REQUIREMENTS FOR ANALYTICAL PERFORMANCE IN CLINICAL CHEMISTRY

AN EVALUATION FROM THE POINT OF VIEW OF THE PRACTICING PHYSICIAN

REQUIREMENTS FOR ANALYTICAL PERFORMANCE IN CLINICAL CHEMISTRY

AN EVALUATION FROM THE POINT OF VIEW OF THE PRACTICING PHYSICIAN

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE GENEESKUNDE AAN DE ERASMUS UNIVERSITEIT TE ROTTERDAM, OP GEZAG VAN DE RECTOR MAGNIFICUS PROF. DR. B. LEIJNSE EN VOLGENS BESLUIT VAN HET COLLEGE VAN DEKANEN. DE OPENBARE VERDEDIGING ZAL PLAATSVINDEN OP WOENSDAG 11 JANUARI 1978, DES NAMIDDAGS TE 4.15 UUR PRECIES

DOOR

WIVEKA E. ELION–GERRITZEN

GEBOREN TE EDE, 11 JANUARI 1938

1978

DRUKKERIJ J.H. PASMANS, 'S-GRAVENHAGE

PROMOTOR:	Prof. Dr. B. LEIJNSE
CO-REFERENTEN:	B.E. COPELAND, associate professor R. VAN STRIK, lector J.H.P. WILSON, lector

Aan vader en moeder,
Frans, Wiveka, Jan Willem en Christine

"Stien, wat wil je worden later?"

"Moeder, net als jij."

TABLE OF CONTENTS

Chapter I

INTRODUCTION

In today's medicine many decisions are based on data from the clinical chemistry laboratory.

Basically laboratory results for a specific constituent serve a dual purpose (35, 65):

a. to recognize the diseased individual by relating the value found to either the range of values observed in a healthy population or to a range of values known to be compatible with a specific condition.
b. to monitor changes in an individual patient.

Consequently measurements in clinical chemistry have to be reliable in two ways. Firstly, they have to be accurate, which is defined (1,34a) as close to the true value or an accepted reference value and secondly, they have to be precise, which means, that replicate test results have to agree closely (1,34a). The question that has often been discussed but hitherto not been answered satisfactory is: how accurate and how precise do measurements in clinical chemistry have to be?

In general in technical disciplines, requirements for accuracy and precision of measurements are set by the consumer needs. For example in steel construction the consequences of measurement errors are carefully considered. If the length of beams is incorrect or shows variations, all or some will not fit. If the actual shapes are smaller than designed shapes or if the strength of the material is insufficient the construction may collapse. If the strength/weight ratio is low the beams will be too expensive to meet the budget. Technical and economical consequences of each deviation from the ideal or required dimensions can be calculated, and, not unimportant, instruments used for measurements are adequate to register these deviations.

In clinical chemistry consumer needs are more difficult to define.

Firstly, ideal or required values are difficult to establish:

- biological differences between individuals exist and apparent identical physiological conditions occur together with constituent levels which differ from individual to individual
- we have to deal with physiological variations within the individual patient. The constituent level in a person may change due to dietary conditions, diurnal variations, body position etc., which are without significance for the condition of the patient or for the conclusion to be drawn from the measurement
- the clinical chemistry result is only part of a series of observations, made by a physician during examination of a patient. In the process of medical decision making the degree of importance of the clinical chemical test result as compared to other data obtained for the same patient varies from case to case.

Secondly, clinical chemistry measurements are seldom carried out with impeccable accuracy and precision. Routine analytical variability or rather variability in the

total process from sample taking to the final analytical result is for many constituents not negligible as compared to either intra-or interindividual biological variations or clinically important differences.

There are a number of chemical, physiological, practical and economical factors interfering with the estimation of in vivo levels of blood constituents, e.g.:

- blood serum levels to be measured may change during sample taking and clotting due to chemical conversion or exchange with intracellular compartments
- the method used for measurement may not be specific and, for example overestimate the constituent by measuring related compounds
- the method used may be sensitive to substances interfering with the chemical reactions
- the technician may not be familiar with proper handling of technical devices or instruments
- the standards and/or reagents used may contain impurities leading to erroneous results, instruments may not function properly.

These factors lead to inaccuracy if their effects are constant from sample to sample, from day to day, from technician to technician or from lot to lot of reagent chemicals. The same factors will lead to imprecision, if they are variable between samples, between days, between technicians or between lots.

This discussion shows why it is difficult to define how accurate and how precise clinical chemistry measurements should be.

Most attempts to attack the question have been made from the theoretical point of view. They are based on the assumption, that the analytical variability should not exceed a certain proportion of the human biological variation.

Tonks (61a,b) states that allowable limits of error should not exceed 1/4 of the normal range, corresponding with a ratio of about 2 to 1 of the biological variation to analytical variation. Zwart Voorspuij and Van der Slik (77) recommend a ratio of "3-5: 1 between physiological variation and analytical variation". Vanko (62) reports that the Center of Disease Control (CDC) uses the 4:1 ratio in its proficiency programs, this being a compromise between the 3:1 to 5:1 ratio suggested by Barnett (4).

Campbell and Owen (9) and Barnett (4) were the first to deal with the problem from the clinical point of view. Campbell and Owen report "acceptable analytical limits in the view of a number of clinicians" and Barnett gives tables of "medically significant values, synthesized from opinions of clinicians and laboratory specialists". More recently Gilbert (27) derived "analytical goals" from Barnett's data.

Cotlove and coworkers (14), in their hallmark paper on intraindividual variation, stressed that tolerable analytical variability should be based on variation within the individual rather than on biological variation between individuals or "on judgement derived from experience with test results that incorporate undetermined degrees of analytical variation". Cotlove's approach has recently been used by Steele c.s. (58) to arrive at requirements for analytical performance.

The IFCC (International Federation of Clinical Chemistry) Expert Panel on Nomenclature and Principles of Quality Control in Clinical Chemistry has issued a provisional recommendation on "Quality requirements from the point of view of health care" (34b) and it is clearly stated that

> "consumer needs – i.e. health care requirements as determined by benefit to patients, clinical practice and cost to the community – must be taken into account, in order to avoid wrong management decisions which might result from reliance on internal laboratory criteria alone; for example, an analytical method may be the best available for a given component, but it may still be not good enough for clinical application; conversely, a method may be more sensitive, specific or costly than justifiable by its use in a particular clinical situation".

We conclude that a close interaction between laboratory and clinician in the field of quality control is indicated to arrive at sensible requirements for analytical reliability.

The difficulties to be encountered in setting these requirements from the clinical point of view are related to (a) the inherent analytical variability of various determinations and (b) the well documented differences among physicians in criteria used in medical decision making (7, 11, 16, 22, 53, 64, 74). Elsom and coworkers (22) report on the "extraordinary independence of objective data displayed by more than 300 physicians in arriving at a diagnosis" and Clark et. al. (11) notes: "It is impossible to determine how an examiner arrived at a diagnosis and how he interpreted laboratory reports". This may well be a too pessimistic picture of the situation since for example, none of these studies took into account the effect analytical variability may have on the recorded differences in clinicians' behaviour towards laboratory results.

The aim of the study described in this report was to get an impression of the present situation: how do physicians use laboratory results and does analytical variability affect medical decision making. In particular we tried to get an answer to the question for which determinations analytical performance is – from the clinical point of view – good, adequate or elegible for improvement respectively.

A questionnaire was designed dealing with various aspects of medical decision making. Specialists in internal medicine in different hospitals were interviewed in person, while simultaneously relevant data from the hospital laboratory were collected. In this report the following subjects will be discussed:

- normal range[1]) values given by the laboratory will be related to those applied by physicians and to "action levels" given by them. Action level is defined as the upper or lower limit of a constituent level that would prompt to an action of the first order (repeat or additional tests, E.C.G, X-ray, change of diet or medication) in the specific situation of the out-patient with ill-defined complaints. The occurance of systematic differences between laboratories can be considered at

[1]) It has been suggested (18,31) that the term "reference values" or "reference interval" should be used instead of "normal range", for the range of values, observed in healthy individuals. The author prefers the latter term at present, for reasons which will be given in chapter III.

the same time, since all laboratories tested samples of the same quality control sera [1]) multiple times.
The difference between limit of normal and respective action level will be considered a measure of the strictness the physician applies at that concentration level for that constituent.

- a medically significant change in an individual patient will be related to analytical variability observed in this study
- in addition clinicians' views on the following points will be reported:
  a. factors affecting variation in results (physiological variation, laboratory error, sample handling)
  b. satisfaction/dissatisfaction with laboratory performance
  c. request of repeat tests and consulting the laboratory in case of an unlikely result
  d. the speed of availability of an analytical result versus its reliability.
  e. the use of molecular units in medical practice
  f. interpretation of results of another laboratory.

[1]) Quality control sera are large pools of serum, subdivided into small portions and stored either frozen or in lyophilized form so to guarantee a minimum change in concentration of the constituents over a stated period of time.
For Hb analyses special liquid quality control samples were used.

Chapter II

DESIGN OF THE PROJECT AND PROCEDURE

2.1 Design of the project

2.1.1 The inquiry

A questionnaire was set up in cooperation with the departments of psychology (Prof. Dr. F. Verhage) and biostatistics (Dr. R. van Strik) of the Erasmus University. After a pilot study including interviews with 4 clinicians minor changes were made in the questionnaire.
Specialists in internal medicine were considered to be – quantitatively and qualitatively – the most demanding clinicians with respect to laboratory tests. A selection was made of senior internists in University hospitals or teaching hospitals affiliated with University hospitals. In the Netherlands there are 30 hospitals with postgraduate training courses in internal medicine ("Interne A opleiding"). The interviewer had the opportunity to interview ten clinicians in three hospitals outside the Netherlands. The following table gives the interviewers by hospital:

type of hospital	country	number of hospitals	number of clinicians interviewed
University hospital	Netherlands	4	33
„	Switzerland	1	3
„	United States	1	4
„	Canada	1	3
Teaching hospital	Netherlands	5	20
		12	63

In the university hospitals in the Netherlands 8 or 9 randomly choosen associate professors and "chefs de clinique" were interviewed. In teaching hospitals all senior specialists in internal medicine working in that hospital were questioned. Four out of 67 clinicians turned down the request for participation.
The age distribution of the participants is given in the following table:

years after MD degree:	0-5	5-10	10-15	15-20	20-25	25-30	30-35
number of clinicians:	1	14	16	14	7	5	2

2.1.2 The survey

The laboratory survey was set up as follows.
Each laboratory tested 20 vials, 10 of a normal and 10 of a pathological quality

control serum in pairs on separate days over a period of 2-5 weeks. The analyses under study were among the most commonly requested tests: sodium (Na), potassium (K), chloride (Cl), calcium (Ca), inorganic phosphorous (P), urea, creatinine (creat.), glucose (gluc.), cholesterol (chol.), total protein (t. prot.), alkaline phosphatase (alk. P-ase, ALP) and lactate dehydrogenase (LDH).
Also 20 vials of hemoglobin (Hb) control material, 10 at each of two levels, were tested.
The lyophilized control sera formed part of the normal and abnormal pools (1974) of the Massachusetts Society of Pathologists regional Quality Control Program (12). Due to an error one laboratory tested samples of the 1975 pools.
No target values for the quality control sera were known to participants. The Hb controls had assigned values printed on package and vial by the manufacturer. For the ten results for each controlmaterial mean, standard deviation (SD) and coefficient of variation (CV) were calculated. Figure 2-1 gives an example of the form used for reporting the results of the survey to laboratory directors: laboratory identification number, constituent tested, mean, SD and CV are listed. In a two-sample plot the individual values and mean value are accumulated. In addendum I it is demonstrated how this cumulative two-sample plot – a modification of the two-sample plot introduced by Skendzel and Youden (51) and Tonks (61a) – can provide valuable information on intra- and interlaboratory variability.

Laboratory directors filled out a questionnaire on normal values:

	Na	K	Cl	Ca	P	urea etc.
Normal values						
Units						

Are the normal values*
- derived from the literature or instrument or kit manual
- determined in your laboratory
- an estimate

If normal values were determined in your laboratory, please answer the following questions:

When were they determined*
- less than 1 year ago
- between 1 and 3 years ago
- more than 3 years ago

Number of persons ♀
♂

Age range ♀
♂

Population*
- blooddonors
- students
- hospital personel
- out-patients
- other

*: please place x at correct answer

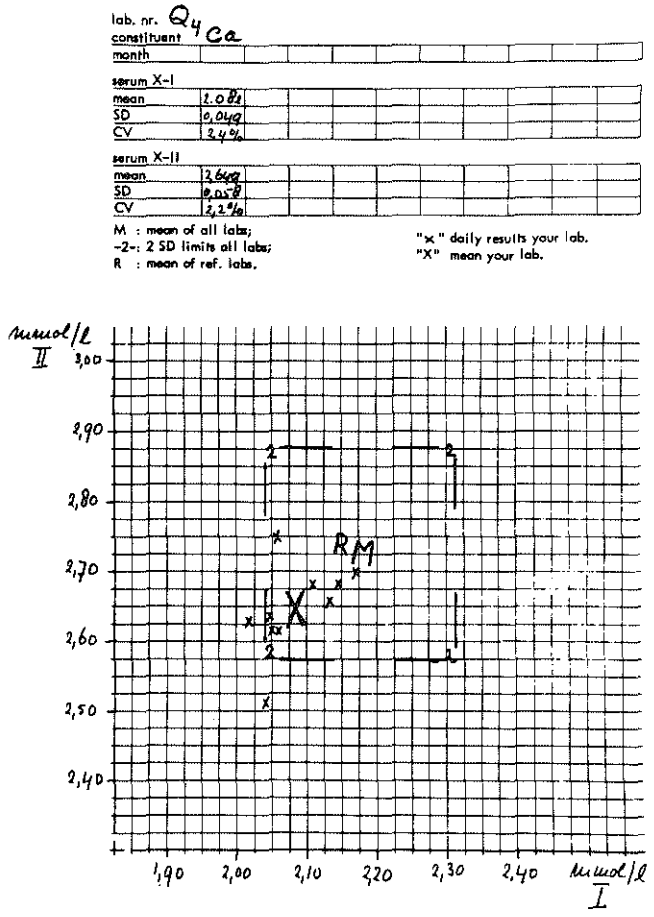


Figure 2-1:

Example of form used for reporting survey results to laboratory directors. Explanation of symbols used in the two-sample plot:

M : mean of all laboratories' mean values in the Massachusetts Regional Quality Control Program
-2- : 2 SD limits of all laboratories' mean values in that program
x : individual results laboratory 4.
X : mean value laboratory 4.

Details of this way of reporting are given in the addendum I.

2.2 **Procedure**

First, an appointment was made by telephone with the director of the hospital laboratory. In a 1-2 hour meeting the organizational plan of the project was explained

and cooperation was asked with the survey. It was made clear that the project would be anonymous, i.c. that no information from the laboratory would be reported to clinicians nor vice versa.
Next, the following letter was written to the clinicians to be interviewed:

> Dear Doctor,
>
> I would like to invite your attention for an inquiry, that has been set up by me in the course of my doctorate studies.
> It is the aim of this project to get an impression of the way the clinician uses laboratory results in clinical diagnosis and the follow-up of patients.
> In consultation with Prof. Dr. B. Leijnse, department of Chemical Pathology, Erasmus University, Rotterdam a questionnaire has been composed, that will be submitted to a number of medical specialists.
> I would like to ask for your cooperation with this project by answering the questions in person. Next week I will contact you by telephone and on that occasion I would like to make an appointment for an interview, that will take about 35 minutes.
> Your cooperation will be greatly appreciated.

After a week an appointment was made by telephone. The appointments with the clinicians in Switzerland, the United States and Canada were in part made through the laboratory director.
The interviews lasted 25-40 minutes, eleven exceeded 40 minutes.
If lack of time was impending question 1c of the questionnaire (see p. 15) on constituent levels prompting to an action of the second order (hospital admission, immediate treatment) and the details on desired degree of precision asked in question IVb and VIb were left out.

The introduction to the interview and the questionnaire read as follows:

"In clinical chemistry much attention has been paid to quality control over the last years. Attempts are being made to improve day-to-day reproducibility in the laboratory and comparability between laboratories.
One question actually has remained unanswered: which accuracy and precision are required from the laboratory to render good service to the clinician? In the clinical chemistry world an answer to this question has been looked for and it has been proposed, that analytical variability should not exceed ¼ of the biological variability. If that criterium is met, one should be able – from the statistical point of view – to distinguish the diseased individual from the healthy population. This is a theoretical approach; when we extrapolate this line of reasoning to the – often small – biological variation within one individual, which is frequently a matter of concern in the practical situation, we arrive at a requirement for example for Na, that reproducibility should be better than ¼% (1 CV). Then the question arises, how many physicians would appreciate this precision. Of course there is also an economical side to the problem.
It is difficult to ask the clinician how accurate and how precise the laboratory should be. His impressions are based on the accuracy he gets from his laboratory. Yet it does seem pertinent to analyse the present situation: when does a clinician react to an abnormal laboratory result, when does he consider a change in a patient medically significant. The tests we will discuss in the first instance are some of the most commonly requested: Na, K, Cl, Ca, P, urea, creatinine, glucose, cholesterol, total protein, Alk. P-ase, LDH and Hb.
Simultaneously we get information from your laboratory on these tests, but no information will be exchanged.
Their data will not be passed on to you, nor will any of your information be relayed to them.

Ia. Which normal values do you apply? Na, K, Cl, Ca etc.
If you apply different normal values for men and women or for people in different age groups, please identify.

b. Are the values mentioned under a.:
- derived from the literature? yes/no
  if so, from which textbook or publication?
- concluded from own personal observation? yes/no
- based on the normal values given by the laboratory? yes/no
- other

c. Suppose you find no abnormalities during a physical examination of a patient with ill-defined complaints. You order a number of laboratory tests.
Please identify for each constituent a lower and an upper limit which would prompt you to action of the first or second order. An action of the first order (Action 1) does include: repeated or additional laboratory tests, ECG, X-ray, initiation or change of diet or medication. An action of the second order (Action II) is defined as: hospital admission, immediate treatment. Na, K, Cl, Ca etc.

II Suppose you find in a patient under your treatment the laboratory result given in the table. You follow the course of this patient and/or the effect of treatment. Peace circle the value, that would represent a significant change – in these cases improvement – (See section 5.1 for details). Na, K, Cl, Ca, etc.

III There are a number of factors responsible for the day-to-day variability of laboratory data. We have the biological variation in the patient, laboratory error, handling of the sample (e.g. technique of sample taking or storage) and possibly other factors. Can you indicate to which extent these factors contribute to this variability: very great: ++, great: +, moderate: ± or none –.
- biological variation in the patient ++/+/±/–
- laboratory error ++/+/±/–
- handling of the sample ++/+/±/–
- other ++/+/±/–

IVa For which constituents do you feel a greater accuracy of the laboratory is needed: in other words, are there laboratory tests for which you would say: I could do better work, if the laboratory did a better job?

b Please identify the degree of accuracy desired (Sate two values, between which you want to distinguish).

Va Do you ever order a repeat test – on a new sample – before making a medical decision always/often/sometimes/never

b Suppose the value reported for a repeat test is substantially different from the first time. How do you proceed?
- you order the test for a third time always/often/sometimes/never
- you consult with your laboratory always/often/sometimes/never
- you disregard the value that is least likely in the given situation always/often/sometimes/never

VI Are there blood constituents – possibly others than the ones named in the table – for which you would prefer a fast semiquantitative result (within an hour) over an accurate value on longer terms. Please state the degree of accuracy desired (two values between which a distinction is to be made.

VII Suppose a patient is referred to you, while laboratory tests data have been determined in another laboratory than your own. How do you proceed?
- you evaluate the results with the normal values you usually apply in mind yes/no
- you ask for a repeat of all tests yes/no
- you ask for a repeat of clinically relevant tests only yes/no
- you relate the data to the normal values of the laboratory in question yes/no
- if these normal values are not directly available, do you inquire? yes/no

VIIIa Suppose a patient is referred to you from another hospital. In the table the laboratory value reported for this patient is given. You ask for a repeat test in your own laboratory and find the other value given in the table (see section 7.1). Na, K, Cl, Ca etc.
Is there a change in the condition of the patient? yes/no

b What would be your answer if the two values were results for one patient both determined in your own laboratory.

IXa a. Does your laboratory report in molecular units? yes/no

b. If so, since when? 1970/1971/1972/1973/1974/1975

c. How long did it take you to get used to the new units?
<3 months/3-12 months/> 1 year

d. Have molecular units enabled you to give better treatment to patients?
yes/somewhat/no

e. Have molecular units given you more insight into biochemical processes?
yes/somewhat/no

f. Do you convert mmol/l into mg% before making a medical decision?
always/often/sometimes/never

g. Do you convert mg% e.g. from the literature into mmol/l for optimal interpretation?
always/often/sometimes/never

Chapter III

THE LABORATORY SURVEY AND THE INQUIRY INTO THE LABORATORIES' NORMAL RANGE VALUES

3.1 Introduction

In order to collect information on intra- and interlaboratory variability of analyses each laboratory tested quality control materials at two different levels on ten different days. Laboratory directors gave information on normal range levels applied in their laboratories. Details of the survey and inquiry are given in chapter II.

3.2 Results

3.2.1 The survey

In table 3-1 the average results ($\bar{X}$) found by each laboratory (numbered 1 through 12) for quality control (q.c.) material I and II, the standard deviation (SD) and the coefficient of variation (CV) of the ten determinations are listed. At the bottom of each table the mean (M), SD and CV of the average results of all laboratories are given. In addition median intralaboratory SD and CV are listed.
In figure 3-3 the average result for quality control material I and II for laboratories 1-12 are given in histograms. If the standard error of the mean (SE) exceeded half the width of the class interval a horizontal line represents the SE. For example for sodium, the scale is graduated: 121, 122, 123 mmol/l etc., the width of the class interval thus representing a variation of ± 0.5 mmol/l. Only SE's larger than 0.5 mmol/l are indicated by horizontal lines.
The upper and lower limits of the normal range reported by laboratory directors are given in the same graph below the concentration scale.
We concluded systematic analytical bias to exist, if for a certain constituent the average result reported by a laboratory ranked high or low for both controlsera in relation to the group. Therefor we ranked laboratories according to their average results for both controlmaterials from 1 to 12 and then added the two rank numbers. The probability distribution of rank sums ranges from 2 to 24, a rank sum of 13 being most likely to occur. The cases where rank sums exceeded 21 or were lower than 5 are listed in table 3-4. This technique has been described by Youden (72). In the same table an indication is given whether or not systematic analytical bias is reflected in the range of normal values. In cases where normal ranges were derived from the literature this is mentioned. Systematic analytical bias was considered to be reflected in the range of normal values when rank sums for the lower and upper limit of normal exceeded 18 or were lower than 8 simultaneously with high or low systematic bias respectively.

SODIUM

LAB nr	control I $\bar{X}_I$	SD	CV	control II $\bar{X}_{II}$	SD	CV
1	140	2,0	1,4	120	1,5	1,3
2	140	1,3	0,9	118	1,4	1,1
3[c]	-	1,4	1,0	-	1,6	1,3
4	144	1,3	0,9	123	1,1	0,9
5	141	1,3	0,9	120	1,5	1,2
6	140	1,5	1,1	119	1,0	0,8
7	138	1,6	1,2	118	1,8	1,5
8	140	0,7	0,5	120	2,5	2,1
9	141	0,9	0,6	120	1,5	1,2
10	143	1,4	1,0	121	0,9	0,7
11	141	0,5	0,3	121	1,0	0,8
12	141	0,8	0,6	120	1,3	1,0
M	140,8	1,3	0,9	120,0	1,5	1,2
sd	1,6			1,4		
cv	1,13%			1,17%		

POTASSIUM

LAB nr	control I $\bar{X}_I$	SD	CV	control II $\bar{X}_{II}$	SD	CV
1	5,6	0,11	1,9	3,3	0,06	2,0
2	5,8	0,05	0,8	3,4	0,11	3,2
3[c]	5,8	0,08	1,4	3,3	0,07	2,0
4	6,0	0,07	1,2	3,4	0,05	1,3
5	5,8	0,13	2,3	3,4	0,07	2,1
6	5,6	0,12	2,1	3,3	0,12	3,7
7	5,8	0,05	0,9	3,3	0,08	2,4
8	5,7	0,05	0,8	3,3	0,08	2,3
9	5,9	0,11	1,9	3,4	0,06	1,8
10	5,9	0,07	1,2	3,4	0,07	2,0
11	5,9	0,03	0,5	3,3	0,04	1,1
12	5,8	0,05	0,8	3,4	0,05	1,5
M	5,80	0,07	1,2	3,35	0,07	2,0
sd	0,12			0,05		
cv	2,06%			1,52%		

CHLORIDE

LAB nr	control I $\bar{X}_I$	SD	CV	control II $\bar{X}_{II}$	SD	CV
1	114	1,6	1,4	99	1,3	1,3
2	115	2,3	2,0	100	1,4	1,4
3[c]	117	1,7	1,5	102	1,2	1,2
4	118	1,0	0,8	102	1,7	1,7
5	115	2,3	2,0	99	1,9	1,9
6	114	2,1	1,8	99	1,8	1,8
7	118	0,5	0,4	100	0,7	0,7
8	115	1,2	1,0	100	2,0	1,9
9	114	1,6	1,4	100	1,3	1,3
10	117	1,3	1,1	100	0,7	0,7
11	114	2,2	1,9	98	2,4	2,4
12	114	1,5	1,3	98	1,0	1,0
M	115,4	1,6	1,4	99,8	1,4	1,4
sd	1,6			1,3		
cv	1,39%			1,28%		

CALCIUM

LAB nr	control I $\bar{X}_I$	SD	CV	control II $\bar{X}_{II}$	SD	CV
1	2,15	0,05	2,5	2,65	0,08	2,9
2	2,26	0,04	1,6	2,82	0,03	1,0
3[c]	-	0,03	1,2	-	0,07	2,8
4	2,08	0,05	2,4	2,65	0,06	2,2
5	2,21	0,06	2,5	2,69	0,11	4,0
6[a]	2,20	0,06	2,6	2,64	0,06	2,1
7	2,20	0,02	0,9	2,68	0,03	1,0
8	2,18	0,05	2,2	2,73	0,07	2,7
9	2,12	0,03	1,3	2,65	0,03	1,2
10	2,28	0,03	1,3	2,80	0,03	1,0
11	2,09	0,04	1,9	2,63	0,04	1,5
12	2,15	0,04	1,7	2,67	0,03	1,3
M	2,175	0,04	1,8	2,692	0,05	1,8
sd	0,064			0,065		
cv	2,95%			2,40%		

INORG. PHOSPHOROUS

LAB nr	control I $\bar{X}_I$	SD	CV	control II $\bar{X}_{II}$	SD	CV
1	1,16	0,10	8,3	2,39	0,09	3,8
2	1,24	0,07	5,4	2,59	0,06	3,0
3[c]	-	0,05	3,7	-	0,09	3,1
4	1,19	0,03	2,6	2,54	0,06	2,4
5	1,15	0,03	2,4	2,49	0,14	5,6
6[a]	1,16	0,05	4,4	2,49	0,06	2,3
7	1,17	0,02	1,5	2,55	0,05	1,8
8	1,17	0,02	2,0	2,43	0,04	1,5
9	1,16	0,06	5,0	2,13	0,09	4,3
10	1,16	0,03	2,3	2,54	0,06	2,3
11	1,14	0,04	3,1	2,51	0,05	1,8
12	1,14	0,03	2,6	2,48	0,08	3,0
M	1,167	0,04	2,9	2,467	0,06	2,7
sd	0,028			0,125		
cv	2,39%			5,07%		

UREA

LAB nr	control I $\bar{X}_I$	SD	CV	control II $\bar{X}_{II}$	SD	CV
1	7,80	0,34	4,4	21,3	0,60	2,8
2	7,30	0,64	8,7	20,9	0,75	3,6
3[c]	7,20	0,23	3,2	20,6	0,31	1,5
4	7,64	0,24	3,1	21,0	0,52	2,5
5	7,28	0,20	2,7	20,9	0,86	4,1
6[a]	6,70	0,36	5,4	21,2	0,88	4,2
7	7,37	0,08	1,1	22,1	0,41	1,9
8	7,24	0,13	1,8	20,9	0,57	2,7
9	7,33	0,19	2,6	20,9	0,19	0,9
10	7,25	0,18	2,5	20,3	0,70	3,3
11	5,66	0,21	3,8	18,4	1,50	7,9
12	6,95	0,21	3,0	20,8	0,36	1,7
M	7,143	0,21	3,1	20,78	0,59	2,8
sd	0,545			0,86		
cv	7,62%			4,15%		

Table 3-1

CREATININE

LAB nr	control I $\bar{X}_I$	SD	CV	control II $\bar{X}_{II}$	SD	CV
1	166	7,1	4,3	450	15,3	3,4
2	155	4,4	3,0	449	9,7	2,1
3[c]	148	5,2	3,4	443	5,3	1,2
4	152	4,8	3,2	437	9,9	2,3
5	166	4,5	2,8	479	26,2	5,5
6[a]	170	7,8	4,6	460	10,2	2,2
7	162	2,0	1,2	471	17,3	3,6
8	159	2,8	1,7	464	7,1	1,5
9	168	11,9	7,1	433	12,6	2,9
10	153	3,2	2,1	442	11,1	2,5
11	151	3,6	2,4	467	24,0	5,1
12	152	7,5	4,9	473	35	7,4
M	158,5	4,7	3,1	455,7	11,9	2,7
sd	7,6			15,3		
cv	4,81%			3,36%		

GLUCOSE

LAB nr	control I $\bar{X}_I$	SD	CV	control II $\bar{X}_{II}$	SD	CV
1	5,60	0,36	6,4	11,6	0,57	4,9
2	5,20	0,20	3,8	12,1	0,27	2,3
3[c]	5,80	0,13	2,2	13,3	0,33	2,5
4	4,93	0,16	3,2	10,3	0,49	4,8
5	5,63	0,29	5,2	11,9	0,59	5,0
6[a]	4,49	0,32	7,1	11,2	0,70	6,3
7	5,36	0,33	6,1	11,8	0,70	5,9
8	5,78	0,27	4,7	12,4	0,33	2,6
9	4,80	0,16	3,3	11,3	0,40	3,5
10	5,05	0,29	5,8	11,1	0,40	3,5
11	6,09	0,26	4,3	12,3	0,68	5,5
12	6,28	0,33	5,3	13,3	0,64	4,8
M	5,418	0,28	5,0	11,88	0,53	4,8
sd	0,541			0,88		
cv	9,98%			73,9%		

CHOLESTEROL

LAB nr	control I $\bar{X}_I$	SD	CV	control II $\bar{X}_{II}$	SD	CV
1	4,08	0,21	5,1	3,69	0,20	5,3
2	3,37	0,14	4,1	-	-	-
3[c]	-	0,12	3,5	-	-	-
4	4,27	0,17	4,0	3,91	0,13	3,3
5	4,07	0,13	3,2	3,93	0,19	4,8
6[a]	3,70	0,18	4,8	3,28	-	-
7	4,16	0,08	1,9	3,93	0,08	2,0
8	4,31	0,13	3,0	3,93	0,14	3,6
9	4,16	0,19	4,6	3,85	0,11	2,9
10	4,39	0,07	1,7	4,25	0,21	4,9
11	4,32	0,13	3,1	4,05	0,15	3,8
12	4,25	0,11	2,6	4,05	0,23	5,7
M	4,098	0,13	3,4	3,887	0,15	3,8
sd	0,305			0,258		
cv	7,43%			6,64%		

TOTAL PROTEIN

LAB nr	control I $\bar{X}_I$	SD	CV	control II $\bar{X}_{II}$	SD	CV
1	66,6	2,6	4,0	57,0	1,3	7,2
2	67,6	1,7	2,5	57,5	1,5	2,6
3[c]	-	1,8	2,8	-	0,9	1,6
4	67,9	1,4	2,1	56,5	1,0	1,8
5	64,2	2,8	4,3	55,1	2,3	4,2
6[a]	66,6	1,8	2,7	53,4	1,5	2,7
7	65,8	0,6	0,9	54,8	0,8	1,4
8	68,9	1,0	1,5	57,5	1,2	2,1
9	67,2	0,6	0,9	56,0	1,3	2,3
10	66,8	0,8	1,2	54,9	1,0	1,8
11	66,1	1,7	2,6	57,5	0,7	1,3
12	68,1	1,0	1,5	58,5	1,8	3,0
M	67,00	1,6	2,3	56,20	1,3	2,3
sd	1,30			1,54		
cv	1,94%			2,74%		

ALK. PHOSPHATASE

LAB nr	control I $\bar{X}_I$	SD	CV	control II $\bar{X}_{II}$	SD	CV
1	52	1,7	3,2	163	6,2	3,8
2	92	6,5	7,1	316	14	4,4
3[c]	-	-	3,1	-	-	3,3
4	89	4,6	5,2	274	14	5,0
5	84	12	14,4	322	40	12,5
6	65	4,3	6,6	191	8,4	4,4
7	46	5,5	12	145	16	11
8	31	1,4	4,5	97	2,7	2,8
9	40	2,0	5,0	122	4,8	3,9
10	87	4,3	5,0	280	3,3	1,2
11	96	3,9	4,1	326	11	3,3
12	76	4,7	6,2	250	13	5,2
M			5,1			4 2
sd						
cv						

LDH

LAB nr	control I $\bar{X}_I$	SD	CV	control II $\bar{X}_{II}$	SD	CV
1	100	7,6	7,6	383	22	5,8
2	80	7,3	9,1	286	26	9,2
3[c]	-	-	3,1	-	-	1,6
4	153	14,4	9,4	605	24	3,9
5	140	17	12	565	70	12,4
6	113	7	6,5	406	18	4,5
7[b]			3,4			4,6
8	211	17	7,9	759	36	4,7
9	166	11	6,3	520	22	4,3
10	151	7,5	5,0	591	25	4,2
11	126	3,8	3,0	475	22	4,6
12	156	12	7,6	592	25	4,3
M			7,1			4,6
sd						
cv						

Table 3-1 cont.

H E M O G L O B I N

LAB nr	control I $\bar{X}_I$	SD	CV	control II $\bar{X}_{II}$	SD	CV
1[b]						
2	5,17	0,19	3,6	6,87	0,18	2,6
3	5,27	0,10	1,9	6,74	0,12	1,8
4	5,30	0,07	1,3	6,77	0,07	1,0
5	5,38	0,20	3,7	6,84	0,25	3,7
6	5,15	0,12	2,2	6,77	0,09	1,3
7[b]						
8	5,23	0,07	1,2	6,71	0,13	1,9
9	5,18	0,09	1,8	6,55	0,14	2,1
10	5,34	0,11	2,0%	6,74	0,11	1,6
11	5,41	0,09	1,7	6,87	0,16	2,3
12	5,47	0,07	1,2	6,98	0,06	0,9
M	5,290	0,10	1,9	6,784	0,13	1,9
sd	0,109			0,116		
cv	2,06%			1,7%		

Table 3-1.

For each constituent and for each hospital laboratory, numbered 1 through 12, the average value ($\bar{X}$) of 10 results for the q.c. materials I and II and their dispersion, expressed as SD and CV are presented. At the bottom of each table the mean, SD and CV of the average values of all laboratories are given. In addition median SD and median CV of intralaboratory variability are listed. Mean of averages and median SD are denoted by M. For the enzymes Alk. P-ase and LDH no summarized data are given for reasons of methodological differences, except for the median CV. Footnotes:

a. In laboratory 6 two types of methodology are available for a number of constituents. Screening by means of a SMA 12/60 is performed on out-patients samples and special methods are applied to samples of in-patients. This laboratory tested the controlmaterials 5 times on the SMA and 5 times with the other methods. The interviewer had admittance to quality control records of this laboratory. In table 3-1 the average values for the SMA are given, since possible systematic analytical bias between laboratories is – in chapter IV – related to physicians answers on action levels in the out-patient situation. The intralaboratory variabilities are given for the special methods since these data will be used with the interpretation of physicians' answers to the question: "what do you consider a significant change within a patient", which relates to the in-patient situation (chapter V).

b. Laboratory 1 did not test Hb controls, since they were not available at the time. Results for Hb controls of laboratory 7 were lost. For this laboratory no mean value for LDH was obtained. Intralaboratory variability (CV) is given for 1975 pools.

c. Due to a mistake laboratory 3 tested the 1975 pools of the Massachusetts quality controlsera. The structure and the targetvalues of these materials do not differ much from the 1974 pools and intralaboratory variability was entered as such (CV). Information for both the 1974 and the 1975 pools was available for one reference laboratory and these were compared to those of laboratory 3. For some constituents a proportional or absolute systematic difference with the reference laboratory was evident and corrected results are entered in the table for laboratory 3. For example for potassium:

	1975		1974	
	serum I	serum II	serum I	serum II
ref. lab.	5.81 mmol/l	3.64 mmol/l	5.83 mmol/l	3.40 mmol/l
lab. 3	5.74 mmol/l	3.56 mmol/l	5.76*mmol/l	3.32*mmol/l

For K, Cl and creatinine a constant systematic difference was observed and brought into account, for urea and glucose a proportional systematic difference.

In table 3-10 intralaboratory variability observed in this survey (median and range) are given together with those of the Quality Assurance Service of the College of American Pathologists (38,48) as a comparison.

3.2.2 The inquiry

In table 3-2 the laboratory directors' answers to the inquiry on normal range values are given. In figure 3-3 the upper and lower limits of normal reported by the laboratory are presented in histograms below the concentration scale while average results for both quality control materials are plotted above this scale.

3.3 Discussion

The data presented in this chapter on dispersion of analytical results within each laboratory and between laboratories as well as the data on the laboratory's normal range values will be considered in the next chapters in relation to physicians' views on medical significance. In this regard we will be particularly interested in the following points:

- is there systematic analytical bias between laboratories? Can observed differences in normal range values be explained by systematic bias?
- which other factors can be responsible for observed differences in normal range values? What is the relevance for this study?
- what is the median intralaboratory variability and is there a significant difference between laboratories as to intralaboratory variability?

3.3.1 Systematic bias between laboratories and differences in normal range values.

If we are to draw conclusions from the survey data as to the existence of systematic analytical bias between laboratories we have to keep in mind, that the survey, although it covered 2-5 weeks of testing, gives a picture of short-term bias. To keep a laboratory determination consistent over long periods of time is difficult (19,26). Although the differences in mean values for survey results are statistically significant (fig. 3-3) in many cases, we did not subject the survey results to extensive statistical analysis and have considered the most evident cases only in combination with the data on normal range values given by the laboratory directors.

To this purpose in figure 3-3 the frequency distributions of the average results for the controlsamples are given together with the lower and upper limit of normal as given by the laboratory director.

From table 3-4 we see, that in about one-half of the cases where systematic analytical bias as compared to the group is present according to the described ranking tests, the normal range does not reflect this bias. This is in fact a disturbing situation since the normal range is often used as a point of reference between laboratories. It is for example common practice in the medical literature to list normal values with case histories (31) so to allow the reader optimal interpretation.

const.	Lab nr 1	2	3	4	5	6
Na mmol/l	135-148 BD/HP, 1-3 200	135-148 HP <1 20 20-30	135-145 HP >3	135-147 OP <3 1500	136-147 HP >3 40 20-35	136-144 BD/HP, 1-3 100 >20
K mmol/l	3, 15-4.45 *	3,5-5.0 *	3,5-5.0 *	3,1-5.1 *	4,0-5.0 *	4,0-4.8 *
Cl mmol/l	98-106 BD/HP/0 >3 100	95-107 *	98-108 *	96-108 *	99-108 *	96-107 *
Ca mmol/l	2,38-2.73 litt.	2,25-2.75 *	2,20-2.60 St/HP 1-3	2,20-2.70 *	2,20-2.60 *	2,25-2.65 *
P mmol/l	0.84-1.35 litt.	0.81-1.65 *	0.81-1.45 HP >3	0.65-1.30 *	1.00-1.60 *	0.81-1.45 *
urea mmol/l	2.9-8.5 BD/HP, 1-3 200	1.8-8.6 *	2.9-7.5 *	3.0-6.7 *	3.0-7.5 *	2.5-7.5 *
creat. umol/l	50-115 >3	45-115 *	80-115 HP <1	50-105 *	50-120 *	70-133 *
gluc. mmol/l	4.3-6.1 BD/HP, 1-3 200	3.4-6.2 *	4.2-6.4 HP >3	3.3-5.5 *	4.5-6.0 *	3.5-5.5 *
chol. mmol/l	3.4-7.3 >3	3.4-6.5 *	3.4-6.8 OP >3	4.5-7.5 litt.	4.0-7.3 *	3.9-7.3 *
t.prot. g/l	62-79 BD/HP, 1-3 200	61-80 *	62-78 HP >3	59-83 BD 1-3	62-78 *	60-80 *
ALP U/I	12-47 *	11-86 *	20-90 BD/OP <1	25-100 *	50-135 *	15-60 *
LDH U/I	96-240 >3	80-120 *	90-190 *	-200 *	-240 *	-160 *
Hb mmol/l	-	8.7-11.2 7.4- 9.3 *	8.7-11.2 7.4-10.0 <1	8.4-11.2 7.2-10.2 litt.	8.5-10.7 7.3- 9.5 *	7.5-11.0 *

* Determined on the same population sample as for the forgoing constituent.

Table 3-2:

Normal range values as reported by laboratory directors. These values were either derived from the literature (litt.) or determined in the laboratory, for example:
"HP <1
20 20-30" means: normal values were determined on hospital personel, less than 1 year ago, number of individuals: 20, age range 20-30.
Other abbreviations: BD: blood donors, OP: out-patients, ST: students, o: others.
1) normal ranges were determined on a population sample as indicated. The figures given in the table have been corrected taking into account the inhomogeneity of the population sample.
2) laboratories 8 and 9 collaborated in a project to establish normal range values.

Lab nr. / const.	7	8 [2]	9 [2]	10	11 [1]	12
Na mmol/l	135-145 BD/HP >3 100 <65	136-148 BD 1-3 100 18-65	136-147 BD 1-3 900 20-65	135-145 HP <1 100 20-65	138-144 BD <1 90♂10♀18-65	136-146 >3
K mmol/l	3.7-5.0 *	3.6-5.1 *	3.6-5.1 *	3.5-5.0 *	3.9-5.0 *	4.0-5.0 *
Cl mmol/l	96-106 *	97-109 *	97-109 *	96-108 HP >3 30 20-65	100-105 *	95-105
Ca mmol/l	2.37-2.79 * 2.12-2.26 Klin.	2.20-2.65 *	2.20-2.65 *	2.20-2.60 HP <1 100 20-65	2.25-2.60 *	2.25-2.75 litt.
P mmol/l	0.60-1.40 *	0.60-1.40 *	0.60-1.40 *	0.70-1.45 *	0.85-1.30 *	0.80-1.30
urea mmol/l	2.5-6.7 *	2.5-8.0 *	2.7-8.2 *	3.3-7.5 *	3.0-6.3 *	3.3-6.7 lab/litt. >3
creat. umol/l	53-106 *	60-110 * <1	55-125 *	55-110 *	60-120 BD 1-3 43♂57♀15-60	-100 >3
gluc. mmol/l	4.0-6.0 *	3.3-6.0 *	3.5-5.5 litt.	3.1-5.6 *	3.7-5.6 BD >3 45♂ 5♀20-55	4.5-6.0 BD 1-3 100
chol. mmol/l	4.0-7.8 *	3.8-8.6 *	3.5-8.5 BD 1-3 900 20-65	3.9-7.8 *	4.7-7.8 BD <1 45♂ 5♀18-65	4.5-7.5 >3
t.prot. g/l	64-80 *	65-80 BD 1-3 100 18-65	65-80 *	60-80 >3	65-75 BD <1 90♂10♀18-65	61-76 lab/litt. > 3
ALP U/I	10-50 *	18-45 *	15-60 litt	35-130 HP <1 100 20.65	-100 *	35-100
LDH U/I	-200	150-350 *	-160 *	-200 *	-175 *	-200 BD >3 40
Hb mmol/l	8.7-11.0 7.4-10.0	8.7-10.5 7.6- 9.7 *	7.5-11.0 *	8.7-10.6 8.1- 9.9 HP >3 200 40	8.5-10.7 7.3- 9.5 N adults 1-3 2500	8.7-11.2 7.5-10.0 BD >3 100

*Determined on the same population sample as for the forgoing constituent.

Table 3-2 cont.

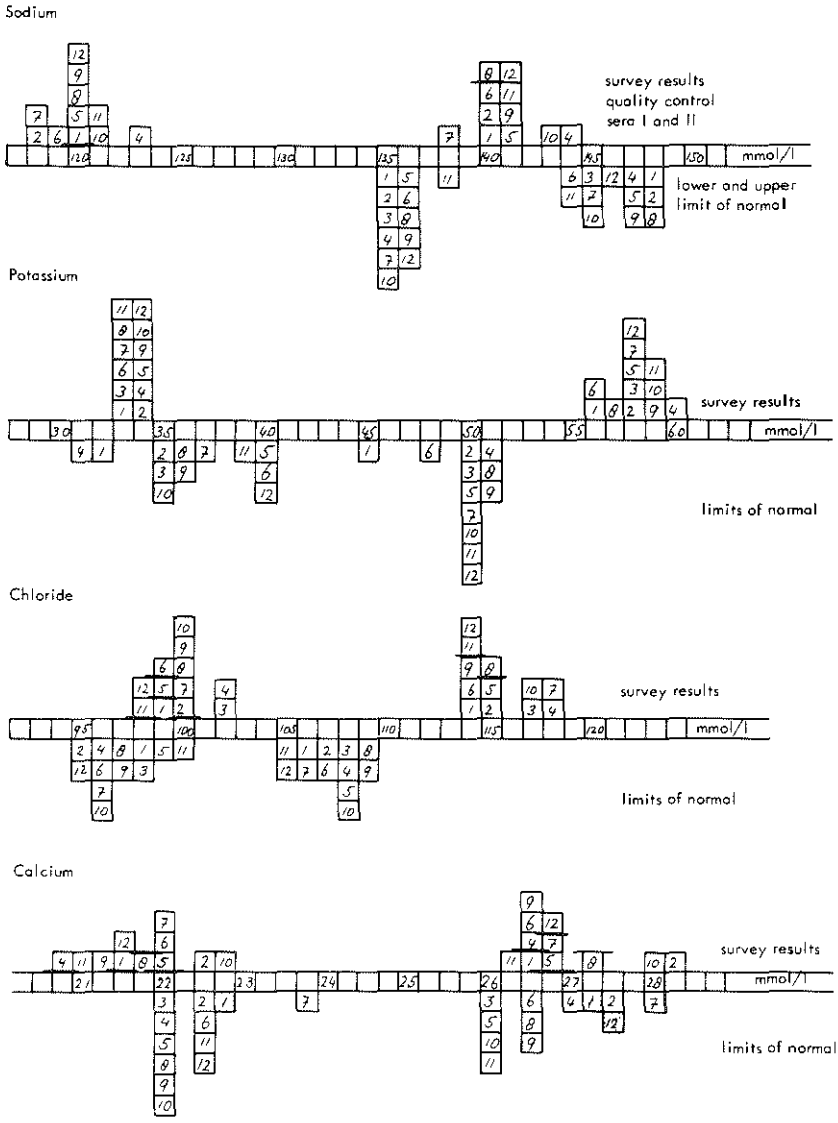


Figure 3-3:

Frequency distributions of average results obtained for quality control materials I and II by laboratories 1 through 12 and frequency distributions of upper and lower limits of normal given by laboratory directors.

▨ normal range values derived from the literature.

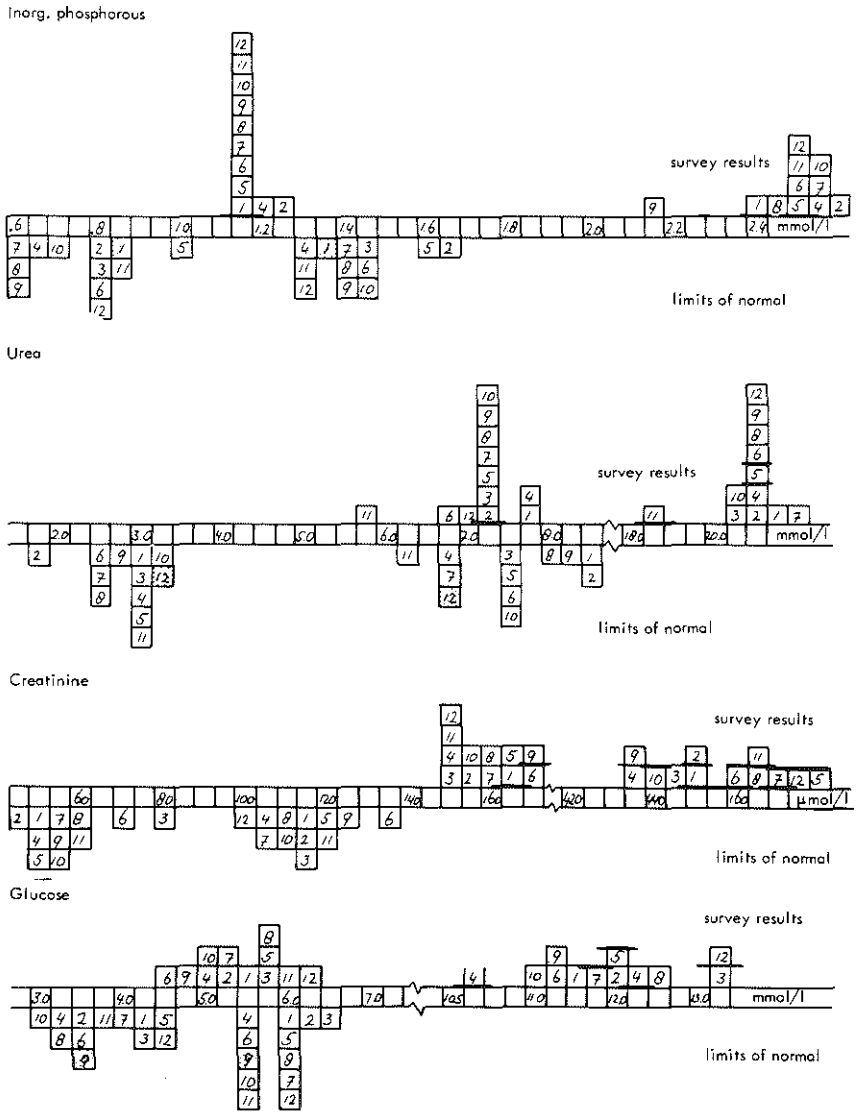


Figure 3-3 cont.

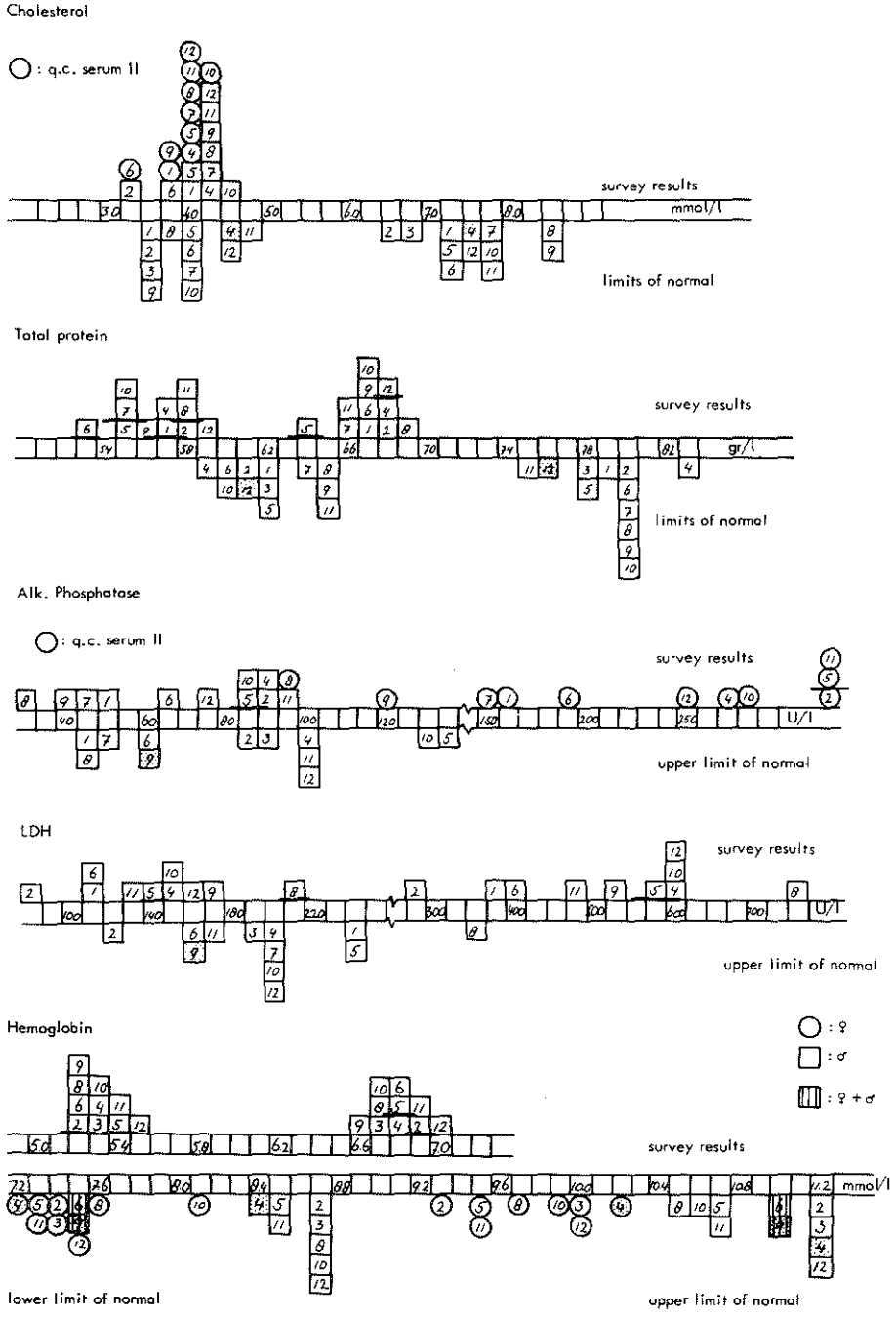


Figure 3-3 cont.

	lab. number of labs. with systematic bias in survey results		is systematic bias reflected in normal range?
	high	low	
Na		7	yes
	10		no
	4		no
K		1	yes
		6	no
	4		no
Cl		11	no
		12	yes
	3		yes
	4		no
Ca	2		yes
	10		no
		11	no
P	2		no
	4		no
urea	1		yes
		11	no
creat.	5		no
gluc.		4	yes
		6	yes
	12		yes
chol.		2	yes
		6	no
	10		yes
t. prot.	8		yes
		5	no
		6	no
		7	no
	12		no (litt.)
ALP		8	yes
		9	no (litt.)
	11		no
LDH		1	no
		2	yes
	8		yes
Hb		8	no
		9	no (litt.)
	11		no
	12		yes

Table 3-4: Laboratories where systematic analytical bias for both quality control materials towards high or low results is observed are given as well as the answer to the question whether or not this bias is reflected in the range of normal values. Criteria are given in the text.

In previous studies we tested the usefullness of the normal range as a point of reference in interhospital communications in a slightly different way and found it negligible (19, 20). Figure 3-5a represents a graph relating for a group of 22 laboratories in the United States and 22 in the Netherlands the average result of four determinations for a controlserum to the upperlimit of normal for calcium (20). Figure 3-5b gives the same graph for the present study.

If systematic analytical bias would be reflected in the range of normal values we would expect a correlation between the value of the normal range plotted and the result for the controlserum, none is seen. This finding is in accordance with a report of Strømme and Eldjarn (59).
Particularly in enzymology methodological differences cause large discrepancies between laboratories and a point of reference is really needed.
In fig. 3-6 the graph relating control serum result to the upper limit of normal is given for LDH, the data being taken from a clinical chemistry survey among dialysis and transplant centers in Europe (20) and little correlation is found. Fig. 3-7 gives similar graphs for alkaline phosphatase and LDH for the present study. We see better results: laboratories reporting a relative low average result for the control serum I in general report a relative low upper limit of normal and vice versa. However, we see also that in some cases extreme values for the control serum do not go together with extreme values for the upper limit of normal as compared to the group. This is the case for laboratory 11 for Alk. P-ase and for laboratory 1 for LDH.
This finding is consistent with table 3-4 where systematic analytical bias, taking into account the results for both control sera, was found to be reflected in the range of normal values in only 3 out of 6 cases of analytical bias for the enzymes.

Conclusion: apparent systematic analytical bias is not consistently reflected in the range of normal values. For the enzymes Alk. P-ase and LDH some correlation is seen for the majority of laboratories between the result for control serum I and the upper limit of normal (Alk. P-ase: $r = 0,8$, $n = 11$; LDH: $r = 0,7$, $n = 10$).

3.3.2 Possible causes of differences in normal range values other than systematic bias

There are a number of other reasons why normal range values may differ from laboratory to laboratory: e.g. regional population differences, selection of the sample of the population for determination of the normal range, intralaboratory variability, technique of sample taking, intraindividual variation and dietary, postural or other differences in the individuals tested.
We will discuss these subjects in view of the data laboratory directors provided and in view of some recent data from the literature.

3.3.2.1. *Intralaboratory variability*

Flynn et al. (25) studied the effect of intralaboratory variability on the width of the normal range. In table 3-8, column 1 the percentage increase in the range of normal values attributable to day to day analytical error is given, with the analytical variability of the test in their study. The results of these experiments agree well with the mathematically derived data of Gowenlock and Broughton (29) in fig. 3-9. Evidently for the electrolytes, particularly for calcium, increased intralaboratory variability will increase the width of the range of normal values substantially. This subject has also been raised in the medical literature (36,72).

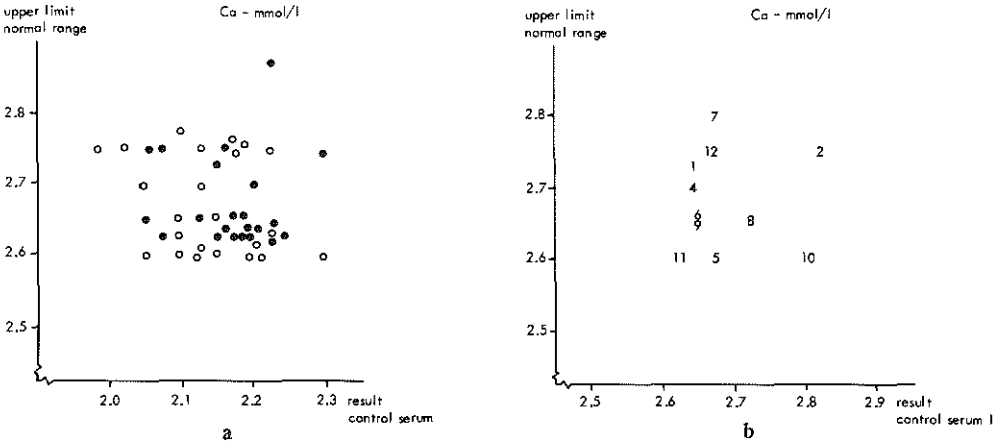


Figure 3-5a:

Graph relating the average result for calcium, obtained by a group of laboratories for a control-serum, determined in duplicate on two different days, to the upper limit of the normal range reported by the laboratory director.
(Survey among 22 laboratories in New England and 22 in the Netherlands (20)).

Figure 3-5b.

A similar graph, relating average result for controlserum I with upper limit of normal for the present survey.

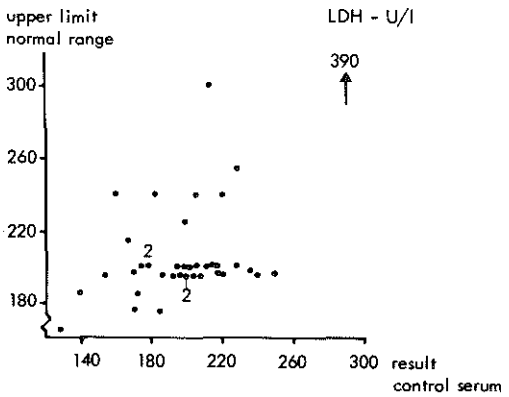


Figure 3-6:

Graph relating the average result for LDH, obtained by a group of laboratories for a controls-serum, determined in duplicate on two different days, to the upper limit of the normal range reported by the laboratory director.
(Survey I among dialysis and transplant centers in Europe (19)).

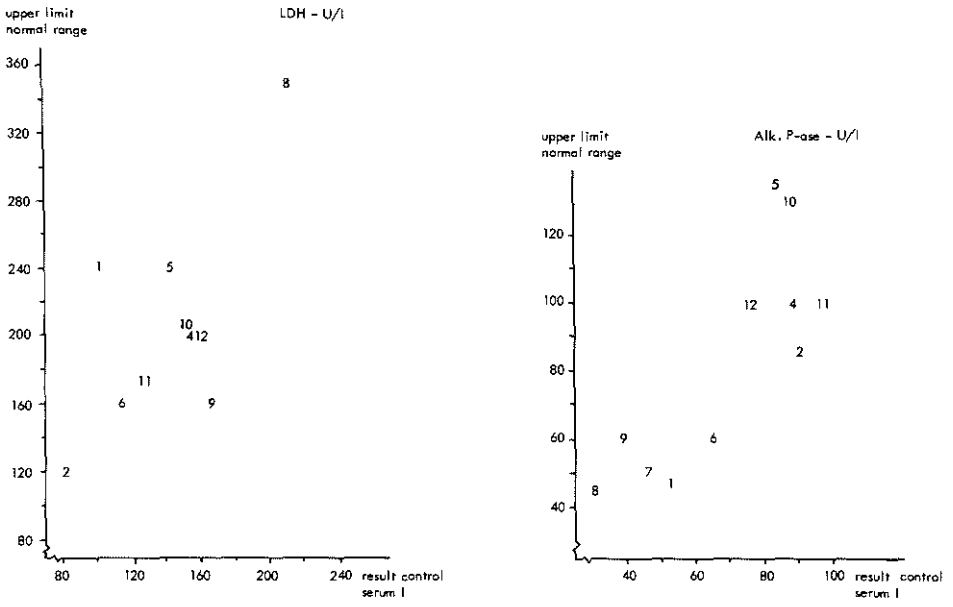


Figure 3-7:

Graphs, relating average result for Alk. P-ase and LDH respectively, obtained for controlserum I by laboratories 1-12 in the present study, to upper limit of normal.

However, in our study we see little of this effect. For example for calcium laboratories 1,4,5 and 8 who show larger intralaboratory variability than the rest of the group do not report wider normal ranges (fig. 3-3). For potassium however, laboratory 11 and 12 show small intralaboratory variability and a relative narrow normal range. On the other hand, laboratory 6 with the most narrow normal range has the largest intralaboratory variability of the group for this constituent. In fact, in 8 out of 26 cases lowest intralaboratory variability goes together with the widest range of normal values or vice versa. Hence, the effect of the intralaboratory variability on the width of the normal range could hardly be demonstrated in our study.

3.3.2.2. *Intraindividual variability*

Total intraindividuality day-today variability has been studied most extensively by Winkel and Statland (68,69) and Young et al. (75). In table 3-8 column 2 and 3 intraindividual variation in the healthy individual, expressed as coefficient of variation (CV), has been compiled from their work. Intraindividual variation may increase the width of the normal range in the same way analytical variability does and from fig. 3-9 we may conclude that – again – sodium, chloride and calcium will be the most sensitive to this influence.

	1 %-age increase due to intralab. var. (CV) (25)	2 intraindiv. variab. (68,69)	3 (CV) (75)	4 age/ sex	5 pop. diff. (70)	6 tourn. appl. (57)	7 lying down stan-ding (57)	8 exercise (56b)	9 eating fasting (56b)
Na	18% (0.6)	0.7%	1.4%	X	X	–	–	–	+2%
K	8% (1.9)	4.3%	4.6%	X	X	−6.3%	−3.9% (L)	−8%	–
Cl	17% (1.0)	2.1%	2.1%	X	X	–	–	–	–
Ca	29% (2.6)	1.7%	1.6%	young ♂ > ♀	weight oral c.	–	−3.1% (L)	–	–
P	9% (3.9)	5.8%	9.6%	young ♂ > ♀	oral c. reg. day	–	+3.6% (S)	+12%	+26%
urea	0% (x)	12.3%	13.6%	age ↑ ♂ > ♀	reg.	–	–	–	–
creat.	8% (x)	4.3%	4.4%	X	X	–	–	+17%	–
gluc.	5% (x)	X	6.5%	age ↑	–	X	X	X	X
chol.	3% (x)	5.3%	4.2%	young ♂ > ♀ age ↑	weight season day	−3.4%	+5.7% (S)	–	–
t. prot.	3% (1.4)	2.9%	2.2%	–	season weight	+4.3%	−6.1% (L)	+3%	–
ALP	3% (x)	4.8%	5.7%	young ♂ > ♀ age ↑	weight reg.	–	–	–	+12%
LDH	X	X	7.3%	–	season	X	X	–	−5%
Hb	X	X	X	♂ > ♀ old ♂ ↓	X	X	X	++	X

X: not studied –: no effect
x: intralaboratory variability cannot be given as data were converted to logarithms.

Table 3-8: Literature data on factors affecting normal range values (see text).
column 1: maximum increase (%) in width of the normal range due to intralaboratory variability. The latter is given in brackets, expressed as CV. (25)
column 2: intraindividual variability, expressed as CV, derived from studies of Statland's group (68, 69).
column 3: idem, derived from studies by Young c.s. (75).
column 4: effect of age and sex on the normal range.
column 5: conditions affecting normal range values (weight, regional differences (reg.), use of oral contraceptives (oral c.), time of sample taking per season or per time of the day (70 a, b)).
column 6: the effect of tourniquet application for 3 minutes on the average value for 11 individuals (57).
column 7: the effect of 30 minutes lying down (L) or 30 minutes standing (S) as compared to 15 minutes sitting down previous to sample taking on the average value for 11 individuals (57).
column 8. the effect of exercise on the average result for 11 individuals (56b).
column 9: the effect of a meal two hours before sample taking as compared to fasting on the average value for 11 individuals (56b).

3.3.2.3. *Population differences and differences in sample of the population.*

In general there seems to be little consistency between hospitals as to the procedure of determination of the normal range. None of the laboratories can be considered to have determined "reference values" as meant by Dybkaer (18) and Grasbeck (31) who introduced this concept: "a set of values of a certain type of quantity obtainable from a single individual or a group of individuals corresponding to a stated description. This description must be spelled out if others are to use the reference values. For each type of quantity a series of reference groups will be necessary taking into consideration: age, sex, race, menstruation, previous diet and excercise, posture etc."[1])

In table 3-8, columns 4 and 5 we have listed the effects of a number of factors related to population differences on the range of normal values as they are reported in the literature.

The most recent extensive study on this subject was performed by Winkelman et al. (70a,b). This group published data on laboratory results for appr. 400 clinically normal men and appr. 550 clinically normal women, ages 20-49, grouped by sex, age, body habitus (weight), regional residence in the U.S. (north/east/south/west, urban/rural), season of sample taking and time of the day. Seventy-four of the women were on oral contraceptives. Most other studies (15,23,37,45,46,60,63) report on age and sex differences only.

For age and sex differences the studies largely agree on the following:
increased serumlevels with age are found for urea, glucose, cholesterol and alkaline phosphatase. The opposite seems to be true for inorg. P.

In young people (up to 40 years of age) calcium, inorg. P, urea, cholesterol and alkaline phosphatase levels tend to be higher in males as compared to females. This tendency persists through higher ages for urea.

Of the constituents considered in this study sodium, potassium, chloride, creatinine and hemoglobin were not studied by the group of Winkelman.

[1]) This is one of the reasons why for this study the term "normal range" has been used. The preference for the term reference interval originated when automated instrumentation became available and large screening projects were carried out. The need for specific reference groups was felt in the interpretation of test result in these projects.
Indeed, if it is our ultimate goal (31) to determine true normal human reference values for each type of measurable property data from a series of reference groups will be necessary. Presently it can be considered misleading systematism to change the name of "normal values" determined in e.g. an unspecified group of blood donors into "reference values".
There is another reason why the term "normal values" is preferred for the present time. In the clinical chemistry vocabulary the terms reference methods and – the derived – reference values are frequently used. The use of one term for two different concepts, i.c. (a) a normal healthy human value and (b) an analytical value measured using a reference method, leads to confusion.
Finally the discussion on what is "normal" is still going on. It is known that blood levels of whole populations may change towards pathological levels (e.g. cholesterol in older people). While these levels are normally, frequently (76) observed in reference groups, clinicians will not consider these levels compatible with optimal health. In time we may want to refer to ideal or optimal values (71).

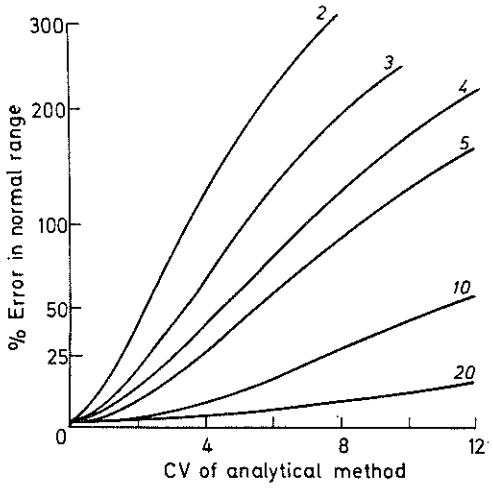


Figure 3-9:

The effect of increasing the C.V. of the analytical method (horizontal axis) on the normal range. Figures at the end of each line correspond to the true C.V. of the normal population. The vertical axis shows the percentage error of the observed C.V. of this normal population (log scale). (Gowenlock and Broughton ref. 29).

The electrolytes were studied with respect to age and sex by Owen and Campbell (46) who found decreasing levels for sodium and chloride with age. For creatinine Files and coworkers (23) found higher levels for males as compared to females in all age groups.

For hemoglobin sex and age differences have been studied by various groups (32,55, 60). All report higher values for males from 20-50 years of age. At higher ages the values for males decrease.

In our study laboratory directors did list the age range of the population sample studied (table 3-2). Few, however, indicated the proportion of males and females. Laboratory 2 and 5 explicitly state a sample of the population of 20-30 and 20-35 years of age respectively. No age effects as described in the literature for urea or glucose can be found in the normal ranges these laboratories report when they are compared with those of laboratories in the group including all age groups in the determination of the normal range. Both laboratories, 2 and 5, report relative high upper limits for phosphorous. It should be noted however, that laboratory 2 also has a systematic analytical bias tending to higher values for this constituent. It cannot be decided whether the low normal range for cholesterol reported by laboratory 2 is due to systematic bias (table 3-4) or to the fact that young people were tested for determination of the normal range.

All but two laboratories give separate normal ranges for males and females for hemoglobin, i.e. for females lower than for males.

Regional and seasonal differences are reported by Winkelman et. al. (70a,b) for inorg. P, urea, cholesterol, total protein, alkaline phosphatase and LDH (table 3-8 column 5). In our study no data are available as to the time of the year normal ranges were determined.
Within the group of Dutch laboratories it is hard to assign regional population differences. There may be some indication in the fact that laboratories 11 and 12, located in the same town in the south of Holland both report low upper limits of normal for phosphorous.
In the group of laboratories cooperating with this study normal range values were in general determined in hospital personel or blood donors. Only laboratory 4 studied a group of 1500 out-patients, which may have contributed to the low lower limit of normal for potassium, the high upper limit for total protein and the wide range for hemoglobin in females.
Also, laboratory 4 is among the group of laboratories reporting upper limits of normal for calcium of 2,70 mmol/l and higher. Of this group the high value of laboratory 2 can be explained by systematic analytical bias. Laboratory 1 and 12 report to have derived normal values from the literature. Laboratory 7 reports its – high – upper limit of normal for calcium explicitly for out-patients. This is interesting in view of reports of Yendt and Gagne (72) and Keating (36) who claim that if meticulous care is given to methodology and sample taking narrower normal ranges can be obtained. Upper limits of normal reported by these authors are 2,58 and 2,60 mmol/l respectively.

3.3.2.4 *Technique of sample taking and patient preparation*

Various aspects of the process of sample taking and a number of factors contributing to intraindividual variability of serum constituent levels have been studied by Statland and coworkers (6,56a,b,57).
Tourniquet application (table 3-8 column 6), 30 minutes standing or lying down – as compared to 15 minutes sitting – previous to sample taking (col. 7), exercise (column 8), eating as compared to fasting (column 9), all may influence test results. It may surprise that long tourniquet application (3 minutes) did not affect calcium determinations in Statland's study. The average result for 11 individuals for samples taken at 11 a.m. (15-30 seconds) and 11.30 a.m. (3 minutes) did not show a significant difference. In a separate experiment eight sequential samples of blood were taken every 40 seconds during continuous tourniquet application in one individual: calcium values rose appr. 3% in the first minute and went up appr. 10% after 200 seconds.
For hemoglobin Sunderman et al. (60) report an increase after exercise.
In our study the relative wide spread in normal range values for inorganic phosphorous and cholesterol observed in this study may not be surprising in view of the many possible influences on these determinations listed in table 3-8. All the more striking is the small dispersion in upper limits of normal of the majority of laboratories for total protein (78-80 gr/l). The same applies for calcium (2.60-2.65 mmol/l)

if we exclude the laboratories 4 and 7 that tested out-patients, laboratories 1 and 12 that derived normal values from the literature and laboratory 2 that showed systematic analytical bias.

In conclusion, differences between hospitals in normal range values can be tentatively explained in some cases on the basis of literature data:
- two laboratories where intralaboratory variability is small for potassium report a relative narrow normal range.
- a low upper limit of normal for inorganic phosphorous is found by two laboratories testing people in the younger age groups.
- one laboratory testing out-patients reports a high upper limit of normal for calcium and total protein, a low lower limit for potassium and a wide range of normal for hemoglobin values in women as compared to the other laboratories.
- there is an indication that the upper limit of normal for calcium should not exceed 2.65 mmol/l. The higher values reported in this study were obtained for out-patients (2 laboratories), were derived from the literature (2 laboratories) or were concurrent with systematic analytical bias (1 laboratory).
- for cholesterol and inorganic phosphorous the relative wide spread of upper limits of normal between laboratories can tentatively be explained by the effects of differences in technique of sample taking and differences in condition of the individuals tested.

3.3.4 Intralaboratory variability

In table 3-10 column 1 the data on intralaboratory variability for control materials I and II, observed on 10 determinations are presented. In column 2 we have listed intralaboratory variability reported by the Quality Assurance Service (QAS) of the College of American Pathologists, a project covering daily quality control of about 1000 hospitals in the United States (38,48).

No statistical comparison between the two sets of data can be made. The general pattern is, that the coefficient of variation for routine intralaboratory variability reported by the QAS is larger than that in our study (except for glucose and hemoglobin), based on sets of 10 determinations in 12 laboratories. Particularly for creatinine precision seems to be better in the group of laboratories cooperating with our study.

Although participating laboratories in our study were not informed as to the target value for the quality control sera, the materials were known as such. There is evidence (29) that in such – not completely blind – surveys intralaboratory variability tends to be lower than the true intralaboratory variability. A recent study by Steele c.s. (58) however, reveals no significant difference between masked and unmasked samples. Effects of this sort may account for differences between our group and the QAS. Within the group however, they are likely to be the same for all participants. The significance of differences in intralaboratory variability among the laboratories of the group under study was tested by means of Cochran's test (17). This test tells

	intralaboratory variability in the normal range (CV)	
	present study	QAS 1975
	median (range)	average (range)
Na	0.9 (0.3-1.4)	1.2 (1.1-1.3)
K	1.2 (0.5-2.3)	1.9 (1.7-2.1)
Cl	1.4 (0.7-2.4)	2.0 (1.6-2.4)
Ca	1.8 (0.9-2.6)	2.9 (2.3-3.5)
P	2.9 (1.5-8.3)	4.3 (2.9-5.7)
urea	3.1 (1.1-8.7)	5.3 (3.3-7.5)
creatinine	3.1 (1.2-7.1)	7.2 (5.6-9.0)
glucose	5.0 (2.2-7.1)	4.0 (3.7-4.5)
cholesterol	3.4 (1.7-5.1)	4.4 (3.6-5.3)
total protein	2.3 (0.9-5.6)	2.4 (2.1-2.8)
alk. P-ase	5.1 (3.1-14)	7.0
LDH	7.1 (3.0-12)	7.4
Hb	1.9 (0.9-3.7)	1.7*(0.9-4.0)

Table 3-10. Intralaboratory variability observed in this study (median and range) compared to average intralaboratory variability (and 2 SD limits) registered by the Quality Assurance Service of the College of American Pathologists, which comprises daily quality control of about 1000 hospital laboratories (38, 48).

*average intralaboratory variability and 95% confidence interval for 100 laboratories (38).

whether or not the variances observed in k groups of n observations differ significantly by testing the quotient of maximum variance and summed variances:

$$\frac{s^2_{max}}{\sum_{i=1}^{k} s_i^2}$$

The results were as follows:

differences in intralaboratory variability were

non-significant $P > 0.05$	for Na, Cl, glucose and cholesterol
significant with $0.01 < P < 0.05$	for K and calcium
significant with $P < 0.01$	for P, urea, creatinine, total protein alk. P-ase, LDH and Hb.

Chapter IV

NORMAL RANGE VALUES AND ACTION LEVELS APPLIED BY CLINICIANS. THE DIFFERENCE BETWEEN LIMIT OF NORMAL AND ACTION LEVEL AS A CRITERION FOR REQUIRED ANALYTICAL PRECISION

4.1 Introduction

This chapter deals with the answers to the following questions in the questionnaire for physicians:

1a. Which normal values do you apply?
If you apply different ranges for men and women or for people in different age groups, please identify.

b. Are the normal values mentioned under 1a:
- derived from the literature yes/no
  if so, from which textbook or publication?
- concluded from own personal observation? yes/no
- based on normal values given by the laboratory? yes/no
- other

c. Suppose you find no abnormalities during a physical examination of a patient with ill-defined complaints. You order a number of laboratory tests. Can you identify for each constituent a lower and an upper limit which would prompt you to an action of the first or second order.
An action of the first order (Action 1) does include: repeated or additional laboratory tests, ECG, X-ray, initiation or change of diet or medication. An action of the second order (Action II) is defined as: hospital admission, immediate treatment.

4.2 Results

4.2.1 Normal range values

When asked for normal range values 31 clinicians out of a total of 63 referred to the normal range given by the laboratory, 16 made corrections for one or more constituents when asked whether they agreed with the ranges given by the laboratory (calcium 8 times, alk. P-ase 7 times, potassium, total protein and LDH 3 times, cholesterol and creatinine 2 times).
Twenty-nine physicians reported normal ranges from memory (5 referred to the laboratory's normal range for 1-3 determinations), but only 17 of them listed action levels at the subsequent question (1c.). In table 4-1 the normal range values for these clinicians are listed. For the remaining 12 clinicians the normal ranges mentioned represent action levels (Action I).

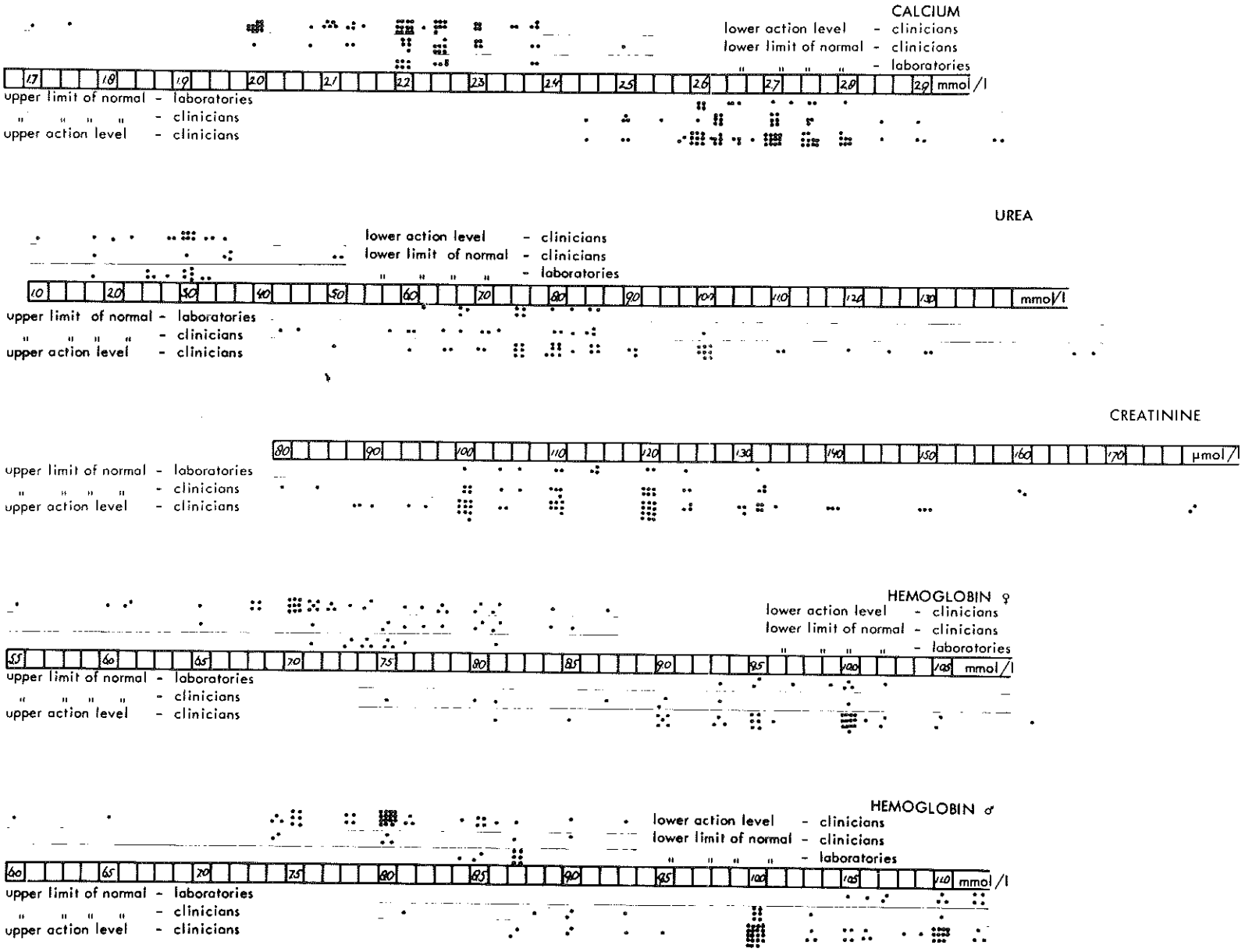


Figure 4-1 cont.

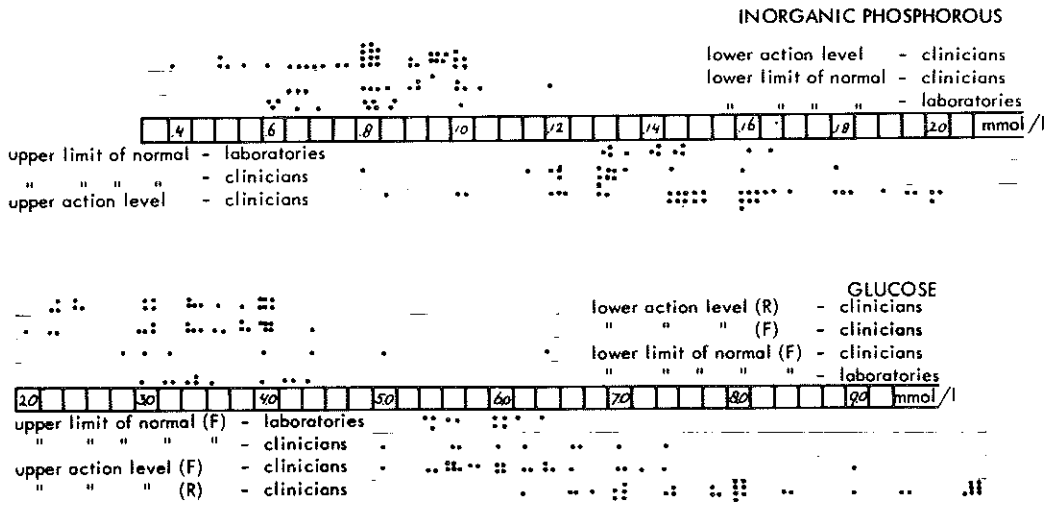


Figure 4-1 cont.

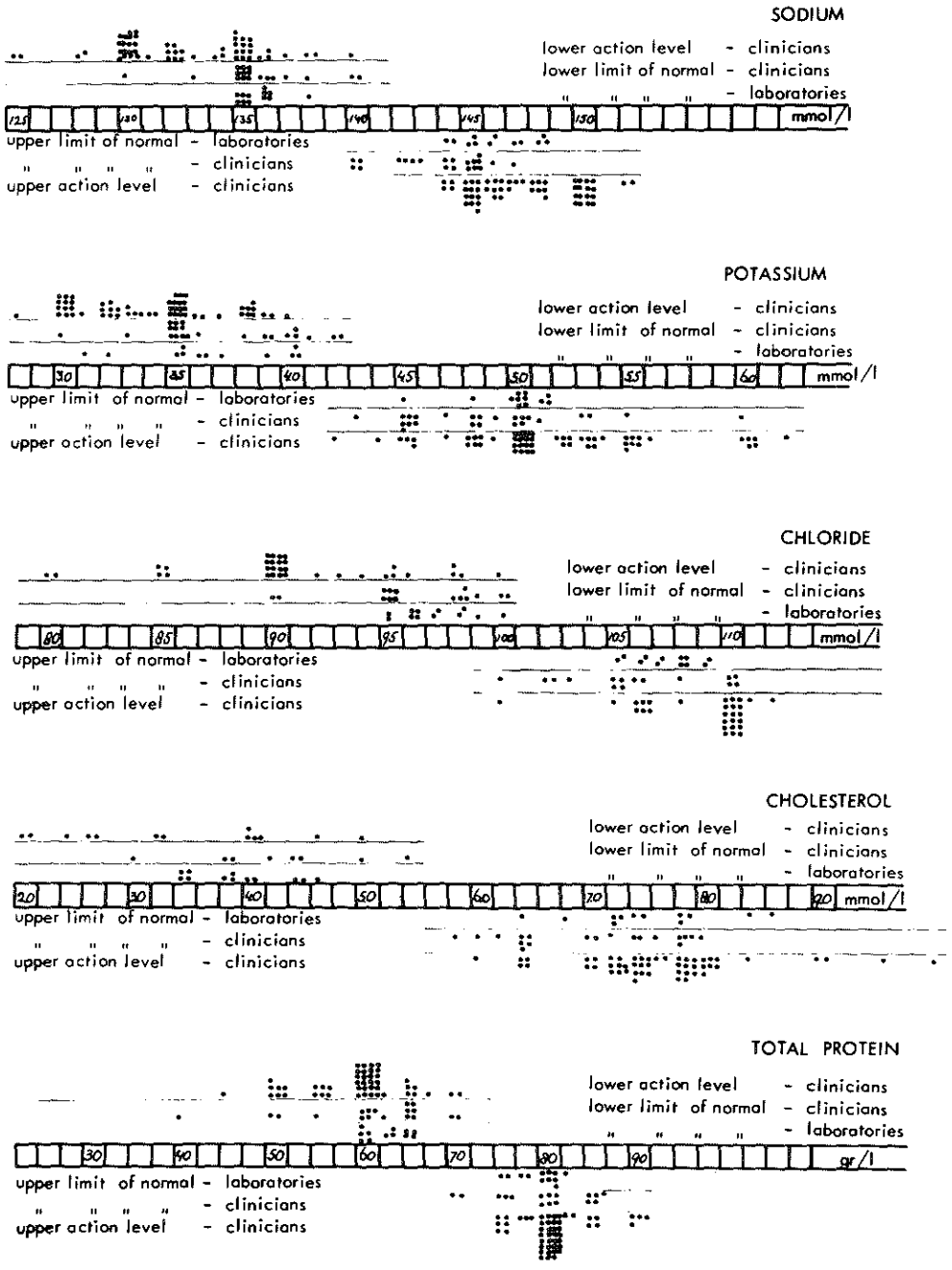


Figure 4-1:

Lower and upper limits of normal given by laboratory directors compared to normal range limits and action levels given by physicians.

		1 normal values laboratories			2 normal values physicians			3 action levels physicians			4 intralab. variability
		M	CV	n	M	CV	n	M	CV	n	CV
		mmol/l			mmol/l			mmol/l			
Na	LLN	136	0.7%	12	136	1.7%	18	132	2.7%	59	0.9%
	ULN	146	1.0%	12	144	1.5%	18	147	1.7%	61	0.9%
K	LLN	3.6	8.5%	12	3.7	8.5%	23	3.4	9.1%	56	2.0%
	ULN	5.0	3.6%	12	4.9	8.1%	24	5.2	7.6%	55	1.2%
Cl	LLN	97	1.6%	12	95	3.7%	17	91	5.5%	35	1.4%
	ULN	107	1.3%	12	107	3.1%	16	109	4.1%	33	1.4%
Ca	LLN	2.25	2.9%	12	2.24	4.9%	21	2.16	6.5%	56	1.8%
	ULN	2.67	2.6%	12	2.67	3.8%	23	2.70	4.0%	61	1.8%
P	LLN	0.76	16.6%	12	0.88	17.8%	15	0.78	21 %	40	2.9%
	ULN	1.42	7.9%	12	1.30	16.2%	16	1.53	19.0%	38	2.9%
urea	LLN	2.8	15.0%	12	3.5	30.3%	9	2.7	27 %	15	3.1%
	ULN	7.5	10.1%	12	7.3	14.7%	18	9.1	27 %	44	3.1%
creat.	LLN	57*	17.8%	12	58*	22 %	7	–	–	0	3.1%
	ULN	115*	8.0%	12	117*	16.7%	24	120*	20 %	57	3.1%
gluc. F	LLN	3.8	13.3%	12	4.2	31 %	7	3.4	19.2%	22	5.0%
PP	LLN				4.2	29 %	6	3.2	21 %	22	5.0%
F	ULN	5.9	5.3%	12	6.3	12.6%	9	6.2	14.3%	25	5.0%
PP	ULN				8.7	12.2%	9	8.2	14.5%	34	5.0%
chol.	LLN	3.9	11.7%	12	4.3	17.6%	8	3.4	28 %	15	3.4%
	ULN	7.6	8.0%	12	7.1	8.4%	19	7.7	10.2%	59	3.4%
t. prot.	LLN	62*	3.4%	12	61	12.1%	17	57*	13.9%	54	2.2%
	ULN	79*	2.7%	12	79*	6.2%	18	80*	4.5%	52	2.3%
Hb	LLN	7.5	3.4%	10	7.7	6.7%	12	7.3	7.4%	60	1.9%
	LLN ♂	8.6	1.4%	10	8.1	6.4%	8	8.0	9.2%	43	1.9%
	ULN ♀	9.8	3.1%	10	8.7	9.7%	8	9.7	6.2%	36	1.9%
	ULN	10.9	2.4%	10	9.7	8.1%	13	10.3	6.2%	52	1.9%

Table 4-2: For each constituent the average upper and lower limit of normal (ULN and LLN respectively) given by the laboratory directors are listed in column 1 as well as the dispersion of these limits of normal expressed as 1 CV. In column 2 the averages and dispersion of limits of normal given by physicians are listed, in column 3 the same parameters of physicians' action levels; n represents the number of laboratories (column 1) and the numbers of physicians (column 2 and 3) respectively. Median intralaboratory variability expressed as CV is given in column 4.

*Mean values in mmol/l except for creatinine (μmol/l) and total protein (gr/l).

Three physicians were asked for action levels only, not for normal ranges.

A number of clinicians take age differences into account.

cholesterol	26/63	inorg. P	6/63
alk. P-ase	21/63	urea	3/63
creatinine	11/63	total protein	1/63
glucose	7/63	Hemoglobin	1/63

Sex differences are taken into account for:

Hb	53/63	
creatinine	38/63	(body stature)
cholesterol	2/63	
alk. P-ase	2/63	
urea	1/63	

Although not specifically mentioned in the questionnaire the factor of time and technique of sample taking was raised for:

glucose (fasting)	39/63
cholesterol (fasting)	7/63
inorg. P (fasting)	4/63
calcium (tourn. appl.)	4/63
t. protein (tourn. appl.)	3/63

In answer to question 1b fifty five clinicians indicated that they use normal values of the laboratory as a guide. Eight derive normal values primarily from own observation. Seven times literature was mentioned as an additional source.

In table 4-1 normal range limits reported by clinicians are given together with the normal range limits reported by laboratory directors and the action levels given by clinicians. In table 4-2 the averages, standard deviations (SD) and coefficients of variation (CV) of upper and lower limits of normal and upper and lower action levels are given.

4.2.2 Action levels

In general limits that would prompt to an action of the first order (repeat of additional laboratory tests, ECG, X-ray, initiation or change of diet or medication) in the given situation could readily be given. A number of physicians interviewed stressed the importance of the clinical picture, but in all but three cases figures were given that "would not pass unnoticed". For one of these three clinicians this question was impossible to answer because the action taken was, for each individual case, weighed against the social, psychological and economical consequences for the patient. The remaining two physicians would not order any tests in the given situation except Hb. When there was some hesitation, the interviewer emphasized: "you did have a reason to order the tests". If the age, sex or stature would come up relative to deciding action versus no-action the physician was asked to consider a normally build man between 30 and 40 years of age.

Due to the limited period of time available for the interview the question as to an action of the second order (hospital admission , immediate treatment) was left out

in many cases. Hence the data are scarce and will not be presented. This represents an area of study of physicians' response to laboratory values which remains to be explored.

In figure 4-1 for each constituent are presented: normal range limits and action levels reported by physicians as well as normal ranges given by the laboratories. Table 4-2 presents the numerical data: average and CV for lower and upper limits of normal and action levels. In the last column of this table intralaboratory variability is given expressed as CV. Figure 4-3 gives action levels listed per hospital. The upper and lower limits of the normal range given by the laboratory director are indicated by arrows.

In this chapter we will consider for each clinician the difference between either limit of normal and the corresponding action level as a measure of the strictness this clinician applies in that particular concentration range and hence as a measure for required precision for that test at the present time.

Table 4-4 gives the median values of the differences between limits of normal and the corresponding action level for each constituent: column 1 gives the median value for the group of physicians who reported both normal values and action levels, column 2 gives the median value for the whole group, thus including the clinicians who reported to use the laboratory's normal values.

In column 3 median intralaboratory variabilities are listed. These are taken from table 3-1, where available in the same concentration range as the upper or lower limit of normal respectively.

For comparison of the median difference between upper or lower limit of normal and respective action levels with intralaboratory variability the latter is expressed as 3 CV. This is the rounded off figure for 2.26 $\sqrt{2}$ CV which can be considered to constitute a significant difference between two individual values, when the variability is calculated from ten independent determinations for a controlserum.

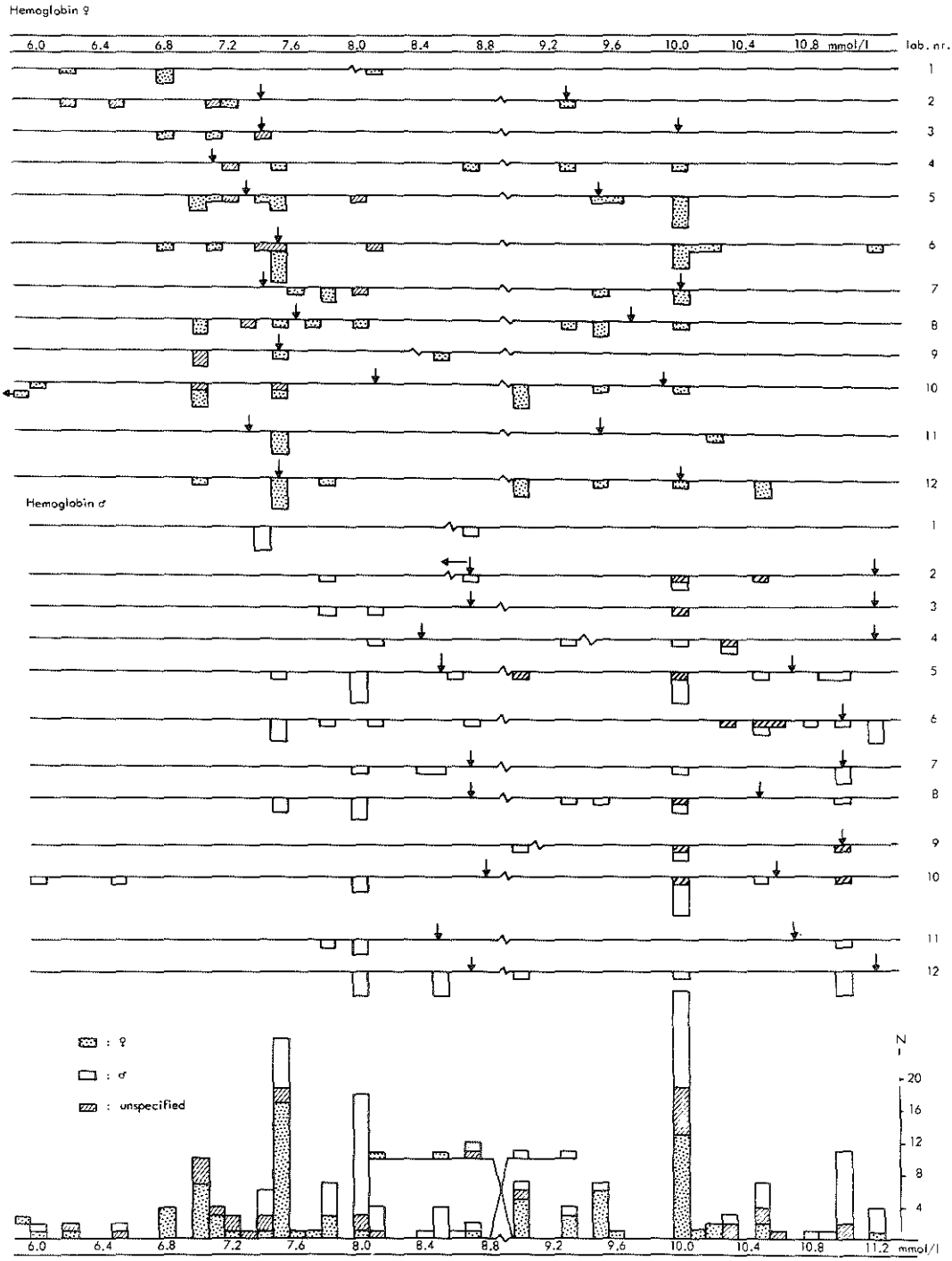
In fig. 4-5 histograms represent the distribution of deviations of action levels from the corresponding limit of normal for each constituent. In these graphs 0 represents the limit of normal, negative values represent action levels within the normal range, positive values action levels outside its limits.

In our attempt to analyze the clinicians' attitude towards laboratory results in the process of medical decision making it will be of interest to know the probability that an action level is comprised in the population of values observed in healthy individuals.

To this purpose we list in figure 4-5 the P-value of the action levels in the graph, assuming the normal range to be the 95% limits of variation of a Gaussian distribution.

We assumed the width of the normal range to comprise for:

Na	10 mmol/l	creatinine	60 μmol/l
K	1.6 mmol/l	glucose	2.0 mmol/l
Cl	10 mmol/l	cholesterol	3.0 mmol/l
Ca	0.40 mmol/l	t. protein	20 gr/l
P	0.80 mmol/l	Hb	2.0 mmol/l
urea	4 mmol/l		


Hemoglobin ♀
6.0 6.4 6.8 7.2 7.6 8.0 8.4 8.8 9.2 9.6 10.0 10.4 10.8 mmol/l
lab. nr.
1
2
3
4
5
6
7
8
9
10
11
12
Hemoglobin ♂
1
2
3
4
5
6
7
8
9
10
11
12
♀
♂
unspecified
N
20
16
12
8
4
6.0 6.4 6.8 7.2 7.6 8.0 8.4 8.8 9.2 9.6 10.0 10.4 10.8 11.2 mmol/l

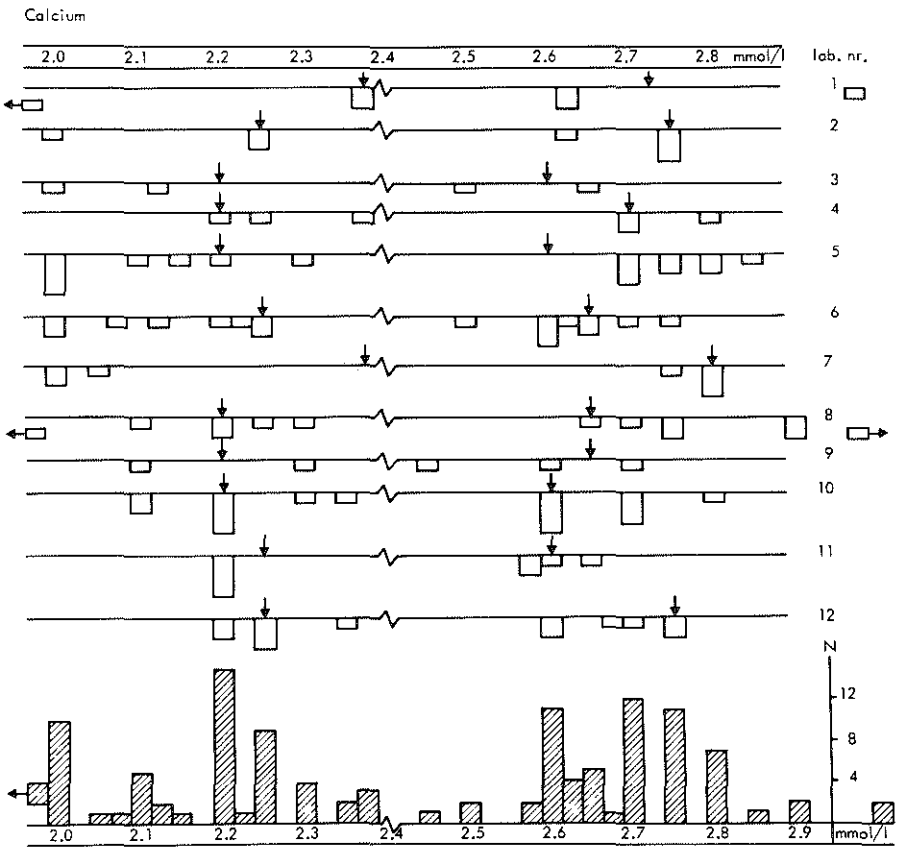


Figure 4-3:

Histograms of action levels given by clinicians per hospital (numbered 1 through 12). The laboratories' upper and lower limits of normal are indicated by arrows.
At the bottom of each figure histograms are given of accumulated data.
The histograms for a number of constituents are presented in addendum II.

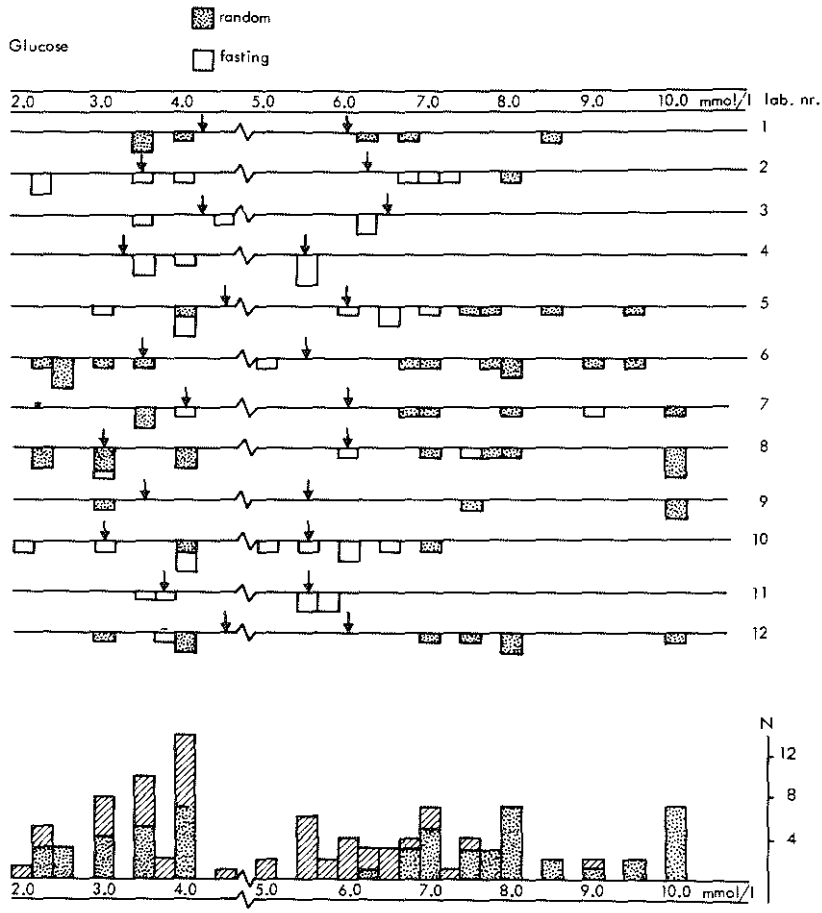


Figure 4-3 cont.

4.3 Discussion

The following points will be discussed respectively:
- the dispersion of normal range values and action levels among clinicians.
- action levels as related to normal range values; the difference between limit of normal and action level as a criterion for required analytical precision.

4.3.1 Dispersion of normal range values and action levels among clinicians.

Figure 4-1 and table 4-2 show us that the dispersion of limits of normal given by the laboratories is in general smaller than the dispersion of clinicians' upper and lower limits of normal.

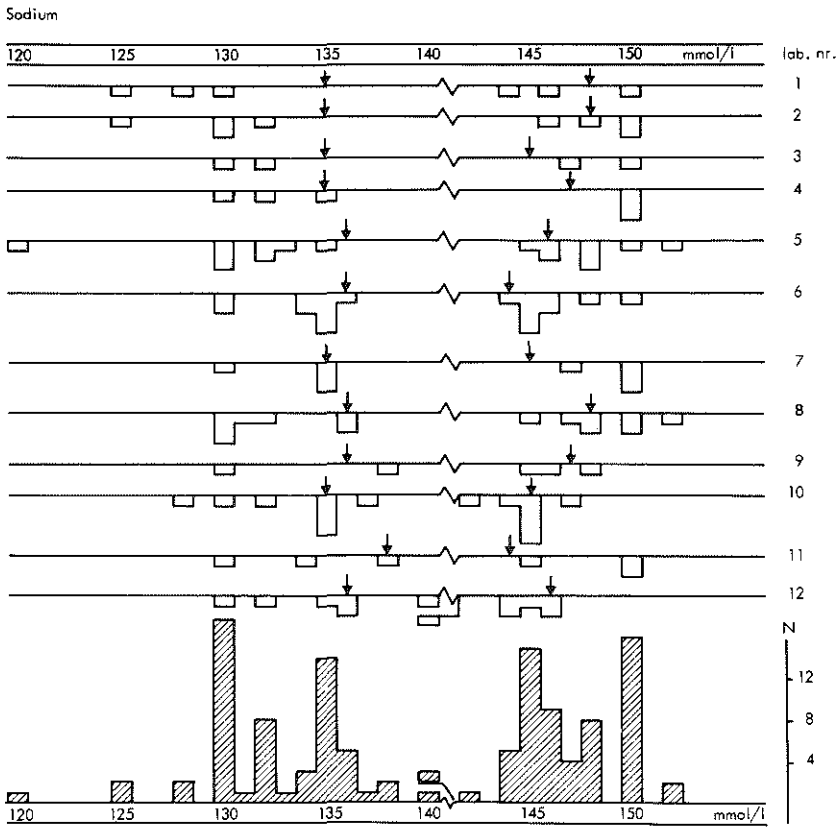


Figure 4-3 cont.

It is striking that the – relatively small – group of clinicians that mentions normal limits as well as action levels tends to apply more strict limits of normal in many cases: particularly for sodium, potassium, chloride and inorganic phosphorous physicians give lower upper range limits than the laboratories (fig. 4-1).
Laboratory results that prompt the clinician to action vary even more among physicians than limits of normal. More than once upper and lower action levels overlap: a value that urges one physician to take action because it is too low, strikes another physician, sometimes in the same hospital, as a disturbing high result. Also we note the clinician's preference for figures ending on 0 or 5 and for even numbers. Evidently there is a need for simplification of the multitude of figures and data that reach the physician in the course of a diagnostic process.
Inspection of fig. 4-3 shows us that variation between physicians within one hospital is often as large as the variation between clinicians of different hospitals. Hence it is extremely difficult to determine whether clinicians respond to the differences in normal range values reported by the laboratories that could be ascribed to factors like systematic analytical bias, differences in samples of the population etc. (Chapter III).

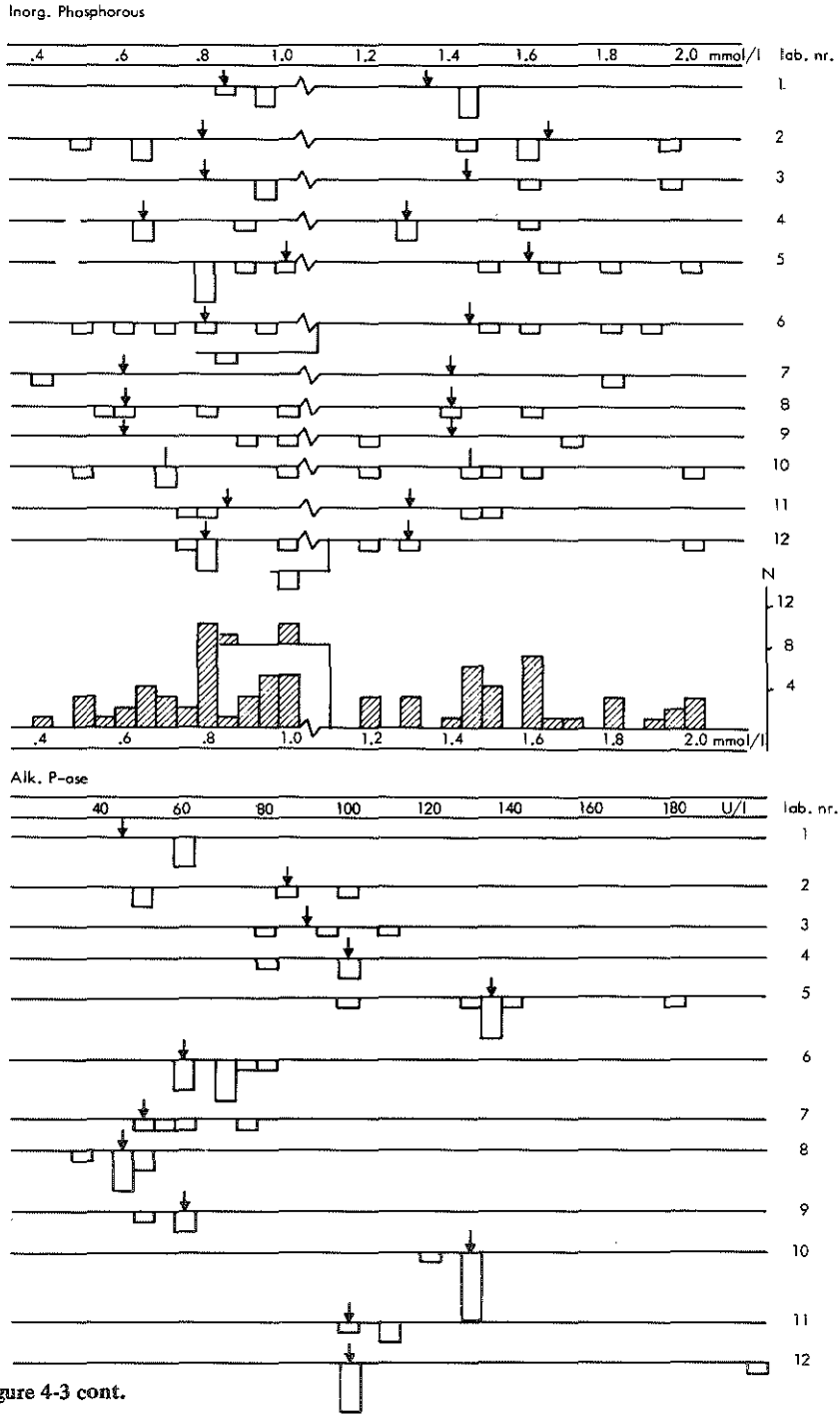


Figure 4-3 cont.

On detection of the large discrepancies in clinician's action levels one question that rises is: can intralaboratory variability account for the differences observed. In other words, is intralaboratory variability such, that the different values reported as action levels in fact cannot be differentiated.
Comparing the CV of the dispersion of clinicians' action levels with the CV of intralaboratory variability (table 4-2) in this group of laboratories it can be seen that for the electrolytes and for the upper limits of total protein and cholesterol indeed intralaboratory variability is even ¼ to ½ of the variability in action levels. There must, however, be other reasons for the observed differences in judgement between clinicians.
Murphy (43) gives seven meanings of the concept "normal":

Paraphrase	Domains of Use	Preferable Term
1. Having probability density function $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$		
(predicated of a metrical character)	Statistics	Gaussian
2. Most representative of its class	Descriptive science (biology, etc.)	Average, median, modal
3. Commonly encountered in its class	Descriptive science	Habitual
4. Most suited to survival and reproduction	Genetics, operations research, quality control, etc.	Optimal or "fittest"
5. Carrying no penalty	Clinical medicine	Innocuous or harmless
6. Commonly aspired to	Politics, sociology, etc.	Conventional
7. Most perfect of its class	Metaphysics, esthetics, morals, etc.	Ideal

The clinician in the process of arriving at a diagnosis "normal" or "abnormal" may have at least 5 of these seven meanings in mind, they are:
- most representative of it's class
- frequently encountered in it's class or not uncommon
- most suited to survival and reproduction
- carrying no penalty or harmless
- most perfect of it's class or ideal

In this regard it is not surprising that normal range levels given by clinicians vary, since each of them will set different priorities and will emphasize different aspects. There is in practice a world of difference between "most commonly encountered", "carrying no penalty" and "ideal".
Apart from the differences in interpretation of the word normal each physician will, as Murphy stresses (43) sample from different populations and for that reason the cut-off points between normal and abnormal will differ. An interesting example is given by Bawkin (3): "...surveyed a group of 1000 children, eleven years of age, from the public schools of New York City and found that 61% of these had had their tonsils removed. The remaining 39% were subjected to examination by a group of physicians, who selected 45% of these for tonsillectomy and rejected the rest. The rejected children were re-examined by another group of physicians, who recommended for tonsillectomy 46% of those remaining after the first examination. When the rejected children were examined a third time, a similar percentage was selected for tonsillectomy, ..."

		1 median differences between lower or upper limit of normal and action levels for physicians naming own normal range				2 median differences between lower or upper limit of normal and action levels for all physicians				3 median intra-lab. variability expressed as 3 CV
					N				N	
Na	LLN	5	mmol/l	3.7%	15	4	mmol/l	2.9%	51	2.7%
	ULN	4	mmol/l	2.7%	16	2	mmol/l	1.4%	51	2.7%
K	LLN	0.35	mmol/l	9.7%	20	0.40	mmol/l	11.1%	50	6.0%
	ULN	0.30	mmol/l	6.0%	20	0.20	mmol/l	4.0%	50	3.6%
Cl	LLN	5	mmol/l	5.2%	11	5	mmol/l	5.2%	26	4.2%
	ULN	4	mmol/l	3.7%	10	3	mmol/l	2.8%	25	4.2%
Ca	LLN	0.10	mmol/l	4.4%	16	0.05	mmol/l	2.2%	49	5.4%
	ULN	0	mmol/l	0 %	23	0	mmol/l	0 %	53	5.4%
P	LLN	0.07	mmol/l	9.2%	10	0	mmol/l	0 %	53	8.7%
	ULN	0.32	mmol/l	23 %	8	0.15	mmol/l	10.6%	32	8.7%
urea	LLN	0.4	mmol/l	14.3%	4	0	mmol/l	0 %	13	9.3%
	ULN	1.6	mmol/l	21 %	14	1.3	mmol/l	17.3%	49	9.3%
creat.	LLN		–	–			–			9.3%
	ULN	0	µmol/l	0 %	19	0	µmol/l	0 %	51	9.3%
gluc.	LLN	0	mmol/l	0 %	11	0	mmol/l	0 %	23	15.0%
	ULN	0	mmol/l	0 %	15	0	mmol/l	0 %	33	15.0%
chol.	LLN		–	–	1	0.25	mmol/l	6.4%	11	10.2%
	ULN	0	mmol/l	0 %	18	0	mmol/l	0 %	51	10.2%
t. prot.	LLN	5	gr/l	8.1%	12	2	gr/l	3.2%	44	6.6%
	ULN	5	gr/l	6.3%	12	0	gr/l	0 %	44	6.9%
ALP	LLN		–	–			–	–		15.3%
	ULN			0 %	21			0 %	51	15.3%
LDH	LLN		–	–	–			–		21 %
	ULN			10.0%	18			9.0%	43	21 %
Hb	LLN	0.3	mmol/l	4.0%	11	0	mmol/l	0 %	46	5.7%
	ULN♀		–	–	2	0	mmol/l	0 %	23	5.7%
	LLN♂		–	–	3	0.5	mmol/l	5.8%	27	5.7%
	ULN	0.5	mmol/l	4.6%	12	0	mmol/l	0 %	43	5.7%

Table 4-4: Median differences between lower (LLN) and upper (ULN) limits of normal and respective action levels expressed in units of measurement and in percentage of the average limit of normal related to median intralaboratory variability observed in this study, expressed as 3 CV (column 3). The median differences between limit of normal and action level (ranges are given in figure 4-5) for physicians naming own normal values are listed in column 1, those for the whole group, including the cases where the laboratory's limit of normal were used for calculations, in column 2. Uncorrected median differences (see text for explanation) for Na (LLN): 3 mmol/l, K (LLN): 0.35 mmol/l, P (LLN): 0.05 mmol/l.

As for action levels the differences in judgment between physicians can be expected to be even more outspoken for the following reasons:

- while for the answer to the question "what do you consider normal" each constituent can be considered separately, the decision to take action is taken only after considering a multitude of aspects. Besides results of physical examination and case history also the consequences and chance of success of the action will play a role.

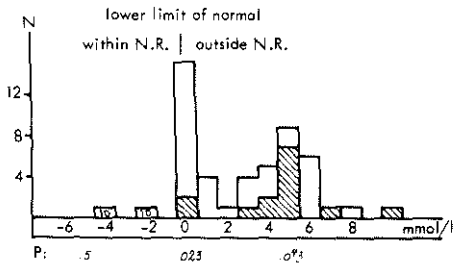
- subspecialization occurs even in the smaller hospitals. It encourages clinicians to be more keen for some determinations and less for others.
- for each clinician the dependence on laboratory results in arriving at a diagnosis is different in relation to the clinical picture. Whereas some see laboratory data merely as an aid to confirm their diagnostic hypothesis, others are convinced that every deviation from the normal, the frequently observed or ideal pattern should be investigated thoroughly as a possible indication of disease.

Although the existence of variations in action levels between clinicians may be plausible, a discussion among physicians on this subject is needed. We refer in particular to the overlap of upper and lower action levels and to the question of importance of laboratory data in relation to other observations.
Differences in interpretation of observations in medical practice have been reported in the literature (7,11,16,22,49,53,64,66,74).
For example, West concludes (64): "Under certain common circumstances, some diabetologists would classify as normal more than half of the one- and two-hour (glucose) values considered to be abnormal by other well qualified diabetologists".
With the advance of computer-aided multivariate analysis of observations important in diagnosis (2, 33, 44, 50, 54) the determination of cut-off points is indicated and agreement should be reached at this point (7).
In the mean time the variations in action levels observed interfere with the conclusion one would want to make from the clinical laboratory point of view.
In fact, this may well be the main reason why this subject has remained untouched for so long. Nevertheless, the data obtained can be usefull to examine the main question of this study: to which determinations should – from the point of view of the practicing physician – priority be given for an effort to improve analytical performance.

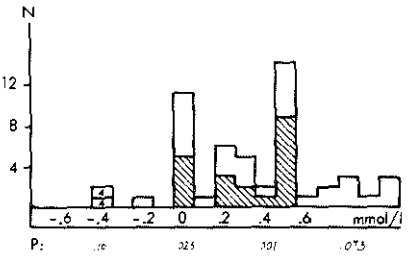
4.3.2 Action levels and normal range limits.

In this study clinicians' action levels are in many cases not similar to the limits of normal set by themselves or by the laboratory (table 4-2, fig. 4-1). In other words, in the process of distinguishing the diseased individual from the normal population the clinician often takes less than the – "usual and convenient" (24) – 1 in 20 chance of erroneously classifying someone diseased as being healthy. This is also demonstrated by the P values in figure 4-5 indicating the probability that an action level forms part of the population of values observed in healthy individuals. In other cases he chooses to be more strict and evidently does not want to take the risk of missing a diagnosis. For a number of constituents his attitude in this respect is different in the upper as compared to the lower decision range. From fig. 4-5 and table 4-4 we see for example, that clinicians tend to be more strict for the upper limits for potassium and calcium than for the lower limit. The opposite seems to be true for inorg. phosphorous. Also, lower limits for creatinine, urea, cholesterol and the enzymes are not of great interest to the majority of this group of physicians.
In fact, Murphy and Abbey (42) express the view, that the use of the normal range is an oversimplification. To put it in extremes "No competent clinician ever makes a

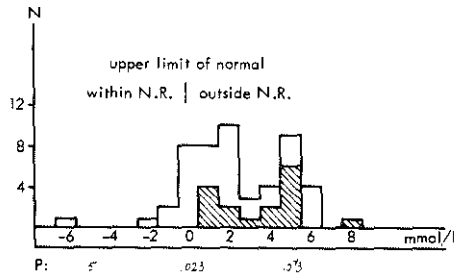
Na - LLN



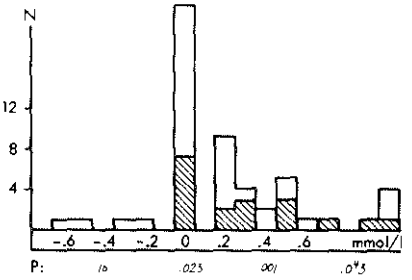
K - LLN



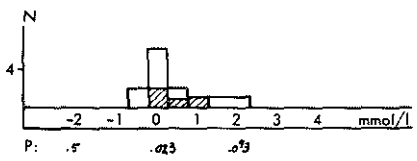
Na - ULN



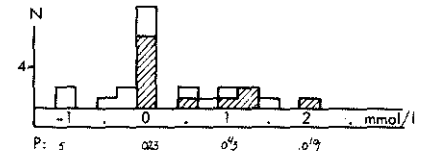
K - ULN



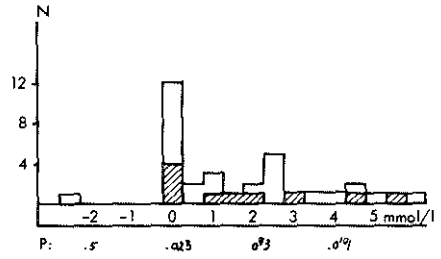
Urea - LLN



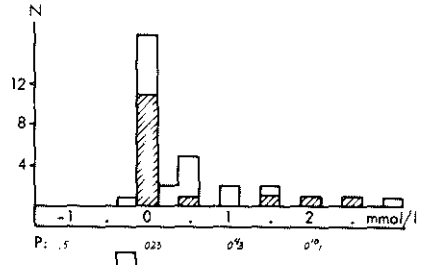
Gluc.- LLN



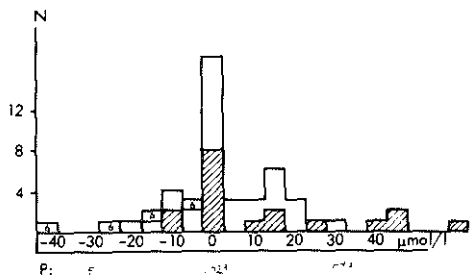
Urea - ULN



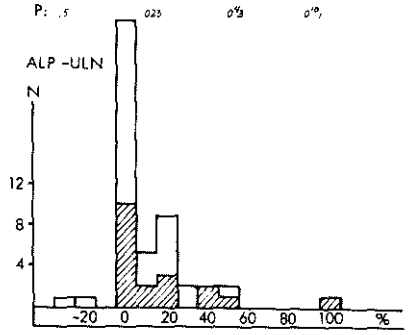
Gluc.- ULN



Creat. ULN



ALP -ULN



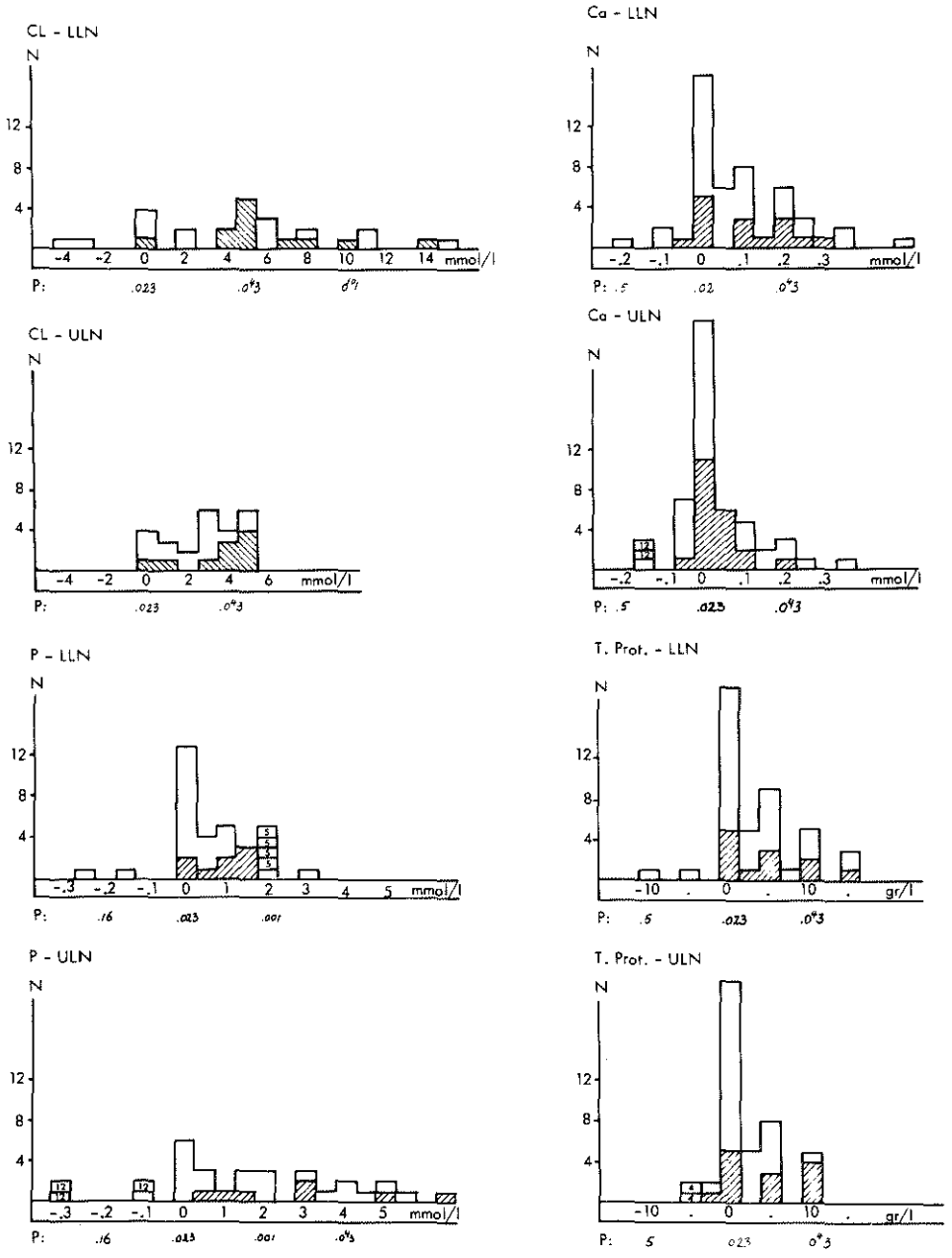


Figure 4-5:

Histograms of differences between lower (LLN) and upper (ULN) limit of normal and corresponding action levels. In each figure "O" represents the limit of normal, negative values represent action levels within the range of normal values, positive values action levels outside these limits. ▨: action levels of physicians reporting own normal ranges.

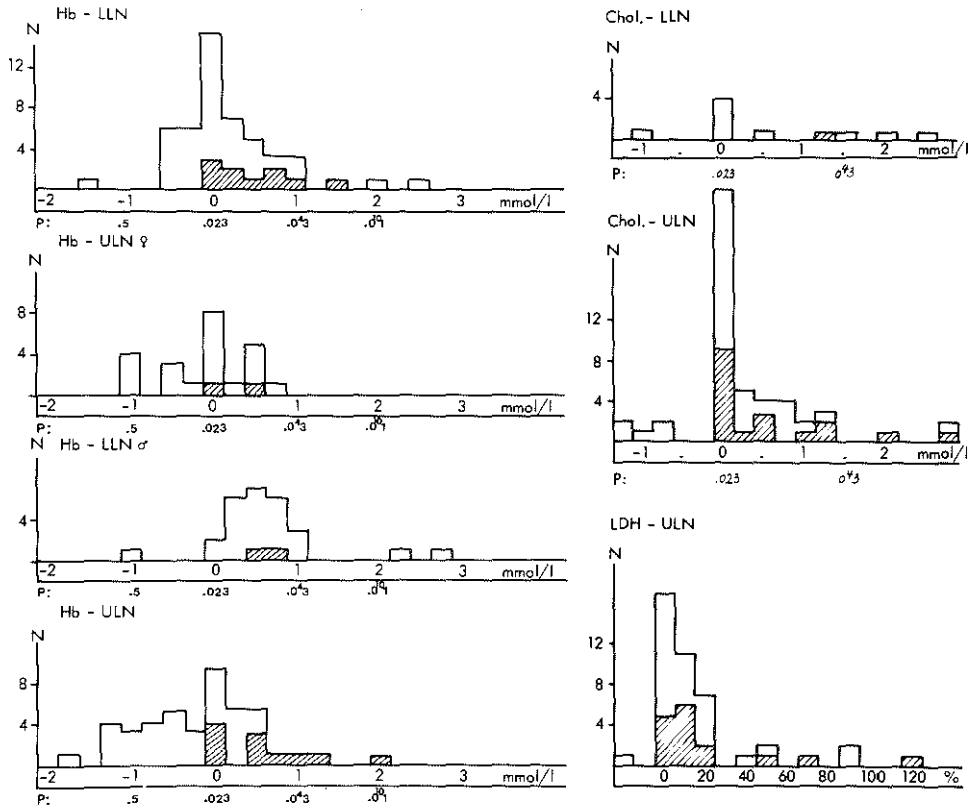


Figure 4-5 cont.

diagnosis on the basis of whether or not the patients value lies within the "normal range" " (42).
Clearly the clinical question is not primarily whether or not a certain value is beyond the limits of normal, but rather whether or not the value found belongs to an alternative to health that calls for action. (39, 42, 47)
The answer to this question is not related to the width of the normal range itself but rather to the location of the range of values in diseased subjects as compared to the range of values in healthy individuals.
Consequently, in order to derive data on clinical requirements for analytical performance in the critical decision ranges between normal and pathological results we propose to consider the difference between limit of normal and the respective action level as a measure.
Laboratory performance, firstly with respect to precision has to be such, that a distinction can be made between action level and limit of normal. According to the foregoing discussion this appears to be a better approach than the requirement that analytical error should not exceed a proportion of the normal range (14, 61a,b, 62, 77). If the latter requirement is met, it only tells us about the laboratory's ability to distinguish the upper limit from the lower limit of normal, which is clinically irrelevant at present.
Also the approach of Young, Harris and Cotlove (75), Statland (69) and Steele (58) to relate allowable analytical error to intraindividual variation is not applicable to the most common situation of wanting to distinguish the diseased individual from the normal population.
The relation of precision to intraindividual variability is, of course, relevant in preventive medicine.
Before drawing any conclusion as to requirements for analytical performance from the differences between limits of normal and respective action levels we will take a closer look at the distributions of these differences first.
While a group of clinicians (shaded in fig. 4-5) gave normal values as well as action levels, the majority indicated to apply the laboratory's normal values. Hence in these cases in the calculation of the differences between action level and limit of normal the laboratory's normal values have been used. The distributions of shaded area's (own normal values) and clear area's (laboratory's normal values) in fig. 4-5 do not differ considerably in most cases and median values (table 4-4) show the same pattern and are fairly close in the two groups. The number of negative values (action level within the normal range), though small, is higher in the group where action levels were related to the laboratory's normal range values.
In chapter III (sections 3.3.1 and 3.3.2) we discussed the differences in normal ranges reported by the laboratories and in a number of cases these differences could be tentatively ascribed to specific causes: systematic analytical bias, sample of the population choosen for determination of the normal range, patient and sample preparation, etc. Some of these causes are relevant for clinicians and others are not. By revisiting the observations of chapter III we may find some explanations for extreme values in the frequency distributions of differences between action levels and limits

of normal of fig. 4-5.
Since in fig. 4-5 action levels are related to the corresponding limit of normal by setting the latter at zero, systematic bias between laboratories can be considered to be cancelled out. For the enzymes, for which systematic analytical bias is considerable, differences between upper limit of normal and action level are expressed as a percentage of the upper limit of normal.
As to the possible regional population difference for phosphate in laboratories 11 and 12 we note that the low upper limits of normal are followed closely by physicians in these hospitals. In hospital 12 three out of four clinicians apply action levels even lower than the upper limit of normal given by the laboratory (see also fig. 4-3). For these cases no extreme values in fig. 4-5 are expected.
Clinicians in hospital 4 ignore however, the laboratory's normal range determined on out-patients for the – low – lower limit for potassium and for the – high – upper limit for total protein.
Four out of six clinicians in hospital 12, which derived its – high – upper limit of normal for calcium from the literature report action levels lower than that limit.
The majority of physicians in laboratory 5, which found a high normal range for inorganic phosphorous in young individuals report action levels far beyond the lower limit of normal.
Two clinicians in hospital 10 report lower action levels for sodium, within the normal range given by the laboratory. This laboratory showed a high systematic analytical bias in the survey, that was not reflected in the normal range (table 3-4).
The upper limit of normal for creatinine in laboratory 6 is high as compared to other laboratories. Five clinicians in this hospital report action levels lower than the upper limit of normal (4 negative values in fig. 4-5).
The three extreme large differences between upper limit of normal and action level for creatinine (62, 45 and 45 μmol/l respectively are reported by American clinicians who evidently apply different criteria.
In the foregoing cases extreme negative or positive values were found for the differences between upper or lower limit of normal and the corresponding action level. They are indicated in fig. 4-5. In fact these values can be omitted. This does however cause changes in the median values of the distributions only in a few cases. In table 4-4 the corrected median values are given. The uncorrected values are mentioned in the legend.
When we now compare the median difference between limit of normal and action level expressed as a percentage of the limit of normal (table 4-2) with intralaboratory variability (table 4-4) we see that for a number of constituents laboratory precision seems to meet the median requirement if 3 CV is used as a measure of intralaboratory variability. This is the case for potassium, inorg. phosphorous and urea. For sodium and chloride laboratories are just short of good performance. The requirements are not met for calcium, creatinine, glucose, cholesterol, alk. P-ase, LDH, total protein and hemoglobin.
The requirements for laboratory performance found are very strict and where they number zero (limited by the number of decimal points) they are extremely hard to

achieve. The results do indicate, however, a pattern of priorities to be set. In chapter VII we will discuss these requirements for the laboratory in the decision range together with those based on a significant change in a patient in the near abnormal range in view of clinicians' satisfaction with test results.

Chapter V

MEDICAL SIGNIFICANT CHANGES IN THE INDIVIDUAL PATIENT RELATED TO ANALYTICAL PRECISION.
POSSIBLE CAUSES OF VARIATION IN LABORATORY RESULTS.
SATISFACTION WITH LABORATORY PERFORMANCE.

5.1 Introduction

In this chapter the answers to the following questions will be discussed:

II Suppose you find in a patient under your treatment the laboratory result given in the table:

Na	K	Cl	Ca	P	urea	creat.	gluc.	chol.	t.prot.	ALP	LDH	Hb
119	2.4	118	2.85	2.52	13.4	230	14.0	9.1	48			5.6
		mmol/l				μmol/l	mmol/l		g/l			mmol/l

You follow the course of this patient and/or the effect of treatment. Please circle the value, that would represent a signicant change, in these cases improvement:

Na	K	Cl	Ca	P	urea	creat.	gluc.	chol.	t.prot.	ALP	LDH	Hb
120	2.5	117	2.80	2.45	12.6	221	13.5	8.6	50			5.9
121	2.6	116	2.75	2.38	11.8	212	12.8	8.2	52			6.2
:	:	:	:	:	:	:	:	:	:			:

Please answer this question for each constituent separately.

III There are a number of factors responsible for the day to day variability of laboratory data. We have the biological variation in the patient, laboratory error, handling of the sample (e.g. technique of sample taking, storage) and possibly other factors. Can you indicate to which extent these factors contribute to this variability?

++;: very great. +: great, ±: moderate, −: none.

- biological ++/+/±/−
- laboratory error ++/+/±/−
- handling of the sample ++/+/±/−
- other ++/+/±/−

IVa For which constituents do you feel a great accuracy of the laboratory is needed; in other words, are there laboratory tests for which you would say: "I could do better work, if the laboratory did a better job".

b Please identify the degree of accuracy desired (state two values between which you want to distinguish).

5.2 Results

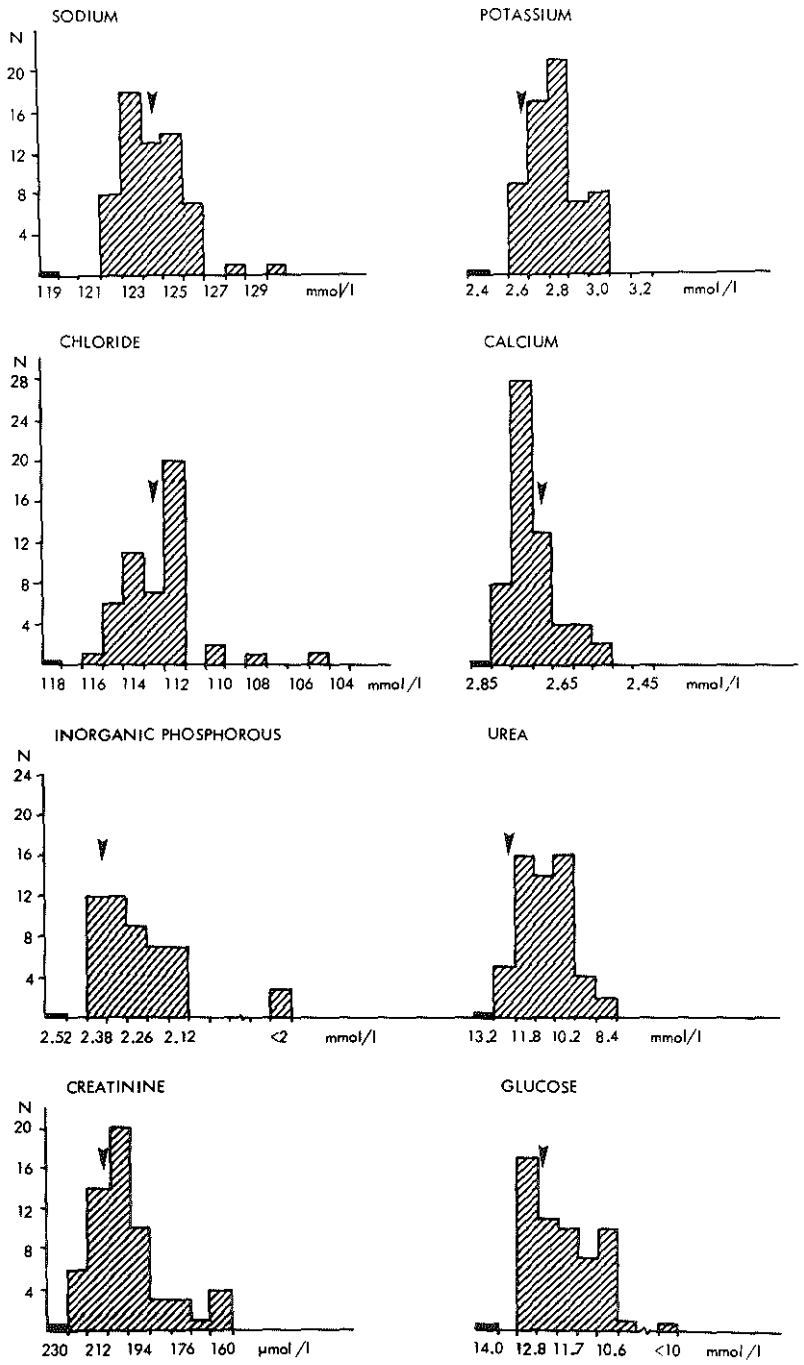
5.2.1 Medical significant changes in a patient

As a starting value in the question "what do you consider a significant change in a

patient" a value was chosen in the near abnormal range, so that:
a. the answer would not be influenced by the nearest limit of normal
b. intralaboratory variability would be known in the same concentration range.
Hence the concentration levels of the controlsera were taken into account.
As a result most starting values were, though not near to the normal range, close to the range of action levels and thus close to the most critical medical decision range, particularly for potassium and cholesterol. The starting values for sodium, inorganic phosphorous and creatinine however, were for most clinicians well outside this range.
Due to methodological differences between laboratories results for enzyme analyses may have numerical values that are considerably different between hospitals and thus a starting value could not be fixed ahead. During the interview a value was chosen 1½ to 2 times the upper limit of normal mentioned at question Ia.
In figure 5-1 the answers given to the question of a medical significant change in a patient are given in histograms. The initial value is at the far left of each graph. An arrow indicates the median intralaboratory variability at this level expressed as 3 SD. This is the rounded off figure for 2.26 $\sqrt{2}$ SD, which can be considered to constitute a significant difference between two individual values when the variability is calculated from ten independent determinations for a controlserum. In table 5-2 the average significant change from the initial value is given in units of measurement as well as in percentage of the initial value, together with median intralaboratory variability (3 CV).
The dispersion of clinicians' answers and intralaboratory variability both expressed as 1 CV are given in column 4 and 5.
In figure 5-3 the medical significant changes are listed per hospital. Here each laboratory's intralaboratory variability (3 SD) is indicated as a bar. The values given by clinicians who, in question IV of the questionnaire reported to be dissatisfied with laboratory performance for that test are shaded.

5.2.2 Possible causes of variation in laboratory results

When asked for the cause of day-to-day variations in laboratory results a general answer was given by a number of clinicians. Many times the question was answered for individual constituents separately.
Biological variation was mentioned 11 times, laboratory error 9 times and sample handling 12 times as most important factor.
Biological variation was mentioned for glucose (16 times), cholesterol (7 times), inorganic phosphorous, total protein and calcium (5 times) and sodium and urea (4 times). Eight other constituents were named 1 or 2 times.
The answers referred to both dietary conditions and physiological variations. Laboratory error was mentioned as the most essential cause of variation for creatinine (9 times) and calcium (4 times). Eleven other constituents were named 1-3 times. For potassium (19 times), total protein (11 times), hemoglobin (10 times), calcium (8 times) and LDH (6 times) sample handling, i.c. tourniquet application, hemolysis


SODIUM
N
20
16
12
8
4
119 121 123 125 127 129 mmol/l
POTASSIUM
2.4 2.6 2.8 3.0 3.2 mmol/l
CHLORIDE
N
28
24
20
16
12
8
4
118 116 114 112 110 108 106 104 mmol/l
CALCIUM
2.85 2.65 2.45 mmol/l
INORGANIC PHOSPHOROUS
N
24
20
16
12
8
4
2.52 2.38 2.26 2.12 <2 mmol/l
UREA
13.2 11.8 10.2 8.4 mmol/l
CREATININE
N
20
16
12
8
4
230 212 194 176 160 µmol/l
GLUCOSE
14.0 12.8 11.7 10.6 <10 mmol/l

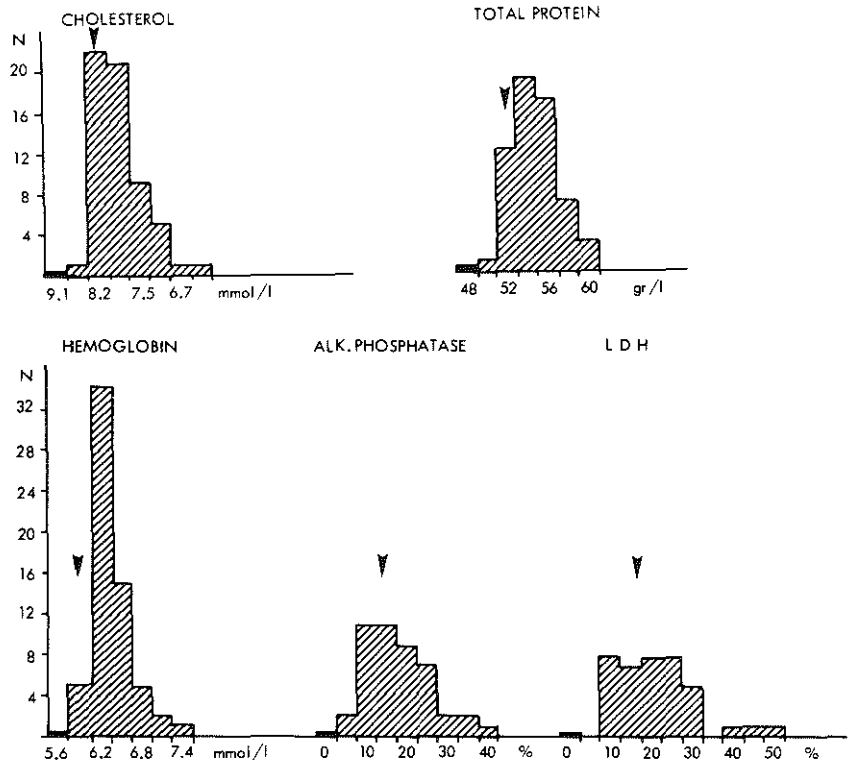


Figure 5-1:

Medically significant change in a patient.
▪: initital value
⩛: median intralaboratory variability (3 SD).

	1 initial value	2 average medically significant change		3 median intra-lab. var. (3CV)	4 dispersion answers clinicians (1CV)	5 median intralab. var. (1CV)
Na	119 mmol/l	5.1 mmol/l	4.3%	3.6%	1.2%	1.2%
K	2.4 mmol/l	0.38 mmol/l	15.8%	6.0%	4.4%	2.0%
Cl	118 mmol/l	5.2 mmol/l	4.4%	4.2%	1.7%	1.4%
Ca	2.85 mmol/l	0.13 mmol/l	4.6%	5.4%	2.7%	1.8%
P	2.52 mmol/l	0.25 mmol/l	9.9%	8.1%	4.5%	2.7%
urea	13.4 mmol/l	2.46 mmol/l	18.4%	8.4%	8.9%	2.8%
creat.	230 umol/l	29.9 µmol/l	13.0%	9.3%	8,0%	3.1%
gluc.	14.0 mmol/l	2.17 mmol/l	15.5%	14.4%	7.9%	4.8%
chol.	9.1 mmol/l	1.33 mmol/l	14.6%	10.2%	5.9%	3.4%
t. prot.	48 gr/l	6.9 gr/l	14.4%	6.6%	4.2%	2.2%
ALP			17.6%	12.6%	7.9%	4.2%
LDH			22.2%	13.8%	10.1%	4.6%
Hb	5.6 mmol/l	0.75 mmol/l	13.4%	5.7%	4.6%	1.9%

Table 5-2: Medically significant change in the individual patient related to intralaboratory variability. Dispersion of clinicians' answers.
column 1: initial value
column 2: average medically significant change expressed in units of measurement and in percentage of the initial value.
column 3: median intralaboratory variability (3 CV).
column 4: dispersion of clinicians' answers (1 CV).
column 5: median intralaboratory variability (1 CV).

etc. were thought to contribute considerably to variation in laboratory results. Six other constituents were mentioned once in this respect.

5.2.3. Satisfaction with laboratory performance.

Fifty-one out of 63 physicians interviewed expressed their concern for the quality of laboratory performance for one or more constituents: Creatinine was mentioned 21 times, hemoglobin 18 times, calcium 19 times, total protein 12 times (protein spectrum 4 times), various enzymes 31 times ("enzymes" 4 times, LDH 8 times, alk. P-ase 6 times, others 13 times), inorganic phosphorous and cholesterol 5 times. Seventeen other tests were mentioned 1-4 times. Several physicians stressed the fact, that sample handling plays a role in the reliability of laboratory results for hemoglobin and calcium.
Thirteen times faulty administrative procedures, clerical errors, delay in reporting, sample exchange and blunders were brought up as a source of aggravation.
These data will also be discussed in chapter VI, section 6.3.1.

5.3 Discussion

The following subjects will be discussed respectively:
- the dispersion among physicians of reported medical significant changes
- medical significant change as related to intralaboratory variability, satisfaction with laboratory performance and causes of variation in laboratory results.

5.3.1. Dispersion of answers among physicians.

When we consider figure 5-1 we see that there is some dispersion in the answers of clinicians to the question: "what do you consider a medically significant change in a patient". This dispersion is of the same magnitude in some cases but in most cases smaller as compared to the dispersion we saw for clinicians action levels (cf. fig. 4-3, totals).
Indeed the decision that a constituent level is medically significant different from a previous value in the same patient seems less complicated than the decision that a value, outside the range of normal values calls for action. In the latter case it is not only a matter of determining that a certain value does not belong to the range of values frequently seen, but more so, the value should fit in a diagnostic pattern and so lead to action.
In the case of a change in an individual patient it will be easier to consider each constituent separately and the decisionmaking variables thought to be responsible for dispersion of action levels in chapter IV (page 48, item 1), i.e. the variety of factors involved in the decision "to take action") will not count as heavily here. The other reasons mentioned in chapter IV for differences in action levels also help to account for the dispersion among clinicians of medical significant changes in a patient:
- subspecialization usually leads to an increased interest in particular tests.
- the dependence on laboratory results as compared to the clinical picture differs from clinician to clinician. Some wait confirmation of their hypothesis only, e.g. the expected effect of their treatment. Others, in a similar setting, will set treatment according to the laboratory result.

Finally, laboratory variability contributes to the variability in answers to the question on medical significant changes.
Particularly with respect to the last point it is interesting to see in table 5-2, column 4 and 5, that median intralaboratory variability is close to the variability of clinicians' answers. In other words, when the majority of clinicians reports for example for calcium, starting value 2.85 mmol/l, values ranging from 2.80 to 2.60 mmol/l as concurrent with a medical significant change we have to realize, that the average hospital laboratory cannot adequately differentiate in this range.

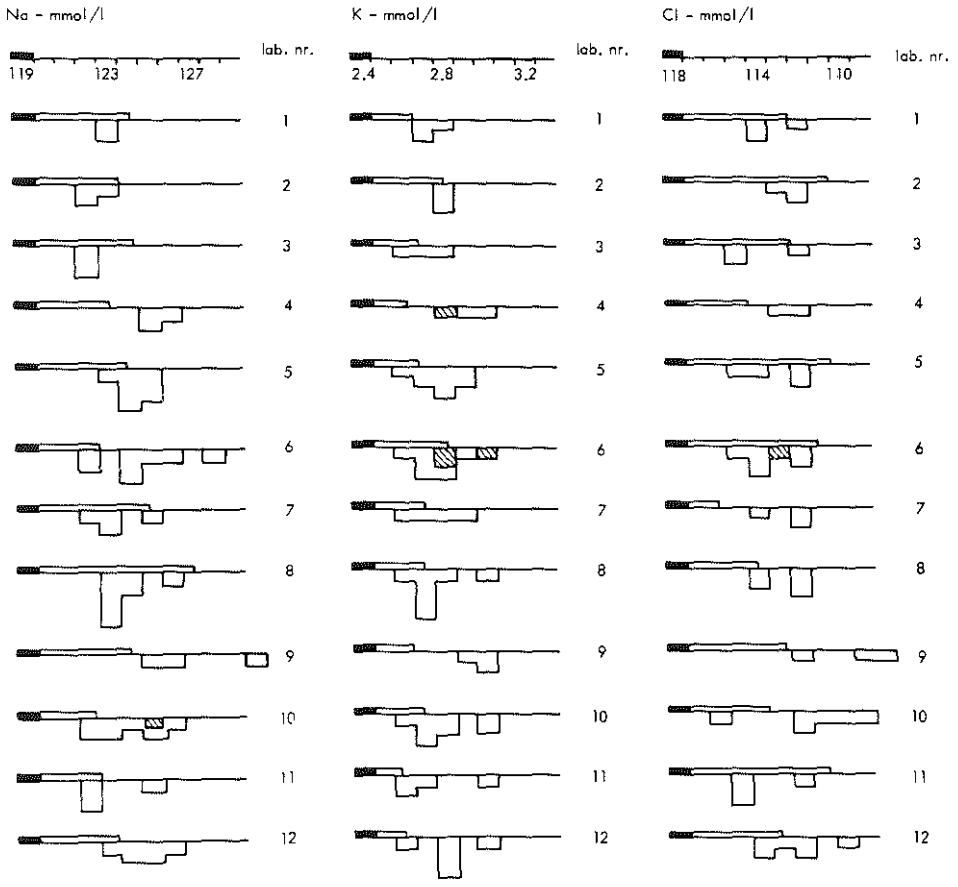


Figure 5-3:

Histograms of constituent levels representing a medically significant change in a patient, given a certain initial value.
: initial value
: answers of physicians not satisfied with the determination
: intralaboratory variability (3 SD).

5.3.2. Average medical significant change and requirements for laboratory performance.

Let us now turn from the dispersion observed among physicians' answers to the average medical significant change. How does the average medical significant change in a patient relate to median intralaboratory variability and how is the situation per hospital?

From table 5-2, column 2 and 3 we conclude, that median intralaboratory precision

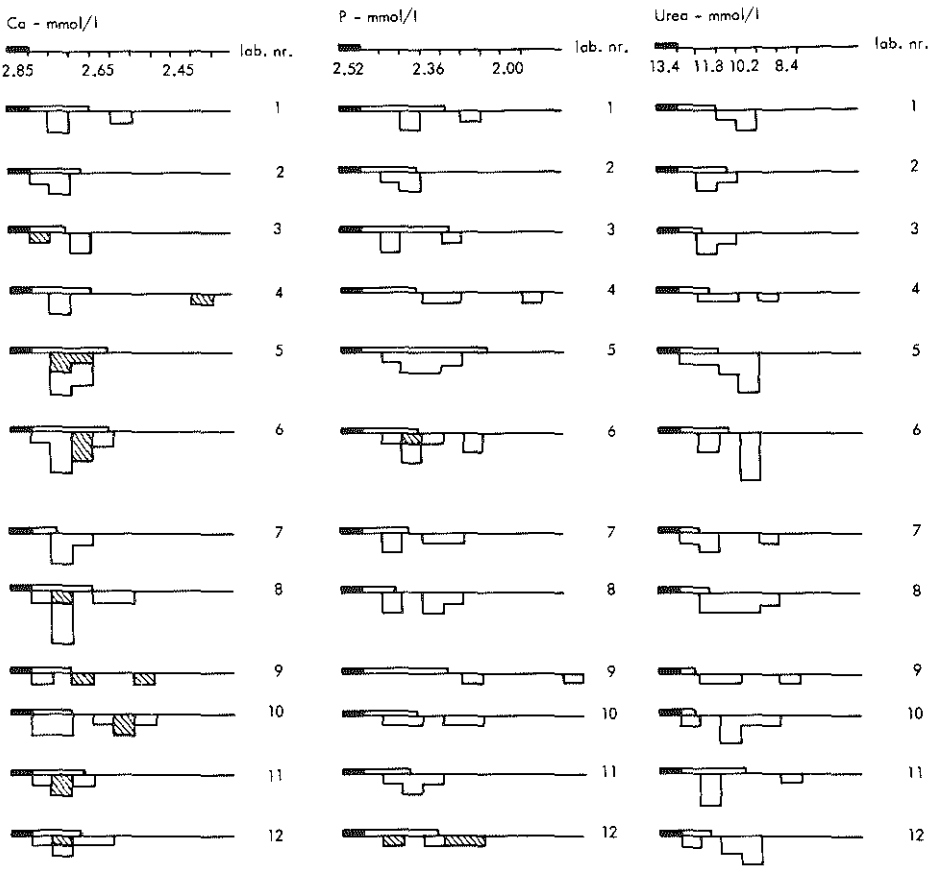


Figure 5-3 cont.

is more than adequate to record changes in the individual patient in the near abnormal range for sodium, potassium, urea, creatinine, cholesterol, total protein, alk. P-ase, LDH and hemoglobin, just adequate for chloride, inorg. phosphorous and glucose and inadequate for calcium.

In figure 5-3 the answers are split up per hospital and the question rises: do clinicians working with a hospital laboratory with small intralaboratory variability respond by reporting small medical significant changes in the individual patient? The fact that clinicians' answers vary to the degree they do does not facilitate the search for an answer to this question.

In chapter III, section 3.3. we found that for all constituents but Na, Cl, glucose and cholesterol the variabilities found in various hospitals were significantly different. For none of these constituents a positive correlation was found between the degree of analytical variability of the laboratory and the average significant change in a patient mentioned by physicians (rank correlation test of Spearman (17)). Yet some remarks can be made when we consider fig. 5-3 and take into account clinicians'

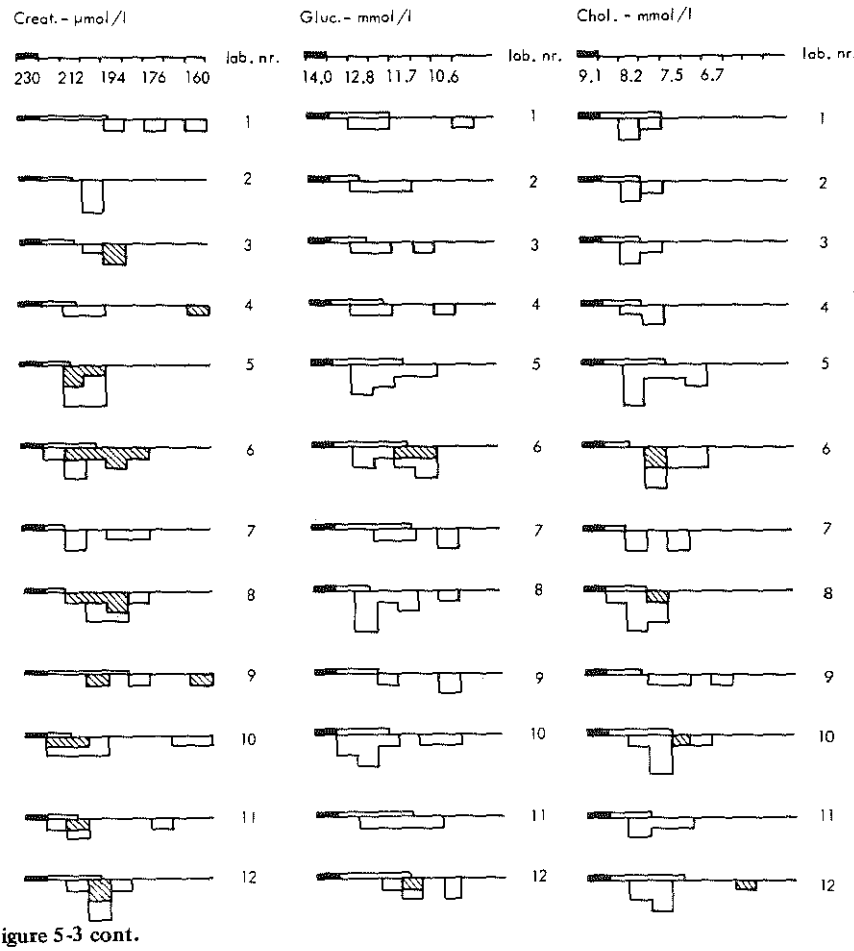


Figure 5-3 cont.

answers to another question of the questionnaire: for which constituents do you feel greater accuracy is needed. Greater reliability of laboratory results was said to be most needed for calcium, creatinine, Hb, total protein and "enzymes".

Three out of four clinicians who are not content with the potassium determination work in hospital 6 with highest intralaboratory variability (SD: 0.36 mmol/l).

In hospital 5 and 6 with highest intralaboratory variability for calcium determinations 3 out of 7 (43%) and 4 out of 9 (44%) clinicians respectively were not satisfied with the quality of calcium analyses. When we rank hospitals according to percentage of clinicians dissatisfied with calcium determinations these two hospitals rank 9 and 10 out of 12. Hospital 7 with lowest intralaboratory variability is one out of three hospitals where none of the clinicians is incontent.

For inorganic phosphorous laboratory 12 ranks 8/12 for intralaboratory precision and 3 out of 4 clinicians who declare to be dissatisfied with phosphorous determinations work in this hospital.

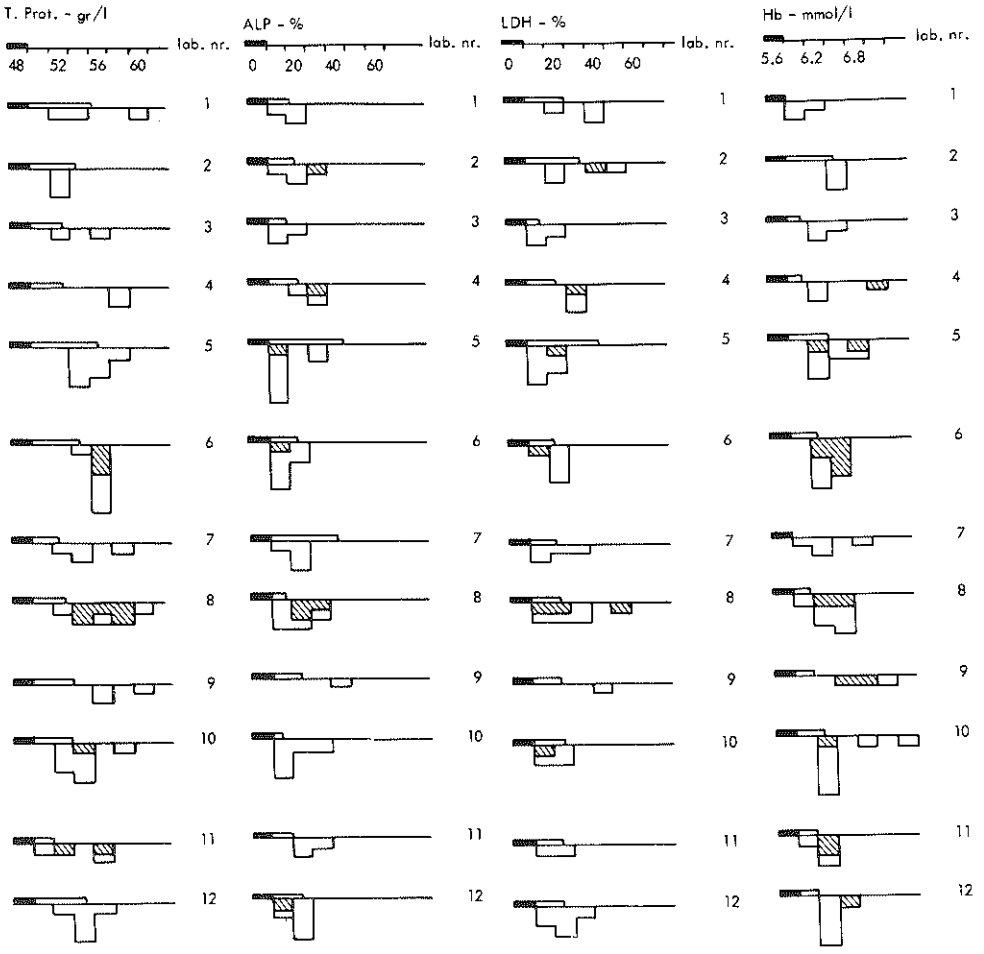


Figure 5-3 cont.

For creatinine we see, that clinicians in hospital 1 and 9, with highest intralaboratory variability (SD: 30 and 36 μmol/l respectively, (other laboratories of the group, SD: 6-23 μmol/l) indicate largest average medical significant changes, 54 and 47 μmol/l respectively as compared to 23-38 μmol/l in the other hospitals. In hospitals 3, 6 and 9 with 5th, third and first highest intralaboratory variability over 60% of clinicians is dissatisfied with laboratory performance of this test. Other hospitals: 0-50%).

No peculiarities as mentioned above were seen for urea, total protein, hemoglobin, Alk. P-ase or LDH.

For hemoglobin, total protein and the enzymes a considerable number of physicians voiced dissatisfaction with the determination, as was the case for calcium and creati-

nine, but no positive correlation could be found for these tests between e.g. the number of dissatisfied clinicians per hospital and intralaboratory variability. This may indicate that dissatisfaction with the laboratory for these tests is caused by other factors e.g. technique of sample taking, patient preparation or unrecognized biological variations.

Summarizing we may conclude that for the follow up of the individual patient in the near abnormal range laboratory performance in general is inadequate for calcium. Strictly taken, also chloride, inorg. phosphorous and glucose determinations need improvement. Differences in precision among hospital laboratories are in some cases discerned by clinicians. In hospitals where laboratories have larger intralaboratory variability clinicians may apply a wider medically significant change and more clinicians may be dissatisfied with the precision of the determination in comparison with clinicians working with more precise laboratories. This situation has been demonstrated for potassium, calcium, inorg. phosphorous and creatinine. Clinicians' dissatisfaction with total protein, hemoglobin, alk. P-ase and LDH determinations may be related to technique of sample taking or patient preparation.
The conclusion with regard to requirements for the laboratory from the clinical point of view derived in this chapter will be related to those derived in chapter IV in chapter VIII.

Chapter VI

CLINICIANS' ATTITUDE WITH REGARD TO ORDERING OF REPEAT TESTS AND CONSULTATION OF THE LABORATORY IN CASE OF AN UNLIKELY RESULT.

FAST SEMI-QUANTITATIVE TEST RESULTS.

MOLECULAR UNITS IN MEDICAL PRACTICE.

6.1 Introduction

In this chapter the answers to the following questions of the questionnaire will be discussed:

Va. Do you ever order a repeat test – on a new sample – before making a medical decision? always/often/sometimes/never

b. Suppose the value reported for a repeat test is substantially different from the first time. How do you proceed:
- you order the test for a third time always/often/sometimes/never
- you consult with your laboratory always/often/sometimes/never
- you disregard the value that is least likely in the given situation always/often/sometimes/never
- other

VI Are there blood constituents – possibly others than the ones named in the table – for which you would prefer a fast semiquantitative result (within an hour) over an accurate value on longer terms?
Please state the degree of accuracy desired (two values between which a distinction is to be made).

IX. a. Does your laboratory report in molecular units? yes/no

b. If so, since when? 1970/1971/1972/1973/1974/1975

c. How long did it take you to get used to the new units? < 3 months/3-12 months > 1 year

d. Have molecular units enabled you to give better treatment to patients? yes/somewhat/no

e. Have molecular units given you more insight into biochemical processes? yes/somewhat/no

f. Do you convert mmol/l into mg% before making a medical decision? always/often/sometimes/never

g. Do you convert mg% e.g. from the litterature into mmol/l for optimal interpretation? always/often/sometimes/never

6.2 Results

6.2.1. Repeat test requests and proceeding in case of a highly unlikely result.

As to the question whether a repeat test is ordered before a decision is made the answers were as follows:

27/63 clinicians order repeat tests always or often.

17/63 clinicians report to always ask for a repeat test if the first result is incompatible with their expectations.

9/63 clinicians indicate that the frequency of repeat tests ordered is primarily dependent on the importance of the decision to be made on the basis of the laboratory result.

10/63 clinicians seldom or never reorder tests.

When a result of a repeat test differs considerably from the first result the procedures followed are:

24/63 clinicians request a third test to determine which of the two divergent results was correct, or reject the least likely result of the first two, depending on the decision to be made.

25/63 clinicians order a third test always and reject one out of three.

8/63 clinicians sometimes or never order a third test and reject the least likely result most of the time.

6/63 clinicians request third tests always or often and reject never, considering all values in their decision.

Consultation with the laboratory in the case of divergent results is always or often done by 40 physicians.

Fourteen clinicians sometimes consult the laboratory, taking into account the importance of the situation (2x). Nine never consult the laboratory, three of these (all working in the same hospital) declare explicitly that there is no use in talking to the laboratory.

6.2.2. Speed of test versus reliability.

When asked for which tests a semiquantitative test would be preferred over a quantitative one, provided that the result would be available in a short turn-around time, clinicians mentioned the following constituents:

glucose	(29 times)	CPK	(4 times)
Hb	(9 times)	Amylase	(4 times)
K	(11 times)	NH_3	(4 times)
creat.	(9 times)	Na	(3 times)
SGOT	(9 times)	eight constituents were named twice,	
SGPT	(9 times)	thirteen others once.	

Some clinicians who voiced their wish for faster results would not accept less accuracy (glucose (2x), K (5x), creatinine (2x), Hb and urea (1x)).

Seventeen clinicians did not have special interest with regard to this question. Seven

of them reported stat-service to be adequate. Four explicitly stressed the importance of accuracy.

6.2.3. Molecular units

Of the 12 hospital laboratories participating in this study 7 changed to molecular units in 1971 (39 clinicians), 1 in 1973 (4 clinicians) and 1 in 1974 (9 clinicians). One clinician changed to a hospital where molecular units were used in 1976 and thus started using these units in that year.

The answers to the question how long it did take to get used to the new units were distributed as follows:

period	less than 3 months	3-12 months	more than 1 year
number of clinicians	15	20	17

In three hospitals (18/52 clinicans) the change to the new units was reported by some clinicians (4/18) to have been planned well. As one interviewer put it: "like the change from left to right traffic in Sweden".

In these three hospitals the answers were distributed as follows:

period	less than 3 months	3-12 months	more than 1 year
number of clinicians	10	7	1

As to the question whether molecular units enabled clinicians to give better treatment to patients the answers were:

better treatment	yes	somewhat	no
number of physicians	4	2	46

The 4 clinicians answering positively to this question mentioned osmolarity (3 times) and electrolytes once.

A larger number of clinicians gained more insight into biochemical processes while using molecular units:

more insight	yes	somewhat	no
number of physicians	5	13	34

Examples given by those who reported to have gained insight were:
osmolarity (6x), "ions" (5x), relationship T_3-T_4 and relationship cholesterol-cholic acids (1x).

A number of clinicians report to still convert mmol/l into mg% before taking a medical decision:

frequency	always	often	sometimes	never
number of physicians	1	4	19(11)	28(5)

The numbers in brackets represent the numbers of physicians that report exceptions to their answer; thus, out of 47 clinicians answering to convert "sometimes" or "never" 16 report to convert mmol/l into mg% "always" or "often" for one or more constituents.

Particularly mentioned were: glucose (15x), urinary glucose (2x), urea (7x), calcium (4x), creatinine (2x), phosphate, Ca. P product, cholesterol and Hb once.

Most physicians are still familiar with the weight per volume units, as becomes apparent from the frequency of converting mg% into mmol/l when these units are presen-

ted e.g. in the literature

frequency	always	often	sometimes	never
number of clinicians	2	5	6	39

6.3 Discussion

6.3.1. Repeated testing and procedure followed in case of an unlikely result.

Fifty two out of 63 clinicians always or often repeat a test before taking a medical decision. Half of them do so in particular when the first result is incompatible with their expectations or when the decision to be made is important.
Repeat tests contribute considerably to increasing workloads in the clinical laboratory and to increasing costs of health care (8) and in this respect it is interesting to discuss the clinicians's habit to order repeat tests.
It is the clinician's responsibility to the patient to repeat tests when results are incompatible with their expectations or when the decision to be made is important. To routinely repeat tests before making a decision may reflect the clinician's doubt about e.g. the dietary status of the patient, the technique of sample taking or the analytical reliability of the laboratory.
In this study we described (section 5.2.3.) for which analyses clinicians reported to think laboratory performance unsatisfactory. We see some correlation between the number of tests per hospital for which dissatisfaction is expressed and the number of clinicians routinely repeating tests. In hospitals 5, 6, 8 and 10, 38-44% of clinicians routinely repeat tests before making a medical decision, while 2.1-4.5 tests per clinician were named that need improvement (mean 3.6). In the other Dutch hospitals: 4, 7, 9 and 11 a dissatisfaction with analyses of 0-3.5 (mean 2.2) per clinician is recorded and in these hospitals 25-33% of clinicians orders retests routinely.
We note the fact that four large hospitals constitute the first group, while the second group represents smaller hospitals. In the hospitals in the United States and Canada 3/4 and 2/3 clinicians report to routinely repeat tests, while they were most satisfied with the analytical performance of the laboratory of all. This incongruency may be due to a different general attitude in these countries towards laboratory results.
In two out of four hospitals where dissatisfaction with laboratory results and frequency of repeat tests is high overall intralaboratory variability ranks* 8 and 9 out of 9 Dutch hospitals respectively and here intralaboratory variability may be responsible for the clinicians' behaviour. In the two other hospitals intralaboratory variability ranks 1 and 3 out of 9 respectively in the Dutch group. There must be another reason for the clinicians' relative dissatisfaction in these hospitals.

*For each constituent the average coefficient of variation for both controlsera is calculated for each hospital. Hospitals are ranked according to average coefficient of variation. For each hospital ranknumbers for all constituents are summed and hospitals are ranked for overall intralaboratory precision according to these ranksums.

In both hospitals 4 out of 8 clinicians (50%) noted, that clerical errors, mix-up of samples etc. were a source of aggravation in response to question IV of the questionnaire as to needed improvement (section 5.2.3). Similar remarks were made by less than 30% of clinicians in 4 other hospitals.
In one of the two hospitals 6/8 clinicians report to never consult the laboratory in case of an unlikely result. Three out of these six note that "there is no sense in talking to the laboratory". No similar remark was heard in any other hospital in this study.
In conclusion, there is an indication that the clinicians feeling of satisfaction with laboratory performance results in a relative low repeat test rate. In 2/4 hospitals, where stated frequency of retesting is high intralaboratory variability is high, in the two other hospitals bad communications and suspected clerical errors seem to be involved. It has been reported by Mc Swiney and Woodrow (41) and by Grannis et al. (30) that clerical errors, mix-up of samples etc. may account for more than 50% of the errors in the clinical chemistry laboratory.
Ten clinicians out of the total of 63 report to seldom or never repeat a test. Their motivation for not doing so may range from relative heavy reliance on the clinical picture and/or a great confidence in the laboratory, to dispair, as one of the clinicians interviewed indicated: "there is no sense in repeating tests. The laboratory will tell you the second value is the same as the first anyhow".
When upon repeated testing a value is reported that is considerably different from the first one 24 physicians choose – depending on the decision involved – between (a) reordering the test for a third time and (b) rejection of the result that is least likely in the given situation. A similar number of physicians (25) orders the test for a third time and rejects one out of the three results. Seventeen out of twenty-five contact the laboratory in this situation. Four out of eight clinicians who reject the least likely result most of the time and retest never, do contact the laboratory. All six clinicians who always order a repeat test, but never reject, consult the laboratory.
Motivations of clinicians to consult the laboratory were: (a) to be informed on possible blunders (mixup of samples, mistakes in calculations etc.) and (b) to inform the laboratory on the situation. Of those who said "sometimes or never to consult the laboratory", 3 did so only if it was an important case, 3 others reported that there was little use in communicating with the laboratory.

6.3.2. Speed of tests versus reliability.

There is a high demand for fast semiquantitative results as compared to a more accurate result on longer terms for glucose (named 27 times). Two clinicians did state that they want fast and accurate results. Out of 19 clinicians who answered the question of desired accuracy by stating two values between which a distinction is to be made, eleven gave figures for glucose of 1 or 2 mmol/l or 10-15% which constitutes about the difference the average laboratory can detect using a quantitative procedure (median 3 CV in this study: 15% (table 3-1)). Six, more realistically,

mentioned differences of 4-10 mmol/l or 30 - 50% at this question. Two more stated to be interested in the differential diagnosis of hyper- or hypoglycaemia.

6.3.3. Molecular units

In 1970 the Dutch Association of Clinical Chemistry decided to recommend to its members the use of SI units (5, 67), the system of units constructed by the Conférence Générale de Poids et Mesures. The association adopted the proposal of the Subcommittee for Standardisation of Units in Clinical Chemistry of the International Federation for Clinical Chemistry (IFCC) to use molecular units where possible to record mass concentration in reports of analytical results in clinical chemistry. The majority of laboratory directors has responded positively to this recommendation and molecular units have been in use for two to five years by 52 clinicians (9 hospitals) cooperating with this study. Since the decision as to the introduction of the SI-system with or without secondary agreements is presently being considered in many countries it is of interest how clinicians have responded to the new units (10) and to this purpose some questions on this subject have been added to the questionnaire.

This study reveals, that 17 out of 52 clinicians needed more than a year to get used to the new units. Remarks such as: "I'll probably never learn" or "I am still not used to them" were heard from 11 members of this group. A total of 21 clinicians still frequently convert for one or more constituents mmol/l into mg% before making a decision. There is some evidence that thorough preparation of the introduction of the new units leads to a decrease of the period of time needed to get adjusted. Seven out of 52 clinicians report to always or often convert mg% e.g. from the literature, into mmol/l for optimal interpretation.

A few clinicians have experienced advantage in the treatment of patients using the mmol. One clinician noted the advantage of molecular units in diabetic coma simultaneously with the impossibility of routine treatment of diabetes with these units. A diabetes specialist, familiar with the mmol from research studies, thought the same unit inconvenient in routine practice.

It is interesting to note, that glucose was mentioned by 17 out of 52 clinicians as being a constituent for which they frequently convert mmol/l into mg%.

In conclusion, in the present state of the art, the molecular unit does not seem to contribute positively to medical treatment. In the case of glucose the objections to the molecular unit seem to be fundamental.

With respect to the procedure of introduction of the new units at least two aspects are to be taken into account:

- There is a true mental inertia comparable to, for example, the lack of immediate response of the individual, familiar with degrees Fahrenheit to a report of a patient running a fever of 40.5 degrees Celsius even if conversion factors are known.
- For a value or numerical figure to become familiar the absolute number is important.

A remark relevant in this respect was "I have nothing against the mmol, but the old units were so much more convenient".

It seems that the more decimal points are introduced, the less popular the figure will become. Unfortunately this is the case for many constituents when mg/% are converted to mmol/l:

calcium:	10.4 mg% = 2.60 mmol/l		cholesterol	250 mg% = 6.5 mmol/l	
glucose:	120 mg% = 6.6 mmol/l		uric acid	9.0 mg% = 0.53 mmol/l	
urea:	30 mg% = 10.8 mmol/l				

Basic didactic techniques are needed to overcome these difficulties.

Chapter VII

INTERPRETATION OF LABORATORY RESULTS FROM ANOTHER HOSPITAL

7.1 Introduction

The answers to the following questions of the questionnaire will be presented in this chapter:

VII Suppose a patient is referred to you while laboratory data have been determined in another laboratory than your own. How do you proceed?
- you evaluate the results with the normal values you usually apply in mind yes/no
- you ask for a repeat of all tests yes/no
- you ask for a repeat of clinically relevant tests only yes/no
- you relate the data to the normal values of the laboratory in question yes/no
- if these normal values are not directly available do you inquire? yes/no

VIIIa Suppose a patient is referred to you from another hospital. In the table the laboratory value reported for this patient is given. You ask for a repeat test in your own laboratory and find the other value given in the table. Is there a change in the condition of the patient? yes/no

b What would be your answer if the two values were results for one patient both determined in your own laboratory? yes/no

	Na	K	Cl	Ca	P	urea	creat.	gluc.	chol.	t.prot.	Hb
			mmol/l				μmol/l	mmol/l		gr/l	mmol/l
1st	136	3.5	108	2.40	0.95	3.3	82	3.5	4.5	73	8.8
2nd	141	4.2	102	2.65	1.25	5.6	100	5.5	6.1	68	7.8

7.2 Results

With regard to laboratory results from another hospital clinicians report to act as follows:

25/63 repeat clinically relevant tests after relating the results to the normal values of the other hospital.

16/63 repeat clinically relevant tests after relating the first results to the normal range of the own laboratory (e.g. electrolytes) and to the normal range of the other laboratory (particularly for enzymes).
8/16 say to repeat only abnormal or unexpected values.

13/63 consider the result with own normal values and the other laboratory's values in mind, but repeat all tests.

7/63 do not pay much attention to the reported result but repeat all tests or clinically relevant tests.
2/63 assume the differences between laboratories to be negligible.

Out of the total of 63 clinicians 8 say that they accept the other laboratory's results particularly if the hospital laboratory is known.
If the normal values of the other laboratory are not known 18/63 declare to seek this information. Ten more clinicians do so only if the situation is acute or important, if there is doubt, if it concerns enzymes or hormones, or if a repeat test is difficult.
In table 7-1 clinicians answers are given to the question if two successive values for one patient given in the table, represent a change in the condition of the patient if (a) the first value was determined in another laboratory and the second one in the own laboratory and (b) if both values were determined in the own hospital laboratory. The two values were chosen to represent successive results for the same patient in the normal range differing about ½ of the range of normal values. In chapter V we discussed which change in an individual patient clinicians would consider significant when given a starting value in the near abnormal range. From the frequency distributions of their answers (figure 5-1) we can deduce how many clinicians would consider the change given in this chapter a significant change in the near abnormal range. In the last column of table 7-1 these numbers of clinicians are given.
In table 7-2 a comparison is made between intralaboratory variability and interlaboratory variability. The latter is derived by calculating the dispersion (coefficient of variation, CV) of the first values of the ten determinations each laboratory carried out for the controlsera.

	1			2	3	4
				number of clinicians considering this difference a change in the condition of the patient if:		
	first value	second value	difference	1st value other lab. 2nd value own lab. (normal range)	both own lab. (normal range)	both own lab. (near AN range)
Na-mmol/l	136	141	5	2	10	39
K-mmol/l	3.5	4.2	0.7	24	35	63
Cl-mmol/l	108	102	6	1	10	46
Ca-mmol/l	2.40	2.65	0.25	29	41	58
P-mmol/l	0.95	1.25	0.30	15	19	33
urea-mmol/l	3.3	5.6	2.3	9	21	22
creat.-µmol/l	82	100	18	15	19	22
gluc.-mmol/l	3.5	5.5	2.0	9	16	32
chol.-mmol/l	4.5	6.1	1.6	19	27	54
t.prot.-gr/l	73	68	5	3	5	13
Hb-mmol/l	8.8	7.8	1.0	35	43	54

Table 7-1: Numbers of clinicians considering a change in constituent level in a patient as given in column 1 indicative of a change in the condition of the patient if:
- the first value was determined in another laboratory and the second one in the own hospital laboratory (column 2)
- both values were determined in the own laboratory (column 3)
- the change as given in column 1 took place in the near abnormal (AN) range (column 3). These data are derived from figure 5-1.

The ratio between interlaboratory variability and median intralaboratory variability is given as well as the number of laboratories where intralaboratory variability exceeds observed interlaboratory variability on this occasion.

7.3 Discussion

Few clinicians accept results from other hospitals as such. The majority repeats clinically relevant tests or abnormal tests only. It can be noted that in the latter case the non-aberrant result is accepted tacitly. A considerable number of clinicians (20/63) repeats all results no matter what.
The greater confidence of clinicians in their own laboratory as compared to another laboratory also appears from table 7-1. Given two hypothetical values for the same patient a change in the condition of the patient is concluded more often if both determinations were carried out in the own hospital laboratory as compared the situation where the first determination was reported by another laboratory and the second by the own laboratory.
In either situation changes for K (3.5-4.2 mmol/l), Ca (2.40-2.65 mmol/l), Hb (8.8-7.8 mmol/l) and cholesterol (4.5-6.1 mmol/l) are considered representative of a change in the condition of the patient most frequently, changes for glucose (3.5-5.5 mmol/l), creatinine (82-100 μmol/l), urea (3.3-5.6 mmol/l) and P (0.95-1.25 mmol/l) in less cases. A change in total protein level of 73-68 gr/l, of Cl of 108-102 mmol/l and of Na of 136-141 mmol/l is considered significant by few clinicians only.
While the values were chosen essentially in the normal range (table 7-1 column 1), it could not be avoided that for a number of clinicians one of the values was near their personal action level for that constituent. This was particularly the case for K, Ca and Hb. When comparing the data of this chapter with those of chapter V on a medically significant change in a patient in the near abnormal range (figure 7-1, column 4) we see that for all but two constituents (urea and creatinine) the same change in a patient is considered significant by more clinicians if the starting value approaches the near abnormal range as compared to the normal range.
In summary, firstly, a change of about ½ of the normal range in an individual patient has – clinically – a different meaning for each constituent. Secondly, the same change for a particular constituent level is differently interpreted if it occurs in a different concentration range. In the near abnormal range clinicians tend to be more strict than in the normal range.

It is generally assumed that reproducibility within the laboratory is 3 to 4 times better than reproducibility between laboratories (29, 52). We do not find this to be true in every case and confirm in this study the recent data of Gilbert (26) who states on the basis of data of the College of American Pathologists Survey Program, that 2/3 to 3/4 of analytical errors are caused by factors within the laboratory. In this study each laboratory tested the controlsera ten times. If we compare the median intralaboratory variability in these ten results with the variability in the

		intralab. var. (CV)	interlab. var. (CV)	ratio inter/intra lab. var.	number of labs. with intralab. var. equal to or exceeding interlab. var.
Na	I	0.9 (0.3-1.4)	1.8	2.0	0
	II	1.2 (0.7-2.1)	2.1	1.8	1
K	I	1.2 (0.5-2.3)	2.7	2.3	0
	II	2.0 (1.1-3.7)	2.8	1.4	1
Cl	I	1.4 (0.4-2.0)	1.6	1.1	4
	II	1.4 (0.7-2.4)	1.6	1.1	5
Ca	I	1.8 (0.9-2.6)	3.8	2.1	0
	II	1.8 (1.0-4.0)	2.5	1.4	4
P	I	2.9 (1.5-8.3)	4.3	1.5	4
	II	2.7 (1.5-5.6)	7.1	2.9	0
urea	I	3.1 (1.1-8.7)	11.0	3.5	0
	II	2.8 (0.9-7.9)	6.2	2.2	1
creat.	I	3.1 (1.2-7.1)	6.1	2.0	1
	II	2.7 (1.2-7.4)	6.2	2.3	1
gluc.	I	5.0 (2.2-7.1)	12.8	2.6	0
	II	4.8 (2.3-6.3)	9.0	1.9	0
chol.	I	3.4 (1.7-5.1)	7.5	2.2	0
	II	3.8 (2.0-5.7)	7.3	1.9	0
t.prot.	I	2.3 (0.9-4.3)	2.5	1.1	6
	II	2.2 (1.3-4.2)	3.4	1.5	2
ALP	I	5.1 (3.1-14)	33	6.5	0
	II	4.2 (1.2-12)	38	9.0	0
LDH	I	7.1 (3.0-12)	25	3.5	0
	II	4.6 (1.3-12)	28	6.1	0
Hb	I	1.9 (1.2-3.7)	3.3	1.7	2
	II	1.9 (0.9-3.7)	2.3	1.2	3

Table 7.2: Intralaboratory variability related to interlaboratory variability.

first reported results for each controlserum by each laboratory (table 7-2) we see that intralaboratory variability is in general 1.1-3.6 (mean 1.9) times larger than median intralaboratory variability. Exceptions are the enzyme determinations, where interlaboratory variability is 3.5-8.6 times larger than the variability within the average laboratory. This stresses the need for the development of a means of communication between hospitals for these determinations in particular (see Chapter III, section 3.3.1).

As mentioned in chapter III intralaboratory variability when studied in the way it was done in this study is, if anything, more likely to be smaller than larger than true intralaboratory variability. Yet, for all constituents, except glucose, cholesterol and the enzymes there is a number of laboratories that exhibit in this survey an intralaboratory variability exceeding the variation between hospitals (Fig. 7-2, last column). In other words, based on this particular set of data, clinicians in these hospitals have no reason to rely more on results from their own laboratory as compared to results from other hospitals in this group.

Chapter VIII

SUMMARY AND CONCLUSIONS

The main question leading to the work described in this thesis is: how accurate and how precise do measurements in clinical chemistry have to be to provide good service to the clinic.
In clinical chemistry required precision and accuracy, or rather allowable variability is hard to establish. Firstly the determinations (inclusive sample taking) have seldom been performed with a variability so small that it did not interfere with the biological variations to be measured. Secondly, medical decision making is a process of pattern recognition in which the laboratory result plays an important role in some cases and a minor one in others.
A project was set up to study the present situation: how do physicians use laboratory results, does analytical variability affect medical decision making and in particular, for which determinations is – from the clinical point of view – analytical performance good, adequate or eligible for improvement respectively.
A questionnaire was designed and submitted to 63 senior specialists in internal medicine, working in teaching hospitals.
The questionnaire focussed on the following subjects:

- normal range values and action levels (action level being defined as a constituent level that prompts the clinician to an action of the first order i.c. repeated or additional laboratory tests, ECG, X-ray, or change in diet or medication)
- medically significant change in the individual patient
- possible sources of variation in test results
- satisfaction with laboratory performance
- a number of other relevant subjects such as: repeated testing, speed of testing versus reliability of the result, the use of molecular units and the interpretation of laboratory results from other hospital laboratories.

The interviews were held in person and lasted 30-45 minutes.
Concurrently the laboratories serving the participating clinicians analysed samples of quality control sera in order to collect information as to intra- and interlaboratory variability. Also, normal range values were reported by the laboratory.
In the addendum to this thesis a form for reporting in interlaboratory surveys, giving simultaneously information on intra- and interlaboratory variability, is described.
The determinations under study were: sodium, potassium, chloride, calcium, inorganic phosphorous, urea, creatinine, glucose, cholesterol, total protein, alkaline phosphatase, lactate dehydrogenase and hemoglobin.

Laboratory survey

The results of the laboratory survey were processed in order to be usefull in the interpretation of clinicians' answers to the questionnaire.
Day to day variability was determined for each laboratory. Significant differences in intralaboratory variability between hospitals were found for all constituents except for sodium, chloride, glucose and cholesterol.
Systematic analytical bias between laboratories was not found to be consistently reflected in the range of normal values given by the laboratory director. In only 16 out of 35 cases of most evident systematic bias with respect to the group of laboratories did the normal range show the same tendency. Consequently, while it is common practice to use the normal range as a point of reference in interhospital communications, this habit finds no justification in this study.
Differences in the population tested for determination of the normal range could in some cases be related to differences in normal range values (outpatients versus blooddonors or hospital personel, age differences).
It was noted that protocols for determination of the normal range differ considerably in the group of hospitals cooperating with this study.

Requirements for analytical performance from the clinical point of view.

In order to derive clinical requirements for analytical performance with respect to day-to-day precision from the data collected in this study three criteria have been used:
- the difference between either limit of normal and the respective action level
- the change of constituent level in the individual patient, given a certain initial value in the near abnormal range, that is considered medically significant
- clinicians satisfaction and dissatisfaction respectively with test results

Previous attempts described in the literature to derive requirements for analytical performance have been predominantly based on the concept, that analytical variability should not exceed biological variability, either between or within individuals. However, to base analytical requirements on biological variation between individuals (the normal range) does not seem to respond to the practical situation:
- in clinical diagnosis it is the difference between the population in health and the population in disease that counts. Upper and lower limits of normal are different entities in this respect and the width of the normal range is irrelevant in this situation.
- in the follow-up of the individual patient the variation between individuals is irrelevant.
- in our study it was found, that a change in an individual of about ½ of the normal range is considered significant for some constituents but much less for others (chapter VII).
- also, in our study clinicians tend to be more strict in the upper decision range as compared to the lower decision range for a number of constituents, for example

for potassium, calcium, creatinine and the enzymes, the opposite is true for inorganic phosphorous (chapter IV).

Biological variation within the individual is a usefull basis for analytical requirements in screening projects in preventive medicine. For daily practice however, this approach does not seem realistic for the following reasons:

- in the most common situation of wanting to distinguish the diseased patient from the healthy population the intraindividual variation does not play a role
- the intraindividual variability has hitherto been determined in healthy individuals. In practice we seldom follow up the healthy individual and intraindividual variation is likely to change in disease. In our study a certain change in constituent level was interpreted differently when the change took place in the normal range, or in the near abnormal range. (chapter V and VII).

Consequently neither approach is complete. In our attempt to establish requirements for analytical performance from the clinical point of view we have considered the two situations of a. to distinguish the patient from the healthy population and b. to follow up the individual patient.
The main difficulty encountered was related to the large discrepancies of clinicians' answers.
The dispersion of answers to the question of a significant change in a patient was of the same order of magnitude as intralaboratory variability. The discrepancies in clinicians' action levels were much larger, differences among clinicians of one hospital often being as large as differences among clinicians of different hospitals. Often action levels in the upper and lower decision range did overlap. Although several good reasons can be named why criteria for taking action may vary a discussion Another factor that should be taken into account upon interpretation of the data was that 31/63 clinicians referred to the normal range given by the laboratory when asked for normal values. While determining for these clinicians the differences between limits of normal and respective action level the laboratory's limits of normal were used for calculations.

In general (except for inorg. P) the median difference between limit of normal and action level was smaller than the average medically significant change in the individual patient. As the initial values for the latter were all in the near abnormal (AN) range, we conclude that the decision range between normal and abnormal is the most critical one where analytical reliability is most needed.

This is also shown in the following table summarizing the data from chapter IV and V on requirements for laboratory performance from the clinical point of view:

	decision range (median difference limit of normal and action level)	near AN range (medically significant change in a patient)	clinicians' opinion (improvement needed)
analytical precision is adequate for:	K, P, urea	Na, K, urea, chol. creat, t. prot., ALP, LDH, Hb	
is just short of good performance for:	Na, Cl	Cl, P, gluc,	
is eligible for improvement for:	Ca, creat, gluc. chol, t. prot., ALP, LDH, Hb	Ca	Ca, creat. t. prot., Hb, "enzymes"

Evidently, improvement of the precision of calcium determinations is a must.
In the decision range improvement of laboratory performance is also indicated for creatinine, glucose, cholesterol, total protein, alk. P-ase, LDH and hemoglobin. This finding largely correlates to the fact that for all but two of these constituents significant numbers of physicians are dissatisfied with test results. This indicates, that the proposed model of deriving criteria for analytical performance from the difference between limit of normal and action level – however limited by the dispersion of answers – is a sensible approach.
For glucose and cholesterol clinicians did not feel improvement of laboratory performance was needed, while according to our data analytical performance was less than adequate in the decision range. Appreciable numbers of clinicians assumed for these constituents biological variations to be the main cause of day to day variations in laboratory results. Probably laboratory variability plays a larger role than presently assumed and improvement of laboratory performance for these constituents may open new horizons.
We have considered median intralaboratory variability and have to remember, that half of the laboratories perform better than reported here, and the other half performs worse. In some cases differences in precision were registered by clinicians. In hospitals where laboratories have larger intralaboratory variability they may apply a wider range as medically significant or more clinicians may be dissatisfied with determinations as compared to clinicians working with more precise laboratories. This situation has been shown for potassium, calcium, creatinine and inorg. phosphorous. It may be relevant, that no such relationships could be demonstrated for total protein, hemoglobin and the enzymes, while considerable numbers of physicians voiced dissatisfaction. Other factors i.c. technique of sample taking or patient preparation may be involved.

Thus having indicated for which constituents improvement of day to day precision in the laboratory is needed table 8-1 shows the desirable degree of precision based on our data compared to those of a number of other studies: Barnett (4), Campbell and Owen (9), Gilbert (27), Cotlove et al. (14) and Steele c.s. (58).
The average medically significant change in a patient in our study is closest to the medically significant values given by Gilbert.
When we consider the conclusions of the study of Steele e.a. (58) evaluating laboratory performance against criteria based on intraindividual variation according to Cotlove and coworkers (14) we find that laboratory performance for sodium and chloride needs to be improved as urgently as that for calcium. This is hard to believe from the point of view of the practicing physician.
The most strict requirements for most constituents are those based on median difference between limit of normal and action level derived from our study. Although hard to meet it is our conviction that these values should be our targets in order to meet present medical needs. They correlate largely with physicians' satisfaction. This is however, not to say, that patient care will automatically improve if laboratories improve their performance to meet these targets.
Other aspects are equally important for example: a discussion of physicians among themselves on the subject of medical significance of laboratory results, a discussion among laboratories and physicians on normal or reference values, standardization of the technique of sample taking and the design of a means of communicating levels of precision – and changes thereof – to the clinician.

Repeat testing

As to the attitude towards repeat testing it was found that the majority of clinicians (52/62) always or often repeats a test before making a medical decision. Half of them declare to do so in particular when the first result is incompatible with their expectations or when the decision to be made is important. Evidence is found, that in the four large hospitals in Holland where the rate of routinely repeat testing is high as compared to the other Dutch hospitals the number of tests per clinician for which greater reliability is desired is comparibly high as well. In 2 out of these 4 hospitals overall intralaboratory variability is high, in the other two hospitals bad communications and clerical errors came up during the interview most frequently of all hospitals.

Fast semiquantitative tests.

Fast semiquantitative results as compared to a more accurate result on longer terms was desired in particular for glucose (27/63 clinicians). When asked for desired precision 11/19 gave figures obtainable with quantitative methods only.

	Requirements for analytical precision (3 SD) based on: Medically significant difference					Intraindividual variability	
	1	2	3	4	5	6	7
	decision range	near AN range indiv. patient	Barnett (4)	Campbell (9)	Gilbert (27)	Cotlove (14)	Steele (58)
Na-mmol/l	3.0	5.1	6	7	4.5	1.5	0
K-mmol/l	0.3	0.38	0.75	0.61	0.45	0.42	0.54
Cl-mmol/l	4	5.2	6	8.6	4.5	2.7	2.6
Ca-mmol/l	0 (0.01)	0.13	0.18	0.11	0.12	0.12	0.12
P-mmol/l	0.10	0.25	0.24	0.13	0.18	0.21	0.27
urea-mmol/l	1.3	2.46	2.1	1.6	2.1	3.2	1.9
creat.-µmol/l	0 (1)	29.9	–	–	27	–	15
gluc.-mmol/l	0 (0.1)	2.17	0.84	0.43	0.51	0.75	0.42
chol.-mmol/l	0 (0.1)	1.33	1.56	–	1.17	1.32	1.02
t.prot.-gr/l	2	6.9	9	–	6	6.3	5.4
ALP-%	0	18	–	–	–	–	–
LDH-%	10	22	–	–	–	27	–
Hb-mmol/l	0 (0.1)	0.75	0.93	–	–	–	–

Table 8-1: Comparison of criteria for analytical performance derived from different studies, expressed as 3 SD. (If analytical variability is 1 SD we may consider a difference of 3 SD to be discernable – see text –). The data in column 1 through 5 represent medically significant differences, i.e. differences between two values between which the physician wants to distinguish. Data in column 1, derived from the present study are based on median difference between limit of normal and action level in the most critical decision range (for details see table 4-4). In cases where median difference is zero the smallest difference recognizable, given the number of decimal points, is placed in parentheses. The data in column 2 represent the average medically significant change in a patient (improvement) starting from an initial value in the near abnormal range (table 5-2).
Data in column 3, 4 and 5 represent "medically significant values", "acceptable analytical limits" and "analytical goals" in the normal/decision range, expressed as 3 SD, as given by Barnett (4), Campbell and Owen (9) and Gilbert (27) respectively.
The criteria for analytical performance in column 6 and 7 are based on intraindividual variability determined by Cotlove et. al. (14) and Steele et. al. (58).

The use of molecular units.

Molecular units do not seem to contribute positively to medical treatment. In the case of glucose the objections to the molecular unit seem to be fundamental.
Difficulties to be encountered with introduction of the new units should not be underestimated.
Seventeen out of 52 clinicians stated to have needed more than a year to get adjusted to the new units and 20/52 reported to after 2-5 years still always or often convert mmol/l into mg% for one or more constituents for optimal interpretation.

Laboratory data from another hospital laboratory.

Few clinicians accept results from other hospitals as such.
The same change of constituent level in a patient is considered a change in the condition of the patient by 2-12 (average 7.6) more clinicians if both determinations were carried out in the own hospital laboratory as compared to the situation where the first result was determined in another laboratory and the second one in the own laboratory.
In our study the average ratio of interlaboratory variability to intralaboratory variability is 1.9 to 1 (range 1.1-3.6 to 1), excluding the enzymes. We conclude, that intralaboratory variability plays an important role in generating differences between laboratories and hence efforts in quality control should emphasize this aspect.

SAMENVATTING

De vraag die geleid heeft tot het werk dat in dit proefschrift beschreven wordt luidt: hoe nauwkeurig, met name hoe reproduceerbaar, moeten klinisch chemische bepalingen zijn om bij te kunnen dragen tot een goede behandeling van de patient.
Het beantwoorden van deze vraag is niet eenvoudig. Enerzijds is de spreiding van dag-tot-dag van laboratoriumuitslagen, waarbij ook andere factoren zoals monstername een rol spelen, zelden verwaarloosbaar ten opzichte van de biologische verschillen die men wenst aan te tonen. Anderzijds is de laboratoriumuitslag slechts één van de vele gegevens, aan de hand waarvan de arts zijn diagnose stelt. De opzet van het onderzoek was na te gaan op welke wijze de arts laboratorium gegevens bij zijn besluitvorming verwerkt en of deze besluitvorming wordt beinvloed door de dag-tot-dag spreiding van de laboratorium bepaling.
In het bijzonder werd gepoogd een antwoord te vinden op de vraag voor welke bepalingen – vanuit klinisch oogpunt bekeken – de nauwkeurigheid van het laboratorium goed, voldoende, dan wel voor verbetering vatbaar is.

Aan 63 internisten in 12 ziekenhuizen met een interne A opleiding, academische en algemene ziekenhuizen, werden vragen voorgelegd over de volgende onderwerpen:
- normale waarden en laboratorium uitslagen die aanleiding geven tot het nemen van actie (actiewaarden). Onder actie wordt hierbij verstaan het aanvragen van herhaald of aanvullend laboratorium onderzoek, ECG, röntgen onderzoek, verandering van dieet of medicatie.
- een klinisch significante verandering van de bloedspiegel van een bepaald bestanddeel bij een patient.
- tevredenheid respectievelijk ontevredenheid met de nauwkeurigheid van het laboratorium.
- voorts: het aanvragen van een herhaling van een test, voorkeur voor een snel semiquantitatief resultaat boven een nauwkeurige uitslag op langere termijn, moleculaire eenheden, interpretatie van uitslagen van laboratorium van een ander ziekenhuis.

De laboratoria van de betreffende ziekenhuizen bepaalden monsters van kwaliteits controle sera. Op deze wijze werden gegevens verkregen omtrent de dag-tot-dag spreiding van resultaten binnen ieder laboratorium en omtrent de spreiding van resultaten tussen de verschillende laboratoria.
Door ieder laboratorium werden de gangbare normale waarden opgegeven met eventuele bijzonderheden.
Bij het bewerken van de gegevens over de spreiding binnen één laboratorium en over die tussen laboratoria werd gebruik gemaakt van een speciale wijze van rapporteren die in een addendum van dit proefschrift beschreven is.
De bepalingen die bij dit onderzoek betrokken waren zijn: natrium, kalium, chloride, calcium, anorganisch phosphaat, ureum, kreatinine, glucose, cholesterol, totaal eiwit, alkalische phosphatase, melkzuur dehydrogenase en hemoglobine.

Vergelijkend laboratorium onderzoek

Uit het vergelijkend laboratorium onderzoek konden de volgende conclusies worden getrokken in het licht van de vragen die aan de internisten gesteld werden:
- dag-tot-dag spreiding binnen een laboratorium verschilde statistisch significant voor alle bepalingen behalve voor natrium, chloride, glucose en cholesterol.
- systematische analytische verschillen tussen laboratoria werden niet steeds weerspiegeld in de normale waarden van de betreffende laboratoria.
- de samenstelling van de groep normale personen, die gekozen werd door de deelnemende laboratoria ten behoeve van het bepalen van normale waarden, verschilde sterk van laboratorium tot laboratorium.

Welke eisen stelt de kliniek aan de nauwkeurigheid van het laboratorium

Door Tonks en anderen is gesteld dat de analytische spreiding van een klinisch chemische bepaling niet groter mag zijn dan een bepaald percentage van de biologische spreiding. Zoals ook uit dit onderzoek blijkt is deze benadering vanuit klinisch oogpunt niet relevant.
In dit onderzoek zijn drie criteria gebruikt om vast te stellen voor welke bepalingen een laboratorium aan de eisen van de kliniek voldoet.
- het verschil tussen boven- of ondergrens van het normale gebied en de betreffende actiewaarde.
- die verandering van de bloedspiegel van een bepaald bestanddeel in een patient welke de arts als klinisch significant beschouwt.
- tevredenheid respectievelijk ontevredenheid met de nauwkeurigheid van het eigen laboratorium.

Ondanks de aanzienlijke spreiding in antwoorden van de internisten die geinterviewd werden konden een aantal conclusies getrokken worden.
Aan de verbetering van de nauwkeurigheid van de bepaling van calcium dient grote aandacht te worden geschonken. In het kritieke gebied tussen normaal en juist abnormaal is ook de nauwkeurigheid van de volgende bepalingen, beoordeeld volgens de twee eerste hierbovengenoemde criteria onvoldoende: kreatinine, glucose, cholesterol, totaal eiwit, alkalische phosphatase, melkzuur dehydrogenase en hemoglobine. Voor deze bepalingen uitten de ondervraagden dan ook het meest frequent hun ontevredenheid, echter met uitzondering van glucose en cholesterol. Waarschijnlijk is de arts geneigd variaties in laboratorium uitslagen voor deze twee bestanddelen toe te schrijven aan physiologische schommelingen onder invloed van bijv. dieet en niet aan het te kort schieten van het laboratorium.
In sommige gevallen kon een relatie gevonden worden tussen de mate van dag-tot-dag spreiding van de bepaling en die gemiddelde bloedspiegel verandering bij een patient, welke door de internisten van dat ziekenhuis als wezenlijk werden ervaren. In andere gevallen bleek een relatie te bestaan tussen de mate van dag-tot-dag spreiding van de bepaling en de mate van tevredenheid van de internisten.

Voor totaal eiwit, hemoglobine en de enzymen konden dergelijke relaties niet aangetoond worden, hoewel relatief veel internisten ontevreden waren met de graad van nauwkeurigheid van deze bepalingen. Aanleiding tot de ontevredenheid over deze bepalingen zijn waarschijnlijk verschillen in monstername techniek of niet herkende physiologische schommelingen.
Het dient te worden opgemerkt, dat de behandeling van patienten niet zonder meer zal verbeteren wanneer het laboratorium aan de nauwkeurigheidseisen, zoals die hier gesteld zijn, voldoet. Andere factoren zijn hierbij eveneens belangrijk, zoals, discussie van artsen onderling over de vraag hoe medische besluitvorming op grond van laboratorium uitslagen tot stand komt, een discussie van artsen en hoofden van laboratoria over de te volgen procedure bij het vaststellen van normale of referentie waarden, standaardisatie van monstername techniek, en het ontwerpen van een wijze van communicatie om de mate van reproduceerbaarheid – en wijzigingen daarvan – door te geven aan de kliniek.

Overige resultaten

Uit de antwoorden op de resterende vragen van de vragenlijst kan nog het volgende worden opgemaakt:

- in twee van de 4 ziekenhuizen, waar dikwijls herhalingen van laboratorium tests worden aangevraagd was de spreiding van de bepalingen in het algemeen groot, in de twee andere ziekenhuizen werden slechts communicatie met het laboratorium en administratieve fouten meer dan in enig ander ziekenhuis genoemd.
- moleculaire eenheden dragen in de huidige situatie niet in belangrijke mate bij tot betere behandeling van patienten of beter begrip van biochemische processen. De practische problemen bij de invoering van deze eenheden mogen niet onderschat worden.
- In het algemeen hechtten de ondervraagden meer waarde aan laboratorium uitslagen bepaald in het eigen laboratorium dan aan die bepaald in een ander laboratorium. Uit het vergelijkend laboratorium onderzoek bleek echter, dat de spreiding binnen een laboratorium in veel gevallen niet onaanzienlijk is ten opzichte van die tussen de laboratoria.

LITERATURE

1. American Society for Testing and Materials. Standard Recommended Practice for use of the terms precision and accuracy as applied to measurements of a property of a material. Designation E 177-71.
2. American Medical Association Advisory Committee on PSRO. Report by the Task Force on guidelines of Care. J.A.M.A. **229**, 166-171, 1974.
3. Bawkin H. New. Engl. J. Med. **232**, 691-697, 1945.
4. Barnett R.N. Am. J. Clin. Path. **50**, 671-676, 1968.
5. Blijenberg B.G., Leijnse B. Ned. T.v. Gen. **112**, 1901-1904, 1968.
6. Bokelund H. Winkel P. Statland B.E. Clin. Chem. **20**, 1507-1512, 1973.
7. Borak, J. J.A.M.A. **237**, 641-642, 1977.
8. Cali J.P. Med. Instrument. **8**, 17-21, 1974.
9. Campbell D.G. Owen J.A. Clin. Biochem. **1**, 3-11, 1967.
10. Clark A.N.G. Lancet Feb. 14, 1976, p. 356.
11. Clark T.W. Schor S.S. Elsom K.O. Hubbard J.P. Elsom K.A. Ann. Int. Med. **54**, 1209-1222, 1961.
12. Copeland B.E. Rosenbaum J.M. Am.J.Clin. Path. **57**, 676-688, 1972.
13. Copeland B.E. personal communication.
14. Cotlove E. Harris E.K. Williams G.Z. Clin. Chem. 16, 1031-1036, 1970.
15. Coulthard A.J. Clin. Chim. Acta **3**, 226-233, 1958.
16. Crosby W.H. Blood **13**, 1198, 1958.

17a De Jonge H. "Inleiding tot de medische statistiek" 2nd ed. 1964. Wolters-Noordhof, Groningen.

b Dixon W.J. Massey F.J. Jr. „Introduction to statistical analysis" 3d ed. 1969 McGraw-Hill, New York.

18. Dybkaer R. VIIIth Int. Congress on Clin. Chem. Copenhagen 1972. Scand. J. Clin. Lab. Invest. **29**, Suppl. 126, p. 19.1, 1972.
19. Elion-Gerritzen W.E. Eurotransplant Clin. Chem. Surveyreports. 1973-1974. Unpublished.
20. Elion-Gerritzen W.E. Copeland B.E. Report of a comparative survey among dialysis and transplant centers in New England and Europe, involving laboratory results important in kidney disease. 1975 Unpublished.
21. Elion-Gerritzen W.E. Am.J. Clin. Path. **67**, 91-96, 1977.
22. Elsom K.A. Ipsen J. Clark T.W. Talerico L. Yanagawa H. J.A.M.A. **201**, 519-526, 1967.
23. Files J.B. Van Peenen H.J. Lindberg D.A.B. J.A.M.A. **205**, 684-688, 1968.
24. Fisher R.A. "The design of experiments", IInd ed. 1937. Oliver and Boyd, Edinburgh p. 13-18.
25. Flynn F.V. Piper K.A.J. Garcia-Webb P.Mc. Pherson K. Healy M.J.R. Clin. Chim. Acta **70**, 179-189, 1976.

26. Gilbert R.K. Am. J. Clin. Path. **61**, 904-911, 1974.
27. Gilbert R.K. Am. J. Clin. Path. **63**, suppl. 960-973, 1975.
28. Glick J.H.Jr. Clin. Chem. **22**, 475-483, 1976.
29. Gowenlock A.H., Broughton P.M.G. Z. Anal. Chem. **243**, 774-780, 1968.
30. Grannis G.F. Grümer H.-D. Lott J.A. Edison J.A. Mc Cabe W.C. Clin. Chem. **18**, 222-236, 1972.
31. Grasbeck R.VIIIth Int. Congress on Clin. Chem. Copenhagen 1972 Scand. J. Clin. Lab. Invest. **29**, suppl. 126, p. 19.2, 1972.
32. Hawkins W.W. Speck E. Leonard V.G. Blood **9**, 999-1007, 1954.
33. Holland R.R. J.A.M.A. **233**, 455-457, 1975.
34. IFCC Expert Panel on Nomenclature and Principles of Quality Control. Provisional Recommendations.
    a. Part 1, stage 2 draft 2. Clin. Chem. **22**, 532-540, 1976.
    b. Part 6, stage 2 draft 2. Clin. Chim. Acta 74, F1-F20, 1977.
35. Johnson E.A. Human Path. **4**, 5-8, 1973.
36. Keating F.R. Jr. J.A.M.A. **178**, 547-555, 1961.
37. Keating Fr. Jones J.D. Elveback L.R. Randall R. J. Lab. Clin. Med. 73, 8 e.v. 1969.
38. Lawson N.S. Ross J.W. personal communication.
39. Lindberg D.A.B. Watson F.R. Meth. Inform. Med. **13**, 151-158, 1974.
40. Massachusetts General Hospital. Case Records. Normal Laboratory Values. New. Engl. J. Med. **290**, 39-42, 1974.
41. Mc Swiney R.R. Woodrow D.A. J. med. Lab. Technol. **26**, 340-346, 1969.
42. Murphy E.A. Abbey H. J. Chron. Dis. **20**, 79-88, 1967.
43. Murphy E.A. Persp. Biol. Med. **15**, 566-582, 1971/1972.
44. New Engl. J. Med. **293**, July 31 issue, 1975.
45. O'Kell R.T. Elliott J.R. Clin. Chem. **16**, 161-165, 1970.
46. Owen J.A. Campbell D.G. Clin. Chim. Acta 22, 611-618, 1968.
47. Ressler N. Whitlock L.S. Clin. Chem. **13**, 931-940, 1967.
48. Ross J.W. Fraser M.D. Am. J. Clin. Path. **68**, suppl. 130-141, 1977.
49. Schneiderman L.J. Desalvo L. Baylor S. Wolf P. Arch. Intern. Med. **129**, 88-90, 1972.
50. Sisson J.C. Schoomaker E.B. Ross J.C. J.A.M.A. **236**, 1259-1263, 1976.
51. Skendzel L.P. Youden W.J. Am. J. Clin. Path. **51**, 161-165, 1969.
52. Skendzel L.P. Youden W.J. Am. J. Clin. Path. **54**, 448-453, 1970.
53. Smyllie H.C. Blendis L.M. Armitage P. Lancet, August 28, 1965, p. 412.
54. Southgate M.T. J.A.M.A. **232**, 515-516, 1975.
55. Spaander J. Helleman P.W. Bibl. Haematol. (Basel) **21**, 129 e.v., 1965.
56. Statland B.E. Winkel P. Bokelund H.
    a. Clin. Chem. **19**, 1374-1379, 1973.
    b. Clin. Chem. **19**, 1380-1383, 1973.
57. Statland B.E. Bokelund H. Winkel P. Clin. Chem. **20**, 1513-1519, 1973.
58. Steele B.W. Schauble M.K. Becktel J.M. Bearman J.E. Am. J. Clin. Path. **67**, Suppl. 594-602, 1977.

59. Strømme J.H. Eldjarn L. Scand J. Clin. Lab. Invest. **25**, 213-222, 1970.
60. Sunderman F.W. McFate R.P. McFadyen D.A. Stevenson G.F. Copeland B.E. Am. J. Clin. Path. **23**, 519-598, 1953.
61. Tonks D.B.
    a. Clin. Chem. **9**, 217-233, 1963.
    b. Z. Anal. Chem. **243**, 760-765, 1968.
62. Vanko M. "Advances in Automated Analysis" I, 1970. Technicon Int. Congress 1969 p. 159.
63. Werner M. Tolls R.E. Hultin J.V. Mellercker J. Z. Klin. Chem. u. Biochem. 8, 105-115, 1970.
64. West K.M. Diabetes **24**, 641-644, 1975.
65. Whitby L.G. Ann. Clin. Biochem. **6**, 104-107, 1969.
66. Wiener S. Nathanson M. J.A.M.A. **236**, 852-855, 1976.
67. Willebrands A.F. Ned. T.v. Gen. **114**, 1615-1620, 1970 (re.: p. 1619).
68. Winkel P. Statland B.E. Bokelund H. Clin. Chem. **20**, 1520-1527, 1973.
69. Winkel P. Statland B.E. Clin. Chem. **22**, 1855-1861, 1976.
70. Winkelman J.W. Cannon D.C. Pileggi V.J. Reed A.H.
    a. Clin. Chem. **18**, 57-66, 1972.
    b. Clin. Chem. **19**, 488-491, 1973.
71. Wright I.S. J.A.M.A. **236**, 261-262, 1976.
72. Yendt E.R. Gagne R.J.A. Can. Med. Ass. J. **98**, 331-336, 1968.
73. Youden W.J. "Statistical techniques for collaborative tests" Association of Official Analytical Chemists 1967. p. 27-29.
74. Young D.M. Drake N. Weir R.J. Can. Med. Ass. J. **98**, 868-870, 1968.
75. Young D.S. Harris E.K. Cotlove E. Clin. Chem. **17**, 403-410, 1971.
76. Zender R. Ann. Biol. Clin. **28**, 15-16, 1970.
77. Zwart Voorspuij A.J. Vander Slik W. Clin. Chim. Acta **9**, 99, 1964.

ADDENDUM I

Reprinted from AMERICAN JOURNAL OF CLINICAL PATHOLOGY, Vol. 67, No. 1, January 1977

Quality Control in Clinical Chemistry

The Two-sample Plot and Improvement of Laboratory Performance

WIVEKA E. ELION-GERRITZEN, M.Sc.

Department of Experimental Surgery, Erasmus University, Rotterdam, The Netherlands

Elion-Gerritzen, Wiveka E.: Quality control in clinical chemistry. The two-sample plot and improvement of laboratory performance. Am J Clin Pathol 67: 91–96, 1977. A two-sample plot is often used to display test results in clinical chemistry external quality control programs. The use of control materials at different concentrations calls for a re-evaluation of the two-sample plot. A modification of the two-sample plot is presented. The axes of the graph are graduated in units of measurement rather than standard deviation intervals. The mean of all laboratories or a reference laboratory's value is entered as the target value. For each laboratory, repeated results for the same or similar control materials are recorded. Visual inspection of the graph gives immediate information as to the extent of intra-laboratory variability, the incidence of systematic errors in the laboratory, and the size and type of systematic errors (proportional or constant) with respect to the target value. (Key words: External quality control; Internal quality control; Youden-Tonks plot; Two-sample plot; Random error; Systematic error; Proportional systematic error; Constant systematic error.)

IN MANY external quality control programs or surveys in clinical chemistry where two control samples are distributed among participants, the two-sample plot as described by Skendzel and Youden[10] or one of its modifications, *e.g.*, the one suggested by Tonks,[12] is used to display test results. Skendzel and Youden point out that the graph displays the results from each participating laboratory and provides an opportunity for the appraisal of each laboratory's performance: "values, consistently high or low suggest a systematic error due to an improperly calibrated instrument, incorrect standard curve, deteriorating reagents or the laboratory's own modification of the recommended procedure. Points near the diagonal line through the median indicate that the laboratory is consistent. Plotted points in other sections of the graph represent values which do not follow a consistent pattern and are attributable to random error, due to improper technic, mix-up of samples, or a clinical mistake."

Received August 11, 1975; received revised manuscript February 9, 1976; accepted for publication February 9, 1976.
Address reprint requests to Dr. Elion-Gerritzen.

While Skendzel and Youden[10] do not specify concentrations of control materials, it is presently assumed that more information on laboratory performance can be derived from the graph if the samples contain constituents in different concentrations.[1,11]

This paper reconsiders the interpretation of the Youden–Tonks graph and presents a modified two-sample plot that can help identify systematic errors.

The Youden–Tonks Graph: Construction and Interpretation

Let us consider an external quality control program where two control samples are being tested. Unless otherwise stated, it is assumed in this discussion that laboratories test the samples once, in the same run. Figure 1 represents a Youden–Tonks graph; the target value (+) usually is defined as the coincidence point of the mean values of all laboratories for both control sera. On each axis equal distances are allotted to the 1 standard deviation (SD) values; the 1 and 2 SD values represent the measure of dispersion of results for all participant values for each control material. According to Skendzel, Youden, and Tonks the 45-degree diagonal line is drawn to allow participants in the program to appraise their performance.

Random Errors

It is relevant to quote here the definition of random error as given in Webster's dictionary:[14] *a statistical*

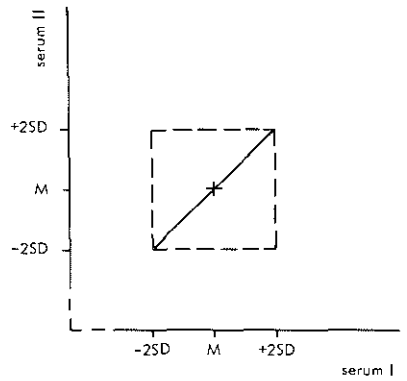


FIG. 1. Two-sample plot as suggested by Skendzel and Youden, modified by Tonks.

error that is wholly due to chance and does not recur—opposed to systematic error. Random errors due to unpredictable variability in the laboratory (*e.g.*, variation in sample or reagent volume) and blunders (*e.g.*, mix-up of samples) are not likely to affect both samples in the same way, and the resulting coincidence point will deviate from the diagonal line in the Youden–tonks graph.

Systematic Errors

A systematic error is a statistical error that persists and cannot be considered as due entirely to chance—opposed to random error.[15] It is our experience that systematic errors can cause plotted points to deviate from the diagonal line in the Youden–Tonks graph. This is particularly the case when control materials contain constituents at different concentrations.

Systematic Errors in the Laboratory

We consider a laboratory using a nonlinear procedure in the sense of less response at high absorbance values, or not sufficiently standardized in the abnormal range. When testing two control sera, one at a low-normal (N) level and one at a high-abnormal (AN) level, this laboratory will report a low value for the high-abnormal serum, and the plotted point will show up as a previously mentioned random error in the Youden–Tonks graph. On repeated testing the coincidence point of this laboratory will always show up under the diagonal line. We have seen an example of this situation in the IInd Eurotransplant Clinical Chemistry Survey.[4] Figure 2 shows the Youden–Tonks graph for urea; laboratories tested both controls twice, in duplicate. In response to a questionnaire, laboratories A–F reported to standardize in the normal range only.

Another example of a systematic error in the laboratory that causes the coincidence point to deviate from the diagonal line is given in Figure 3. Laboratory G tested two control sera (1974 pools of the Massachusetts Society of Pathologists Quality Control Program, MSPQCP)[1] on four different days for cholesterol. G is the mean value; the dots represent the individual results. Repeatedly a too-high value for serum II was found. Most likely the high bilirubin level in serum II systematically interfered with this cholesterol procedure, which did not call for a sample blank.

The systematic errors described here are persistent; they are predictable and avoidable once the cause is known.

Proportional and constant systematic errors

When control materials in the *same concentration range* are distributed, incorrect standardization or differences in methodology between laboratories will result in values higher or lower to the same extent for both samples. The coincidence point will be located along the diagonal line of the Youden–Tonks graph. When control materials contain constituents in *different concentrations* it is important to realize that systematic errors or systematic differences between laboratories can be constant or proportional.[8,13]

A distinction between these two types can readily be made in a graph where the axes are graduated in

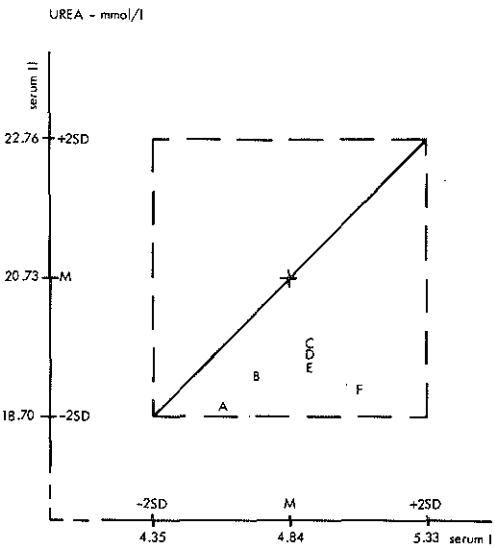


FIG. 2. Youden–Tonks graph for urea (Second Eurotransplant Clin Chem Survey). Laboratories A–F reported to standardize in the normal range only.

units of measurement instead of SD. For example, in Figures 4 and 5 results for an N level control sample are plotted on the abscissa and results for an AN control sample on the ordinate. If only *proportional systematic errors* are involved (Fig. 4), increasing with increasing concentrations of the constituent, values for N and AN control samples will be higher or lower than the respective target values by differing amounts proportional to concentration. The target area, defined by M ± 2 SD, will be an upright rectangle, the slope of the diagonal line will be given by

$$\frac{\text{M high control sample}}{\text{M low control sample}} \text{ or } \frac{\text{SD high control sample}}{\text{SD low control sample}},$$

and the diagonal line will pass through zero. If *constant systematic errors* are involved, values for both samples will be higher or lower than the respective mean value by the same amount, irrespective of the concentration of the constituent (Fig. 5). The diagonal line of the square target area is a 45-degree line. Many nonspecific procedures, for example for glucose, creatinine or cholesterol, measure a fixed amount of background, as compared with specific procedures for these constituents.

In a group of laboratories constant and proportional errors will occur. In the Youden–Tonks graph the target area and the position of the diagonal line are

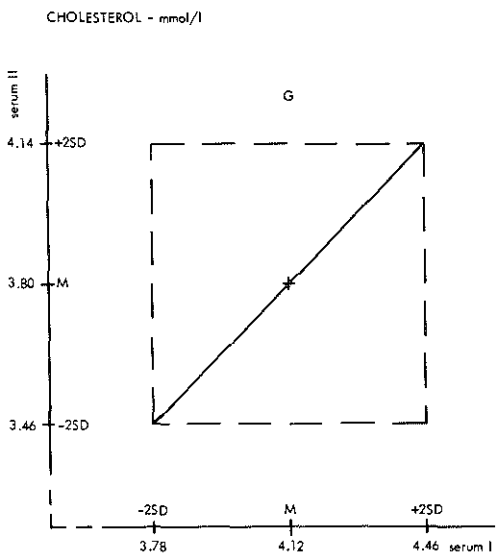


FIG. 3. Youden–Tonks graph for cholesterol (MSPQCP—1974 pools). Laboratory G shows systematically high values for serum II, probably due to bilirubin interference.

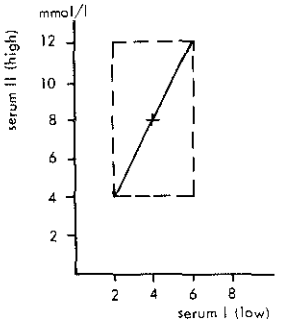


FIG. 4. Hypothetical two-sample plot, axes graduated in units of measurement, of a group of laboratories, showing proportional differences from the target value (+).

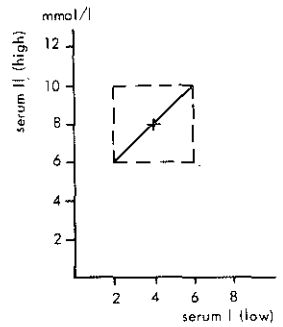


FIG. 5. Hypothetical two-sample plot, axes graduated in units of measurement, of a group of laboratories, showing constant differences from the target value (+).

defined by M ± 2 SD, the SD being derived from the combined data. If the majority of laboratories show constant systematic differences from the target value, the coincidence point of a laboratory showing proportional systematic error will be interpreted as being due to random error.

An example is given in Figure 6. In the Youden–Tonks graph for cholesterol (MSPQCP, 1975 pools), axes graduated in SD (Fig. 6A), laboratories H, J, and K show random errors according to conventional interpretation. From the graph in units of measurement (Fig. 6B), it can be concluded that most laboratories show proportional systematic errors, while laboratories H, J, and K show a constant systematic difference from the mean of all laboratories; this finding correlates with the fact that these laboratories use a method with extraction.

A similar situation is illustrated in Figure 7. In a two-sample plot graduated in units of measurement for creatinine (MSPQCP, 1974 pools) the average results

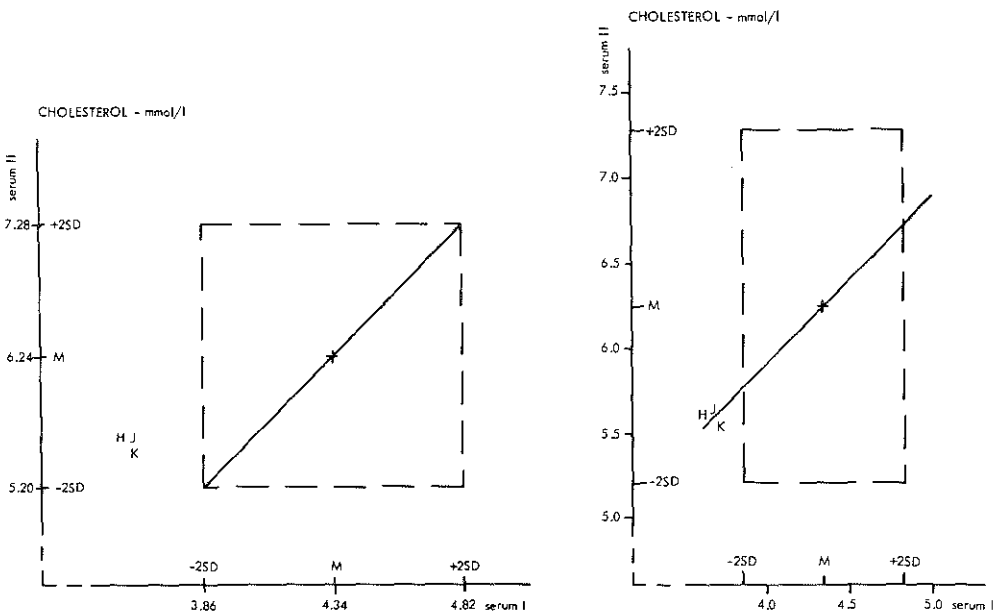


FIG. 6. *A (left)*, Youden–Tonks graph for cholesterol (MSPQCP—1975 pools). Axes are graduated in SD. Laboratories H, J and K (employing a method with extraction) show random errors according to conventional interpretation.

B (right), two-sample plot for cholesterol, axes graduated in units of measurement. From this graph it can readily be concluded that laboratories H, J and K show constant systematic differences from the mean of all laboratories, while the majority of laboratories show proportional systematic errors.

(capitals) and daily values (small print) for laboratories X and Y are entered. R is the mean value for four reference laboratories.* The mean value for laboratory X is situated on the line with a slope

$$\frac{\text{ref. value serum II}}{\text{ref. value serum I}}$$

through the reference value, suggesting a proportional systematic difference from the reference value. The difference was negligible after renewal of the standard. Laboratory Y appears to have a constant systematic error with respect to the reference value, although the dispersion of the daily results indicates large random variability. (It is interesting that additional information can be derived from the "dot-cloud" of daily results. The dispersion of laboratory X's plotted points shows a trend, and we conclude that there is a small proportional systematic error in the laboratory due to, *e.g.*, instability of the standard. Similarly, we can expect in a procedure with baseline problems a dot-cloud along a 45-degree line).

*It is interesting that for the 1974 pools of the MSPQCP, for all components where the concentration in the AN serum is more than 2.5 times that in the N control serum (*e.g.*, urea, glucose, creatinine), the reference laboratories' mean value for the AN serum is significantly higher than the mean for all laboratories. Nonlinearity may be a common error, causing a low value for the mean of all laboratories for the AN serum.

Conclusion

We summarize this discussion of the interpretation of the Youden–Tonks graph as follows: in the present situation in a single external quality control survey, where control materials contain constituents in different concentrations, a point deviating from the diagonal line can be attributable to (1) random intra-laboratory variability, (2) systematic error in the laboratory, (3) constant systematic error if proportional systematic errors prevail in the group under study, or (4) proportional systematic errors if constant systematic errors prevail.

A point on the diagonal line indicates only that, on this occasion, the laboratory shows the same type of systematic differences from the target value as the majority of laboratories.

Proposal

More information can be derived from a two-sample plot in external quality control programs if the program is designed with the possibility of collecting multiple results for the same sets of control materials.

Plotted in a two-sample plot graduated in units of measurement, the position of the mean value will give information about persistent proportional or constant systematic differences from the target value. By inspection of the "dot cloud" of accumulated results, the presence of recurring systematic errors in the laboratory affecting day-to-day variability, such as instability of the standard or baseline problems, can be detected.

A format for reporting in external quality control programs based on these premises has been in use for more than a year in a group of six laboratories in the Netherlands using the quality control serum pools of the MSPQCP as a primary or secondary control. The form (Fig. 8) gives the following information:

1. A table stating the laboratory's monthly mean, SD, and coefficient of variation (CV). SD and CV are calculated as soon as a total of 20 results is received.
2. A graphic display of daily results (arabic numerals)

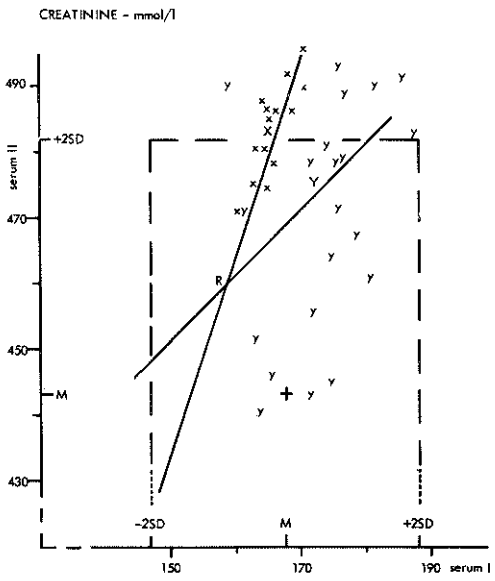


FIG. 7. Two-sample plot for creatinine, axes graduated in units of measurement. X = mean value, laboratory X; x = daily results, laboratory X; Y = mean value, laboratory Y; y = daily results, laboratory Y. For explanation see text.

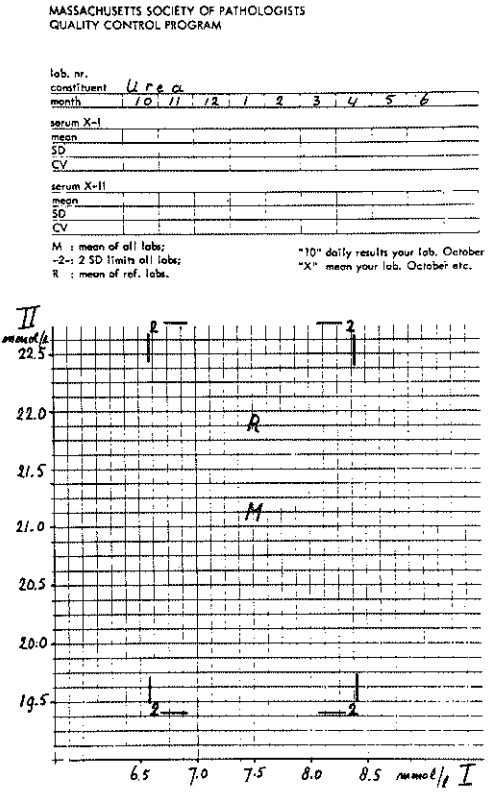


FIG. 8. Example of the suggested form for reporting in external quality control programs. Two lines may be drawn through the target value, one at a slope of 1, the other at a slope of

$$\frac{\text{target value serum II}}{\text{target value serum I}}$$

and monthly means (roman numerals) in a two-sample plot with axes graduated in units of measurement.

Discussion

It has been demonstrated that the Youden–Tonks plot, when used to picture results surveyed, may fail to distinguish between systematic and random errors when control materials in different concentrations are used. One of the consequences is that numerical data expressing the difference of an individual result from the target value as a proportion of the SD observed for all laboratories (SD interval,[9] variance index[16]) give an evaluation of laboratory performance that is not really useful.

The existence of systematic errors causing plotted points to deviate from the diagonal line in the Youden–Tonks graph has been proven by the fact that the laboratory in question consistently reported the same deviating result. In the quotation of Skendzel and Youden (see introduction) the word "consistent" is mentioned twice in correlation with systematic error. It should be realized that single testing of two samples covers within-run variability or consistency; multiple testing is needed to check the influence of day-to-day variability. This is particularly important in view of recent reports[3,5] that intra-laboratory variability can be considerable and may obscure systematic differences between laboratories.

The proposed two-sample plot has been helpful in identification of sources of error in the laboratory and in improving day-to-day precision and compatibility with the target value. Intra-laboratory quality control with a two-sample plot (Twin-plot®) has been described by Griffin and Greeban.[7] In the Twin-plot, however, the center point of the graph is the laboratory's own mean value instead of a target value, which eliminates the possibility of simultaneous external quality control.

As regards the interpretation of the proposed two-sample plot, the following remarks are in order:

1. Before a conclusion as to proportional or constant systematic differences from the target value can be reached, it is necessary to consider the possibility of systematic errors in the laboratory such as, for example, nonlinearity. If, in addition to the control materials in the different concentrations, a 1:1 mixture of the two is tested, it will be possible to determine whether the system shows constant or proportional systematic differences over a range of concentrations (*cf.* operational line, Grannis[6]). The quality control serum value/absorbance standard graph has also been suggested[2] for detection of this type of systematic error in the laboratory.

2. For a number of constituents, the difference between N and AN levels is relatively small. Here, a difference between proportional and constant systematic differences will be difficult to detect. As an example, however, we would like to mention how a differentiation can be made for potassium, if the N level target is 4.0 mmol/l and the target value for the AN level 6.0 mmol/l. A constant systematic difference of 0.4 mmol/l would bring the values up to 4.4 and 6.4 mmol/l, respectively; a proportional systematic difference of 10% would bring them to 4.4 and 6.6 mmol/l. If 1 SD of ten determinations is 0.09 (1.5%) at the 6 mmol/l level, the difference ($6.6 - 6.4 = 0.2$ mmol/l) will be significant at $p < 0.01$, the Student's t value being 5.0. If 1 SD is 0.18 mmol/l (3%), t will be 2.5 and the distinction between proportional and constant differences with the target value can still be made with a p between 0.05 and 0.02.

The graphic display of survey results was proposed by Skendzel and Youden for easy visual interpretation.[10] Presently, there is a tendency towards a large output of numerical data in quality control programs.[9] The suggested individual cumulative two-sample plot in units of measurement is suited for both manual and computerized data handling, and will provide participants with much useful information from visual inspection of the graph.

Acknowledgments. Dr. Bradley E. Copeland, Boston, offered advice and encouragement. Dr. G. T. Berends, Dr. P. Brombacher, Dr. Y. S. L. Heck, Dr. A. R. Helbing, and Dr. H. Oldenziel participate in the experiment. The Steering Committee of the Massachusetts Society of Pathologists Quality Control Program and Mr. R. K. Malkin, Dade Division, American Hospital Supply Corporation provided help essential for the realization of the project.

References

1. Copeland BE, Rosenbaum JM: Organization, planning and results of the Massachusetts Society of Pathologists Regional Quality Control Program. Am J Clin Pathol 57:676–688, 1972
2. Elion-Gerritzen WE: The absorbance of the standard, a valuable tool in quality control of spectrophotometric methods. Am J Clin Pathol 60:493–498, 1973
3. Elion-Gerritzen WE: Attempt at reduction of inter-laboratory variability of clinical chemistry results in a cooperative clinical investigation program, using a common reference serum (abstr). Z Klin Chem Klin Biochem 12:253, 1974
4. Elion-Gerritzen WE, Schippers HMA: Report of two Eurotransplant clinical chemistry surveys (unpublished)
5. Gilbert RK, Chir B: The size and the source of analytical error in clinical chemistry. Am J Clin Pathol 61:904–911, 1974
6. Grannis GF, Grumer ND, Lott JA, et al: Proficiency evaluation of clinical chemistry laboratories. Clin Chem 18:222–236, 1972
7. Griffin DF, Greeban LB: Practical charting technics in quality assurance programs. Monograph by Dade Education, 1971
8. Leijnse B: Kwaliteitscontrole in de klinische chemie, pericula in mora. Lecture at the occasion of the 25th anniversary of the Dutch Association for Clinical Chemistry, November 21, 1973
9. Quality Assurance Service of the College of American Pathologists
10. Skendzel LP, Youden WJ: A graphic display of inter-laboratory test results. Am J Clin Pathol 51:161–165, 1969
11. Stamm D: Ringversuche in der Klinische Chemie. Schweiz Med Wochenschr 101:429–437, 1971
12. Tonks DB: A study of the accuracy and precision of clinical chemistry determination in 170 Canadian hospitals. Clin Chem 9:217–233, 1963
13. Vogel AI: A Textbook of Quantitative Inorganic Analysis. Third edition. London, Longmans, 1962, p 1121–1129
14. Webster's New International Dictionary of the English Language. Third edition, unabridged. Springfield Mass., G. and C. Merriam, 1967, p 1880
15. Ibid: p 2323
16. Whitehead TP, Browning DM, Gregory A: A comparative survey of the results of analyses of blood serum in clinical chemistry laboratories in the United Kingdom. J Clin Pathol 26:435–445, 1973

ADDENDUM II

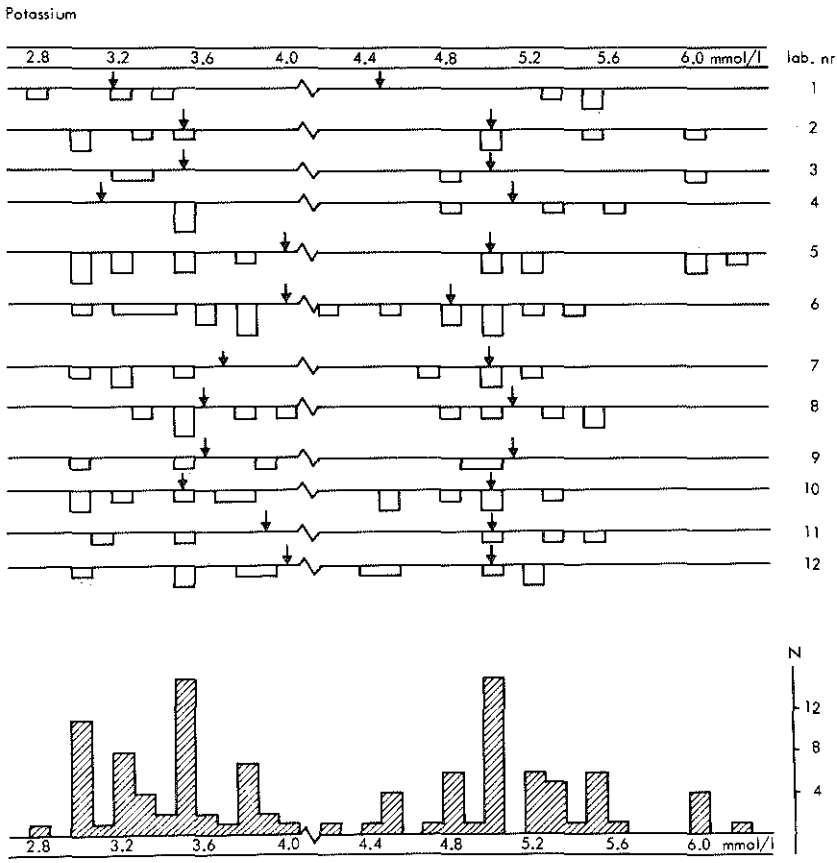


Figure 4-3:

Histograms of action levels given by clinicians per hospital (numbered 1 through 12). The laboratories' upper and lower limits of normal are indicated by arrows.
At the bottom of each figure histograms are given of accumulated data.

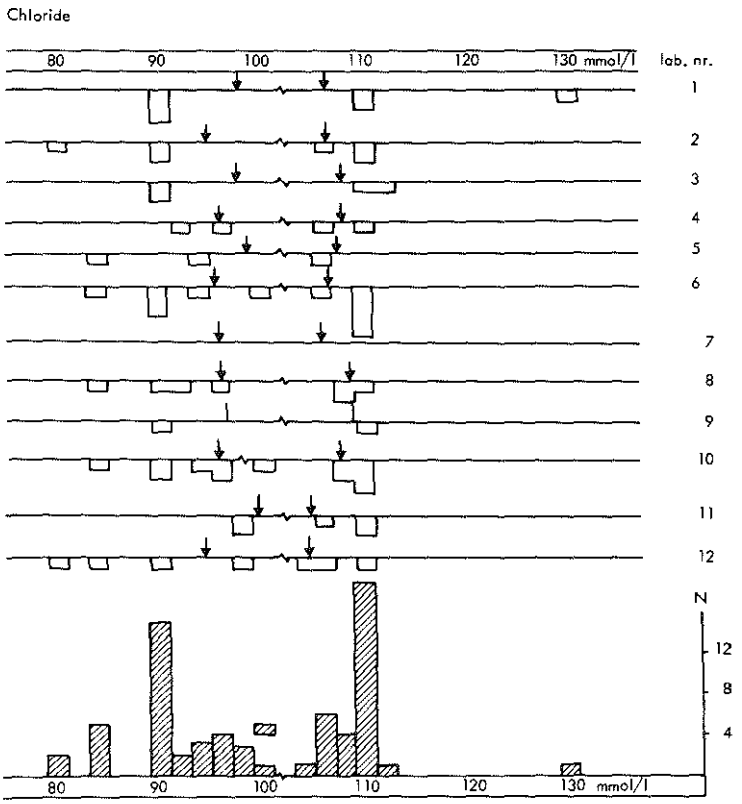


Figure 4-3 cont.

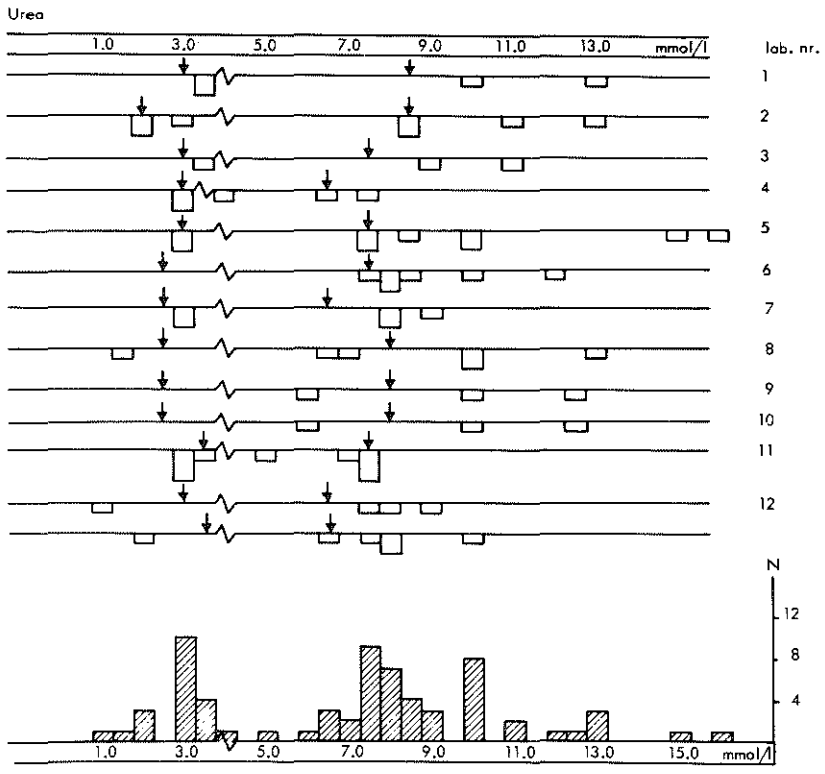


Figure 4-3 cont.

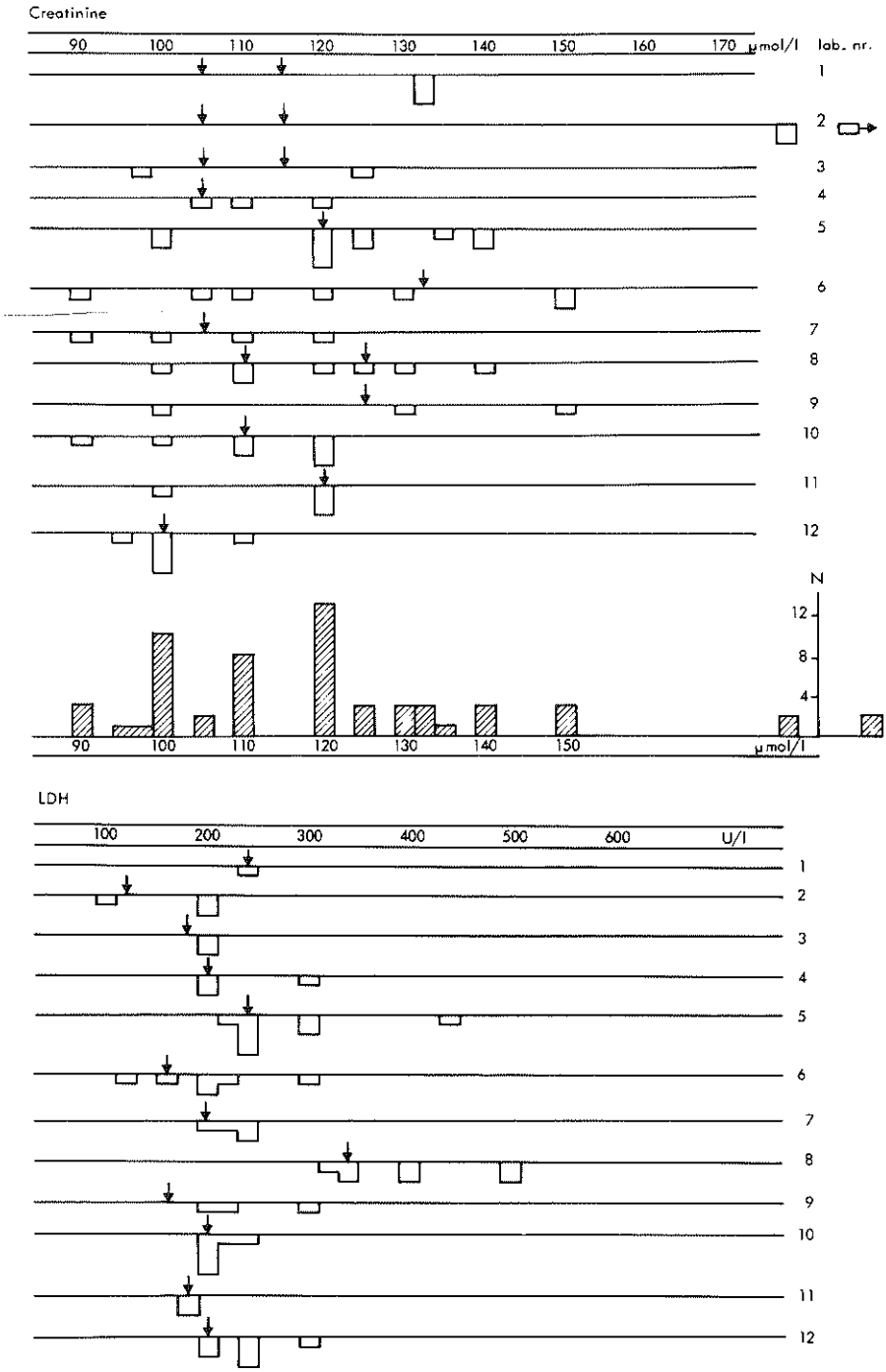


Figure 4-3 cont.

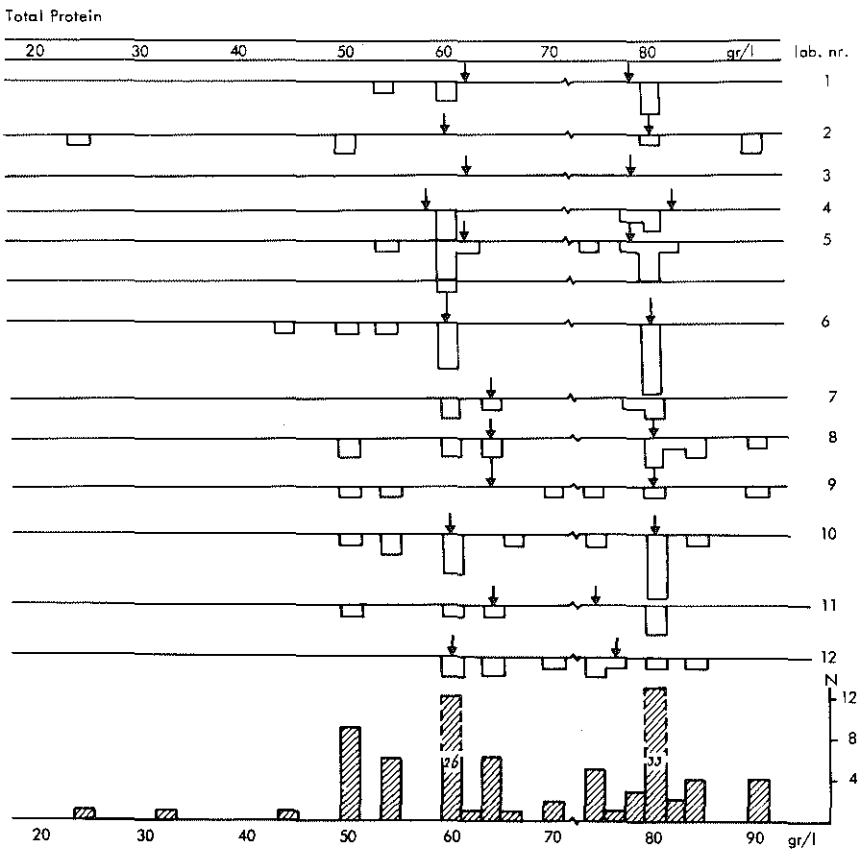


Figure 4-3 cont.

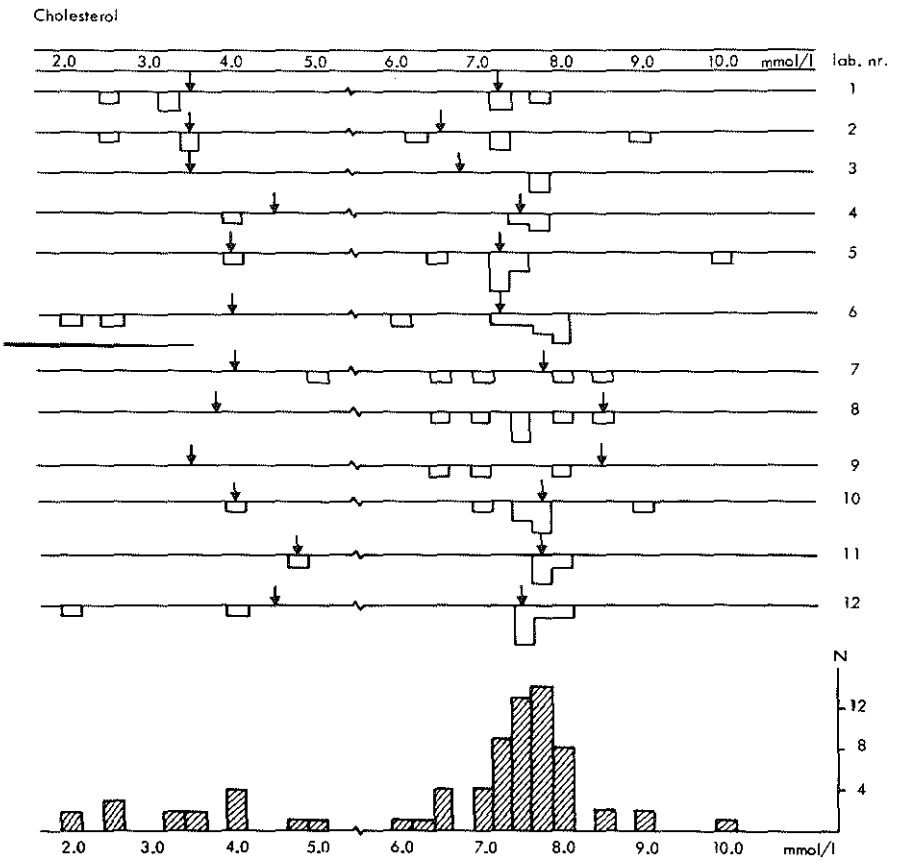


Figure 4-3 cont.

ACKNOWLEDGEMENTS.

In fact much of the work described in this thesis has been done by people other than the author. It is an unsatisfactory situation that laboratory directors and physicians cooperating with this study can not be named in person. Their contribution to this work was indispensible.
Much support was received from Prof. Dr. B. Leijnse, Dr. B.E. Copeland, Dr. R. van Strik and in a later stage from Dr. J.H.P. Wilson. If this thesis in any way meets the expectations raised by its title I owe this largely to our constructive and pleasant discussions.
This work was effected during a period of part-time employment at the department of Experimental Surgery of the Erasmus University.
Dr. D. Westbroek stimulated this work to great extent by giving me the freedom to choose this subject and to allocate my time to this project not directly related to experimental surgery.
The department of Audiovisual Services, in particular Mr. H.M. Kneefel, carefully and patiently provided for the many pictures of this thesis.

CURRICULUM VITAE.

1944-1946	Openbare lagere school te Lunteren.
1946-1949	Eerste Leidsche Schoolvereniging te Leiden.
1949-1955	Rijnlands Lyceum te Wassenaar. Eindexamen gymnasium β.
1956-1964	Rijks Universiteit te Leiden. Candidaatsexamen scheikunde (F'). Doctoraalexamen biochemie met bijvakken medische chemie en farmacologie.
1962-1965	Afd. Medische Chemie, Rijks Universiteit Leiden. Candidaats/doctoraal assistente.
1964-1967	Derwent Publications Ltd., London. Literatuuronderzoek.
1967-1968	Martin Place Hospital, Madison Heights Mich., U.S.A. Medewerkster bij de afdeling Special Chemistry.
1968-1970	Highland Park Hospital, Highland Part Ill., U.S.A. Biochemica bij het klinisch chemisch laboratorium.
sinds 1971	Laboratorium voor Chirurgie, Erasmus Universiteit Rotterdam. Wetenschappelijk medewerkster.