556

California State Department of Health
ATTN: Dr. Howard L. Bodily
2151 Berkeley Way
Berkeley, California 94704

Center for Disease Control
ATTN: Dr. Joseph H. Boutwell
Atlanta, Georgia 30333

Hartford Hospital
ATTN: Dr. George W. Bowers
Hartford, Connecticut 06115

University of Oklahoma Medical Center
ATTN: Dr. Kurt Dubowski
P. O. Box 26901
Oklahoma City, Oklahoma 73104

University of Kansas Medical Center
ATTN: Dr. Russell Eilers
Rainbow Boulevard and 39th Street
Kansas City, Kansas 66103

University of Alabama Medical Center
ATTN: Dr. John W. Foft
619 South 19th Street
Birmingham, Alabama 35233

Columbia-Presbyterian Hospital
ATTN: Dr. S. Raymond Gambino
630 West 168th Street
New York City, New York 10032

National Institutes of Health
ATTN: Dr. Harvey R. Gralnick
Clinical Center
Bethesda, Maryland 20014

University of Washington Hospital
ATTN: Dr. Alex Kaplan
Chemistry Department
Seattle, Washington 98105

State Laboratory of Hygiene
ATTN: Dr. Ronald Laessig
437 Henry Mall
Madison, Wisconsin 53706

University of California
San Francisco Medical Center
ATTN: Dr. H. Loken
Clinical Laboratory
San Francisco, California 94122

National Institutes of Health
ATTN: Dr. James MacLowry
Microbiology Department
Clinical Center
Bethesda, Maryland 20014

Illinois Department of Health
Division of Laboratories
ATTN: Dr. Robert Martinek
Laboratory Evaluation Section
Chicago, Illinois 60612

Jackson Memorial Hospital
ATTN: Dr. Thomas Noto
1700 NW. 10th Avenue
Miami, Florida 33136

William Pepper Laboratory
ATTN: Dr. Royden Rand
3400 Spruce Street
Philadelphia, Pennsylvania 19104

Georgetown University Hospital
ATTN: Dr. Martin Rubin
3800 Reservoir Road, NW.
Washington, D.C. 20007

Yale-New Haven Hospital
ATTN: Dr. David Seligson
789 Howard Avenue
New Haven, Connecticut 06504

National Institutes of Health
ATTN: Dr. Donald Young
Clinical Chemistry Service
Bethesda, Maryland 20014

Appendix C. Differences in Laboratory Performance When Analyzing Dialyzed and Serum Pool Specimens

In Section 3.5 of the main report, the conclusions regarding method performance were considerably weakened by the inclusion of a caveat about possible nonrepresentativeness of results obtained with dialyzed test specimens. In this Appendix, the representativeness of such results will be investigated. The question of interest is whether some methods perform differently in analyses of dialyzed and serum pool specimens. To simplify the ensuing discussion, it is convenient to define the relative bias of the technique mean with respect to the target value as $(\bar{x} - T)/T$ where $\bar{x}$ is the technique mean and T is the target value. Using this definition, the hypothesis to be tested can be stated as follows: the relative bias of the technique mean of the dialyzed determinations with respect to the target value for the dialyzed specimen will not differ in a statistically significant manner from the relative bias of the serum pool technique mean with respect to the target value for the pooled serum specimen.

For each constituent, a t-test can be used to examine the statistical significance of the difference between the serum pool and dialyzed relative biases. However, there are two difficulties associated with defining the appropriate test statistic. The first difficulty arises because the levels of some constituents in the pooled serum specimens are very different from the levels in the dialyzed specimens. For example, the dialyzed normal bilirubin sample has more than twice the concentration of the serum pool normal sample. Since most laboratory methods for bilirubin are sensitive to changes in concentration, it would be unrealistic to expect all methods to have the same relative bias in analyzing both samples.

The second problem which arises is in the choice of a target value. Several possibilities exist: the mean laboratory analyses (MLA), the mean reference analyses (MRA), the mean analyses by the techniques which were best applied in the dialyzed analyses (MABAT), the manufacturer's average assays (MAA), and the weigh-back values (WBV). Two of these possibilities, the MRA's and WBV's, were immediately rejected because they were not available for the pooled serum specimens. There is no readily apparent rationale for a choice among the remaining possibilities. However, a criterion can be established which may reduce the number of possibilities: any target is unsatisfactory if its use leads to the conclusion that a substantial number of techniques performed differently in the first and second hematology shipments. This appears to be a logical criterion because the same process was used to prepare the two hematology batches. Thus, there should be no differences in the bias of technique means for the two batches if the target value is properly chosen.

Application of this criterion to the three available target choices was rather disconcerting. None of the targets satisfied the criterion. Table C1 shows the number and percentage of technique means (for a normal or abnormal analysis) with relative biases which differ significantly at the 90% level of probability for the two shipments of hematology specimens.

Table C1 Number and Percent of Techniques with Significantly Different Relative Biases for Various Targets

Target	# of Techniques	# of Biased Techniques	% of Biased Techniques
MLA	46	18	39
MABAT	56	18	30
MAA	46	28	61

All of these percentages exceeded 35%. Furthermore, a high rate of technique bias occurred in analysis of each constituent. Thus it would appear that the data from the two shipments were not sufficiently consistent to support the analysis initiated in this Appendix, and there is no choice but to label the results as inconclusive. For reference purposes, basic data on the serum pool specimen analysis reports is displayed in Table C2. The table contains spearate entries on sample size (n), mean ($\bar{x}$), and standard deviation (s) for each technique used by 15 or more study participants.

Table C2 Pooled Serum Specimen Analyses

Constituent	Technique	Normal Specimen			Abnormal Specimen		
		n	$\bar{x}$	s	n	$\bar{x}$	s
Glucose	Ferricyanide/AutoAnalyzer	32	78.4	6.05	30	241	13.5
	O-Toluidine/Manual	57	80.4	6.31	56	234	14.2
	Copper Neocuproine/SMA 12-60	38	81.9	9.76	39	247	12.2
Urea Nitrogen	Diacetyl Monoxime/Manual	39	10.6	1.75	39	48.1	5.07
	Diacetyl Monoxime/AutoAnalyzer	34	10.5	0.92	34	51.2	3.07
	Diacetyl Monoxime/SMA 12-60	43	10.3	0.53	43	51.1	1.49
	Berthelot Reaction/Manual	22	11.0	1.50	22	48.1	5.47
Calcium	Calcein Fluorimetric/Manual	15	8.2	0.73	15	12.1	1.00
	Cresolphthalein Complex/AutoAnalyzer	16	8.4	0.38	16	12.3	0.82
	Cresolphthalein Complex/SMA 12-60	49	8.1	0.23	50	12.3	0.45
	Other/Manual	30	8.5	0.93	29	12.4	1.11
Total Bilirubin	Diazo-Alcohol/Manual	70	0.4	0.21	71	4.7	0.83
	Diazo-Other/Manual	32	0.7	0.42	34	4.7	1.24
	Diazo-Other/SMA 12-60	47	0.4	0.12	48	4.9	0.54
Cholesterol	Lieber-Burchard-w/o Extr./Manual	65	174	11.1	66	173	14.6
	Lieber-Burchard-w/o Extr./SMA 12-30	40	172	6.5	41	171	8.4
Uric Acid	Phosphotungstate/Manual	75	5.3	0.48	77	9.1	0.89
	Phosphotungstate/AutoAnalyzer	26	5.3	0.34	26	9.3	0.76
	Phosphotungstate/SMA 12-30	48	5.6	0.23	49	9.4	0.32
	Phosphotungstate/SMA 12-60	17	5.4	0.53	17	8.9	1.16
Sodium	Flame Photometer/Manual	112	135	3.2	114	129	4.3
Total Protein	Biuret/Manual	71	6.2	0.34	70	5.3	0.33
	Biuret/AutoAnalyzer	30	6.0	0.27	30	5.2	0.26
	Biuret/SMA 12-30	45	6.1	0.17	44	5.2	0.15
	Refractometer/Manual	24	5.8	0.37	24	5.1	0.38

Appendix D. Medical Usefulness Plots

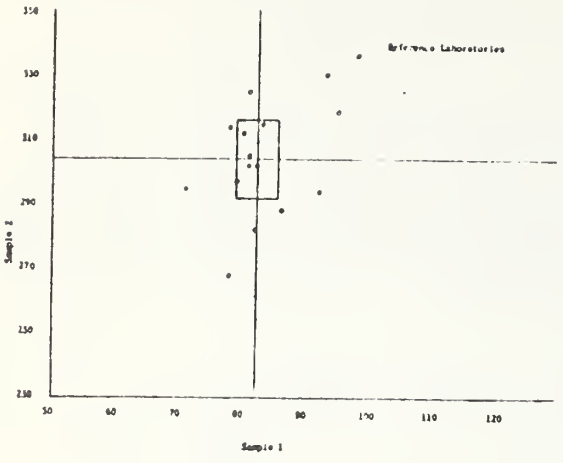
This Appendix contains the plots used in estimating the probability that a laboratory in any given group would "correctly" assay one of seven clinical chemistry constituents. Details on the estimation procedure are given in Section 3.6.2 (page 49).

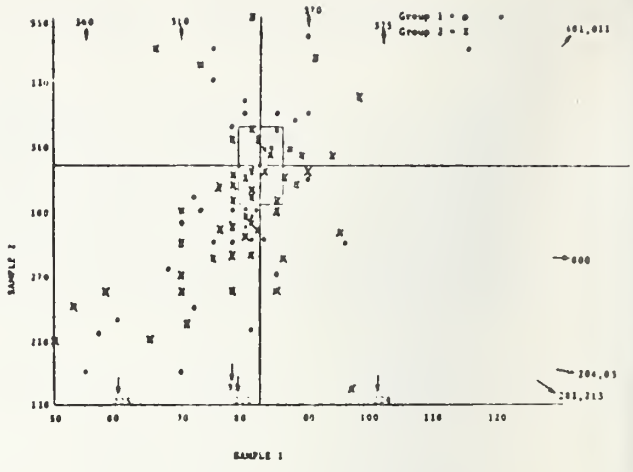
Each page displays plots for a single constituent. Table D1 gives a key to the group numbers on the plots and the corresponding group names used in the report. On each plot, sample 1 is the normal sample and sample 2 is the abnormal sample. Each point indicates the report of a single laboratory. Points beyond the scaled axes of the graphs are indicated in two manners. In cases where both the normal and abnormal values are off-scale, an arrow and two numbers are shown. The first number gives the value of the reported normal analysis and the second gives the value of the abnormal analysis. If only one value is off-scale, an arrow and a number appear. The tail of the arrow lies on the on-scale value and the head points in the direction of the off-scale value.

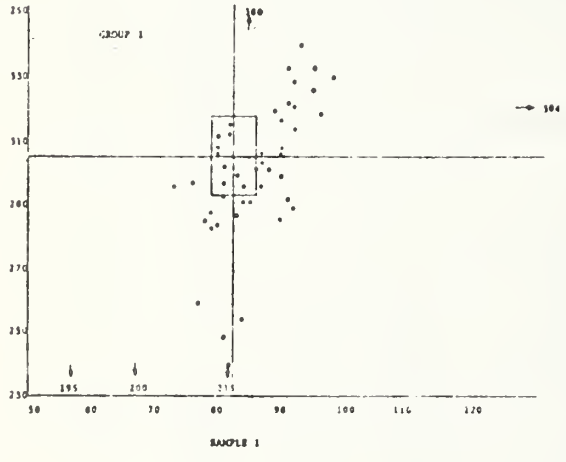
Table D1

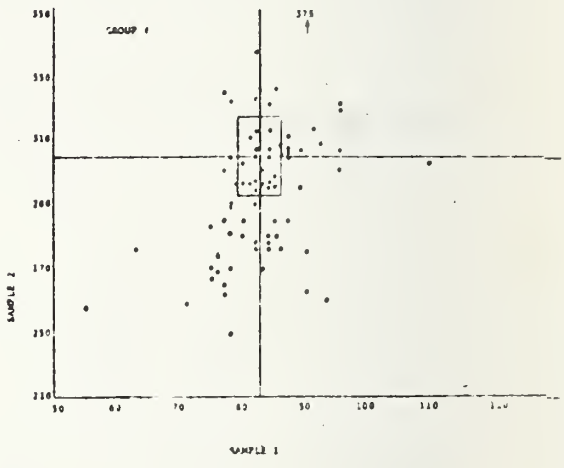
Group Number	Group Name
1	AAFP
2	ASIM
3	Interstate
4	JCAH
5	Medicare Hospitals
6	Medicare Independent Labs

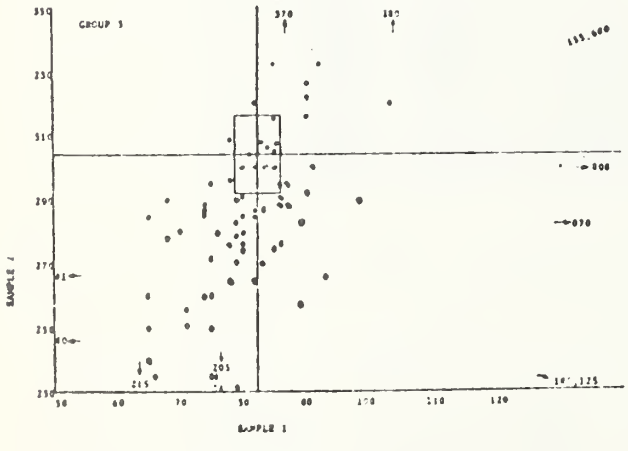
GLUCOSE
(mg./100 ml.)

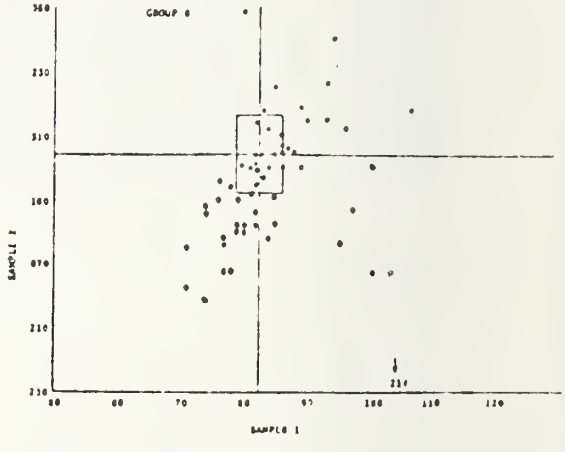








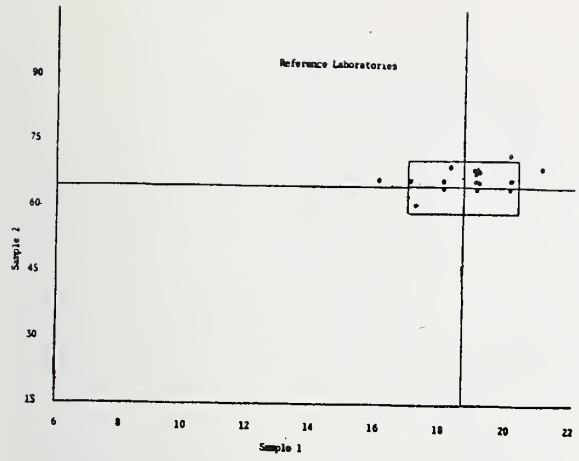


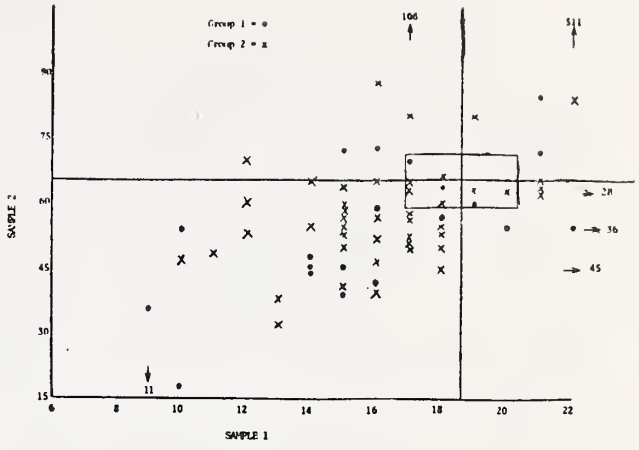


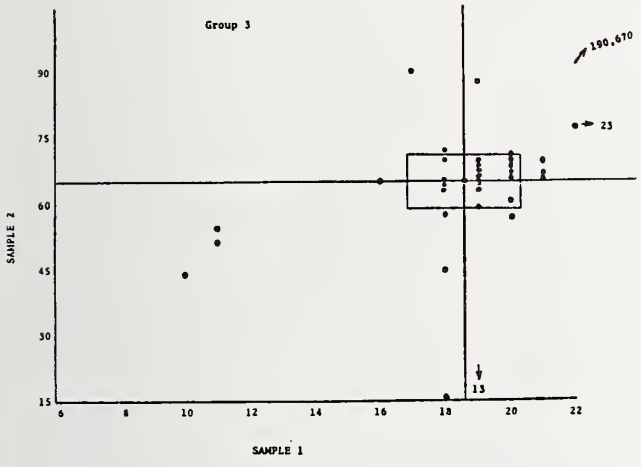
Requests for larger copies of these plots should be addressed to

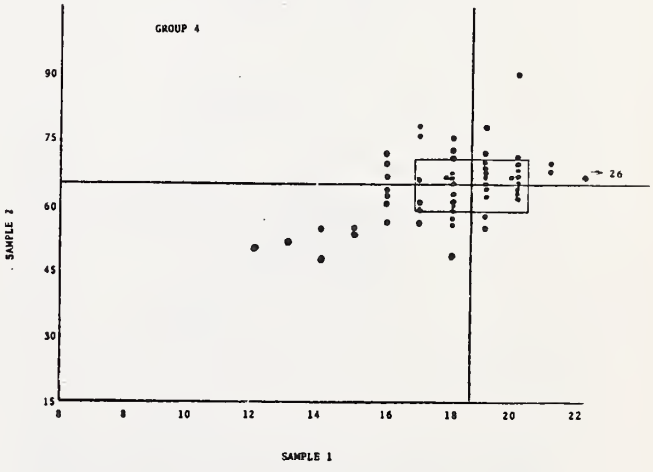
Public Policy and Program Analysis Area
Division 431.00, National Bureau of Standards
Building 220, Room A360
Washington, D. C. 20234

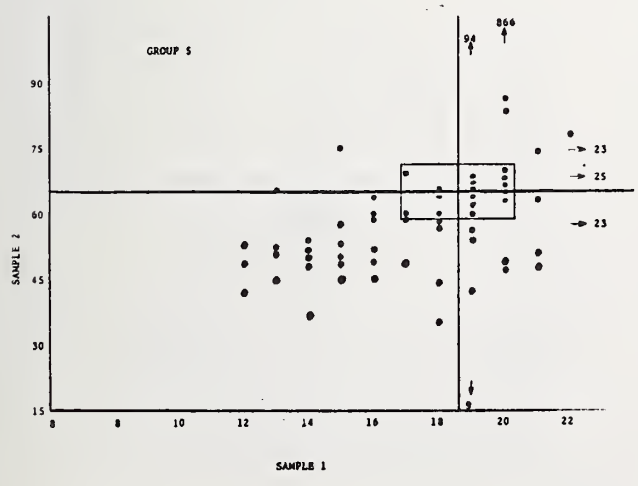
UREA NITROGEN
(mg./100 ml.)

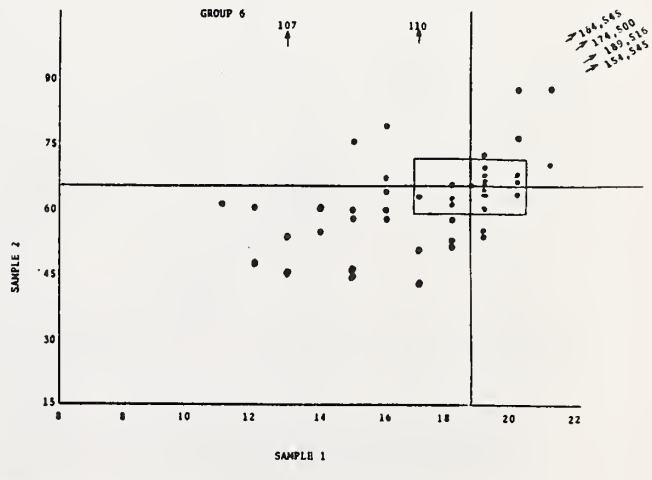




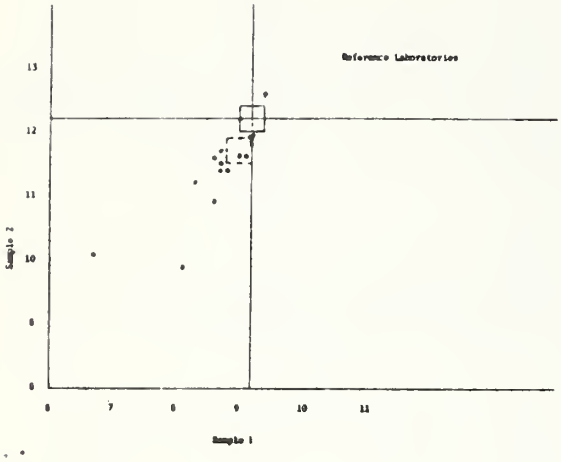


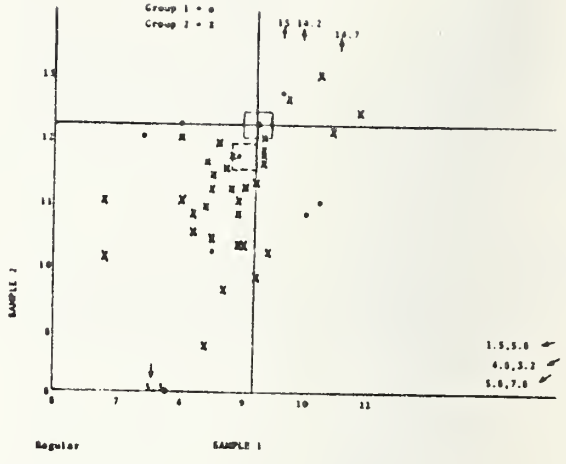


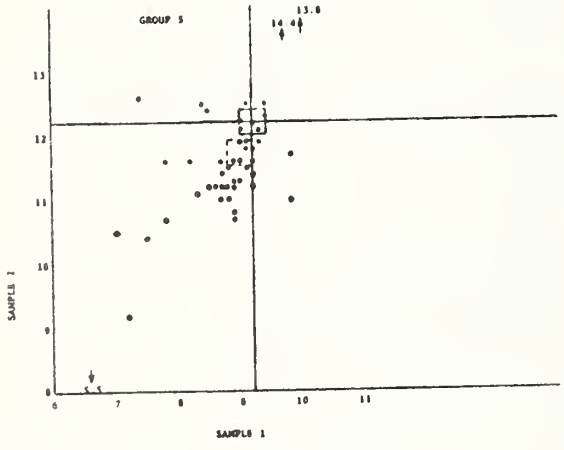


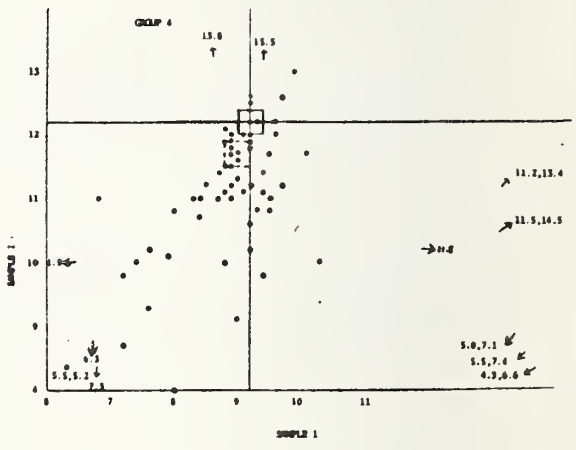


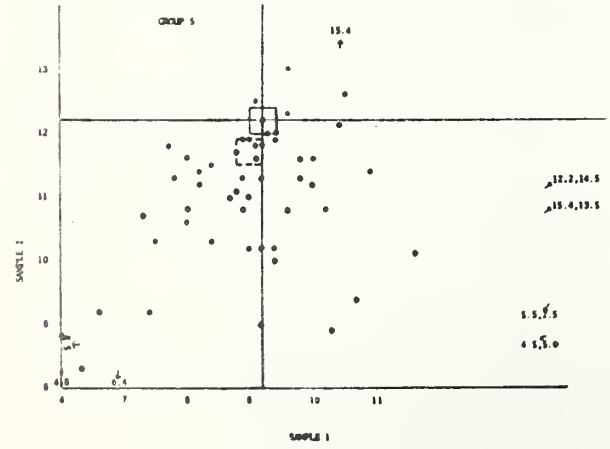
CALCIUM
(mg./100 ml.)

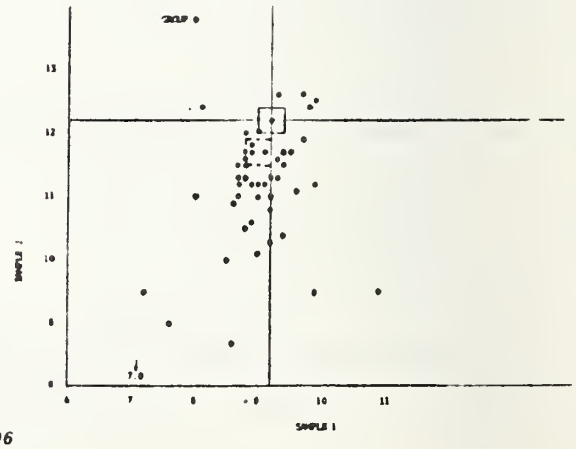




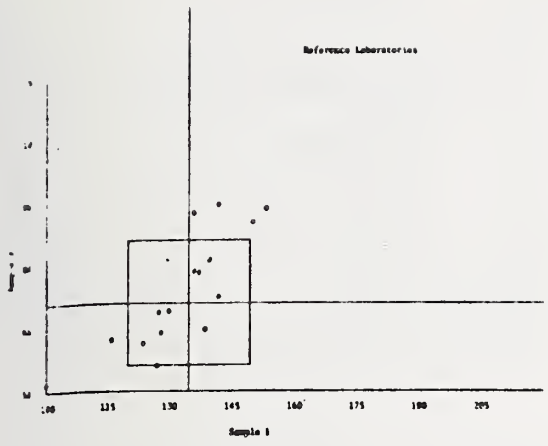


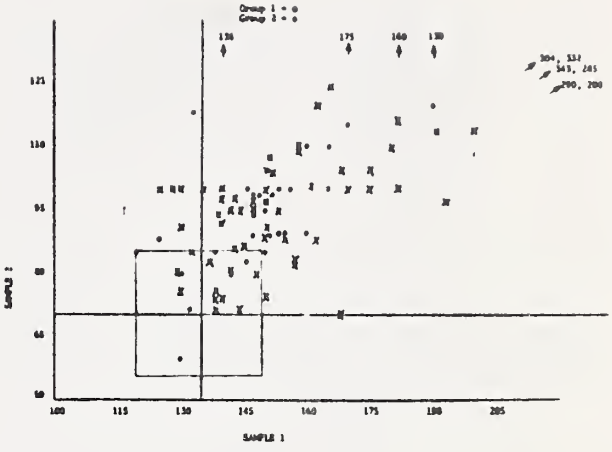


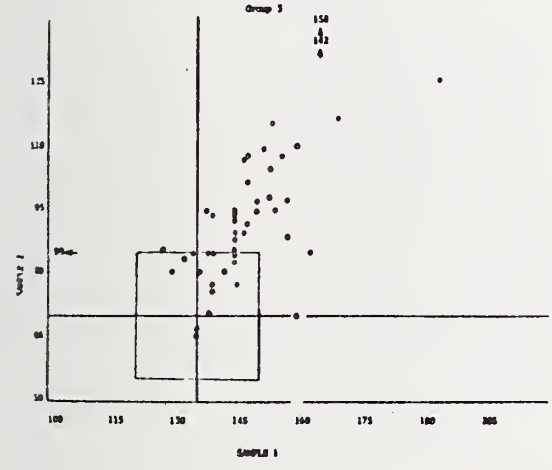


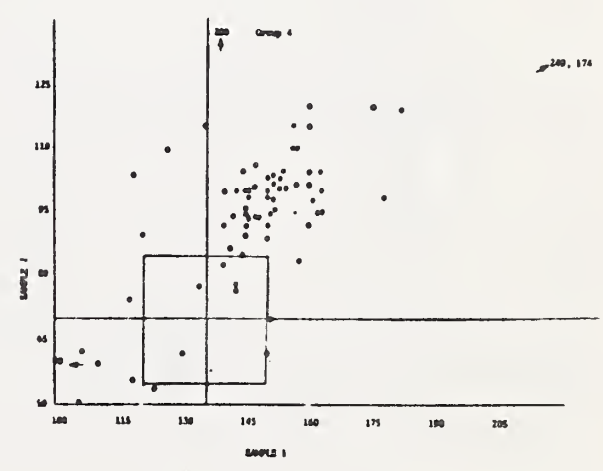


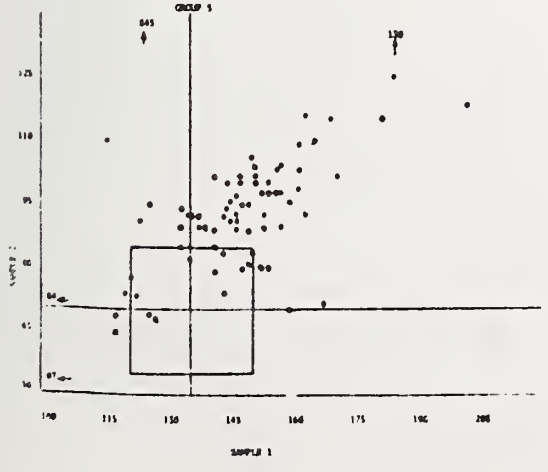
CHOLESTEROL

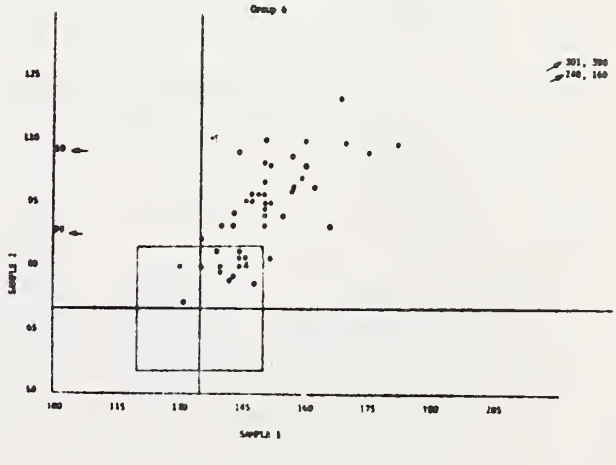




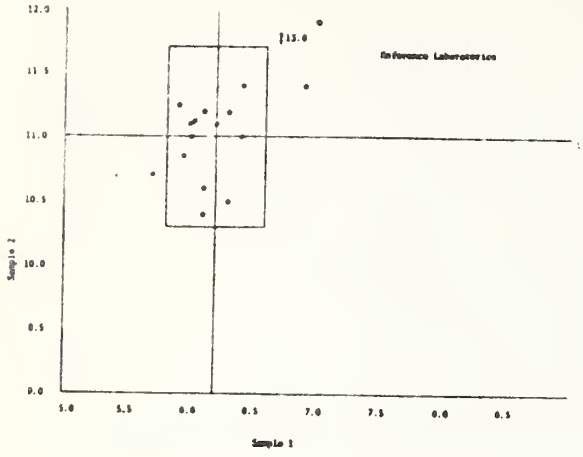


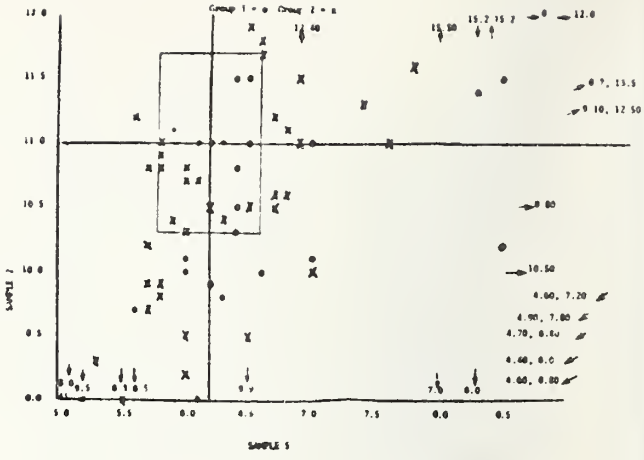


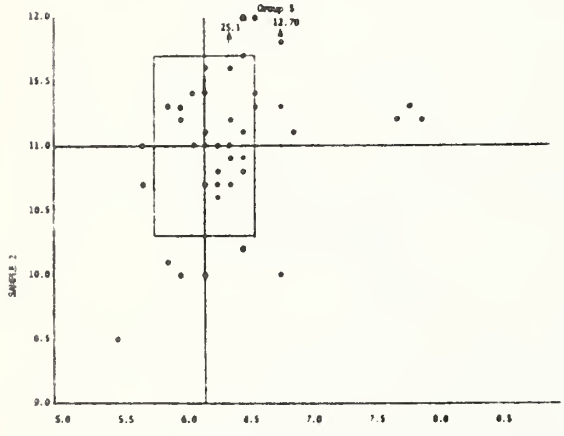


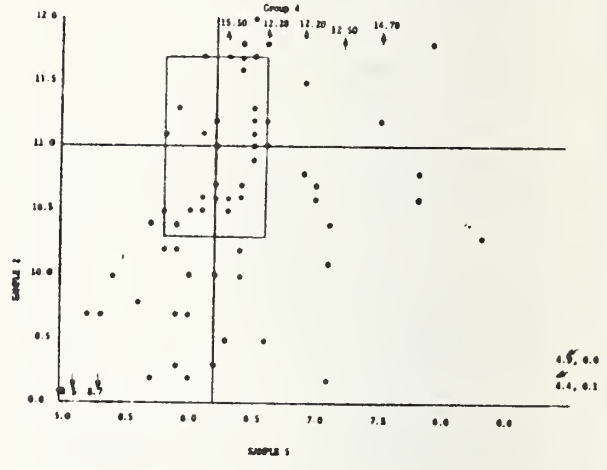


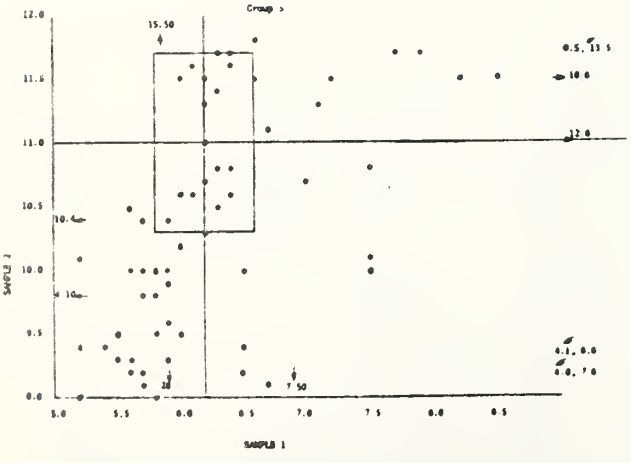
URIC ACID

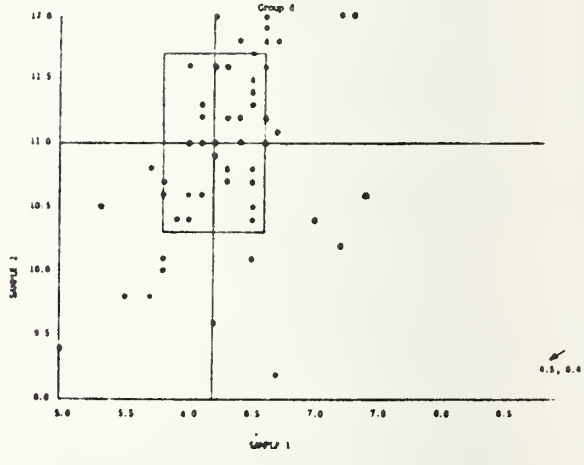

Reference Laboratories
Sample 2
Sample 1


Group 1 = o Group 2 = x
SAMPLE 2
SAMPLE 1

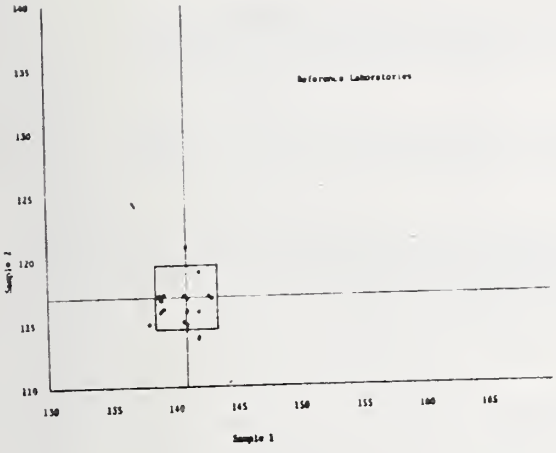

Group 5
SAMPLE 2

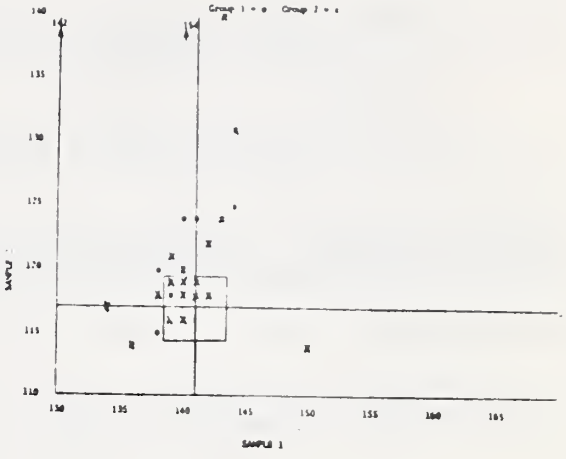

Group 4
SAMPLE 2
SAMPLE 1

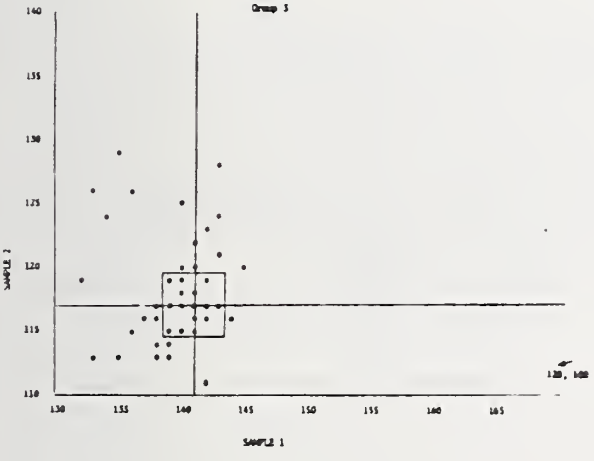

Group 3
SAMPLE 2
SAMPLE 1

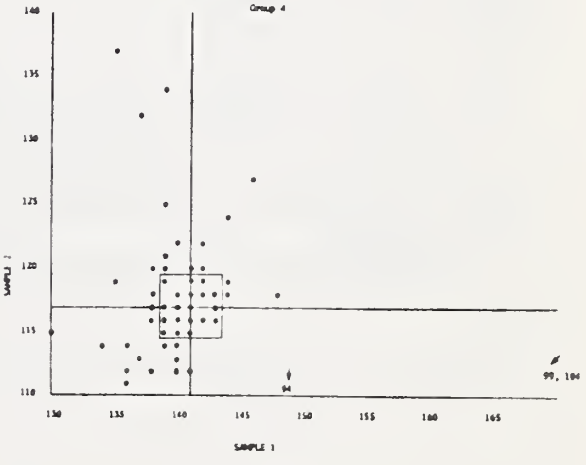

Group 6
SAMPLE 2
SAMPLE 1

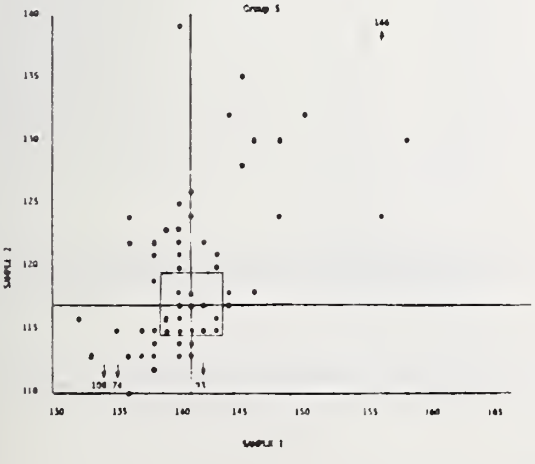
SODIUM

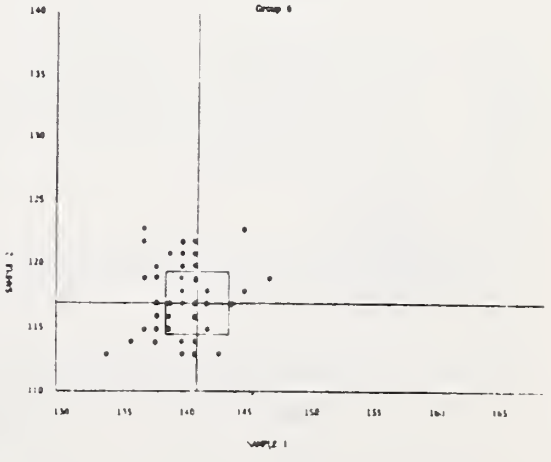

Reference Laboratories
Sample 2
Sample 1

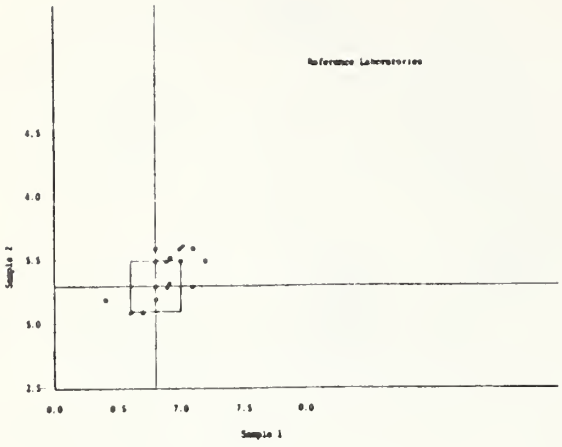

Group 1 = o Group 2 = x
SAMPLE 2
SAMPLE 1

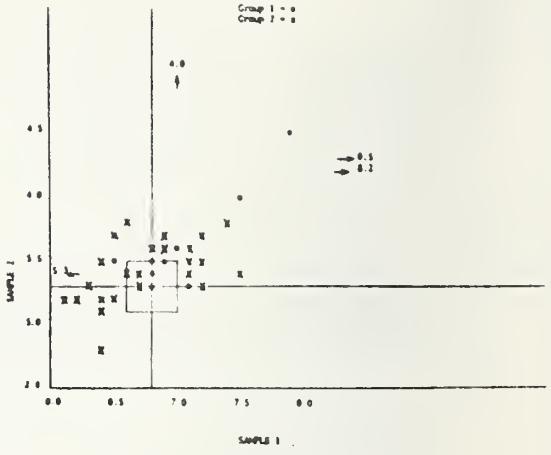

Group 3
SAMPLE 2
SAMPLE 1

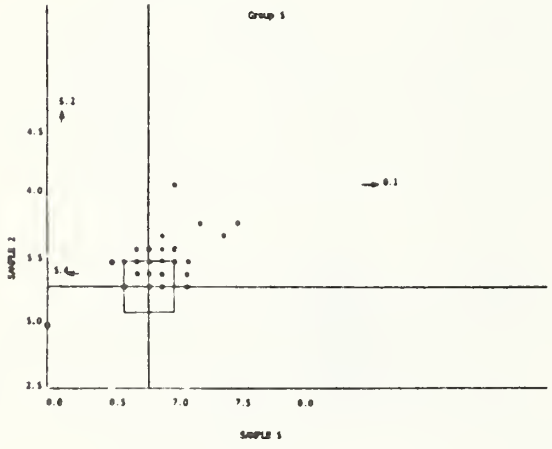

Group 4
SAMPLE 2
SAMPLE 1

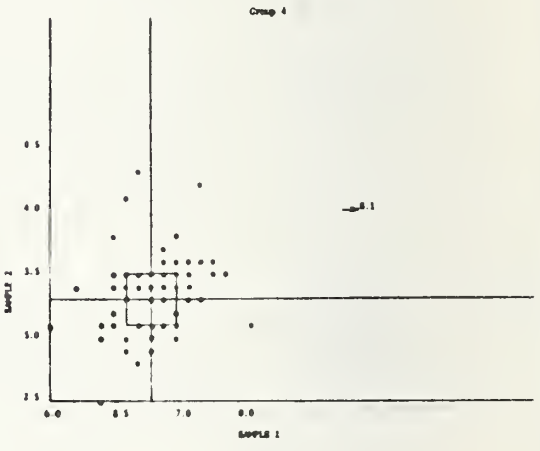

Group 5
SAMPLE 2
SAMPLE 1

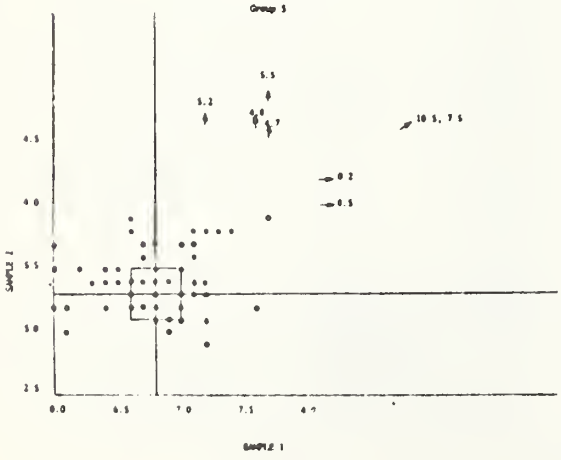

Group 6
SAMPLE 2
SAMPLE 1

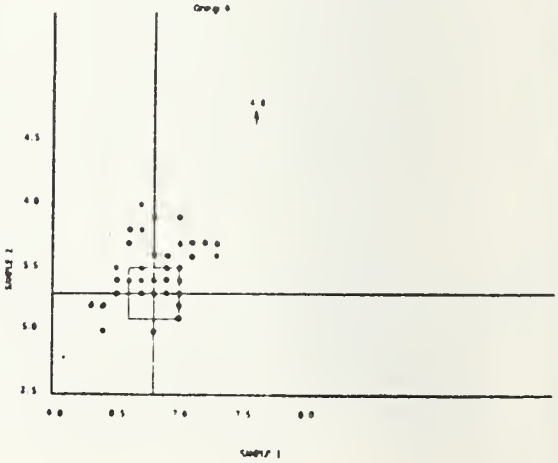

Reference Laboratories
Sample 2
Sample 1


Group 1 = a
Group 2 = b
SAMPLE 2
SAMPLE 1


Group 5
SAMPLE 2
SAMPLE 1


Group 4
SAMPLE 2
SAMPLE 1


Group 3
SAMPLE 2
SAMPLE 1


Group 6
SAMPLE 2
SAMPLE 1

Appendix E. References for Tables 3.8.1 and 4.2.1

1. CDC Analyses - P. Finkel and J. Rowen, Clinical Laboratory Performance Analysis Using Proficiency Test Statistics, NBS Report 10 685, 1972.

2. CAP 1949 and 1969 - L. Skendzel, D. Hanson, and W. Civin, "The 1949 College of American Pathologists Survey Revisited," American Journal of Clinical Pathology, Vol. 54, No. 3, September 1970, pp. 493-495.

3. AAB - various including American Association of Bio-Analysts, 1971 Programs - Proficiency Testing Service, 1970 and the AAB publication Bulletin, Vol. 13, No. 10, October 1970.

4. Canada - D. Tonks, "A Study of the Accuracy and Precision of Clinical Chemistry Determinations in 170 Canadian Laboratories," Clinical Chemistry, Vol. 9, No. 2, April 1963, pp. 217-235.

5. Scandinavia, Australia - J. Stromme and L. Eldjorn, "Surveys of the Routine Work of Clinical Chemistry Laboratories in 116 Scandinavian Hospitals," Scandinavian Journal of Clinical Laboratory Investigation, Vol. 25, 1970, pp. 213-222.

6. New Zealand - F. Desmond, "A Clinical Chemistry Proficiency Survey," New Zealand Medical Journal, Vol. 63, 1964, p. 716.

7. LA Physicians - I. Schoen, G. Thomas, and S. Lange, "The Quality of Performance in Physicians' Office Laboratories," American Journal of Clinical Pathology, Vol. 55, No. 2, February 1971, pp. 163-170.

8. Italy - R. Saracci, "Factors Affecting Accuracy and Precision in Clinical Chemistry," American Journal of Clinical Pathology, Vol. 52, No. 2, August 1969, pp. 161-166.

9. CAP 1965 - J. Koepke and J. Eilers, "Survey of Red Cell Parameters in 1600 United States Hospitals," Bibliotheca Haematologica, Vol. 24, 1966, pp. 137-144.

10. American Society of Clinical Pathologists - W. Sunderman, "Status of Clinical Hemoglobinometry in the United States," American Journal of Clinical Pathology, Vol. 43, No. 1, January 1965, pp. 9-15.

11. CAP 1972 - R Gilbert, The 1971 Chemistry Survey Program, College of American Pathologists, 1972.

☆ U.S. GOVERNMENT PRINTING OFFICE 1973—542-652/98

FORM NBS-114A (1-71)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBSIR 73-163	2. Gov't Accession No.	3. Recipient's Accession No.

4. TITLE AND SUBTITLE

A Proficiency Test Assessment of Clinical Laboratory Capability in the United States

5. Publication Date

6. Performing Organization Code
NBSIR 73-163

7. AUTHOR(S)
Peter W. Finkel and Ted R. Miller

8. Performing Organization
/

9. PERFORMING ORGANIZATION NAME AND ADDRESS

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

10. Project/Task/Work Unit No.
4314457

11. Contract/Grant No.

12. Sponsoring Organization Name and Address
Division of Health Evaluation
Office of the Assistant Secretary for Planning, and Evaluation
Department of Health, Education, and Welfare
Room 5526 HEW North Building, Washington, D. C. 20201

13. Type of Report & Period Covered
Final 3/71 - 4/73

14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

16. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)

The proficiency of a selected sample of physician, hospital and independent laboratories was assessed with respect to their ability to analyze clinical chemistry and hematology samples and to identify microbiological organisms. For the assessment of clinical chemistry and hematology proficiency, the laboratories were grouped and determinations of group accuracy and group precision were made. Further analyses were performed to determine relative accuracy and precision of the techniques presently applied by these groups. There was no significant difference at the 95% confidence level in the accuracy achieved by the various laboratory groups involved in clinical chemistry and hematology analysis. In clinical chemistry, the Medicare-Certified Independent Laboratories, CDC Tested Laboratories and JCAH-Members generally proved more precise than Physicians' Office and Medicare-Certified Hospital Laboratories. However, none of the laboratory groups were sufficiently accurate to permit the monitoring over time of variation in an individual patient's constituent concentrations. It would appear that poor selection of techniques was an important contributor to this low performance level. In hematology the Physicians' Office Laboratories proved to be the least precise of the groups. There was no noticeable difference in precision between participants in the CDC proficiency testing program and non-participants. With respect to microbiology, 7.6% of the identifications by laboratories participating in the CDC testing program were incorrect, while 19.4% of all other identifications were incorrect.

17. KEY WORDS (Alphabetical order, separated by semicolons)
Accuracy; clinical chemistry; hematology; medical usefulness; microbiology; proficiency testing

18. AVAILABILITY STATEMENT

[XX] UNLIMITED.

[ ] FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NTIS.

19. SECURITY CLASS (THIS REPORT)
UNCLASSIFIED

21. NO. OF PAGES

20. SECURITY CLASS (THIS PAGE)
UNCLASSIFIED

22. Price

USCOMM-DC 66244-P71